IMPERIAL COLLEGE LONDON

THE COMPUTATIONAL ENHANCEMENT OF AUTOMATED NON-DESTRUCTIVE INSPECTION

by

Nicholas James Standish Brierley

A thesis submitted to Imperial College London for the degree of

Engineering Doctorate

Department of Mechanical Engineering

Imperial College London

London SW7 2AZ

January 2014

This page intentionally left blank

Declaration of Originality

This thesis is my own work, produced under the supervision of Professor Peter Cawley. Where I have built on the work of others, I have fully acknowledged this, providing appropriate references.

Nicholas Brierley

31/01/14

This page intentionally left blank

Copyright

This page intentionally left blank

Abstract

In industrial NDE it is increasingly common for data acquisition to be automated, driving a recent substantial increase in the availability of data. The collected data need to be analysed and currently this is largely done manually by a skilled operator - a rather painstaking task given how rarely defects occur. Moreover, in automated NDE a region of an inspected component is typically interrogated several times, be it within a single data channel due to multiple probe passes, across several channels acquired simultaneously or over the course of repeated inspections. The systematic combination of these diverse readings is recognised to offer an opportunity to improve the reliability of the inspection, for example by enabling noise suppression, but is not achievable in a manual analysis. Hence there is scope for the inspection reliability to be improved whilst reducing the time taken for the data analysis by computational means. This thesis describes the development of a software framework providing a partial automation capability, aligning then fusing the available experimental data to declare regions of the component defect-free to a very high probability whilst readily identifying indications, thereby optimising the use of the operator's time. The framework is designed to be applicable to a wide range of automated NDE scenarios, but the focus in development has been on two distinct, industrial inspections: the ultrasonic inspection of power station turbine rotor bores and the ultrasonic immersion inspection of aerospace turbine disks. Results obtained for industrial datasets from these two applications convincingly demonstrate the benefits of using the developed software system.

This page intentionally left blank

Acknowledgements

I would like to express my heartfelt gratitude towards my supervisor Professor Peter Cawley of Imperial College, not only for years of patient advice but also for accepting me into his Non-Destructive Testing Group in the first place. Moreover, his demanding yet relaxed supervision style but has contributed both to my productivity and well-being these last years.

I am most grateful to my colleague, Trevor Tippetts, for hours discussion, thousands of lines of code, occasional frivolity and teaching me to program in Python. Without his contributions I would have been lost in more ways than one and I cannot imagine a more constructive and enjoyable partnership.

I would like to thank Paul Crowther and Chris Ward of RWE npower for their support and industrial input, including their willingness to answer my awkward questions. Similarly, I thank the inspection staff of RWE npower's Ferrybridge site for putting up with me whilst I hovered around them armed with a clipboard as they worked.

I am indebted to David Wright and Tony Dunhill of Rolls-Royce (Aerospace) for their support, knowledge and continued enthusiasm concerning the company's application of the project work.

Within the EngD program I am thankful for Professor Chris Scruby's help, and for the company of my colleagues of the 2009 cohort, Ed and Chris.

In the Imperial Non-Destructive Testing Group I am grateful to all past and present members for creating a stimulating and enjoyable work environment. The following deserve a special mention: Tim H. and Peter H. for putting up with my coding questions, Robin and Jacob for being entertaining desk neighbours, Anders, Tristan and Daniel for being the most reliable of lunch partners.

The wider Research Centre in Non-Destructive Evaluation (RCNDE) has provided many opportunities for the useful exchange of ideas for which I am thankful. I am especially grateful to Simon Pickering of the University of Bath for many stimulating programming conversations.

I would like to extend thanks to the Engineering and Physical Sciences Research Council (EPSRC) and the Royal Commission for the Exhibition of 1851 for supporting me and this project financially.

Last but not least, I would like to thank my parents David and Ulrike for taking a continued interest in my education and my sisters Charlotte and Clarissa for supporting me through thick and thin.

Contents

List of Figures

List of Tables

This page intentionally left blank

Nomenclature

Registration

ϕ_0	scan start circumferentially
ρ	probe angle, in plane normal to z-axis
σ	probe skew
θ, r, z	cylindrical polar physical coordinates
d	radial distance, along A-scan
d_0	scan start radially, along A-scan
d_{vs}	speed of sound in component
f_s	data sampling frequency
g	gain factor
R	bore radius
t, u, v	3D index space, with basis vectors $\bar{t}$, $\bar{u}$, $\bar{v}$
u_{rev}	number of A-scans per revolution
w	index along helical probe path
x, y, z	Cartesian physical coordinates
z_0	scan start axially
z_{rev}	helix pitch, axial increment per revolution

Data fusion and detection

δ	*Dirac* delta function
$\mathcal{F}$	*Fourier* transform
ω	*Fourier* space variable
f	number of occurrences of r
$H_{01}, \ldots, H_{0k}$	null hypotheses for k independent tests
H_0	null hypothesis
H_1	alternative hypothesis
k	number of independent tests or samples
m	number of distributions considered
n	number of levels in distribution
P	probability
p	probability density or mass
P_{cdf}	cumulative distribution function, c.d.f.
P_{sf}	survival function, s.f.
q	*p-value*
$q_1, \ldots, q_k$	*p-values* for k independent tests
r	minimum of $q_1, \ldots, q_k$
w_a	weight factor associated with x_a
x	random variable
x_a	observed value of random variable x
U(0,1)	uniform distribution over [0,1] interval

Chapter 1

Introduction

1.1 Non-Destructive Evaluation

Non-Destructive Evaluation (NDE), also termed Non-Destructive Testing (NDT), is a field of applied physics and engineering concerned with assessing the structural integrity of load-bearing components without causing damage. It is not only a field of active scientific research but also an important area of industrial practice, both for quality assurance in high-value manufacturing, for example in the aerospace industry, and maintenance of plant, for instance in power generation. NDE encompasses a wide range of sensing techniques, from thermography to radar and ultrasonics to radiography [1], and there is some overlap with the fields of both medical imaging [2] and remote sensing [3]. The particular choice of technique is very application specific, dependent not only on the material of the component to be examined, but also on the nature of the defects one wishes to guard against. It can in fact be very advantageous to use more than one sensor system, with the potential to improve coverage, inspection speed and / or sensitivity.

There are significant costs associated with conducting industrial NDE. Staff time, inspection equipment and consumables, and outage or production time, for mid-manufacture and in-service inspections respectively, all contribute. Given these costs of inspection, NDE is only applied where the costs of failure and component replacement are very high, but the probability of a defect necessitating the replacement of a part is low. For in-service inspection, if these conditions are not met, a component is likely to be scrapped

and replaced without inspection at the end of its service life, calculated at the time of manufacture based on the loads the component is expected to experience during the course of its use. A materially more efficient approach that also allows extensions beyond the originally calculated service life is condition-based maintenance where NDE is used to inform the maintenance schedule. For inspection prior to operation, during manufacture, NDE can be used to place an upper bound on imperfections present, enabling longer service life values to be computed, with important implications for operating costs.

Because NDE is important for managing the risk to life, limb and property associated with highly load-bearing components, in many industries the use of NDE is mandated by regulation, often in response to a catastrophic failure [4]. Especially the nuclear and aerospace industries are highly regulated [5]. While ensuring minimum standards such regulation does limit the ability of industries to adopt new inspection methodologies. Additionally, any new approach brings with it risks and initial costs, such as those associated with re-training staff. Nonetheless, new approaches can bring significant benefits, typically in the form of reduced costs, either due to the inspection itself being less expensive or an improved inspection allowing the lowering of running costs.

One general inspection approach that industries have been increasingly adopting is the automation of the data collection in an inspection. Automated inspection essentially involves a mechanical scanning system that moves one or more sensors across the component systematically, as is explained further in Chapter 2, together with the two examples that provide the industrial focus of this thesis. The first application is an in-service inspection of a power station component, the second a mid-manufacture inspection of an aerospace jet engine part. In both cases the component is of forged metal. The NDE technique that is central to both applications is pulse-echo bulk wave ultrasonic testing (UT) [6]. This is based on using a transducer, typically made of a piezoelectric material, to excite a high-frequency, low amplitude stress wave in the component via a coupling medium such as water or gel. This stress wave propagates through the component but interacts with encountered discontinuities, including defects such as cracks and non-metallic inclusions, giving rise to waves that travel back to the transducer where they can be detected as a change in voltage. The time-trace of the transducer voltage is known as an A-scan, and is the starting point for all bulk wave UT NDE [1].

1.2 NDE Reliability

As the following sections will explain, one of the ultimate aims of the project was to improve the reliability of certain NDE inspections. It is therefore worth examining what is meant by NDE reliability. There are slightly different models describing this. Rummel [7] identifies three components: reproducibility, repeatability, and capability. These broadly refer to using a suitable equipment calibration, having inspectors follow a well-written procedure, and the ability of the procedure, in principle, to detect defects of the type sought, respectively. The Federal Institute for Materials Research and Testing (BAM) of Germany [8] have proposed a perhaps more comprehensive model, consisting of the intrinsic capability of the inspection, application parameters and human factors, all linked by an overarching organisational context within which the inspection takes place. The first term encompasses the physics-based limitations of the inspection, the second variables such as those describing the defect and processing. The third element relates to all causes of human imperfection in the role of an inspector, ranging from inexperience to uncomfortable working conditions and "social loafing" [9]. In either of the two models, it is clear that NDE reliability can only be assured if several key components are present in equal measure, and it is not adequate to rely entirely on the physical ability, perhaps quantified statistically, of the chosen NDE technique. Within each of the key elements of NDE reliability, there are numerous contributions and considerations that may be adjusted to improve overall performance [10], some of which we will return to in later sections of this thesis.

1.3 Industrial Background

The author's Engineering Doctorate was directly supported by RWE npower. RWE npower is one of the major integrated energy companies of the UK. The organisation has several branches, including a retail company that supplies over five million residential and business customers with energy products, and a power generation organisation that operates numerous coal, gas and biomass power stations [11]. Power station operations and maintenance services, including NDE, are part of RWE Power International, a fur-

ther part of the organisational structure. These maintenance services are provided both to RWE npower's own generation sites and external clients. RWE npower therefore has significant operational NDE experience and an interest in deploying new, improved NDE techniques [12].

The other industrial sponsors of the project were Rolls-Royce (Aerospace) and Tenaris. Rolls-Royce is a cutting-edge manufacturer of gas turbine engines, serving civil and defence aerospace, marine and power sectors world wide. Tenaris is a leading manufacturer of steel tubes and related products for the global energy industry. RWE npower and Rolls-Royce provided the two specific applications considered in the project, described further in Chapter 2. Tenaris did not provide a project application, as the company's high speed production environments were recognised to be excessively demanding for initial development purposes. Instead, the company's involvement was limited to financial support, granted as the project output was envisaged to find applications within the company in the long-term. This large project was completed in close collaboration with another research student at Imperial College, Trevor Tippetts. While this thesis describes the progress made by the author in fulfilling the above objectives in the context of the RWE npower application, Tippetts' PhD thesis [13] focuses on the Rolls-Royce application. Section 3.4 explains the division of responsibilities in the project. Key individual contributions are highlighted and acknowledged throughout the thesis.

1.4 Project Objectives

As will be discussed further in Sections 2.1 & 2.5, while automated NDE has a number of advantages over conventional, manual testing, it is not a fully mature technology and brings some new challenges that are yet to be addressed. Specifically, the increased availability of testing data can readily become overwhelming, and vast quantities of data are currently not exploited to the fullest extent possible. This project sought solutions to some of these challenges, to enable automatic NDE to reach its full potential and facilitate the further deployment in industry. The overall objective was to improve the inspection reliability (see Section 1.2) and speed of data analysis for automated NDE inspections, focussing on data from bulk wave ultrasonic testing. This was to be achieved through the

design of a piece of software that:

- exploits the inspection data, including historic data from a previous inspection if available, to the fullest extent possible, especially making use of multiple inspections of a given component volume
- is as general as possible, capable of handling data from a range of automatic inspection types and NDE sensors
- applies a detailed understanding of the physics underlying the inspection
- leverages the extensive calculation capabilities of modern computers, but can run on a relatively standard, stand-alone desktop computer, appropriate for use in an industrial shop-floor environment
- identifies and segments regions of the inspected part that are defect free with a very high probability, whilst flagging up regions that are of questionable integrity
- allows experienced operators to focus on those regions of the inspected part where their skills are best applied, thereby reducing fatigue and associated human errors

The project also set out to identify recommendations for improving the inspection procedures to ensure more reliable data acquisition, in the context of the outlined partially automated data analysis. It should be emphasised that the decision not to aim for a fully automatic analysis system, such as one based on artificial intelligence (A.I.) [14, 15], was itself motivated by reliability considerations: such systems are inevitably limited by the quality of the training data available and a high degree of automation is liable to encourage excessive trust in the system, negatively affecting overall inspection reliability [9]. There are practical reasons for this decision, too, as the limited amount of physics involved in an A.I. evaluation and the "black-box" nature of the analysis means that such systems are unlikely to be adopted by industry, especially given regulatory constraints.

1.5 Thesis Structure

Although the software is meant to be highly general as stated in the preceding project objectives, necessarily the development work focussed on specific examples of automated

NDE, provided by the sponsor companies. Chapter 2 provides details of the two automated inspections considered: an in-service inspection of a power station turbine rotor, as used by RWE npower, and the mid-manufacture inspection of an aerospace jet engine disk, as used by Rolls-Royce. The similarities and differences between these two inspections are emphasised, as are the advantages of automated NDE over traditional approaches and shortcomings of the current procedures.

Chapter 3 gives an overview of the designed software system and explains the division of responsibilities between Tippetts and the present author. Some of the implementation challenges faced, and the solutions developed are also described.

In Chapter 4 we consider the approach taken to register, that is align to a common coordinate system, data from different channels in the rotor inspection. A formal data acquisition model is introduced, and the processing stages of registration are expounded: feature extraction, feature matching, quantifying the quality of alignment through the evaluation of error metrics and optimisation. This chapter features some of the author's contributions to knowledge.

Following the methodology of the preceding chapter, Chapter 5 presents the results of testing the registration scheme using an industrial dataset.

Chapter 6 describes the development of a novel and highly general framework for the probabilistic combination of signals acquired in different data channels but relating to the same spatial location in the component. The basic processing stages are data decorrelation, local data modelling to convert amplitudes to probabilities taking into account local data statistics, then fusion of the different probabilities associated with a spatial region of interest using a consensus test. This chapter contains the author's primary claim to novelty and academic merit.

The results of extensively testing the data fusion system of Chapter 6, using data from both the automated inspection applications considered, are presented in Chapter 7. The novel fusion is shown to offer dramatic gains in indication detection capability and analysis reliability over conventional analyses.

Finally, Chapter 8 offers some concluding remarks as well as suggestions for further work.

Chapter 2

Automated Inspection

2.1 Introduction

As mentioned in the preceding chapter, the automation of the data acquisition in industrial NDE is increasingly common. This essentially involves a mechanical scanning system moving one or more sensors across the component systematically whilst data is collected. Such automation is attractive to industry as the variability in the data collection is significantly reduced compared with manual approaches. The systematic nature of the scan ensures coverage and consistency across different parts, and reduces the scope for human error to compromise the inspection (see Section 1.2). The higher scanning speeds and positional accuracy achievable by mechanical means also allow more data to be collected from a part in less time than in traditional approaches, especially if multiple sensors are deployed simultaneously. In principle, more data bring the promise of greater defect sensitivity. Automated inspection also relieves staff of the most tedious work, often in cramped or otherwise uncomfortable and potentially dangerous environments. Moreover, compared with manual UT, for example, a complete and permanent (subject to data storage choices) record of the inspection can be obtained, suitable for off-line analysis, easy reporting and insurance purposes.

This chapter introduces in detail the two examples of automated inspection that are the focus of this work. First, we examine the in-service power station rotor bore inspection used by RWE npower, that serves as this author's main software application. Next we

consider the mid-manufacture inspection of aerospace engine disks used by Rolls-Royce that is the focus of Tippetts' work [13]. We then highlight the similarities and differences in these inspections, as well as the shortcomings of the current procedures, setting the scene for this project and subsequent chapters.

2.2 Rotor Bore In-Service Inspection

This inspection relates to the hollow bore found on many of the steam turbine (and generator) rotors operated by RWE npower - see Fig. 2.1a. These bores were cut out of the ferritic steel rotors to remove the bulk of the inclusions from the forging-based manufacturing process that serve as crack initiation sites when operating at high temperatures and speeds. Bore sizes vary, but 150mm diameter is typical, stretching the length of the rotor - around 6m. Despite the bore, any remaining inclusions are most likely to be found in the region at the centre of the rotor, and machining marks from the boring can themselves act as crack initiation sites. Possible cracks are furthermore of the greatest concern near the axis of the rotor as this is where the largest overall tensile stresses occur, taking into account thermal, centrifugal and residual stress contributions [16]. Cracks tend to be the result of mechanical fatigue or, in the case of higher pressure turbines, creep and thermal fatigue [17]. The size and energy of the rotors means failure is catastrophically damaging, both physically and financially, giving rise to an application for NDE.

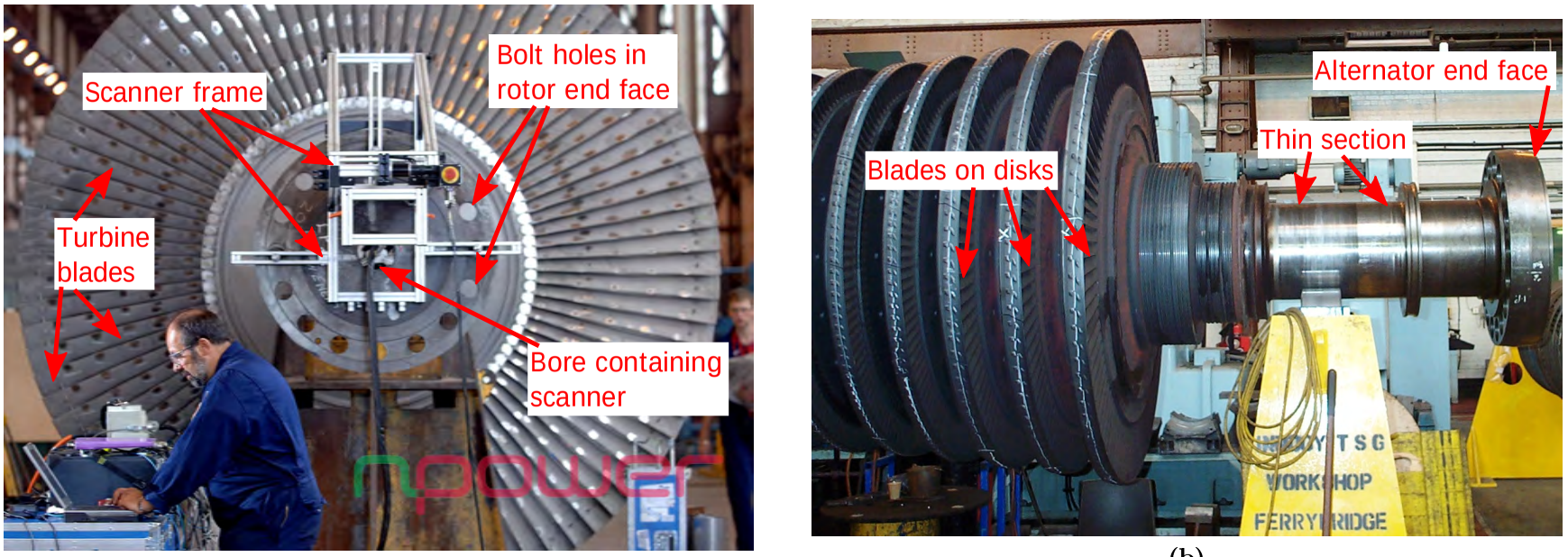
2.2.1 The scanner

The rotors are inspected during routine major power station outages, approximately every ten years. The rotors are removed from their housings to allow access to the bore. The bore surface must be prepared prior to inspection as a hard scale accumulates during operation which must be removed by blasting or honing [18]. The quality of the surface finish may be confirmed by an initial visual inspection of the bore, completed using a boroscope. Then the rotor bore scanner system made by Phoenix Inspection Systems Limited [19] is mounted on one end face of the rotor, with the scanner head in the bore - see Figs. 2.1a & 2.1b. The choice of end, either steam or alternator, varies from

inspection to inspection depending on access opportunities. The scanner head is gradually withdrawn as the head rotates, the movement being controlled by separate motors, giving rise to a helical path. The position of the scanner is monitored by encoders. The rod used to withdraw the probe consists of joined parts and the axial scanning is completed in sections equal to the length of the rod parts, with the scan sections overlapping slightly to ensure good coverage.

2.2.2 The scanner head

The scanner head, illustrated in Figs. 2.2a & 2.2b, has two spring-loaded probe pans, offset by 180°, that push up against the bore surface. The smooth movement of the probe pans is facilitated by ball rollers. The probe pan system allows the scanner to fit a range of bore sizes and follow the bore surface even in bores with slightly changing diameter. Each sensor is held in position on the probe pans by a spring-loaded holder to further ensure consistent surface contact. In the standard configuration [20], each probe pan has two ultrasonic testing (UT) probes with Rexolite® wedges of a shape appropriate to the bore diameter. Irrigation channels through the wedges ensure the mineral oil couplant is injected where needed. Two twin-element probes provide normal-to-surface compression (longitudinal) waves, but focussed to different depths. A further two twin-element probes provide shear (transverse) waves at 60° to the surface normal, in a plane normal to the bore axis, either with or against the direction of scanner rotation. The probes used are summarised in Table 2.1. One of the probe pans additionally carries a seven-coil eddy current (EC) array [21] operating at a 360kHz excitation frequency, but only five coils are used in practice due to redundancy. Overall this sensor system enables simultaneous data collection across a total of nine channels (counting the used EC coils separately). The arrangement of probes effectively only allows for the detection of axial cracks, although the EC array has some sensitivity to circumferential surface flaws. As the focus of this work is the data from the volumetric ultrasonic inspection we will not dwell on the eddy current further.



(a) (b)

Figure 2.1: The power station rotor bore inspection. In the photograph (a) the scanner frame is seen mounted on the end face of a low pressure turbine rotor, with the scanner head inserted down the central bore of the rotor, while an inspector monitors the inspection output on a laptop. In (b) a photograph of the alternator end of an intermediate pressure turbine rotor, photographed side-on, is shown. The bore that is scanned runs the length of the rotor, typically about 6m long. *Adapted from photos provided by RWE npower.*

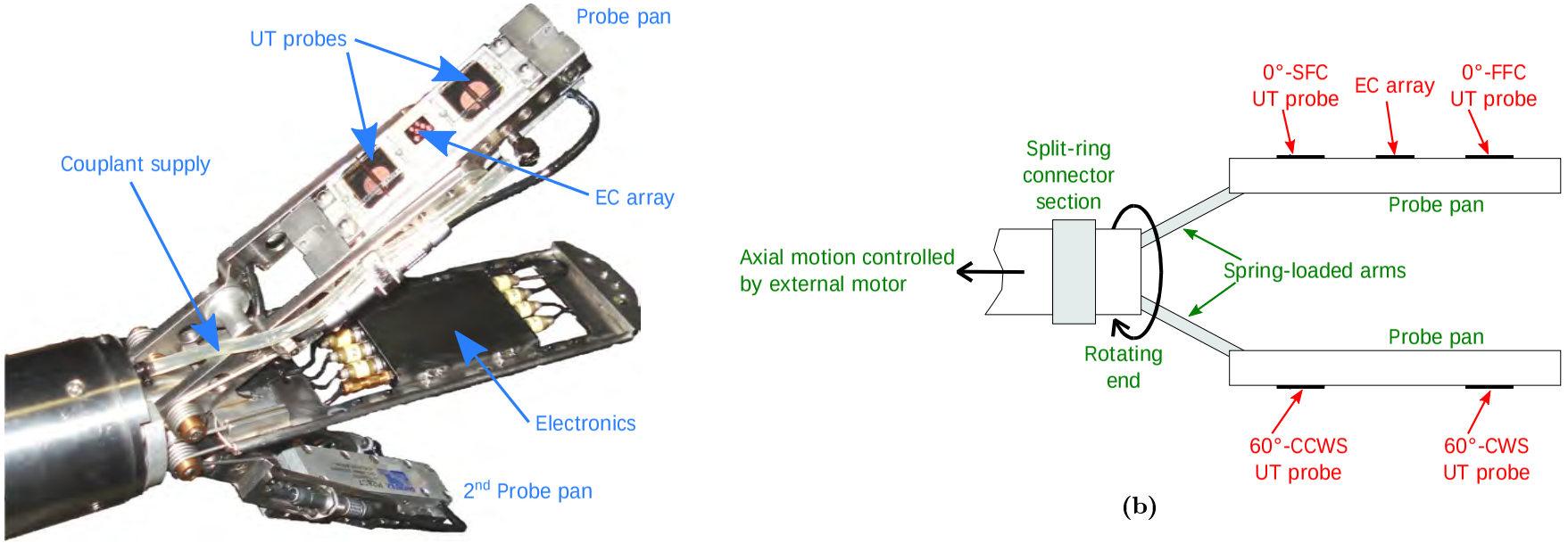


Figure 2.2: The head of the Phoenix Inspection Systems Limited rotor bore scanner [19] as used by RWE npower. In (a) this is presented as an annotated photograph - as an indication of size, note the face of the EC array has 1cm long edges. *Photo taken by author at RWE npower's Ferrybridge site.* In (b) a corresponding simplified schematic is shown. Probes are held in the probe pans by spring-loaded holders to ensure good surface contact, and probe wedges contain irrigation channels for the mineral oil couplant. There are also ball rollers on probe pans to facilitate smooth omni-directional motion across the surface.

Number	Name	Abbreviation	Frequency	Details
1	60° clockwise shear	60°-CWS	4MHz	Pointing circumferentially, in direction of head rotation
2	60° counter-clockwise shear	60°-CCWS	4MHz	Pointing circumferentially, opposed to direction of head rotation
3	0° short-focus compression	0°-SFC	5MHz	Focussed to 0-15mm depth
4	0° far-focus compression	0°-FFC	5MHz	Focussed to 10-60mm depth

Table 2.1: The 4 ultrasonic data channels of a rotor bore inspection, shown with assigned channel number, abbreviation and details of the probe specification..

2.2.3 Data collection

During an inspection, each probe moves over the bore surface along a helical path, but given the axial and rotational offsets associated with the arrangement of the probes on the scanner head (as seen in Figs. 2.2a & 2.2b), these helical paths are offset relative to each other. A visualisation of the paths of two probes is shown in Fig. 2.3. The pitch, that is the axial increment per revolution, of the scanner helical path is 5mm, and probes are typically set to acquire 720 A-scans per revolution. Probe signals are sampled at 25MHz, envelope-detected, digitised into 8-bit samples (corresponding to 256 amplitude levels) and recorded on a computer using Zetec UltraVision® [22]. The scanner moves at 25° per second circumferentially, 0.35mm/s axially [20]. The surface speed of the probe (for a 150mm diameter bore) is therefore 33mm/s and the scanner has 20ms per surface sample position during which to acquire data. These figures are to be compared against the speed of sound in the component, taken to be 3230m/s and 5900m/s, for shear and longitudinal waves respectively, and the typical total (full-skip) time of an A-scan of 60µs. Given the orders of magnitude between the speeds and times it is clear that there will be no problems due to the fact the scanner does not stop to collect A-scans.

The procedure [18] requires the ultrasonic inspection to only cover a depth of 100mm from the bore surface. Consequently, there is no expected return signal for any of the ultrasonic channels (with the exception of the far-focus compression wave channel in the thinner end regions of the rotor - such as those seen in Fig. 2.1b) if the part is defect-free, as the radius for which data is collected falls short of the outer features of the rotor. This is problematic for generally ensuring the quality of the data collected, and also for our data processing purposes here, as we will discuss in Chapter 3. To help address this problem the scanner is set up immediately before a bore inspection using a standard calibration

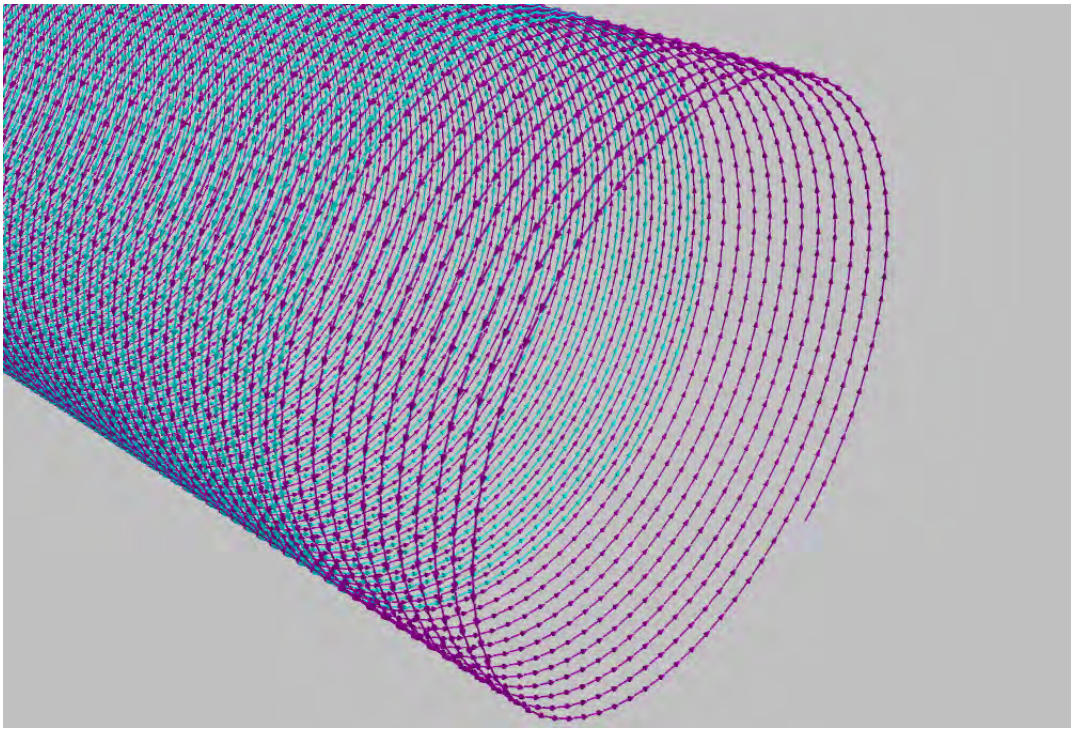
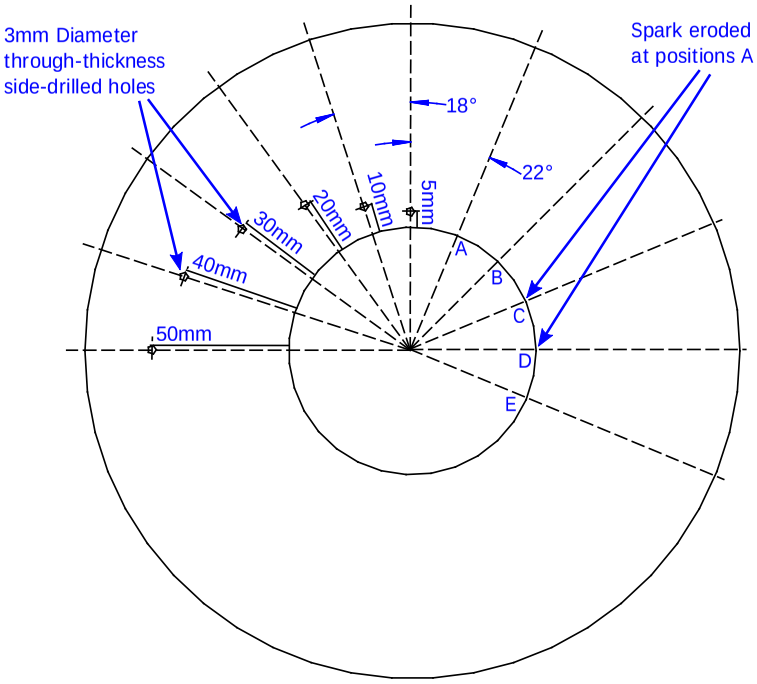


Figure 2.3: A screen shot of a 3D rendering of the helical motion of two probes during a rotor bore inspection. Purple and cyan arrows correspond to the shear clockwise and shear counter-clockwise probes of this inspection, respectively. The helices shown have a 5mm pitch, each arrow spans 10 A-scan sample locations on the bore surface - a cylinder of 153mm diameter. Note how the paths of the two probes are offset and interwoven.

block with machined defects for the UT (and EC) probes to detect. The calibration piece is typically mounted to act as an extension of the rotor, allowing the scanner to continue directly on to the rotor after set up. Such a test piece is illustrated in Figs. 2.4a & 2.4b, showing the side-drilled hole UT target reflectors. These test defects are used to construct Distance Amplitude Correction (DAC) curves for all probes to allow the effects of beam spreading and attenuation to be compensated for [6]. In some inspections this DAC is integrated into the data recording, by the application of Time Corrected Gain (TCG). The gain settings used to construct suitable DACs (taking into account on-screen height of the observed signal according to [18]) from the calibration scan are then supplemented with a further 6dB of gain when moving the scanner to the rotor bore, in the hope of compensating for example for any loss of coupling going from one surface to the other. The data acquisition on the rotor itself, henceforth known as the main scan, as opposed to calibration scan, then proceeds.

(a)



(b)

Figure 2.4: A rotor bore calibration test piece. In (a) a photograph of such a test piece is presented. The disk section containing the target reflectors (side-drilled holes for ultrasonic testing) and the guide tube are clearly identifiable. A similar guide tube also extends out of the far-side of the disk section. *Photo taken by author at RWE npower's Ferrybridge site.* In (b) a schematic diagram is shown, viewing the component down the bore axis - not to scale. A number of ultrasonic and eddy current targets, taking the form of through-thickness side-drilled holes and spark eroded slots, respectively, are indicated. The test piece disk shown has guide tubes attached axially to extend the length of bore. *Diagram derived from [23].*

2.2.4 Current data analysis

The current data analysis is based on investigating all ultrasonic signals that fall at or above the DACs set from the calibration scan [18]. Such signals are to be investigated by an examination of the echo dynamics (i.e. the way the A-scans change as the probe passes over the responsible reflector), either within the originally saved data or possibly by revisiting the responsible section of the bore for a second, higher resolution scan. Such methods should permit indications to be classified. Indications of a measurable spatial extent should then be sized using a sizing technique appropriate to the determined flaw type [24] and recorded. The gathered information is then used by fracture mechanics and component lifing experts to sentence the part on behalf of the station manager who is responsible for the overall plant safety and accountable to the Health and Safety Executive (HSE). Rotor rejections are very rare. The author is aware of only one RWE npower rotor scan after which significant remedial work was required, and the responsible defect in that case was a large surface-breaking crack that could readily have been identified from just a visual inspection.

2.3 Turbine Disk Mid-Manufacture Inspection

This inspection concerns the inspection of titanium aerospace jet engine disk forgings, as used at Rolls-Royce. The integrity of disks is essential to the safety of jet engines because the kinetic energy of a disk when the engine is under load is so high that failure containment by the engine housing is impossible. Given access constraints in an assembled engine, disks are only inspected during manufacture and then assigned a service life, after which the part is replaced. The inspection therefore seeks to identify tiny inclusions or similar imperfections in the disk that could serve as crack initiation sites and then lead to failure before the end of the calculated service life. Because the final shape of these disks is extremely complicated, including for instance slots to hold blades along the outer circumference, disks are inspected in a mid-manufacture stage when the disk has a rectilinear cross-section. A photograph of a typical disk is shown in Fig. 2.5a, next to the disk cross-section in Fig. 2.5b. The relatively simple shape, at least compared with

the final product, greatly simplifies the inspection as there are fewer interfering reflection and diffraction effects. However, such an inspection is wasteful in the sense that regions of the inspected part will be machined away in later manufacturing stages so the presence of defects in those regions is largely irrelevant. As this inspection is the focus of Tippetts' thesis [13], only a shortened description is provided here.

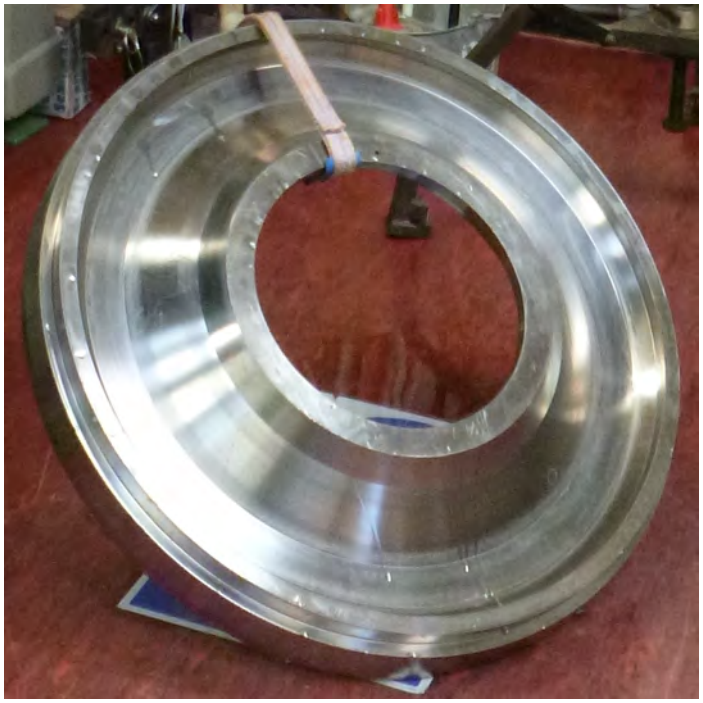
2.3.1 The scanner & data collection

The disks are inspected in a water-filled immersion tank system built by ScanMaster [25]. A disk is lifted onto a turntable that then spins whilst a manipulator arm above moves a single ultrasonic probe in a radial plane. The scanner is shown in Figs. 2.6a & 2.6b. The probe is moved to scan each surface of the disk rectilinear cross-section (see Fig. 2.5b) in turn, with the probe taking an axial and / or radial step after every turntable revolution. In fact each surface is scanned three times; normal to surface, and at $+5°$ and$-5°$ to the surface normal, in a radial plane. Half-way through the inspection, the disk is manually flipped over to allow access to the other side of the disk. The manipulator is set-up to maintain a constant water path to the component surface, such that the focus of the 5MHz focussed probe lies just beneath that surface. Note that refraction and possible mode-conversions must be considered in such situations, although in practice the refracted longitudinal (compression) wave is by far the most significant [13]. This travels at around 6170m/s in the titanium forging.

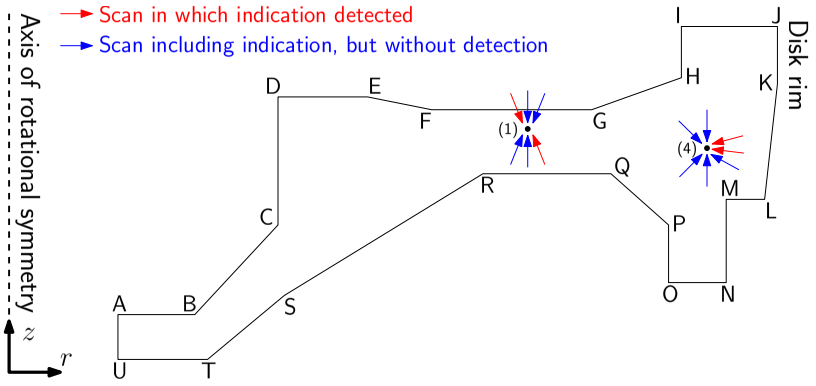
Before the start of the disk scan, the probe is calibrated using a set of flat-bottom hole test pieces down the side of the tank, seen in Fig. 2.6b. A Distance Amplitude Correction (DAC) is computed and this is then applied in hardware, so that all recorded data already include a compensation for beam spreading and attenuation.

2.3.2 Current data analysis

Given the DAC applied in the data acquisition, a global amplitude threshold, set at -18dB down from the flat-bottom hole calibration, is applied. Any signal that reaches or exceeds this threshold is investigated manually - in practice the scanning system will usually be set up to stop when this happens, requiring the immediate attention of a human operator.



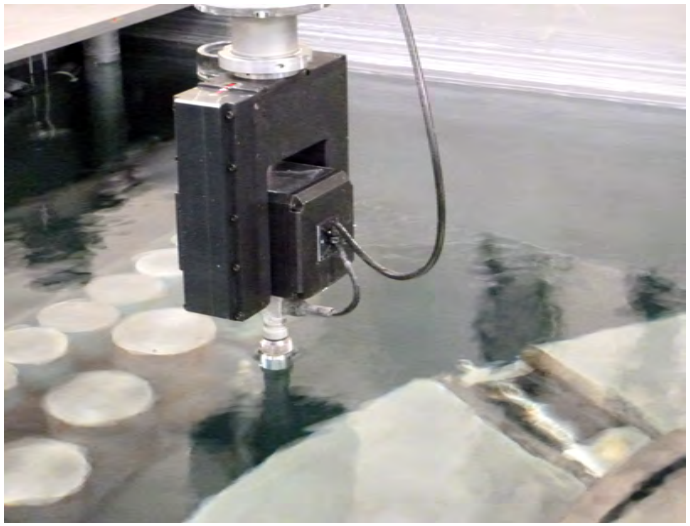
(a)



(b)

Figure 2.5: A titanium aerospace jet engine disk to be inspected in its mid-manufacture stage. In (a) a photo is presented, showing the underside of the disk, clearly revealing the different surfaces of the rectilinear cross-section that are to be scanned. The disk is approximately 90cm in diameter. *Photo taken by author at the Rolls-Royce site in Derby.* In (b) the corresponding schematic cross-section is shown. Each individual surface is identified from the bounding two letters, and each surface is scanned at at least one angle. Each scan then is identified by the surface insonified and the probe angle to the surface normal. The points marked 1 & 4 correspond to the known indications of seeded defect disk 5, a Rolls-Royce test piece used in the later detection results, Chapter 7. For these two points the diagram indicates from which directions the indications might be seen, and in which scans these indications are detected by the current data analysis procedure. *Diagram derived from one provided by Rolls-Royce.*

(a)



(b)

Figure 2.6: The ScanMaster [25] immersion tank disk inspection system used by Rolls-Royce. In (a) the tank, including the turntable loaded with a disk and the manipulator arm used to position the ultrasonic probe, is shown. Photograph (b) offers a close-up of the inspection head. The focussed probe is seen at the water surface. The manipulator arm offers 6 degrees of freedom for positioning the probe. Circular test blocks used in the equipment set-up are seen down the left of the picture, while the turntable is seen bottom-right. *Photos taken by author at the Rolls-Royce site in Derby.*

Unfortunately, this threshold may be tripped quite frequently as the scan is complicated by spurious high amplitude signals. Such signals can be the result of microstructural noise, given that these titanium parts tend to contain neighbourhoods of large, closely aligned grains [26, 27], and surface machining imperfections, for example. Real defects, inclusions of the sort the inspection is designed to guard against, are extremely rare, but a significant number of disks do not pass the inspection and at least need some further work prior to further manufacturing stages.

The Rolls-Royce inspection system does have the benefit of several seeded defect disks, containing realistic defects after having been forged from contaminated billet. These may be used to test detection performance, both of current systems, and of the software system described in this thesis. Specifically, data from the seeded defect disk 5 is used as the basis for Tippetts' results, some of which are reproduced in Section 7. The two known indications examined there are indicated in Fig. 2.5b. This figure also reveals how the described amplitude threshold detection system only identifies the known indications in a small fraction of the available scans that had the opportunity to detect at those locations.

2.4 Similarities and Differences

Comparing and contrasting the two automated inspections of Sections 2.2 & 2.3, it is clear that there are both interesting similarities as well as striking differences. Both are ultrasonic, volumetric inspections that feature multiple interrogations of the sample volume. However, in the case of the rotor bore inspection, four different UT probes are used, providing four distinct data channels, while the disk inspection uses a single probe, providing many scans, identified by the surface insonified and the probe angle to the surface normal. The number of views of a given sample can be very high for the disk inspection, higher than ever achievable in the case of the rotor bore, and the range of view angles is also higher for the disk inspection. The fact the rotor bore inspection is an in-service inspection means that the possibility of a comparison with a past scan exists. On the other hand, RWE npower only examines a few rotor bores a year, whilst Rolls-Royce has hundreds of disks to work through in that time, potentially allowing for a population-based analysis. Both data acquisitions have significant but differing complicating factors.

For example the rotor bore inspection is exposed to coupling problems, especially as gravity in fact causes the couplant to run and pool at the bottom-most section of the bore, whilst the disk inspection is sensitive to how the sound water path is accounted for. From a data processing perspective, in both cases the collected datasets are vast. For example, in just one set of rotor bore main scan data files, originating from just a 50cm axial bore section, there are over 2.9×10^5 A-scans, 3.8×10^8 saved amplitude samples, corresponding to 2.8GB of data (if saved as uncompressed 8-bit samples). However, the data of the rotor bore is conveniently broken down into axial sections (around 50cm long), meaning that the volume of data to be considered at once is not as great as for the disk inspection. Moreover, because all data relating to a given axial stretch of the bore is collected at once, it might be conceivable to fully process some of the data whilst the data acquisition is still ongoing. It is unlikely that such a scheme could ever be implemented for the disk inspection given the sequence in which scans are collected. On the other hand in the rotor bore inspection only the signal envelopes are saved, whilst the disk inspection system saves full RF (Radio Frequency) signals, which is very advantageous should one wish to use frequency-domain processing techniques.

2.5 Shortcomings

The two automated inspections of Sections 2.2 & 2.3 also both have significant shortcomings, even though automation of the inspection brings numerous advantages, as seen in the chapter introduction. The data volumes collected are truly overwhelming, containing millions of A-scans, exceeding anything a human operator could look through in its entirety. The set amplitude thresholds used for detection purposes in both inspections are arbitrary in the sense that they are based on idealised target reflectors, and it is unlikely that any real defect would resemble one of these. Moreover, such thresholds are unable to take into account variations in the microstructural noise across a component, strong return signals from the vicinity of geometric component features, or the amplification of electronic noise due to higher effective gain at greater A-scan depths. Related to this, the high number of false calls experienced due to signals exceeding the set amplitude threshold slows down the inspection considerably, and is draining for a human operator to deal with. Lowering

the amplitude threshold used for detection to enable the detection of smaller indications, and hence for instance longer service life ratings or maintenance intervals, would make the number of false calls too high to manage. The procedure for assessing signals exceeding the threshold is also less than systematic and distinctly subjective. No meaningful attempt to consider signals from different scans or channels simultaneously is made, even when these relate to the same sample volume. In the case of the rotor bore inspection, available past inspections will only ever be incorporated in the analysis in a qualitative sense, by the inspectors reviewing the previous inspection report. Furthermore, humans are fundamentally poor at assessing random behaviour and correlations [28], compromising both the assessment of signals in high noise environments and across channels, if this were attempted. Overall these shortcomings mean that the inspection reliability is at the very least not as good as it could be given the available data. The software system described in this thesis seeks to address all the identified problems to allow automated inspection to reach its full potential.

2.6 Summary

This chapter has presented details of both the two automated inspections that are used as example applications in this work. Similarities and differences between the two have been assessed, and shortcomings in the current data analysis procedures have been highlighted. Together with the advantages of automated inspection outlined in Section 2.1, these shortcomings provide the the motivation for the project, feeding into the project objectives outlined in Section 1.4. Undoubtedly there is significant scope for the inspection to be improved using a partially automated data analysis system. Given the large number of available scans or channels, the wide range of view angles into component regions, the point-like nature of the defects sought, the relatively high false call rate currently and large number of similar inspections, especially the disk inspection application is considered to hold great potential for computational enhancement. In the next chapter we consider the implementation of the data analysis software.

This page intentionally left blank

Chapter 3

Software Approach

3.1 Introduction

Following on from the project objectives of Chapter 1 and the description in Chapter 2 of the applications focussed on in this thesis, this chapter gives the first details of the methodology used. The focus here is on implementation tools, whereas later chapters (specifically Chapters 4 & 6) focus on the science and application, offering few implementation details.

First, some of the computing and programming challenges are highlighted. The key operations in the execution of the program are detailed. Then the broad, overall structure of the software and the division of responsibilities between the author and his colleague, Tippetts [13], is outlined. This is followed by an explanation of specific tools used to address the identified computing challenges. Finally, an insight is provided into the current program's practical use.

3.2 Challenges

There are some considerable software design and computing challenges associated with the project which have to be overcome. From a software engineering perspective, the program should be as general, flexible and extensible as practically possible to allow the software to operate on data from a range of inspections. The code also has to be

organised in a modular manner to allow cooperative working, given Tippetts' and the author's collaboration, to permit each contributor to work independently and have well-defined responsibilities. From a computing point of view, most of the challenges are associated with the limited computing hardware available - by design, as for the software to be useful in industry it needs to be able to run on a stand-alone, relatively standard if perhaps high-specification, machine. So while for instance high performance, cluster computing facilities are available, a reliance on these would undermine the likelihood of the software ever being adopted - see also the project objectives of Section 1.4. Instead, the computer described in Section 3.2.1 was used. The major computing hurdles then concern both the data and calculation volumes. The first is significant, as when dealing with the data saved by automatic NDE systems such as those of the previous chapter, it is not unusual for even a single scan or channel file to be multiple Gigabytes in size, such that the quantity of available RAM memory is rapidly exceeded just trying to open one or two files. Calculation volumes are a concern as some of the computations explained in the later chapters are of considerable complexity, placing significant strain on CPU resources. This situation is compounded by the demands placed on the memory, limiting the ability to, for instance, store intermediate results to reduce the required processing. Moreover, to be practically usable in industry, the computing time must be compatible with the data acquisition time and not result in many hours of waiting for staff or components.

3.2.1 Specifications of computer

The specifications of the computer that all calculations were performed on are detailed below.

- 2.83GHz Intel® Core™Quad 2 CPU (Central Processing Unit)
- 8GB RAM (Random Access Memory)
- 500GB SSD (Solid State Drive)
- 1TB HDD (Hard Disk Drive)
- 64-bit Windows™7 Professional OS (Operating System)

The computer was a state-of-the-art PC in early 2010 - the SSD was a later upgrade, elaborated on in Section 3.5.2. Note the choice of OS type was based on what might be used in an industrial (rather than academic) setting, and the 64-bit architecture was critical for using larger quantities of RAM.

3.3 Key Software

Python [29,30] was chosen as the primary programming language. This choice was guided by several considerations:

- Python is designed for Object Oriented Programming [31], which means it uses concepts such as class *inheritance* and variable *encapsulation* to structure code, making it very suited to collaborative work on a large program.
- It is a high-level, user-friendly language, allowing rapid prototyping and testing, for efficient development.
- It is open-source, meaning that a stand-alone program for industrial use can be developed, that essentially all code can be examined and fixed by the user, and that an extensive user community can be consulted for help.
- The language has an extensive set of available software libraries, and also lends itself well to linking in code written in other languages, for instance for speed-optimisation.
- Tippetts was already an experienced Python user.

The key software packages, all included in the chosen Python distribution (provided by Enthought [32]), are listed here:

- NumPy™, for array-based computations [33]
- SciPy™, for additional scientific functions [34]
- PyTables, for handling large data files [35]

- Matplotlib, for producing 2D plots [36]
- Mayavi, for generating 3D renderings [37]

3.4 Software Structure

Tippetts [13] devised and implemented a software structure which would both allow the program to be flexible and extensible, and facilitate close cooperation, with minimal interference and duplication. The structure chosen consists of a network of so-called *Operators*, quite self-contained units of code with specific tasks and consistent interfaces. The *Operators* interact though a directed network: each exposes its available outputs to the downstream *Operators* for them to request as an input to their own calculations. The network can readily be adapted and extended, providing the necessary code flexibility and growth-potential: for example, if a further data channel becomes available, this typically just calls for a few extra *Operators* to be added and linked in appropriately. Moreover, the nature of *Operators* means that they can readily be developed by different people, interact efficiently and, in places be substituted for each other to handle processing specific to a given application. A further description of the network and example network illustrations representative of the overall program developed are to be found in Appendix A.

The close collaboration between Tippetts and the author was further enabled by the use of a shared code repository, as well as code testing and integration tools. However, a clear division of responsibilities, detailed next, remained central throughout.

3.4.1 Basic operations and division of responsibilities

At the highest level the processing can broadly be split into registration, aligning different data channels to a common coordinate system, followed by data fusion and detection, combining different signals to assess the probability of a location in the component being worthy of further investigation. Registration for the rotor bore and disk inspections is described in Chapter 4 and Tippetts' thesis [13], respectively. The data fusion system is developed in Chapter 6. The agreed allocation of responsibilities between Tippetts and the author is described next, interwoven with a basic, linearised overview of the processing

stages involved in the program. A schematic summary of the division of responsibilities is provided in Fig. 3.1, based on a simplified representation of the program's data flows.

Raw data access

First the original data, as saved by the current inspection equipment, must be accessed, and converted into a universal format. Given the proprietary nature of the software used to record the data, and the commercial concerns of the involved inspection software companies, this was found to be a painstaking task, and necessarily very application specific. Therefore Tippetts and the author each developed their own *Operators* for this task.

Data pre-processing

In both considered applications some initial processing is required. In the case of the rotor bore scan this includes gating to eliminate the probe ring-down at the start of A-scans - see Fig. 3.2. For the disk inspection this stage is more involved, featuring gating, re-sampling and envelope detection of the data. Consequently, *Operators* for these tasks had to be developed separately by Tippetts and the author.

Data acquisition model

A parametric data acquisition model, describing the coordinate system of the captured data, is established. This is again rather application specific, so was developed separately for both applications of Chapter 2. A comprehensive description of this model for the rotor bore inspection is provided in Section 4.2, while the equivalent for the disk inspection features in [13].

Feature extraction

As will be described extensively in Section 4.3, feature extraction is a critical part of registration. Signals from significant geometric reflectors are identified to line up different data channels with. Unfortunately, the features that can be used for registration in the

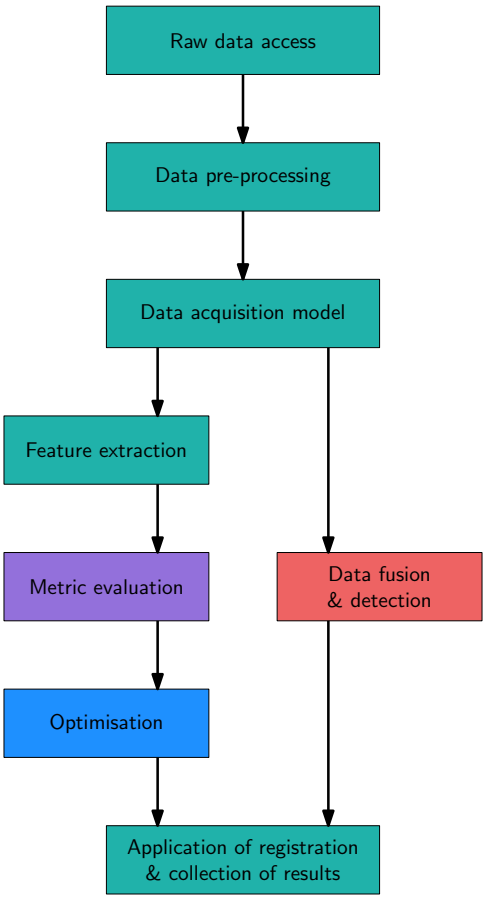


Figure 3.1: Schematic diagram of the approximate data flows, indicated by arrows, though the main processing stages of the software, shown as boxes. The box colours indicate the breakdown of responsibilities between the author and his colleague, Tippetts. Turquoise shading identifies software modules that were predominantly developed independently by both, specific to the applications considered. Tippetts had sole responsibility for the registration optimisation, shown in blue. The present author in turn had sole responsibility for the fusion and detection module, shown in red. Both parties contributed to the purple metrics module. The data fusion and detection is shown parallel to the registration processing as computationally these two sides of the software are independent. *Diagram adapted from equivalent used by Tippetts [13].*

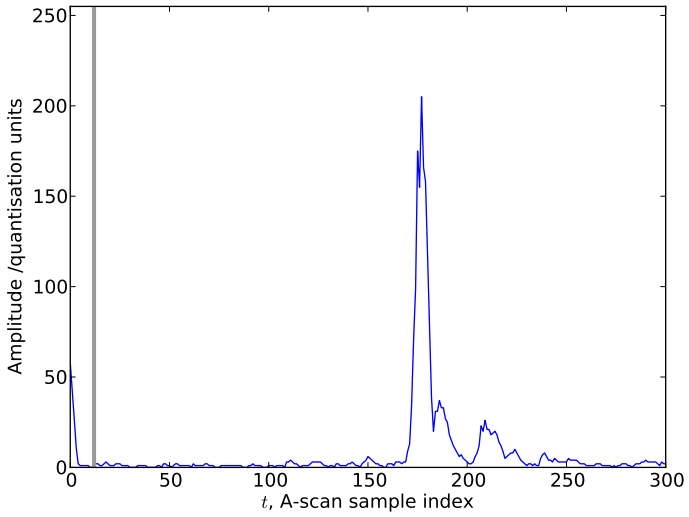


Figure 3.2: An A-scan from the clockwise shear channel of a calibration scan. This particular A-scan features a strong reflection from one of the side-drilled calibration piece targets at around $t = 175$, and only the first 300 points are shown. The vertical grey line marks the gate applied to eliminate the tail of cross-talk between the twin crystals on the probe visible here near $t = 0$. The extent of this unwanted part of the signal varies depending on how the trigger on the data recording is set up. Note the absence of a front-wall reflection, due to the arrangement of the angled probe.

two applications differ significantly, so this is a further type of *Operator* for which two distinct versions were produced.

Registration metric evaluation

Registration error metrics must be computed to quantify the quality of the alignment between features of different channels, so that the fit may be improved by optimising the metric value. As the signals extracted by the feature extraction in the two applications are of the same data format, this *Operator* was jointly developed, with inputs from both Tippetts and the author.

Registration optimisation

The optimisation that is used to align data channels by the minimisation of the registration error metrics was developed exclusively by Tippetts - see [13] for details. This *Operator* is highly general and could be used without modification to help process data from other automated inspections, too.

Data fusion and detection

The data fusion and detection framework for the combination and evaluation of multiple signals once registration has been completed was developed by the present author and is described in Chapter 6. While probably not quite as general as the optimisation, this module is applicable to a very wide range of data inputs, so could readily be used to help process data from other examples of automatic NDE.

General note

It is clear that all the *Operators* for which two versions were developed, one for each of the two applications, are significantly application dependent. That the differences in the two applications proved to be so extensive, limiting the amount of code used by both significantly, is unfortunate, but does mean that the flexibility of the program network structure has been tested. Moreover, if the capability to process data from a further inspection type were desired, there is a reasonable chance that some helpful commonality with one of the current applications could be identified, allowing application-specific sections of the code to be generalised and reused.

3.5 Data Handling

Having in Section 3.4 examined the approach taken to tackle the software engineering challenges of the project, in this section the key tools for addressing the computational hurdles identified in Section 3.2 are described.

3.5.1 On-demand evaluation

Because after the initialisation of the *Operator* network described in the last section the *Operators* only calculate requested outputs, this corresponds to an efficient, demand-based, "lazy" evaluation scheme [38]. Such a scheme can radically reduce the number of unnecessary calculations. This scheme is supported by a caching capability, also implemented by Tippetts [13], that ensures that if the same *Operator* output is called for twice, the second time the previously computed answer can be recovered from memory. As such,

this evaluation scheme helps to address the challenge of the processor load, at the cost of some memory for the cached intermediate outputs.

3.5.2 Solid state drive

As indicated in Section 3.2.1, the computer was upgraded with a large solid state drive (SSD) to help tackle the data access and memory problems. SSDs have only in recent years become available in large capacities and at reasonable cost. Compared with conventional hard disk drives (HDDs), they offer extremely fast and consistent data access, related to the fact that there are no moving parts [39], and even experience only negligible slow-down if accessed by more than one process at a time, ideal for parallelised operations. In modern high-performance computers, SSDs are usually set up to hold OS files, to help boost start up and shut-down times. However, in this project the SSD is used to store intermediate data files, elaborated on in the next section, from the developed program. Thus the SSD provides a means of storing huge quantities of data in a very rapidly accessible manner, providing an ideal compromise between the large capacity but slow data access of an HDD and the limited size but very high data access speeds of the available RAM, and reducing the demands placed on HDD access and RAM capacity.

Additionally, this author placed his computer's Windows page file on the SSD, moving it from the HDD. The page file is broadly used as an overflow from RAM, allowing the virtual memory available to be increased as underused data from RAM is set aside in this special system file. Unfortunately, running out of physical RAM, as is likely to happen occasionally working on a data-intensive project such as this regardless of precautions taken, given the presence of other processes on the computer, is usually likely to cause the computer to seize up. The reason is that "thrashing" occurs, where the computer has to rapidly dump large quantities of data to the page file, and this process is limited by the slow disk access of the HDD [40]. Placing the page file on the SSD means that this problem is essentially eliminated, even when the used memory substantially exceeds the physical RAM installed. This is therefore very helpful in addressing the memory challenges of the project, and provides an efficient alternative to installing very much more RAM.

3.5.3 Efficient intermediate files

To complement the efficient solid state disk storage described in the last section, an efficient format for storing intermediate outputs was required. The flexible and powerful HDF5 file format was chosen [41]. In the implementation by the author, this is accessed through the library PyTables [35], offering some further advanced tools and enhancements. Specifically, the data is compressed and decompressed "on the fly" using a novel compression algorithm designed for rapid compression and decompression, rather than just a high compression ratio, providing the means to accelerate memory-limited processes. Making adjustments to the intermediate file format has been shown to dramatically affect overall program performance, and a limited optimisation of the settings was completed. Thus the careful choice of the intermediate data file format further helps to address the computational challenges of the project.

3.5.4 Custom array structures

Further to the *Operator*-based demand-driven computation and the efficient intermediate file format already described, the author designed and implemented several custom array-like data structures. These "fake" arrays all behave externally as standard arrays, and can be indexed as expected for a normal data array, but internally do not contain merely a set of values, held in RAM. Furthermore, these objects (in an Object Oriented Programming [31] sense) may be used together and nested, to produce more complex structures. They also can be thought of as providing the link between the three coordinate systems to consider:

1. A 2D index-space representation of the data, recorded directly as a sequence of 1D A-scans, giving essentially no positional information.

2. A 3D index-space representation of the data which is obtained by rearranging the 2D data taking into account the periodicity of the scan. For the case of the rotor bore inspection this involves cutting the 2D sequence of A-scans into strips corresponding to the A-scans acquired during a single probe revolution, then stacking these. The indices used are denoted by (t,u,v), relating to A-scan time (quasi-radial), probe

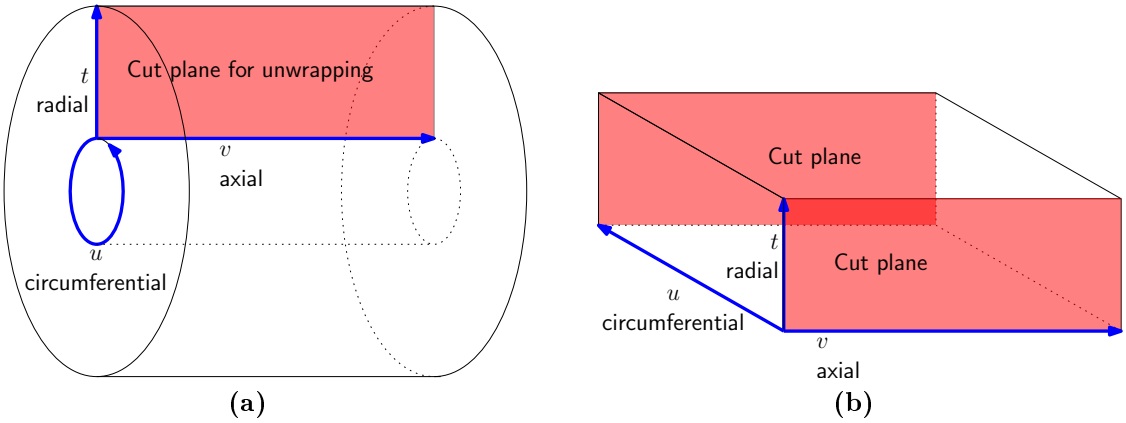


Figure 3.3: An illustration of how the physically wrapped data volume is unwrapped in index space by introducing a cut parallel to the component / data axis. The indices used are (t,u,v), relating to A-scan time (quasi-radial), probe circumferential and axial positions, respectively. In (a) the wrapped array is presented, similar to the physical data collection for an axi-symmetric component (such as the rotor bore and jet engine disk). In (b) the corresponding cuboid, representing a conventional 3D data array, is shown, having split (a) along the cut plane indicated.

circumferential and axial positions, respectively. This coordinate system allows spatial proximity within a dataset to be easily assessed, as nearest neighbours in index space correspond to nearest neighbours in physical space - the exception to this being data points physically adjacent but split across the cut introduced to unwrap the data from the 2D index-space format. This point is illustrated in Fig. 3.3 and elaborated on later in this section.

3. A physical, spatial coordinate system, specified in terms of Cartesian or cylindrical polar coordinates. This coordinate system is built up from a data acquisition model (see Section 4.2), and provides the link to the inspected component, as well as the means to compute physical distances between data points.

Three key types of the custom array structures are explained below.

HDF5 container

This structure is essential for the programmatically elegant use of the intermediate HDF5 files. It acts as a wrapper for the intermediate file on disk, handling all data access issues internally when indexed. Therefore it is possible to pass around a vast array that behaves

exactly as if it were held in RAM, when in fact the underlying data is entirely inside a linked HDF5 file on the SSD. This therefore allows RAM use to be conserved.

On-demand array

This structure offers a lower-level, demand-based, "lazy" evaluation [38] scheme to the *Operator*-based scheme described in Section 3.5.1. In a manner similar to the HDF5 container structure, this object computes the output called for at the time of indexing. Therefore while the structure may appear to be a vast array, it in fact only contains a function with inputs, necessary for computing any requested elements. This is again very memory saving, and critical for much of the processing. Specifically, the challenge of the data volumes is made several times worse by the fact that for much of the processing the coordinate 3D locations of the data must be considered. In a naïve implementation, these coordinate values would require at least three times the memory of the raw amplitude data - and as even just the raw data can readily exceed the available memory, the coordinate data would be completely overwhelming. Instead the forward mapping of the data acquisition model of the next chapter (Section 4.2) is implemented efficiently through this array structure, minimising both the processor and memory loads.

Wrapped array

As mentioned and illustrated in Fig. 3.3, placing the data for either of the two inspections considered into 3D index representation involves introducing a cut parallel to the component axis through the data. This is not in general accounted for in the 3D data array. A simple but very inefficient solution would be to pad the 3D data array in the circumferential direction across the cut with a copy of the data physically found across the cut. A much more elegant way to incorporate the physical wrapping into the representation is offered by a custom array structure. When indexed, this performs calculations on the provided indices before actually accessing the underlying array (which is itself likely to be "fake"). It thereby becomes possible to index chunks of data that lie across the introduced cut. Moreover, the data structure can handle both helical and circular wrapping, appropriate for the rotor bore and disk inspection applications, respectively.

3.5.5 Chunk-wise processing

The motivation for this processing approach is to allow computations on full, massive arrays (which are likely in fact to be "fake", i.e. of the form described in Section 3.5.4) to be completed without running out of RAM. The most natural way to achieve this is to break the computations down into chunks, here termed *subsets*, processed sequentially [42]. This is precisely what has been implemented for various key calculations of the program, in places directly integrating with the custom array structures of the preceding section. While a small processor penalty is incurred, this is more than outweighed in practical use by the reduction in memory required.

3.5.6 Structured interpolation

The need to be able to evaluate the amplitude field specified by the very large data and coordinate arrays at arbitrary locations is a particular processing challenge, even if the basic memory problems have been dealt with using the above tools. This is related to the fact the acquired samples describing the amplitude field do not lie on a regular spatial grid, due to the complex inspection geometry (see Section 4.2). The evaluation of such fields at arbitrary target locations must usually be handled using sparse interpolation, requiring extensive computationally demanding triangulation. Yet interpolation is of critical importance both for the registration processing (see Chapter 4) and the data subtraction that may feature in the later data fusion processing (see Section 6.2). The solution adopted, pioneered here by Tippetts [13], is to exploit the inverse mapping of the data acquisition model (see Section 4.2) to obtain 3D floating value indices for the target physical coordinate locations, and then apply a function for structured, spline-based interpolation provided by the SciPy [34] library. This approach is very efficient computationally, and circumvents the need to resort to hardware accelerated approaches [43, 44].

3.6 Practical Use

The top level of the code used to set up the network of *Operators* (see Section 3.4) is currently quite separate for the two applications of the preceding chapter, though some

further integration between the work of Tippetts and the author could be achieved there. However, at least for within the rotor bore application, the program at that level is very general and capable of automatically configuring the required *Operator* network for a wide range of circumstances based on only a minimum of inputs, as all control parameters have default values. Certain advanced testing does necessitate some lower level configuration to access hidden options, but once the network is initialised, itself involving some preliminary calculations, any output may be called for from any *Operator*, launching the main calculations. For example, the registration optimisation results of Chapter 5 merely required the optimisation *Operator* to be identified and the optimisation output to be called for.

It should be emphasised that at this stage even this top level is not suitable for anyone but an expert and software developer to use. Therefore, for any kind of industrial use at least this top-most level, and the presentation of the final outputs, would need to be wrapped in a user-friendly graphical user interface (GUI). Other important implementation work that would need to be completed to facilitate industrial trials in practice includes a revision of the raw data file access, as for both applications this involves more manual intervention than would be acceptable in an industrial setting. This may also be aided by partial integration with the software currently used for data acquisition, or cooperation with the company offering that program. Additional speed optimisation would also be important in practice. Furthermore, it might be desirable to also implement a set of rapid, basic checks to be completed on the uploaded data prior to more advanced calculations. This would give an inspector immediate feedback on the quality of the data acquired, ensuring that no re-scan, for instance due to a loss of coupling, is required. In the presence of such initial feedback it is likely that the time the computations could be permitted to take in a practical, industrial setting would be increased, to the extent that an overnight calculation may well become permissible.

3.7 Summary

This chapter has set the scene for more detailed explanation of the calculations in subsequent chapters, and has described the key software engineering and computing chal-

lenges to be addressed in this project. An overview of the main processing stages to be completed has been presented, highlighting the division of responsibilities between Tippetts and the author. The developed tools for overcoming the identified main implementation challenges have been detailed, and a description of both current and potential practical use has been offered. Next we turn our attention to the specifics of the data registration.

This page intentionally left blank

Chapter 4

Registration

4.1 Introduction

As described in Chapter 2, it is nowadays common for multiple data channels or scans to be acquired during an automated NDE inspection. For example, the rotor bore inspection that is the focus of this thesis features four ultrasonic data channels. For convenience these are summarised in Table 4.1. It should be emphasised that for our purposes these are all of different modalities - defining channels to be of different modalities not only if they are based on different physical phenomena (e.g. pulse-echo ultrasonics vs. eddy current), but more broadly if any element of the acquisition hardware set-up is different, for instance due to a different probe angle. This chapter examines the alignment of the available data channels to a common coordinate system. This processing step, known as registration, is essential for any subsequent comparison or combination of data from different channels in a joint analysis as it facilitates spatial association across channels. While much of the work presented is relevant to a range of applications, we focus on the rotor bore application. An equivalent analysis for the disk inspection is provided by Tippetts in his thesis [13].

In broad practical terms we need to be able to identify the recorded amplitudes across all channels relating to any given spatial location. For this we need to mathematically describe where physically the amplitude samples originate from. The model used will be characterised by a set of parameters in each channel. These parameters are of uncertain values and are only meaningful relative to each other in the same coordinate system -

Number	Name	Abbreviation
1	60° clockwise shear	60°-CWS
2	60° counter-clockwise shear	60°-CCWS
3	0° short-focus compression	0°-SFC
4	0° far-focus compression	0°-FFC

Table 4.1: Summary of the 4 ultrasonic data channels of a rotor bore inspection, together with assigned channel number and abbreviation.

for example errors in setting up the probe positions mean every channel's physical scan start location is specified relative to a different, arbitrary origin. Therefore the recorded amplitudes themselves must be used to try to recover the correct parameter sets for aligning all channels in a single coordinate system. To achieve this, features identifiable across channels must be selected and matched up. Then, the parameter combination computed to bring these features into alignment is considered to be the correct parameter set to use for spatial association of data throughout the data volume.

Much previous work on such registration has been done, in a range of fields, but mostly focussing on (2D) image registration. A substantial overview of the subject is provided by [45]. The authors identify four major steps in most typical registration procedures that also apply to our application: feature detection (termed extraction here, to avoid confusion given subsequent processing to detect possible indications), feature matching, transform model estimation, and image transformation. Another sizeable survey of techniques, but drawn from medicine, is provided by [46]. While there are parallels to medical imaging in this project, especially in terms of the dimensionality and potential multi-modality of the data [47], NDE test subject variability is low and the types of possible distortions limited compared with the human body [48, 49]. Therefore in Section 4.2 we develop a physical model of the data collection to describe physically conceivable distortions, in contrast not only to much work in medical imaging but also some algorithms in computer vision [50].

After the description of the data model, Section 4.3 considers feature extraction. Features are matched according to Section 4.4, and metrics for quantifying the quality of alignment between matched features are outlined in Section 4.5. These metrics are used in an optimisation detailed in Section 4.7, having first considered the mechanics of aligning more than a pair of channels simultaneously in Section 4.6.

Parameter	Explanation
R	bore radius
z_{rev}	helix pitch, axial increment per revolution
z_0	scan start axially
u_{rev}	number of A-scans acquired per revolution
ϕ_0	scan start circumferentially
ρ	probe angle, in a plane normal to the bore axis
d_{vs}	speed of sound in component
d_0	scan start radially, along the A-scan
g	gain factor, a linear amplitude scaling

Table 4.2: The parameters of the data acquisition model for every individual ultrasonic channel. The first 5 completely describe the path of the probe along the bore, the next 3 the samples along the main beam of probe for a given location on the bore surface, and the final parameter is the only one to change the amplitudes directly.

4.2 Data Acquisition Model

A parametric data acquisition model is required to allow benign differences between channels to be compensated for by the careful choice of parameter values. Such a model is also central to determining the spatial coincidence of the signals from different channels in subsequent data fusion - see Chapter 6. For each channel of the rotor bore inspection, the data acquisition model developed consists of a description of the probe position as the probe moves along the helical scan, and a description of the probe main beam along which the saved amplitudes were collected, or are projected to build up a spatially accurate three-dimensional data field. A simple amplitude scaling factor also accounts for the effect of different gain settings. The parameters of this model are detailed in Table 4.2.

4.2.1 Forward mapping

The forward mapping concerns the translation of index space locations into physical, spatial coordinates, moving from the domain of (t,u,v) to (x,y,z) or (θ,r,z), for Cartesian and cylindrical polar coordinate systems respectively (see Section 3.5.4). The relevant diagram is presented in Fig. 4.1. First, it is convenient to define w as the index along the helix, a composite of u and v:

$$w = v\,u_{rev} + u \tag{4.1}$$

Then, using the defined parameters, the Cartesian coordinates of a probe along its path on the bore surface are given by:

$$\begin{aligned} x_s(w) &= R\cos(\phi_0 + \frac{2\pi w}{u_{rev}}) \\ y_s(w) &= -R\sin(\phi_0 + \frac{2\pi w}{u_{rev}}) \\ z_s(w) &= z_0 + w\frac{z_{rev}}{u_{rev}} \end{aligned} \tag{4.2}$$

The minus is necessary so that the direction of rotation corresponds to the physical rotor bore scan, while remaining consistent with the datums used in the procedure [20] of the inspection considered: the axial datum is provided by the alternator end face of the rotor (see Fig. 4.2), the rotational datum by blade 1 (identified by a stamp), and the scanner moves in the manner of a left-hand screw ("clockwise") down the bore. In the described coordinate system the axial datum is at $z = 0$ and the rotational datum on the x-axis, and the scan proceeds up the positive z-axis.

The distance along an A-scan d may also be computed, defining f_s to be the data sampling frequency:

$$d(t) = t\frac{d_{vs}}{2\,f_s} + d_0 \tag{4.3}$$

The factor of a half is required to ensure d is the distance to a reflector rather than the "full-skip" distance the sound travels. The components (s_x,s_y,s_z) of the unit vector, at the probe angle ρ to the bore surface normal, in a plane normal to the bore axis (see Fig. 4.1), are given by:

$$\begin{aligned} s_x(w) &= \cos(\phi_0 + \frac{2\pi w}{u_{rev}} + \rho) \\ s_y(w) &= -\sin(\phi_0 + \frac{2\pi w}{u_{rev}} + \rho) \\ s_z(w) &= 0 \end{aligned} \tag{4.4}$$

Again, the minus sign is required for correspondence with the inspection procedure, treating a "clockwise" probe angle as a positive ρ. Then, the Cartesian coordinates for a general

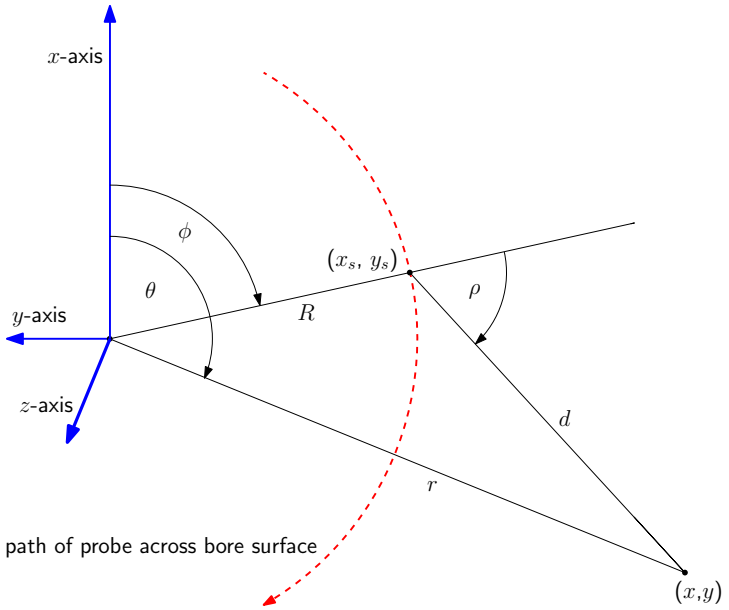


Figure 4.1: The angles, lengths and coordinate system relevant to the scan geometry, in a plane normal to the z-axis and hence the bore. The probe moves axially in the direction of the positive z-axis, out of the page towards the reader. The angle ρ shown is appropriate for the clockwise shear channel, and $\phi_0 = 0$ is used.

point in the data volume, denoted by its index location (t,u,v), are:

$$x(w,t) \quad = \quad x_s(w) + d(t)s_x(w) \tag{4.5}$$

and similarly for y and z. The equivalent coordinates for cylindrical polar axes follow naturally. The model is presented graphically in Fig. 4.1.

This geometry model used, with effectively nine parameters per channel is considered to account for almost all physically conceivable global distortions between channels. The most significant parameter that is excluded is one describing probe skew σ - the angle of the plane the probe beam lies in relative to a plane normal to the bore axis. This is only relevant to probes that are not axi-symmetric about the probe normal, and a further justification for this omission is provided by Fig. 4.3. Based on the small angle approximation (the first terms of a Taylor expansion), the dominant effect of a non-zero skew is in the axial direction, and (over a limited range of r) this will approximate to a change in axial offset z_0 across the dataset.

Note the model makes no effort to compensate for local distortions, on a scale smaller than

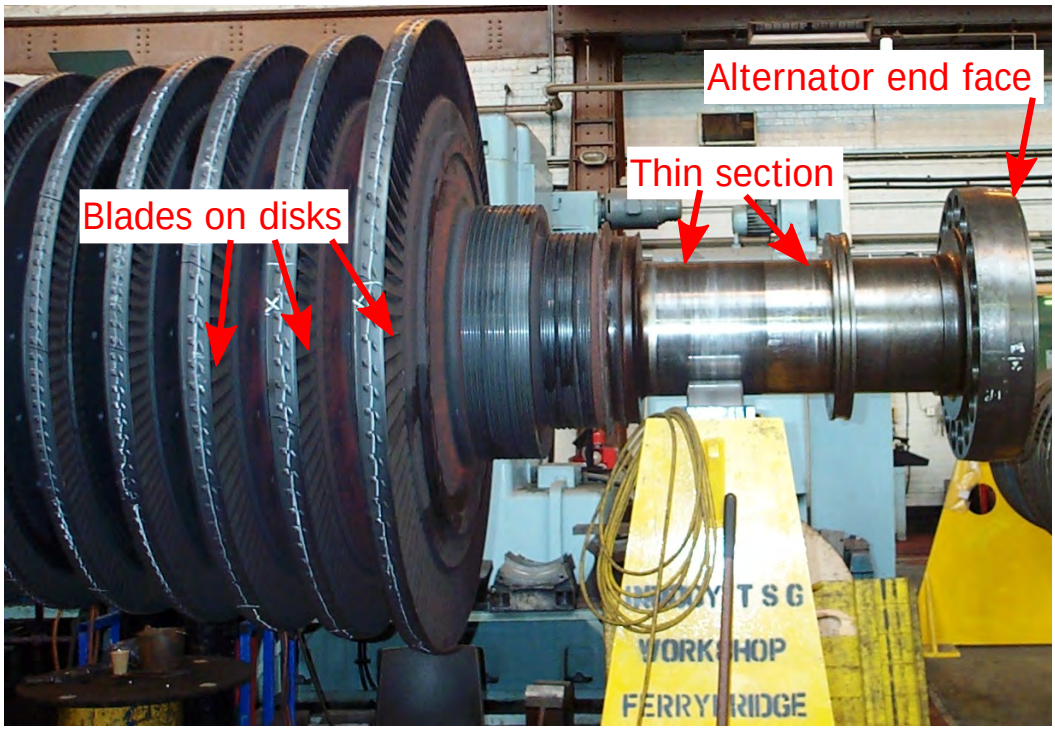


Figure 4.2: The alternator end of an intermediate pressure turbine rotor, photographed side-on. An end-on view of the inspection was provided in Fig. 2.1a. The bore that is scanned runs the length of the rotor (typically about 6m long), with the face plate of this end (on the right) providing the axial scan datum. Note the step changes in rotor thickness in this axial range and how thin the rotor is in sections compared with the length along which blades are fitted. The raw data underlying later results calculations is drawn from a section of the bore in the central, bladed, region. *Adapted from a photo provided by RWE npower.*

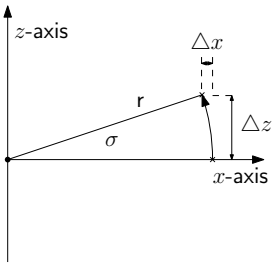


Figure 4.3: The effect of a non-zero probe skew for an angled probe, such that the main beam is not aligned with the circumferential axis, taken here to be parallel to the x-axis. Using small angle approximation the distortion introduced for a given point, at a distance r from point the beam enters the sample, in axial and circumferential directions respectively, is $\triangle z = r\sigma$ and $\triangle x = r\sigma^2/2$. Hence the dominant effect is in the axial direction, and (over a limited range of r) this will approximate to a change in axial offset z_0 across the dataset.

the length of a bore in one (approximately 50cm long) segment of the inspection for which the data are analysed together. This does mean that effects such as probe slip and coupling variations are not accounted for in this model and resultant distortions will largely remain uncompensated. Similarly, the effect of temperature is only included in the model through the variation of the single speed of sound parameter d_{vs}. Therefore, complex temperature effects giving rise to local data distortions will not be compensated for. It remains to be seen how problematic such limitations of the model are in practice. While the effect of temperature is of the greatest importance in structural health monitoring (S.H.M.) where registration is termed baseline optimisation (for subsequent baseline subtraction) there are practical reasons for assuming the effect of temperature is less significant in this project [51–53]. For example, temperatures can be expected to be more uniform and consistent in the applications considered than in the sort of large structures, exposed to both variable weather and operating conditions, to which S.H.M. is applied. Additionally, the interaction of guided waves (as used in most S.H.M.) with complex structures such as pipe supports in pipeline monitoring (a typical S.H.M. application) is more complex than the bulk wave reflections considered in this project, and the wave propagation distances, over which the effect of temperature changes might accumulate, are shorter in automated NDE [54].

4.2.2 Inverse mapping

The inverse mapping concerns moving from the domain of (x,y,z) physical Cartesian coordinates (or (θ,r,z) cylindrical polars) as provided by the forward mapping back to the (t,u,v) index space. This is of critical importance for the structured interpolation (see Section 3.5.6) that facilitates the rapid re-sampling of the amplitude field, necessary for data subtraction, as in the evaluation of some error metrics (described in Section 4.5), as well as 3D rendering (demonstrated later, in Figs. 4.12a & 4.12b).

This inverse mapping is non-trivial. In particular it is not appropriate to try to solve the Equations 4.5 simultaneously for a test point (x,y,z) by first solving for w from z and then finding t. This can effectively only work for test points that coincide with a data point and hence integer (t,u,v) index values - the approach disregards the fact that the data is not

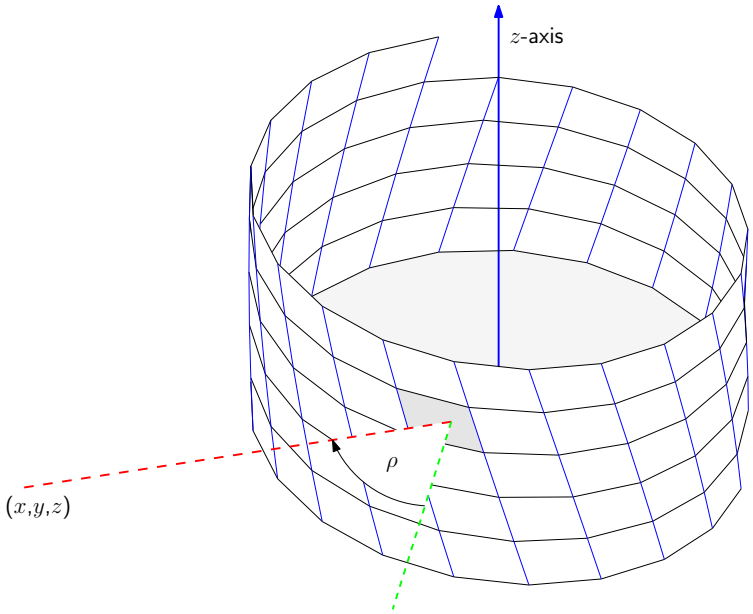


Figure 4.4: Schematic diagram of the back-projection (in red) of a test point to the cylindrical bore surface, at angle ρ to the surface normal (in green) in a plane normal to the bore axis required for the inverse mapping, from physical coordinates (x,y,z) to index space (t,u,v). The parallelograms on the cylinder show the adjacency of A-scan acquisition locations as considered in the 3D index representation of the data.

sampled on a regular grid spatially and there is only a single A-scan for every axial step along the bore. Instead, it is necessary to back-project the test point to the cylindrical bore surface, along a line that makes an angle equal to the probe angle with the surface normal at the point where it touches. The principle is illustrated graphically in Fig. 4.4. Coordinate transforms on the surface of the cylindrical bore surface ultimately provide a perfect inverse of Equations 4.5, even for the especially complex case of a non-integer u_{rev} value [55].

4.3 Feature Extraction

Feature extraction is an important processing step in the registration, both to substantially reduce the computational cost of the subsequent optimisation, and to reduce the sensitivity to differences in channels not accounted for by the data acquisition model, such as data noise. The features identified should correspond to reflections from the component geometry as these can be extracted with a high degree of consistency between channels and the data model should be able to account for distortions in these to a high degree

of accuracy, at least for channels of the same type. Additionally, there are considerable dangers in registering on the basis of features that might have been caused by defects, as the registration will attempt to minimise the defect residual. Therefore for any kind of subtraction-based processing (such as evaluating the r.m.s. residual described in Section 4.5) the registration is only safe in the presence of major signals known for certain not to be from defects [54].

The fact that target reflectors will give quite different signals in channels of differing modalities complicates the alignment of multi-modal channels [44, 56]. Analytic or finite-element (F.E.) models could potentially be used to relate feature signals in different channels. However, such an approach is not only extremely complicated and computationally intensive, but also dependent on the geometry of the reflector being well-known and relatively simple.

Unfortunately, in the rotor bore application there is a severe lack of significant geometric features in almost all the main scan files (the exception being the far-focus channel in the thinner end regions of the rotor - see Fig. 4.2), as the radius for which data is collected falls short of the component edges. As described, this critically undermines not only mono-modal registration, but also the already very challenging multi-modal alignment. Consequently, one major recommendation of this chapter is that the inspection procedure be modified to provide meaningful geometric features in the data, in all channels and all axial bore sections of the inspection. Probably the easiest solution to this would be to increase the length of the collected A-scans, such that elements of the rotor's external features (see Fig. 4.2) are captured, though the directivity of potential reflectors relative to the probes would be critical, and the pulse repetition frequency (p.r.f.) might need to be reduced.

The approach adopted under these suboptimal circumstances is to extract noise-like features in the main scan files, in the hope that these are caused by the microstructure and are therefore structural, coherent. This has the potential to be adequate for mono-modal registration, appropriate for the alignment of two channels of the same type but from two different acquisitions, even though the robustness will not be able to match that achievable with significant geometric features in the data. Noise-like features are entirely unsuitable for multi-modal registration, as required to ensure the alignment of the different channels

acquired during a single acquisition, as microstructural features are highly likely to give unrecognisably different signals in channels of different modalities. For multi-modal registration we therefore instead exploit the fact that the probes are held in a fixed relationship to each other on the probe pans (see Figs. 2.2a & 2.2b) during an acquisition, including both main and calibration scans, to focus on the calibration scan. This is advantageous as the calibration piece offers several clear and simple geometric features by design - see diagram in Fig. 4.5. The implicit assumption in doing so is that the probes are indeed fixed exactly and no changes (intentional or not) are made to the set-up from calibration scan to main rotor scan. This is a very reasonable assumption given that the two scans are normally done immediately adjacent to each other. As will be explained further in Section 4.6, in effect the calibration scan data are exploited to confirm the alignment of the probes on the probe pans for a given acquisition, while the main scan data are used to ensure alignment between acquisitions. If a modification to the inspection procedure, as suggested above, provided meaningful geometric features in the main scan data a robust mono-modal registration would be facilitated as noise-like features would not be needed. For the multi-modal alignments, the convenience of not having to use the calibration scan data would need to be weighed up against the complexity of the features in the main scan data, relative to the side-drilled hole reflections of the calibration scan data - use of the calibration scan as described might still be the best option for dealing with the complications of multi-modality.

In Sections 4.3.1 and 4.3.2 the developed feature extraction algorithms, for calibration and main scans, respectively, are described. In both cases each feature takes the form of a small subset of the full amplitude data array. This feature type was chosen as it seemed more appropriate for 3D data and easier to define than the sort of point features widely used in 2D image processing [50]. Additionally, it is believed that such features retain the appropriate high-resolution data to enable a precise calculation of alignment.

Both feature extraction algorithms also operate primarily on the amplitude data in its 3D index form (see Section 3.5.4), disregarding the underlying exact physical locations of data points. This is both computationally attractive and sensible, given that the physical locations are subject to change as the model parameters are changed during the registration. Nonetheless, doing so does make an assumption about the relative proximity

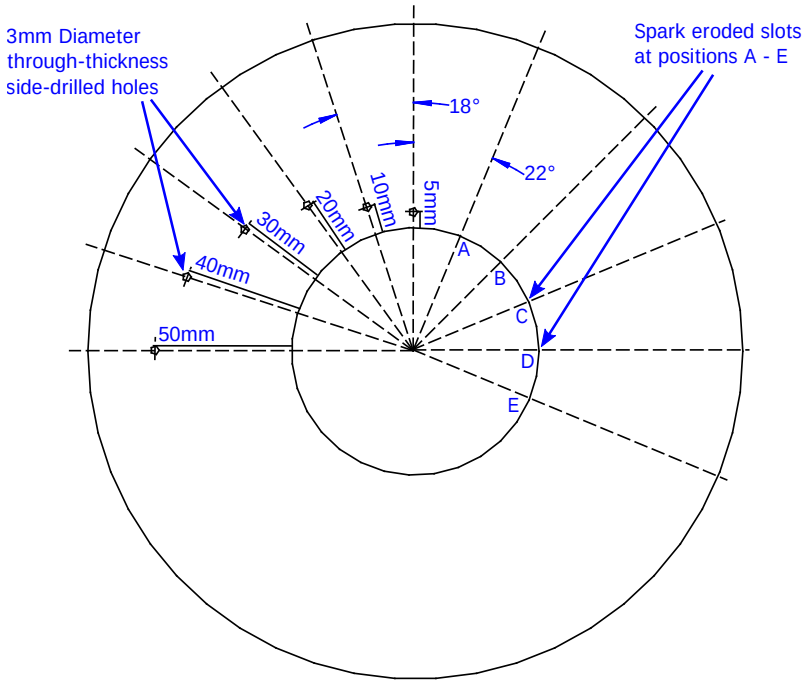


Figure 4.5: A rotor bore calibration test piece, viewed down the bore axis - not to scale. Test piece disk shown has guide tubes attached axially to extend the length of bore. A number of ultrasonic and eddy current targets, taking the form of through-thickness side-drilled holes and spark eroded slots, respectively, are indicated. *Diagram derived from [23].*

of points and requires the index space data form to be a good representation of which data values are near each other. Given the described data model of the last section, this will hold across a range of parameter values - with one parameter exception: the number of A-scans per revolution, u_{rev}, acts as the wrapping period of the data, and even a small change in this can cause dramatic distortions in the 3D arrangement of the data, especially as the effect is cumulative over the length of the scan. As described in Section 4.8.1, the author found a means to pre-compute analytically this parameter, allowing it to be eliminated from the registration optimisation, thereby overcoming the problem associated with 3D index space.

The index space used is illustrated in Fig. 4.6, indicating the important axes and projections that will be referred to in the algorithm descriptions. The diagram also specifies the approximate size of the data array - the data volume provides a challenge in itself (see Sections 2.4 & 3.5).

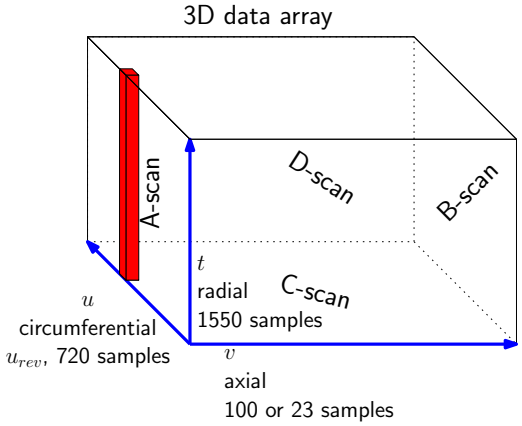


Figure 4.6: The 3D index space of a data array. The index space axes ($\bar{t}$,$\bar{u}$,$\bar{v}$) are shown, with an indication of which physical directions they approximately correspond to and the typical edge lengths of the array. The two lengths shown for the v-axis pertain to a full scan file and a calibration scan, respectively. An A-scan makes up a column of the array, as shown. Additionally, the three standard data projections [1] are labelled, so for instance a C-scan collapses the array along the t-axis. Note that physically points located at the end of the u-axis are close to each due to helical wrapping in that direction, and that there is effectively a cut plane in the D-scan plane at $u = 0$ and / or $u = u_{rev}$.

4.3.1 Calibration scan

The calibration piece, as seen in Fig. 4.5, provides a sequence of side-drilled holes as targets for the ultrasonic channels. Of particular importance to the multi-modal registration for which we seek to use the signals from these targets is the fact that the signals should be readily identifiable in all the four ultrasonic channels available (Table 4.1). Moreover, given the simple, symmetric nature of the reflector type, the differences in the signals of different channels should be limited despite the use of a range of inspection modalities.

The calibration scan feature extraction is highly specific to the rotor bore application, developed by combining a number of image processing operations in a manner determined by extensive experimentation. The algorithm involves the following stages:

1. The maximum t (quasi-radial depth) index of the side-drilled hole signals is computed conservatively, including safety margins, from the approximate data model parameters. This is used to eliminate data at greater depths from the analysis that follows. This step for example removes some deeper multi-reflections from the data volume under consideration - see Fig. 4.7a.

2. A complex C-scan projection is built up. First the (remaining) array is filtered in

$\bar{u}$ (using the bar to denote a vector direction, here along the u-axis, as opposed to index), applying a local threshold and normalisation to each circumferential vector followed by edge detection [50]. This step has the effect of eliminating signals of significant circumferential extent, applying a distance-amplitude correction and emphasising sharp features (such as those from the target reflectors). Then the maximum projection along $\bar{t}$ with prior smoothing in that direction gives an image in the ($\bar{u}$,$\bar{v}$) plane that highlights signals that are not only relatively large and sharp in $\bar{u}$ but also correlated in t. An example of this stage is shown in Fig. 4.7b.

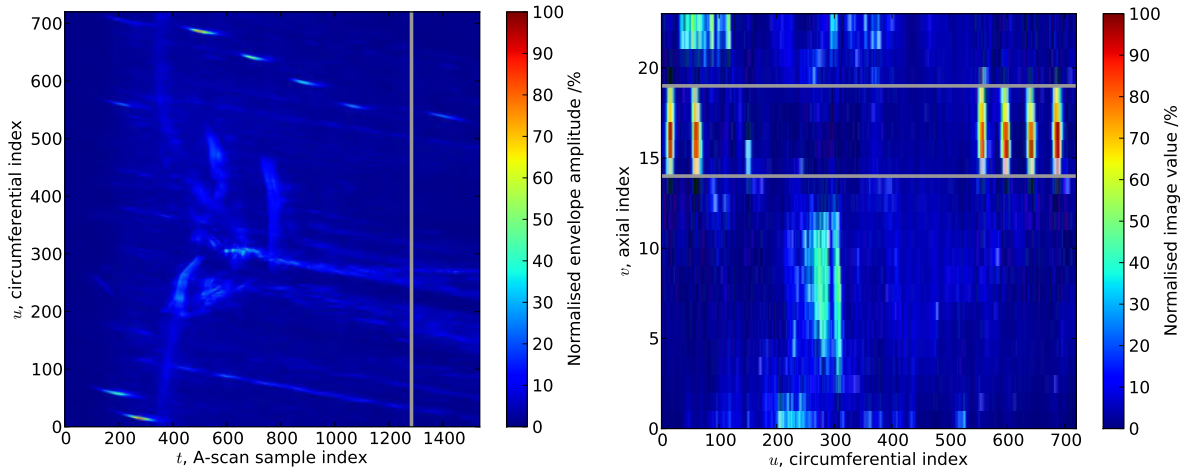
3. The axial range of the data volume containing the target signals is determined by segmentation of a circumferential ($\bar{u}$) mean signal projection of that C-scan - the result is marked on Fig. 4.7b. Having identified an index range of interest in both $\bar{t}$ and $\bar{v}$, the remaining data volume is significantly reduced, making calculations easier. The remaining task is to identify bounds in (t,u) indices appropriate to each side-drilled hole.

4. A complex B-scan projection of the remaining raw data volume is created. First it is filtered in $\bar{u}$, by local thresholding and normalisation to each circumferential vector to again provide distance amplitude-correction. Then filtering in $\bar{t}$ by local thresholding embodies the knowledge that only the largest reflectors in depth are of interest. A median amplitude projection in $\bar{v}$, prefixed by smoothing in the perpendicular plane gives the desired projection, exploiting the fact that the side-drilled holes are axially aligned.

5. The B-scan is thresholded using *Otsu*'s method [57] - see Fig. 4.8b, to be compared with Fig. 4.8a.

6. The binary mask obtained is subject to morphological erosion and dilation operations, as well as filtering to enforce a minimum feature size and the presence of a single reflector along $\bar{t}$, to eliminate spurious small signals and all but the first reflection from the targets.

7. The signals remaining are compared against the number of side-drilled hole reflections expected, and some eliminated if required. The first outlier rejection scheme



(a) (b)

Figure 4.7: Illustrations of the first stages of the calibration feature extraction, operating here on an input from the clockwise shear channel. In (a) a maximum amplitude B-scan projection of the full data file is shown, together with a vertical grey line marking the computed upper bound for the region of interest in the t index. Plot (b) shows a complex C-scan image, where horizontal grey lines mark out the calculated axial range of interest. Note the six prominent signals in that range, corresponding to the side-drilled hole reflections.

considers the amplitudes and sizes of the candidate signals, as these attributes should each be similar (after distance-ampltude correction) across the correct reflections. The second assumes that the radial and circumferential positions of the target reflectors is monotonically related, as is typically the case.

8. Finally, the bounds in (t,u,v) of the remaining regions are found.

The effectiveness of the scheme developed is demonstrated by Fig. 4.9, for clockwise shear and far-focus compression channels, in (a) and (b) respectively. In both cases the algorithm has correctly isolated the primary reflections from all the target reflectors of the calibration piece.

A final adjustment to the features is undertaken prior to comparison with others, to help address the multi-modality of the channels being aligned: the data associated with each feature is spatially displaced along the length of the A-scans by the radius of the side-drilled holes (in practice, 1.5mm). The effect of this is that the reflections from different probes from one target should all coincide in the calculations along the line where physically the centre line of the side-drilled hole target lies. An illustration of

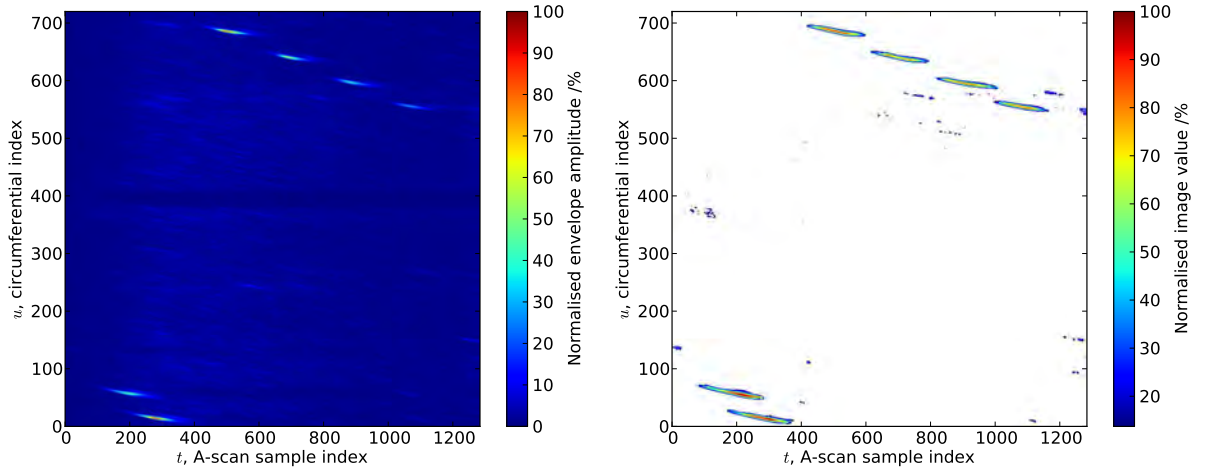


Figure 4.8: An illustration of the intermediate stage of the calibration feature extraction, operating here on an input from the clockwise shear channel. Plot (a) provides a conventional, maximum amplitude B-scan projection of the data in the computed region of interest. In (b) a complex B-scan equivalent is shown that has also been thresholded, setting the eliminated pixels to white. Note the differences in the remaining signals in (b) compared with their equivalents seen in (a).

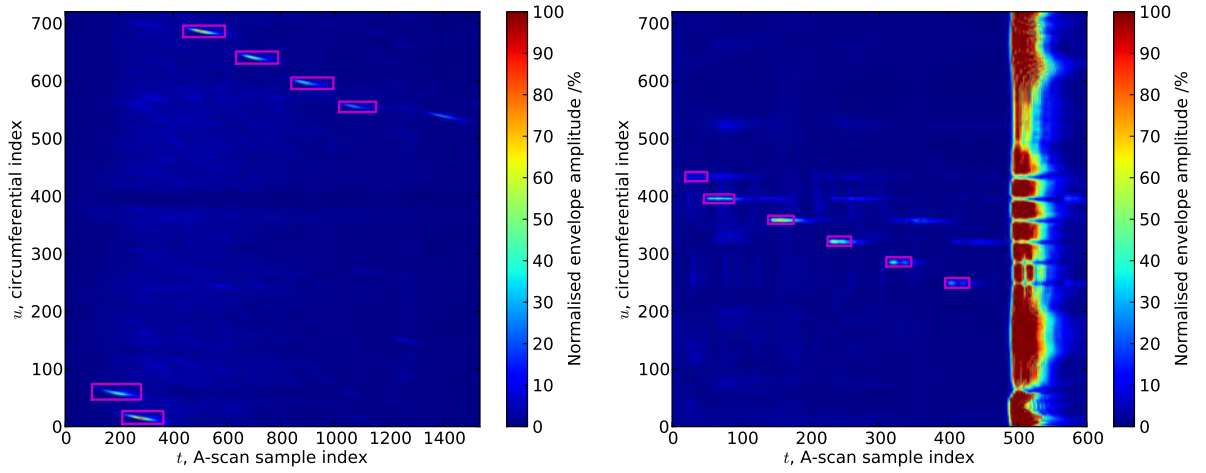


Figure 4.9: B-scans showing the overall action of the calibration scan feature extractor. Both plots are maximum amplitude projections, with the axial extent of the data having already been limited to the computed domain of the side-drilled holes, and magenta boxes are used to demarcate the identified features. Plot (a) is for the clockwise shear channel (providing the final output for the intermediate stages shown in the previous figures), while (b) relates to the far-focus compression channel. In both plots the feature extractor has correctly isolated the primary reflections from the 6 target reflectors, ignoring similar noise signals, repeat reflections and the very high amplitude back-wall echo seen around $t = 500$ in the latter plot.

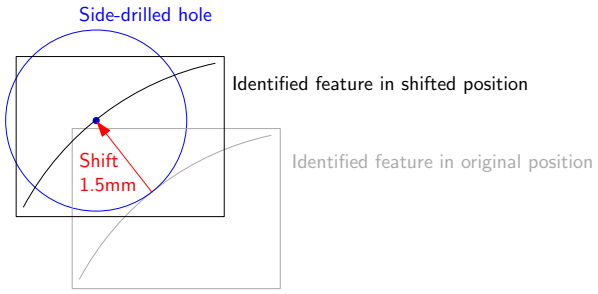


Figure 4.10: Schematic of a side-drilled hole and the associated registration feature identified in the data, in an axial, B-scan, projection. The feature here is a rectangle around the arc-like reflection of the side-drilled hole seen in the data as depicted. For the purposes of registration, this feature is shifted by the radius of the side-drilled hole, so that reflections from different directions coincide at the centre of the hole.

this is provided by Fig. 4.10. While this simple approach to accounting for the multi-modality of the data is inevitably an approximation that could potentially undermine the data model, such problems seem unlikely given the small displacement involved.

4.3.2 Main scan

As described, the main scan feature extraction is forced, in the absence of meaningful geometric reflectors, to consider noise-like regions in the hope that these will be stable across repeat inspections. Compared with the calibration scan feature extraction of Section 4.3.1, not only is the feature stability less assured across the channels being aligned (even if of the same modality, as here), but the desired output of the feature extraction considerably less well defined. Broadly, features should merely be consistent across scans for good subsequent feature comparisons, constitute a small fraction of the total data, for computational reasons, and be well-spaced to provide information about the alignment across the component. Key additional challenges of this feature extraction are the large data volume to process and the possibility that features may be split in the index representation across the wrapping cut (see Fig. 4.6, and Fig. 3.3 from the previous chapter). The first of these is addressed using a C-scan projection to identify regions of interest initially, the latter by extending that C-scan projection across the wrapping cut with a shifted copy of the original - though this does then mean that features can appear twice (once in each copy) and the duplicates need to be eliminated.

The main processing steps of the algorithm, based on several image processing operations,

are:

1. A complex C-scan projection is built up: filtering and normalising in $\bar{u}$ (the vector direction along the u-axis), before smoothing and taking the maximum along $\bar{t}$.

2. A number of iterations are used to adaptively set a threshold across the C-scan, to leave a suitable (by number and fraction of the data) set of regions after morphological erosion and dilation operations [50].

3. The full data column associated with the identified regions in the ($\bar{v}$,$\bar{u}$) plane is extracted and subject to a simplified "blob" detection routine [50], based on the *Laplacian of the Gaussian*. The filter is applied for a range of smoothing parameters, normalising for the selection of that filter width, and then selecting the filter that gives the best response.

4. The bounds in (t,u,v) of the "blob" are determined.

5. Finally, a user-specified number of features is selected sequentially, maximising the spatial separation between features.

An example of the operation of this scheme is presented in Fig. 4.11. It is hard to assess the quality of the features identified, both because there is no correct output (as there is for calibration scans) and because of the significant axial extent of the data, which has been flattened to produce the B-scan projection shown. Nonetheless, the output is not obviously wrong, and the large circumferential reflectors have all been ignored.

4.4 Feature Pairing

Having identified features appropriate for registration in two channels, these features need to be matched up to enable a meaningful comparison. The matching must pair-up features from the *base* channel, considered invariant during the registration, with those from the *float* channel that will be adjusted by the optimisation. The registration is then able to proceed by quantifying of alignment between these pairs of features by the evaluation of error metrics (Section 4.5), and then optimising one or more of these (Section 4.7).

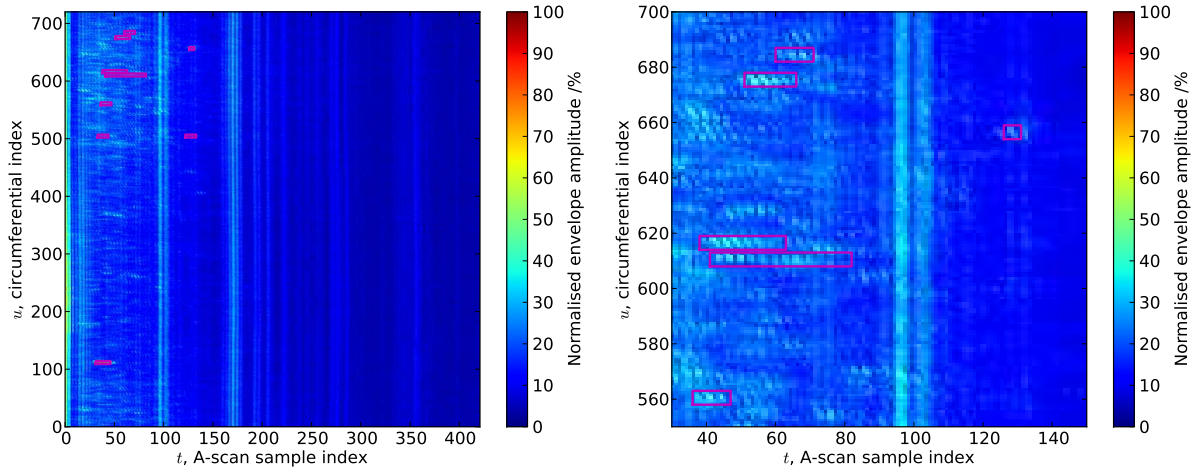


Figure 4.11: A B-scan projection of a full main scan file, for the clockwise shear channel, showing the action of the general feature extractor. Plot (b) offers a magnified, restricted section of the full file seen in (a). In both, magenta boxes demarcate the identified features. In the absence of meaningful geometric reflectors, the feature extractor is forced to rely on noise features. Note that the feature extraction has disregarded even large reflectors spanning the full circumference.

Significantly, an incorrect matching will undermine the ability of the optimisation to ever converge to a good registration, especially if the matching is not recomputed at any stage. Note also that while all the parameters of the model (see Table 4.2) that the registration may vary are continuous quantities, the correspondence of features should be considered Boolean or binary. This means that if the matching is recomputed during the course of the optimisation and allowed to change, discontinuous and multi-modal effects are liable to appear.

For the disk inspection application (see Section 2.3), where surface echoes can be used as the registration features, the orientation of different scans is enough to identify unequivocally the correct feature matchings [13, 58]. However, for more general features, such as those in the rotor bore application (see previous section), a more complex means of determining feature correspondence is required, yielding less certain matchings. Therefore the work here and the resultant effects are not found in Tippetts' thesis.

There is a considerable body of literature on feature matching. However, the literature is overwhelmingly based on 2D image processing, which differs in several respects from the feature matching required here. Firstly, the number of features typically considered in standard image processing is at least an order of magnitude greater than the number

considered here [59,60]. This means that whereas in standard 2D image processing it may be acceptable to return a low percentage of correct matches and possibly a significant percentage of false matches, both these scenarios must be avoided here. Secondly, the features here are not just points but 3D segments of data. Thirdly, the distortions that may occur between corresponding features of different datasets, with which the matching algorithm must cope robustly, are more complex than those in typical 2D image feature matching.

Nonetheless, three principles of general feature matching can be identified [61]: the *principle of proximity*, the *principle of similarity* and the *principle of exclusion*. The first of these expresses that, in the absence of other differences, the spatially closest potential match should be chosen. The second means that all else being equal, the most similar potential match should be preferred - the similarity of features may be assessed in terms of appropriate descriptors. The third principle formalises the fact that only one-to-one matches are permitted.

4.4.1 Correspondence matrix

A matrix expressing the quality of match for all possible pairings serves as a starting point. Working from the seminal paper [62], and that many papers, including [63], have built on, the *principle of proximity* is expressed mathematically as:

$$G_{ij} = \exp(-\frac{a_{ij}^2}{2\alpha^2}) \tag{4.6}$$

where G_{ij} is the pair weight for the features i and j, separated by the distance a_{ij} in space. The adjustable distance scale parameter α governs the decay of the expression and makes it scale invariant, by normalising a_{ij}. The pair weights vary between 1 for identical features to near 0 for very disparate features. This restricted range gives an intuitive physical interpretation to particular values and makes comparisons easier. Given the geometry of the rotor bore application, it was decided to measure the feature separation a in the circumferential direction along circular arcs, rather than in a conventional Euclidean sense, to down-weight features separated by large rotations. Additionally, an upper distance limit was imposed to reduced the number of potential outliers.

The *principle of similarity* can potentially be merged into the weight matrix G_{ij} with an addition term [61], using feature descriptors to assess the similarity of features [64]. Such descriptors must be distinctive but largely invariant to the expected between-feature-set transformations, allowing discrimination between different possible pairings. Effectively, one wishes to "maximise the ratio of between-class scatter to that of within-class scatter" [65], describing a feature and all its distorted correspondences as being of the same class. As an example, one of the most effective and widely used feature descriptors in image processing and computer vision is provided by the Scale Invariant Feature Transform (SIFT) algorithm [59, 66]. Unfortunately, such descriptors are not appropriate to the features used here, and while for instance [67] does propose a descriptor appropriate for 3D feature matching, experiments to develop and incorporate a similarity assessment into the pair weights were unconvincing. Therefore the currently implemented matching algorithm does not actually include the *principle of similarity* - so this makes an obvious suggestion for potential future work.

Having established the matrix G_{ij} describing the quality of match for all possible matches, the next step is to use this to select as many good matches as possible, while avoiding outliers. A traditional approach to combinatorial optimisation is the *Hungarian* algorithm [68]. This will find the set of matches that maximise the total weight of pairs, taking into account trade-offs to avoid the pitfalls of a greedy selection scheme. However, the algorithm does not consider the possibility of outliers, so it is unsuitable in the situation here. Instead we continue the processing based on [62], applying Singular-Value-Decomposition (S.V.D.):

$$G = TDU \tag{4.7}$$

where T and U are unitary matrices (of different sizes if G is non-square) and D is a (general) diagonal matrix of non-negative singular values. The method then calls for these singular values to be replaced by unity, to give another (general) diagonal matrix E. The following matrix multiplication then gives the correspondence matrix P:

$$P = TEU \tag{4.8}$$

The method uses matrix orthogonality to neatly incorporate the *principle of exclusion* and steer the solution towards a set of one-one matches.

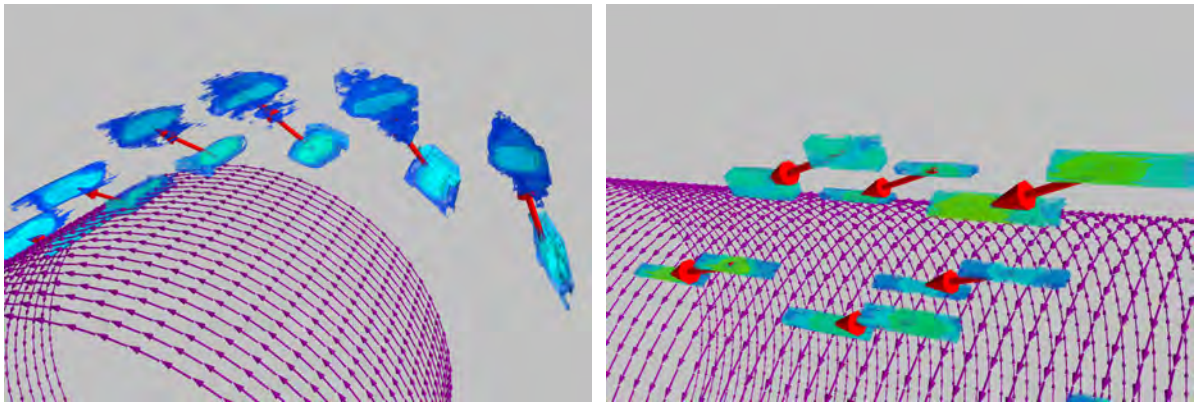
4.4.2 Selection of pairings

For an ideal case with one-to-one matching of all features, P will be a binary matrix, with exactly one entry of unity in each row and column, from which the required pairings are obvious. However, if there are n features competing to match a feature of the other dataset, there will be n non-zero values in the relevant row or column, each of magnitude $1/\sqrt{n}$, such that the row / column remains normalised to unity. In practice therefore some thought must be given to how the correspondence matrix is interpreted.

The simple scheme adopted for the selection of one-to-one matches relies on the fact that, because of the described matrix normalisation, any entry larger than $\sqrt{1/2}$ will be dominant across both the row and column. Therefore clear one-to-one matches can be identified by thresholding P at just above this critical value, and retaining the remaining entries.

More complex schemes are certainly conceivable and may offer significant advantages - this too is a proposed area for further work. For example, it may be beneficial to permit matchings other than one-to-one in the case of ambiguity in the pairing matrix, as this may allow features that have been inadvertently split during feature extraction to be handled correctly. Additionally, a sequential selection scheme could check the consistency of the vector displacement associated with candidate matches against those already selected, making the assumption that the distortion field relating the two channels is smooth.

Despite the simplicity of the implemented selection scheme and the correspondence matrix computation, the achieved performance with the currently used features (from Section 4.3) is good. This is illustrated using screen shots from the interactive visualisation tool in Figs. 4.12a & 4.12b, for calibration scan features from two different channels and main scan features from two copies of the same channel, respectively.



(a) (b)

Figure 4.12: Screen shots from the interactive visualisation of the feature pairing. This tool provides a 3D rendering of a probe path, shown here with purple arrows, that each span 10 A-scan sample locations on the bore surface, together with renderings of identified registration features for two channels visualised as isosurfaces (the levels of which are loosely related to amplitude, but computed for visual effect), and pairings indicated by (red) arrows between features of the two channels. In (a) the pairing of calibration scan features from clockwise shear and far-focus compression channels is shown (see also Fig. 4.9), in (b) the pairing of main scan features for two copies of the clockwise shear channel (see Fig. 4.11) is presented. Significant distortions were introduced between the channels to be aligned for demonstration purposes. In both images the computed feature matching is qualitatively convincing, taking into account feature shapes and positions.

4.5 Registration Error Metrics

Having found pairs of features to compare, the quality of alignment between these must be quantified. There are a number of registration error metrics that may be defined for this purpose. In most optimisations a single such metric is chosen to serve as the objective function to minimise, but as will be explained (see Section 4.7), here more than one metric is used. The two metric calculations used in practice are described in the subsections that follow. Nonetheless, there are many potential alternative metrics that could be defined. In particular, the author's colleague Tippetts examined a metric based on mutual information in detail, before abandoning it [13, 69, 70]. While such information theoretic metrics are considered especially useful for multi-modal data, they are complex to compute.

4.5.1 RMS residual

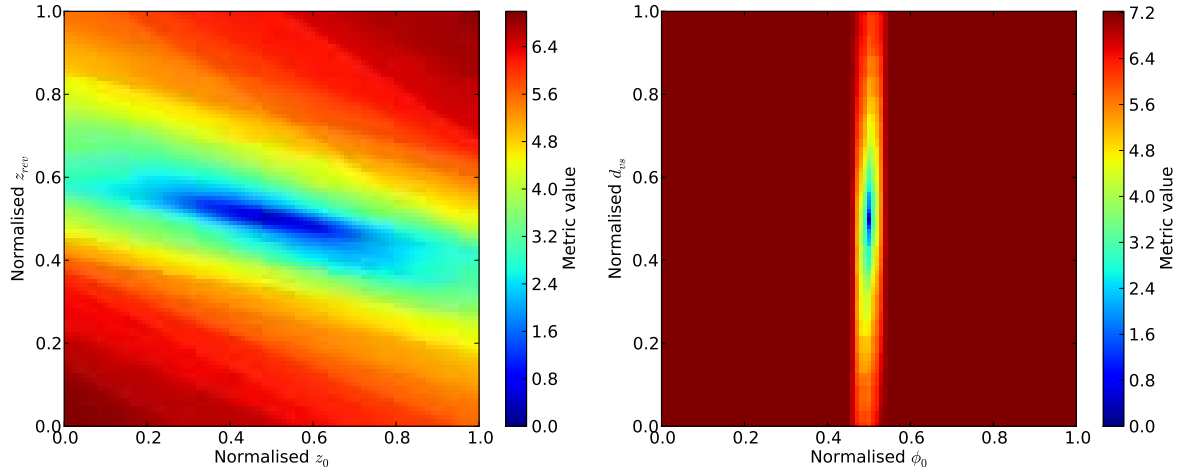
This metric is calculated by finding the sum of root-mean-square (r.m.s.) residuals between the amplitude data of paired features interpolated to a common spatial coordinate grid. This metric may also be described as the L_2 norm of the residual of two vectors, where these vectors contain the interpolated feature amplitudes.

In the implementation some care must be taken to handle appropriately incomplete overlap between paired features, as well as the possible existence of unpaired features. This then ensures that the registration metric is lowered for more complete overlap of paired features, and also reduced if more overlapping paired features are identified.

The variation of this metric across sections through the higher-dimensional parameter space of a simple example registration in two planes is shown in Fig. 4.13. The parameters are normalised based on bounds placed on the range of values each parameter can take, so that each spans the interval (0,1), as required for the optimisation to follow. In (a) z_0 and z_{rev} are varied, in (b) ϕ_0 and d_{vs}, holding parameters not varied at their correct values. The correct normalised value for each of the varied parameters is 0.5.

The utility of the metric is confirmed by the fact that in both plots the correct parameter combination corresponds with the global minimum, and the minimum is well defined. However, plot (a) also exhibits a slant, suggesting, as might have been expected from the data acquisition model, that an increase in the pitch z_{rev} can to some extent be compensated for by a decrease in the axial offset z_0. Such parameter interactions are interesting but also make the optimisation difficult. A further problematic feature of this objective function space is illustrated by plot (b): the minimum at the centre is (necessarily related to the somewhat arbitrary normalisation of the parameters) very sharp, especially in ϕ_0, and surrounded by an extensive plateau of exactly zero gradient. The plateau is caused by no paired features overlapping so that the r.m.s. residual error is maximal, taking a value based on subtracting zero from the features of the *base* channel.

An objective function space with these features is very hard to optimise over, and any gradient-based solver would inevitably fail if it ever found its way onto a plateau as seen in plot (b). Therefore an alternative metric was sought.



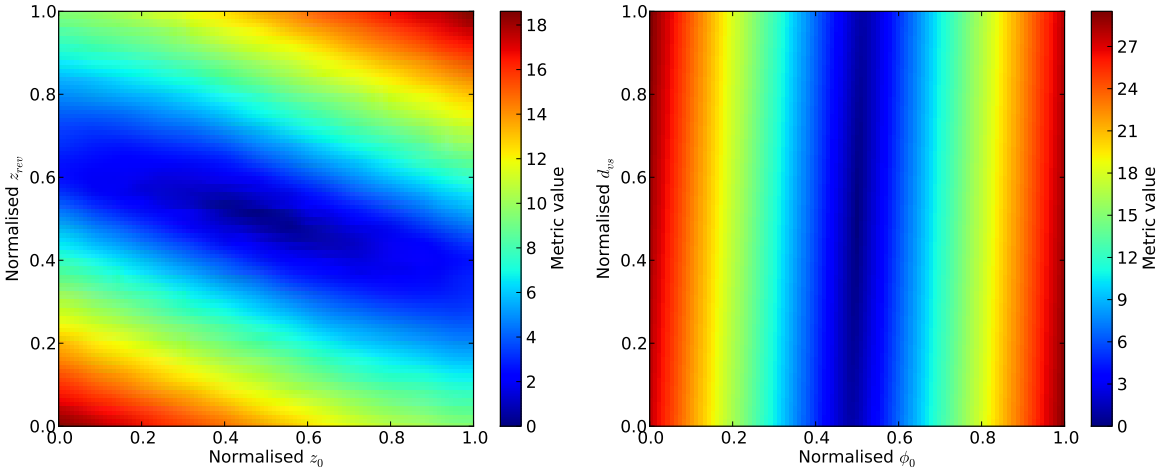
(a) **(b)**

Figure 4.13: Example cross-sections through the parameter space for a simple registration showing the variation of the r.m.s. residual metric, varying z_0 and z_{rev} in (a), and ϕ_0 and d_{vs} in (b). Parameter values are normalised to (0,1) as for the optimisation. For both plots the parameters not shown were held constant at their correct values, and the correct parameter combination is (0.5, 0.5).

4.5.2 Feature centroid displacement magnitude

This metric is given by mean Euclidean distance between the centroids of paired features, computing centroids as a "centre of mass" and treating amplitudes in a manner analogous to masses. Hence, this calculation does not require any interpolation of amplitude data and is consequently less computationally expensive than the r.m.s. residual metric of Section 4.5.1. Readers familiar with 2D image processing will recognise that the feature centroids computed here could be considered to be point features, with the 3D segment of the amplitude field around each serving as a feature descriptor.

The variation of this metric across sections through the parameter space for the same optimisation considered in Fig. 4.13 are presented in Fig. 4.14. The plots (a) and (b) are directly comparable with those for the r.m.s. residual metric. Again, the global minimum of the plots is at the correct (0.5, 0.5) parameter combination. However, here the objective function space provides far smoother optimisation surfaces than before and the metric is not dependent on spatial overlap between the paired features to obtain a meaningful metric value. On the other hand, the feature centroid displacement metric is insensitive to the exact amplitude distribution within a feature and therefore less capable



(a) **(b)**

Figure 4.14: Example cross-sections through the parameter space for a simple registration showing the variation of the feature centroid displacement magnitude metric, varying z_0 and z_{rev} in (a), and ϕ_0 and d_{vs} in (b). Parameter values are normalised to (0,1) as for the optimisation. For both plots the parameters not shown were held constant at their correct values, and the correct parameter combination is (0.5, 0.5).

of differentiating between small parameter variations close to the optimum combination than the r.m.s. residual.

Given that both metrics considered complement each other well, it was decided to investigate using both simultaneously. Tippetts took this forward to develop a multi-objective optimisation, described further in Section 4.7 and [13].

4.6 Objective Functions and Parameters for Multi-Channel Registration

Up to this point we have only considered a pair of channels to be aligned. However, often there are more than two channels available. For instance, in the most extensive scenario for the rotor bore inspection there are two acquisitions of four channels each to be registered to the global coordinate system provided by a user-specified *base* channel. The definition of appropriate objective functions must be considered, and also what parameters should be varied during the optimisation, given that some parameters are likely to apply equally to more than a single channel.

While the material that follows is specific to the rotor bore application, the underlying framework is general and applicable to a wide range of multi-channel and multi-acquisition data analyses, and the diagrams that follow can readily be adapted for a different number of data channels, for example. The author is not aware of a similar framework in the literature, at least not in NDE, and claims some novelty for the development work.

4.6.1 Objective functions

As registration relies on comparing the matched features of two channels, the complete alignment of multiple channels must be determined by a set of pairwise comparisons. The channels to compare to achieve registration should be selected carefully: they must be comprehensive and, at least indirectly, link to the *base* channel that defines the global coordinate system, but should not include redundant comparisons for computational reasons. Given eight channels, the minimum number of comparisons to achieve full registration is seven. This still leaves some choice over which seven channel pairs to compare, for instance the channels could all be compared against the *base* channel directly or "daisy-chained" together. Understanding the application allows the most appropriate selection to be made. In the case of the rotor bore inspection it is advisable to avoid cross-acquisition, cross-modality comparisons. This is because such comparisons are not only more difficult and more uncertain than within-acquisition, multi-modal or cross-acquisition, mono-modal comparisons, but the chosen approach for dealing with the challenges of feature extraction (see Section 4.3) in this application would not be applicable.

On that basis, two possible registration schemes have been drawn up. Scheme 1 is illustrated in Fig. 4.15, scheme 2 in Fig. 4.16, arbitrarily choosing Channel 1 (clockwise shear, in practice) of acquisition 1 to be the *base* channel. In both diagrams coloured arrows indicate the pairwise channel comparisons required, distinguishing between within-acquisition, multi-modal and cross-acquisition, mono-modal comparisons. As no comparisons are simultaneously cross-modality and cross-acquisition, it is theoretically possible to achieve complete registration using calibration scan data for the within-acquisition comparisons and noise-like features from the main scans for the cross-acquisition comparisons, in line with Section 4.3. In practice, for the case of the rotor bore inspection, scheme 1

corresponds to aligning all four channels of one acquisition based on the calibration scan data, and registering each channel of the second acquisition against its equivalent in the first acquisition using main scan data. This does not use the calibration data for the second acquisition at all. Scheme 2 on the other hand calls for the four channels of each acquisition to be aligned, using the calibration scan data of each acquisition, and the two acquisitions to be registered together using the main scan data for a single channel of each acquisition.

The second scheme is attractive in principle if the within-acquisition comparisons make a negligible contribution because the recorded probe pan settings that govern the default, pre-registration alignment of the channels within an acquisition are so accurate that adequate, complete registration can be achieved with a single mono-modal, cross-acquisition comparison. Additionally scheme 2 is preferable to scheme 1 if the main scan file for exactly one channel is known to contain clear geometric features to use for registration (e.g. the far-focus compression channel obtaining reflections from the outer rotor surface) so that good full registration is achievable despite a lack of useful geometric features in the remaining channels' main scan files. However, scheme 2 does involve additional degrees of separation between equivalent channels of the two acquisitions, such that when these are later subtracted (or similar) their alignment hinges on a sequence of comparisons in the registration having correctly steered the optimisation. This then entails more risk than the more direct scheme 1, borne out by the results of the next chapter.

Having chosen one of the two registration schemes, feature pairs are computed for each comparison. Note that some of these comparisons entail neither linked channel being the *base* channel, counter to the two-channel explanation of feature pairing of Section 4.4. The two registration metrics of Section 4.5 are defined for each comparison. This means that in the most extensive rotor bore inspection scenario, featuring eight channels and seven comparisons, a total of 14 metrics, and hence objective functions for the subsequent optimisation, are defined. As this is a very high number and liable to cause computational difficulties for the multi-objective optimisation (see Section 4.7), a metric collator was introduced, providing the option to merge the metrics of a given type and hence reduce the number of objective functions in the optimisation to two. In the current implementation, the merging operation is completed by taking the arithmetic mean.

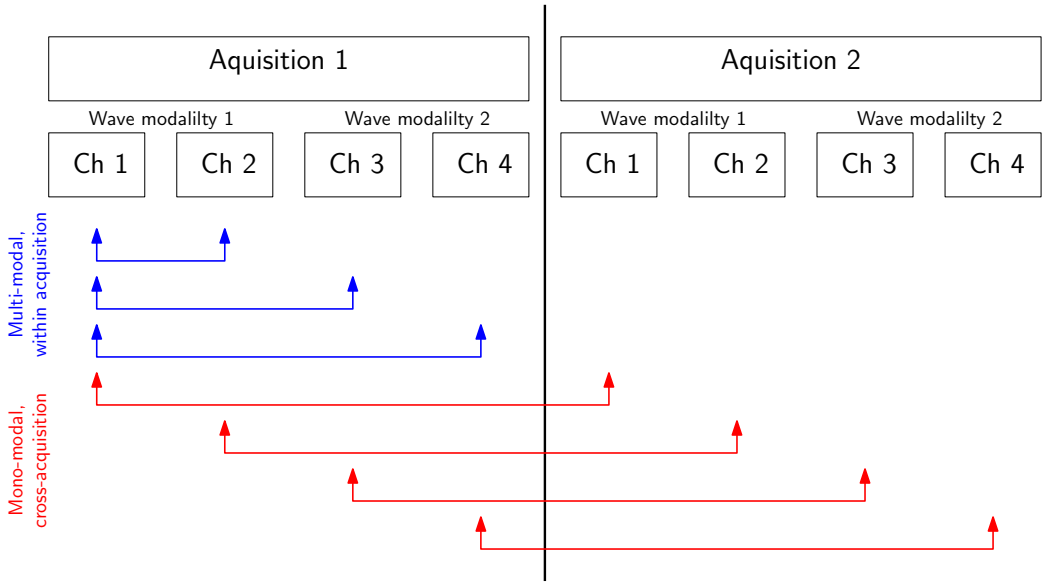


Figure 4.15: Diagram of registration scheme 1, showing the pairwise channel comparisons required for the complete registration of 2 acquisitions of 4 data channels each. The left-most channel is the global *base* channel, against which all others are to be aligned. Channels using the same type of interrogating wave type are labelled as being of the same wave modality. In practice, for the rotor bore application, the blue (multi-modal) comparisons rely on the calibration scan data for that acquisition, the red (mono-modal) comparisons use the main scan data.

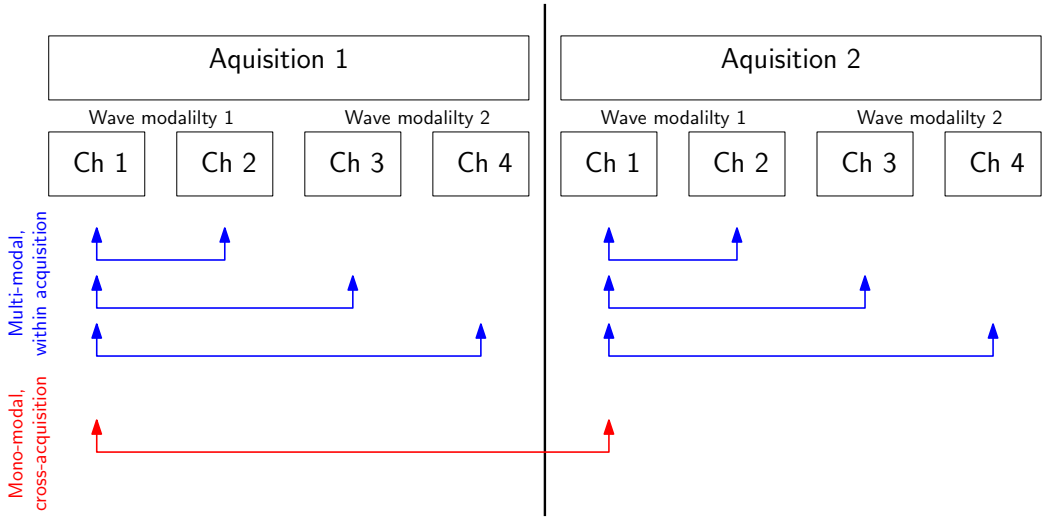


Figure 4.16: Diagram of registration scheme 2, showing the pairwise channel comparisons required for the complete registration of 2 acquisitions of 4 data channels each. The left-most channel is the global *base* channel, against which all others are to be aligned. Channels using the same type of interrogating wave type are labelled as being of the same wave modality. In practice, for the rotor bore application, the blue (multi-modal) comparisons rely on the calibration scan data for that acquisition, the red (mono-modal) comparisons use the main scan data. Compared with registration scheme 1 (see Fig. 4.15) this here has fewer cross-acquisition comparisons, but there are additional degrees of separation between equivalent channels of the two acquisitions. For example, the relative alignment of the two channel 3 instances hinges on each being aligned with channel 1 of their acquisition, and then those two channels being registered correctly.

4.6.2 Parameters

We must next consider what parameter should be varied during the optimisation. In Table 4.2 the parameters in the data acquisition model of Section 4.2 for describing the data acquisition in a single channel were presented. However, it would be physically wrong to vary all of these for each of the eight channels (in the most extensive rotor bore inspection scenario) independently. This is because some parameters must be the same across more than one channel, for instance the bore radius is the same for all channels of an acquisition. Additionally, the fact that the multi-modal comparisons will rely on the calibration scans rather than the main scans (as used for the mono-modal, cross-acquisition comparisons) must be taken into account, given that the parameters for the calibration scan may differ from those for the subsequent main scan.

To address this accurately, and in a generally applicable manner, the parameters for the single channel data model are classified according to whether they apply only to a single channel, an inspection wave type or all channels of an acquisition. The parameters are also split according to whether they can be considered the same across calibration and main scans in a channel, or are unique to each scan. This two-dimensional classification is illustrated in Table 4.3, indicating, for example, that each probe has three unique offsets, but that these values hold across both the main scan and the calibration scan, while the scan pitch (z_{rev}) is the same for all channels (given that the probes are held together on the probe pans) during any given acquisition. It should be emphasised that while the wave speeds (d_{vs}) for shear and compression channels of a single acquisition are physically related, the dependence on the (temperature dependent) material properties means that without an analytic expression connecting two, setting one does not specify the other in registration. This explains why d_{vs} falls in the wave modality, rather than acquisition, parameter class. Note that the classification shown nonetheless embodies several assumptions in an effort to keep the total number of parameters that need to be solved for to a minimum. Not only is it assumed that the probes are indeed fixed exactly and no changes (intentional or not) are made to the set-up from calibration scan to main rotor scan (other than potentially a change in a gain settings that does not affect relative gain settings), but also that the wave speed (material and temperature dependent) in the

	Main & calibration scans	Unique to main & calibration scans
Channel	z_0, ϕ_0, d_0, ρ, g	
Wave modality	d_{vs}	
Acquisition	z_{rev}, u_{rev}	R

Table 4.3: The classification of the data model parameters from Table 4.2 according to whether they apply to a single channel, wave modality or acquisition. Additionally, parameters are split according to whether or not they can be set to hold the same value across both the main scan and the preceding calibration scan for a particular channel. As explained in the text, the classification shown embodies several assumptions.

two components is the same. These assumptions allow the same values for all parameters, other than the bore radius, to be used across the two scans in each channel.

Based on these classifications, a diagram showing which parameters should be varied can be drawn up, as in Fig. 4.17. All parameters, regardless of class, associated with the *base* channel (shown left-most) are not varied during the optimisation, by the definition of the *base* channel. Therefore the diagram indicates, for example, that to register one rotor bore acquisition of four channels against a similar baseline acquisition requires three variables for the speed of sound, d_{vs}, to be determined by optimisation: taking the clockwise shear channel of acquisition 1 to be the *base* channel, one instance is needed for the compression wave channels of the same acquisition, and then one for each of the two wave modalities (shear and compression) of acquisition 2.

As will be explained in Section 4.8.1, in practice u_{rev} is not optimised over, leaving eight single channel data acquisition parameters to vary. If all these were used in the registration of two acquisitions of 8 channels as in the diagram, this would result in a 42-dimensional parameter space: five different channel parameters, over seven channels, plus one wave modality parameter, over three wave modalities, plus two acquisition parameters, over one acquisition, plus one acquisition parameters specific to each of the two calibration scans (i.e. $5 \cdot 7 + 1 \cdot 3 + 2 \cdot 1 + 1 \cdot 2$). This is a very high dimensionality for our purposes, and much higher than what Tippetts uses for the registration in the disk inspection application [13].

Therefore to help address this, a subset of the data acquisition parameters considered most worthwhile optimising over in practice was defined: z_{rev}, the helix pitch and axial increment per revolution; z_0, the scan start axially; ϕ_0, the scan start circumferentially; d_{vs}, the speed of sound in component; d_0, the scan start radially, along the A-scan (exclud-

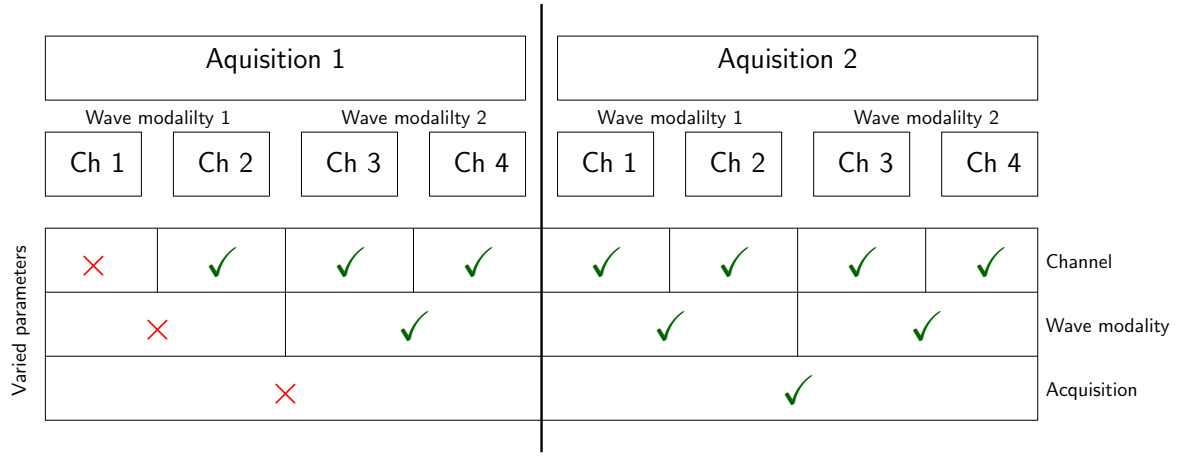


Figure 4.17: Diagram of which parameters to vary during the registration optimisation, depending on the class of the parameters: channel, wave modality or acquisition. The channels across which a parameter of a certain class applies are indicated by the column widths of the table. Green ticks mean that the corresponding parameters should be varied, red crosses that these parameters are held constant. Note that all the parameters associated with the left-most channel, which provides the global *base* coordinate system, are invariant.

ing u_{rev}, the number of A-scans acquired per revolution, as before). The probe angle ρ was eliminated since manufacturing standards and expected beam spread mean the expected deviation from the nominal value is small and the practical effect smaller still. The bore radius R was excluded as this quantity can readily be measured and should be very well defined, so no adjustment is expected to be necessary. Finally, the gain adjustment g was eliminated as this is of very limited significance in the post-registration processing (see Chapter 6). Using this reduced set of parameters, the dimensionality of the parameter space for the full registration is reduced to 25 (deduced as $3 \cdot 7 + 1 \cdot 3 + 1 \cdot 1$). It is conceivable that for instance using the data from the encoders fitted to the scanner (see Section 2.2) the number of parameters required in the optimisation could reasonably be further reduced.

4.7 Multi-Objective Optimisation

This section highlights some features of the optimisation used to minimise the registration error metrics (see Section 4.5) defined in line with Section 4.6.1. The optimisation is used to select parameters in the parameter space set-up based on Section 4.6.2, to facilitate the transform of the data channels into a common coordinate system. Given the division of

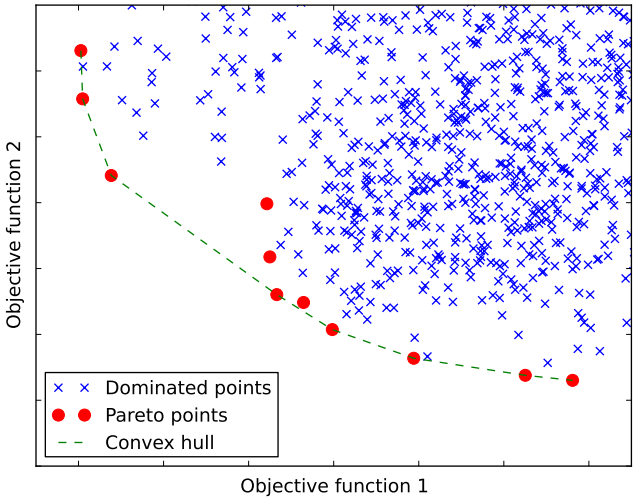


Figure 4.18: Example output of a multi-objective optimisation with two objective functions to minimise. The *Pareto* points marked in red are those samples which cannot be improved upon in one objective function without incurring a penalty in the other. The remaining samples are *dominated* and shown as blue crosses. The convex hull of the *Pareto* points is shown in green - note that some *Pareto* points lie above the line and are therefore outperformed by points on it.

responsibilities outlined in Section 3.4, beyond some initial investigations [71,72], work on the optimisation was exclusively done by Tippetts, as presented in detail in his thesis [13].

To the author, the optimisation used is largely a "black-box" algorithm, barely distinguishable from other global optimisation algorithms [73–80]. Almost the only difference in use is that this optimisation algorithm accepts a number of objective functions, revealing that it is in fact a multi-objective optimisation [81]. Associated with this, the output computed is not a single parameter combination considered to be the global optimum, but rather a set of such parameter combinations. They are a subset of the points in parameter space sampled, and are optimal in the sense that it is not possible to find a point in the set of samples that improves on one of the objective functions without incurring a penalty in another. This is known as *Pareto* optimality, and these output parameter combinations a termed *Pareto* points, making up a *Pareto* front. Figure 4.18 provides an example: the red dots are *Pareto* optimal, whereas the other samples, in blue, are described as being *dominated*. As a further enhancement, the plot also marks out the convex hull, made up of a subset of the plotted *Pareto* points, used in some applications, exploiting the fact every point on the hull plotted may be reached by stochastic interpolation between the end points of that line segment [82].

The fact that the optimisation yields a set of output points allows it to thereby represent the physical uncertainty in the registration that will exist: each *Pareto* parameter combination will map a test point to a different location in physical space, so each recorded amplitude can in fact be considered to physically lie in a point cloud. These sort of considerations are very important for later data fusion processing (see Chapter 6) as this relies significantly on the spatial coincidence of different signals. Interestingly, multi-objective optimisations may be biased to steer the evaluation to a more comprehensive *Pareto* front or representation in a response quantity (here physical coordinate location) [58].

As with any other optimisation, the dimensionality of the parameter space being optimised over affects how long the optimisation takes to reach a stopping criterion. Broadly, the more parameters there are, the longer it takes for that space to be sufficiently explored as more samples are required. In the case of this optimisation, additional consideration should be given to the dimensionality of the objective function space. Using a greater number of objective functions will also tend to slow down the optimisation, primarily because the hypervolume integral used by the implemented optimisation becomes more computationally demanding [13]. However, on occasion, an additional objective function may provide an efficient pathway to the basin of the global optimum, helping the optimisation converge (however that may be defined) more rapidly. In practice, the termination of the multi-objective optimisation used depends solely on the user-specified number of objective function space evaluations it is required to complete. As this optimiser in fact processes a fixed-size sample population at a time, the total number of evaluations is an integer multiple of the number of samples in one such generation (20, here).

4.8 Parameter Pre-Computation

This section describes how analytic calculations may be used to compute certain parameter values prior to the described optimisation, so as to reduce the load on that complex calculation. First, we examine how the problem of a varying wrapping period u_{rev} was overcome before considering how the calibration scan can be further exploited.

4.8.1 The wrapping period

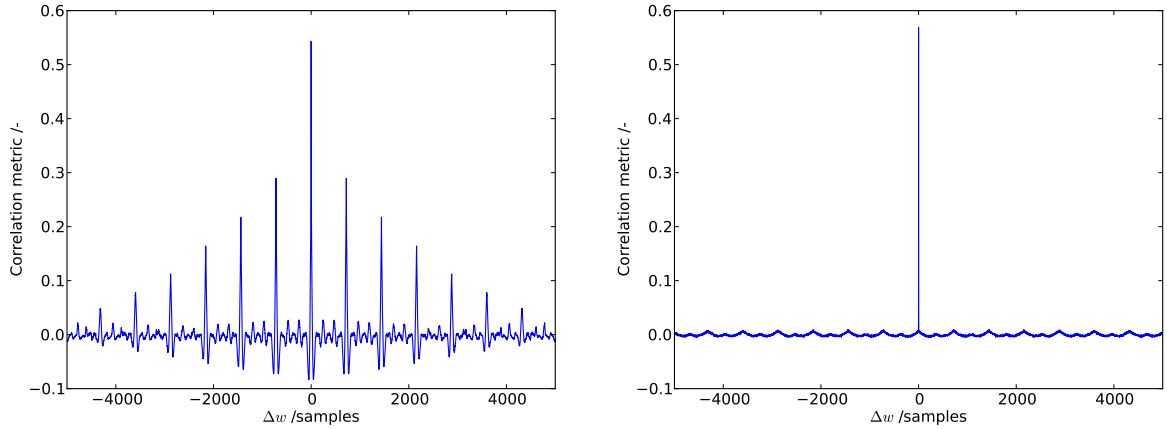
As described in Section 4.3, because the wrapping period u_{rev} affects the 3D index space representation of the data and hence the relative spatial proximity of different amplitude samples, changes in this parameter can undermine the feature extraction, which in turn is liable to undermine the entire registration. This is also related to the fact that it is not possible to recover from an inappropriate u_{rev} value without recomputing the features, which will only occur if u_{rev} changes enough to cause a change to the value rounded to nearest integer: while physically u_{rev} may take a float value, when used to define the number of samples along one axis of a data array it must take an integer value (the structure of the data was shown in Fig. 4.6). Therefore while the registration is very sensitive to some u_{rev} changes (causing a change to the rounded version), the optimisation has no sensitivity at all to changes near the rounded value of the parameter. This is a significant challenge to the registration [55].

Fortunately, the author discovered an innovative solution to this problem: pre-computing u_{rev} from the data and then eliminating that parameter from the set of parameters optimised over. The pre-computation is based on examining the periodicity of the data in the helical direction, along w (see Equation 4.1). The inherent periodicity in the sequence of A-scans, revealing the wrapping period of the data, is probed using the autocorrelation in this direction. In practice this is computed in the frequency domain at a number of A-scan depths (t indices) in the frequency domain, as in Equation 4.9:

$$a_f(\omega) = \mathcal{F}_f^{\star}(\omega)\,\mathcal{F}_f(\omega) \tag{4.9}$$

where a_f is the autocorrelation of function f, ω is the (spatial) frequency, $\mathcal{F}$ denotes the *Fourier* transform, and * the complex conjugate.

Figure 4.19 illustrates an example of the output obtained by evaluating an autocorrelation of the sort described, in plot (a) for a calibration scan and in (b) for a main scan. The autocorrelation is by definition symmetric about and maximal at the zero shift location. Both plots exhibit a sequence of regular peaks, which are extremely sharp in (a) and less pronounced but still detectable in (b). The separation of any two neighbouring peaks directly gives the sought wrapping period u_{rev}. The specific implementation adds a few



(a) **(b)**

Figure 4.19: The mean (across several t indices) autocorrelation of the data along the length of the helix (for select A-scan depths), used to compute the circular wrapping period from the data. By the definition of the autocorrelation, these plots are symmetric about zero shift. Both plots relate to the clockwise shear channel, but (a) is based on the scan of a calibration test piece, with large reflectors (side-drilled holes) aligned with the bore, while (b) is based on a main inspection scan, containing no meaningful geometric features. In the former very large, regular peaks readily and correctly identify the wrapping period (720 samples), and while the peaks are not nearly as prominent in (b), they are sufficient for the calculation to succeed.

refinements, such as outlier rejection, given that multiple estimates of the parameter can be obtained, and has been shown to be both extremely accurate and robust. In practice, the computation is just used to confirm that the nominal u_{rev} value is correct, and allows u_{rev} to be safely eliminated from the variable parameters of the registration optimisation.

The calculation does make the implicit assumption that the features in the data responsible for the dominant peaks in the autocorrelation are on average axially aligned. This is in practice certainly the case for the calibration scan, given the axial orientation of the side-drilled hole target reflectors, but not necessarily the case in general. A misalignment with the component axis could lead to the computed data feature period not corresponding to the wrapping period sought. However, such a possible misalignment can be compensated for being applying a misalignment factor, effectively an average angle of features to the axis, computed in the *base* data channel of the registration.

4.8.2 Using the calibration scan

In addition to using the calibration scan data during the multi-modal registration comparisons (as described in Sections 4.3 & 4.6), these files can be used to compute some parameters analytically, given how well-known the target reflectors are.

In the current implementation of the software only the A-scan start parameter d_0 is computed prior to registration in this way, allowing the optimisation to be initialised (see also Section 4.9) with a broadly correct value, despite this particular parameter not being explicitly specified in the original data files and potentially depending on the analogue-to-digital converter trigger configuration in relation to the probe design. This calculation simply extracts the depths of the observed side-drilled hole reflections and fits these against the target hole positions.

However, more complicated calculations are conceivable and may be beneficial. For example, it should be possible to calculate several parameters from the *base* channel calibration scan and thereby confirm that the parameters for this *base* channel (including for the main scan) are indeed consistent, as currently assumed. Additionally, given analytic sound propagation and probe models [6] it may be possible to extract a more comprehensive model of the beam (beyond just considering the main beam as a perfect line) using an iterative fitting scheme (i.e. a further optimisation) [83]. Beam descriptors such as main beam width could then be exploited during a later processing stage using the Synthetic Aperture Focussing Technique (SAFT) (see Section 6.2 and Appendix B), for example.

4.9 Initialisation

This section outlined how the optimisation described in Section 4.7 should be initialised to improve performance (for a given number of iterations). Broadly, the optimisation should be started at the "best guess" of the parameter values, and allowed to explore from there. However, as the optimisation in fact works in terms of generations of points, a set of start locations, corresponding to a full generation, may in fact be specified.

In the current implementation, a set of parameter combinations making up a generation of optimisation points is built up using a biased Latin hypercube [84]. The biasing places

the selected points around the parameter space point of the best available estimates of the parameter, taking into account pre-computations (see Section 4.8) where available but defaulting to the nominal parameters provided by the original data files.

The selection of the initial parameter combinations with which to launch the main optimisation could potentially be improved with the outputs of a short, preliminary optimisation over a reduced objective and parameter space [85], or an attempt to compute a subset of parameter values analytically [55]. A further option is to exploit the encoder data acquired during the main scan to inform the initial parameter selection.

4.10 Summary

This chapter has presented a comprehensive system for the registration of two rotor bore data acquisitions, of four different ultrasonic channels each, to a common coordinate system. Such a system is of critical importance to the subsequent comparison and fusion of data from different channels, detailed in Chapter 6. The system as described is specific to the rotor bore inspection, with a number of features differentiating it from that presented by Tippetts for the case of the disk inspection application [13], but much of the underlying logic and framework is applicable to a great range of multi-channel registration situations. While the registration has the optimisation developed by Tippetts at its core, the framework of Section 4.6 and method for overcoming the challenge of the wrapping period in Section 4.8.1 developed by the author represent original contributions to knowledge. The next chapter examines the performance of the described registration system, for a range of test scenarios.

This page intentionally left blank

Chapter 5

Registration Results

5.1 Introduction

After explaining the motivation for data registration and the deployed mechanics in Chapter 4, in this chapter we examine the performance of the registration operating on inputs of the sort typical in industry. The results here complement those presented in the thesis of the author's collaborator, Tippetts, for the disk inspection application [13]. Whilst most of the novelty and academic merit claims of the author in this work pertain to the data fusion side of the work, and the registration approach heavily relies on the work of Tippetts, the unique challenges of the rotor bore inspection application described in the previous chapter compared with the disk inspection application mean that there are elements of novelty and academic merit on the registration side, too. Some of these have already been detailed in the registration method, but others will be highlighted in the results, as the complexity of the registration in the rotor bore application gives rise to some interesting features.

The registration performance is analysed in terms of the positional error of test points that have a (at least relatively) well known correct position. Given that the cross-acquisition, mono-modal and within-acquisition, multi-modal alignments have different properties, and primarily affect baseline subtraction and single acquisition fusion, respectively, these are analysed separately throughout. We start by considering test cases for these two types of alignment in isolation, building up to full multi-channel, dual-acquisition registration.

Note that the effective registration accuracy requirements are primarily determined by the spatial scale, informed by the registration uncertainty, of the data association between channels in the fusion processing - see Chapter 6.

All results presented in this chapter are based on the data for the inspection of a low pressure stage 2 turbine rotor at RWE npower's Didcot A coal-fired power plant, Forging C001 993 Y21657, inspected in January 2007 as job number GBF4201. The files used in the analysis here all relate to the section of the bore 2.4 - 2.9m axially from the alternator end of the rotor bore, given that the 6m long bore is inspected in 50cm sections (plus a little overlap).

Unless otherwise stated, all computations of this chapter involved running the optimiser described in Section 4.7 and developed by Tippetts [13] for 2000 evaluations of the objective function space, corresponding to 100 generations of 20 parameter combinations. Parallelised sections of the code ran on three cores of the CPU on the desktop computer described in Section 3.2.1. Computations for Sections 5.2 and 5.3 took around 45min and 90min, respectively, those for Section 5.4 considerably longer. Figures A.1, A.2 & A.3 provide an example of the sort of *Operator* (see Section 3.4) configuration relevant to the calculations of this chapter.

5.2 Cross-Acquisition, Mono-Modal Registration

In this section two copies of a single channel's main scan data (the counter-clockwise shear channel was chosen for this section) are registered against each other, having first distorted the parameters relating to the second copy away from their correct values. The objective of the registration is then to recover these known parameter values, and the success of this is assessed in terms of the spatial error remaining after registration at a number of test locations, this being of greater practical significance than the error in parameter space. Given that the required mapping is known for any point in the data domain and the parameter distortion, due to the complexity of the geometry, may lead to significantly varying positioning errors across that domain, the test locations used were 20 points uniformly spread across the domain by Latin hypercube pseudo-random sampling [84].

Parameter	Explanation
z_{rev}	helix pitch, axial increment per revolution
d_{vs}	speed of sound in component
z_0	scan start axially
d_0	scan start radially, along the A-scan
ϕ_0	scan start circumferentially
ρ	probe angle, in a plane normal to the bore axis
R	bore radius
g	linear amplitude scaling
u_{rev}	number of A-scans acquired per revolution

Table 5.1: The parameters that could potentially be varied to control the dual-acquisition alignment of the two copies of a particular type of channel. The top five are considered the most significant for practical purposes. The next three may be also be varied to increase the complexity of the model and dimensionality of the parameter space. The final parameter (u_{rev}) is included for completeness but never varied, only computed from the data to confirm the nominal parameter value - see Section 4.8.1.

Out of the possible eight parameters that could potentially be varied to control the dual-acquisition alignment of the two copies of a particular type of channel in the model of the preceding chapter, here the computations and introduced distortions are restricted to the five parameters of Section 4.6.2 considered to be most significant in practice. These are detailed in Table 5.1. As a single channel comparison is used, for which two objective functions are defined, the optimisations in this section feature a two-dimensional objective function space over a five-dimensional parameter space.

5.2.1 Small distortion

First, a small distortion, corresponding to an axial shift, rotation and modified sound speed, was introduced to the parameters of the second acquisition's channel. The effect of this distortion on the test locations is shown in Fig. 5.1 as a histogram. The registration samples the two-dimensional objective function space as shown in Fig. 5.2. This plot also illustrates the way the root-mean-squared residual metric (see Section 4.5.1) reaches a plateau when the registration features (see Section 4.3) do not overlap at all, and shows that the definition of the metrics means that the correct parameter combination does under these perfect conditions (with two identical inputs) correspond to the origin of the objective function space. Mapping the test locations to the different parameter combina-

tions corresponding to *Pareto* points identified in the objective function space of Fig. 5.2 allows a point set for each location to be identified. These point sets are summarised in two perpendicular projections in Figs. 5.3a and 5.3b, plotting the centroids and standard deviations, computed using Tippetts' *Pareto* density correction to adjust for the uneven spread of points on the sampled *Pareto* front discovered by the optimisation [13]. An alternative presentation of these results is offered in Fig. 5.4, reducing the positional errors in different directions to a single Euclidean distance - both for all *Pareto* points together and for the *Pareto* point with the lowest r.m.s. residual error metric value. Without close examination, this particular *Pareto* point would by default be considered the best single output, as the second metric, the feature displacement metric, largely serves to draw the features together so that they overlap (to permit the calculation of the residual-based metric) and depends significantly on the definition of the feature centres between which the distance is computed. However, in this case, that parameter combination is seen to contribute some of the highest errors, providing further evidence of the benefits of defining multiple objective functions. Nonetheless, the reduction in error in these results histograms over the equivalent pre-registration plot in Fig. 5.1 demonstrates the efficiency of the registration.

5.2.2 Large distortion

Now we consider the introduction of a larger parameter distortion and attempt to apply registration to recover from that. Figure 5.5 shows the positional errors of the test locations prior to registration. The objective function space of the optimisation is displayed in Fig. 5.6, indicating the four *Pareto* points identified. The associated point sets from the test locations being mapped using the parameter combinations underlying the *Pareto* points are shown in two projections in Figs. 5.7a and 5.7b. These positional errors post-registration are summarised in the histograms of Fig. 5.8. As previously, the *Pareto* point of the lowest residual metric value happens to have some of the highest associated positional errors, so relying on a single objective function, even at the end of the optimisation, is risky. Nonetheless, the reduction in error over the starting point (Fig. 5.5) is

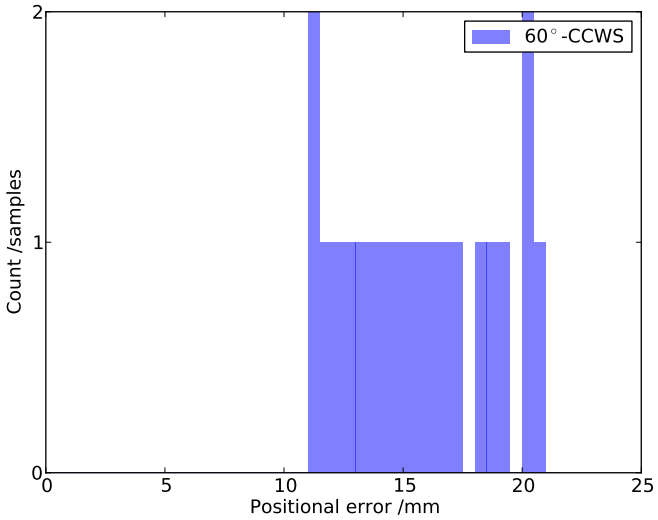


Figure 5.1: The effect of the imposed small parameter distortion on the test locations relative to their positions when the correct parameters are applied, prior to registration. In practice such an error would be due to the probe for the channel in question not being set-up for a second acquisition to match the first acquisition. The error bins of the histogram are arbitrarily chosen. The plots of this sort in this section are based on registering 2 copies of the 60° counter clockwise channel (CCWS) against each other.

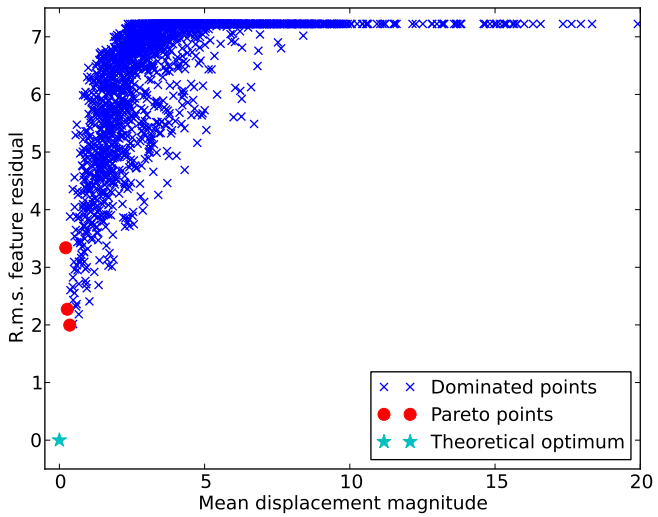
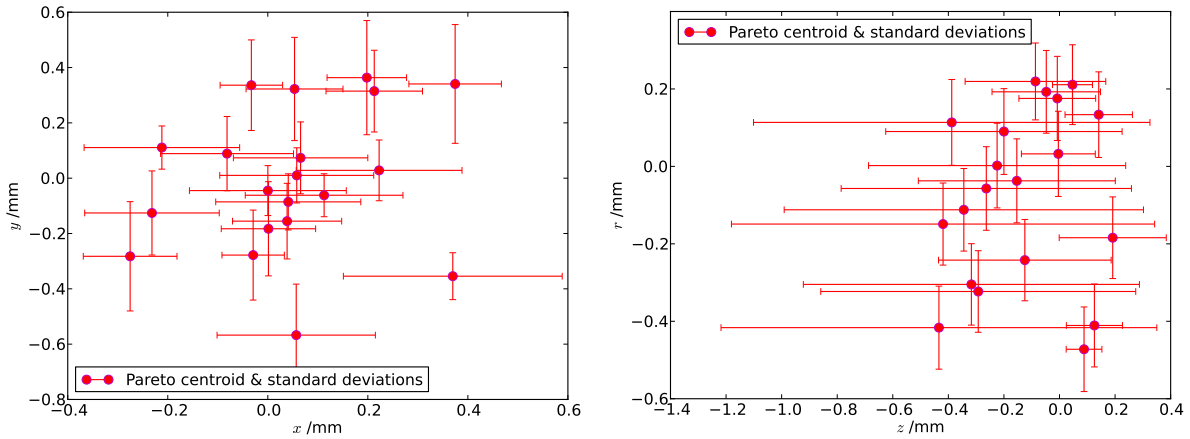
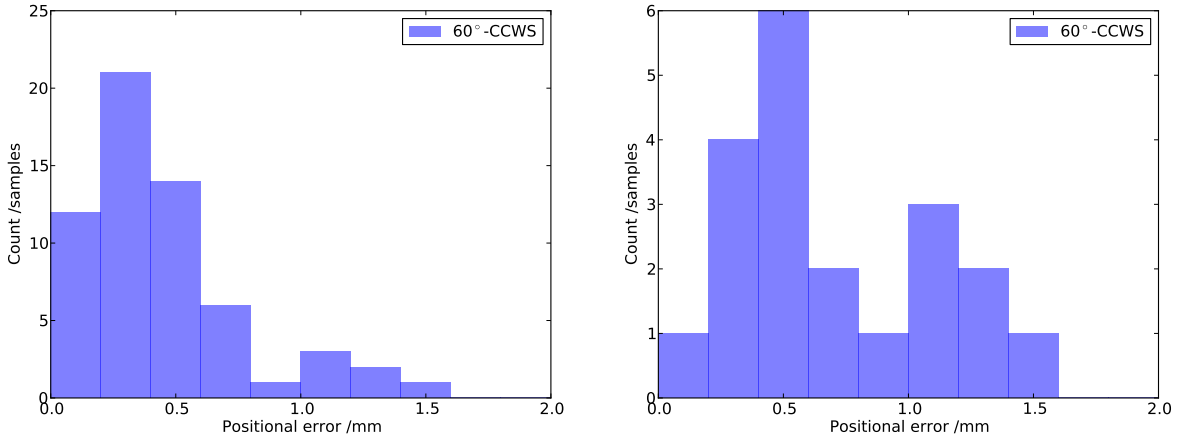


Figure 5.2: The 2-dimensional objective function space for the optimisation applied for the registration of 2 copies of one channel to compensate for a small applied distortion. 3 *Pareto* points are identified, jointly minimising the 2 objective functions. Note the plateau feature of the metric plotted on the vertical axis, the result of the registration features having no spatial overlap at those parameter values, as well as the zero objective function values at the correct parameter combination.



(a) (b)

Figure 5.3: The centroids and standard deviations of the point sets resulting from mapping the test locations with the parameters underlying the optimisation's *Pareto* points shown in Fig. 5.2. In (a) the point sets are projected into the plane normal to the bore axis, in (b) into the plane normal to a circumferential vector. While in (b) some bias towards negative axial errors is evident, all centroids indicate sub-millimetre errors.



(a) (b)

Figure 5.4: Histograms of the test location omnidirectional positional error in the registration of two copies of one channel. In (a) all the identified *Pareto* points contribute, whilst in (b) only the parameters associated with the *Pareto* point of the lowest r.m.s. residual error metric value are used in the mapping of the test points. Note that (b) features most of the larger errors shown in (a), indicating that the mappings associated with the other *Pareto* points have lower associated errors.

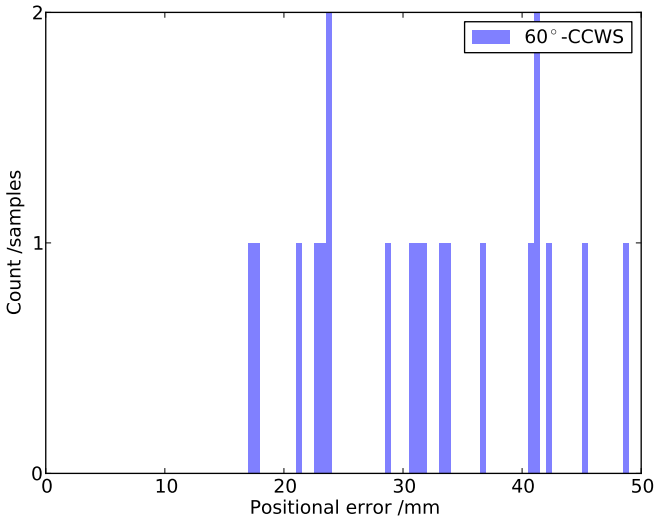


Figure 5.5: The effect of the imposed large parameter distortion on the test locations relative to their positions when the correct parameters are applied, prior to registration. The error bins of the histogram are arbitrarily chosen. This plot, like those before, is based on registering two copies of the 60° counter clockwise channel (CCWS) against each other.

dramatic, again illustrating the efficiency of the registration process.

It is worth noting that the errors at the end of the optimisation are comparable to those for the previous case of a smaller distortion. Similarly, Tippetts found that the positional errors of the parameter combinations converged to were comparable over a great range of introduced distortions, suggesting that these errors reflect physical uncertainty in the data, a conclusion supported by the correlation lengths (see next chapter) of the data being of a similar scale [13].

5.2.3 Discussion

The results presented in Sections 5.2.1 & 5.2.2 have demonstrated the ability of registration to effectively recover the correct channel alignment for some test distortions applied to one of two copies of the same channel. While some parameter distortions are likely to be more difficult to recover from than others (given pathways through the objective function space to global, multi-objective optima), potentially requiring more optimisation generations to reach a suitable output, a separate greater concern in practice is the validity of the assumed model, in contrast to the consistency of that model. Distorting the parameters as described to provide a test input only provides a subset of the possible

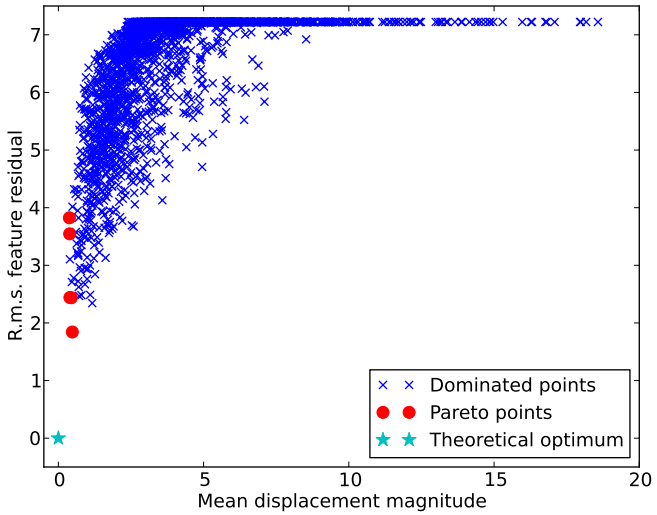
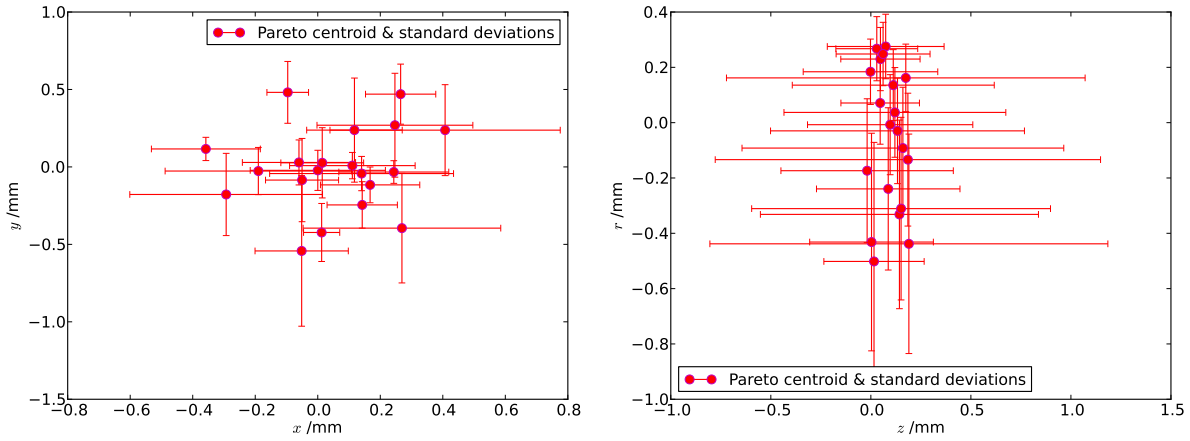


Figure 5.6: The 2-dimensional objective function space for the optimisation applied for the registration of 2 copies of one channel to compensate for a large applied distortion. In this case 4 *Pareto* points are identified. As in Fig. 5.2, a plateau in the r.m.s. residual metric is evident.



(a) **(b)**

Figure 5.7: The centroids and standard deviations of the point sets resulting from mapping the test locations with the parameters underlying the optimisation's *Pareto* points shown in Fig. 5.6. In (a) the point sets are projected into the plane normal to the bore axis, in (b) into the plane normal to a circumferential vector. In (b) the centroids are all very near zero error axially, but with significant uncertainty - much greater than in the radius direction, for example. However, the errors plotted are all very small.

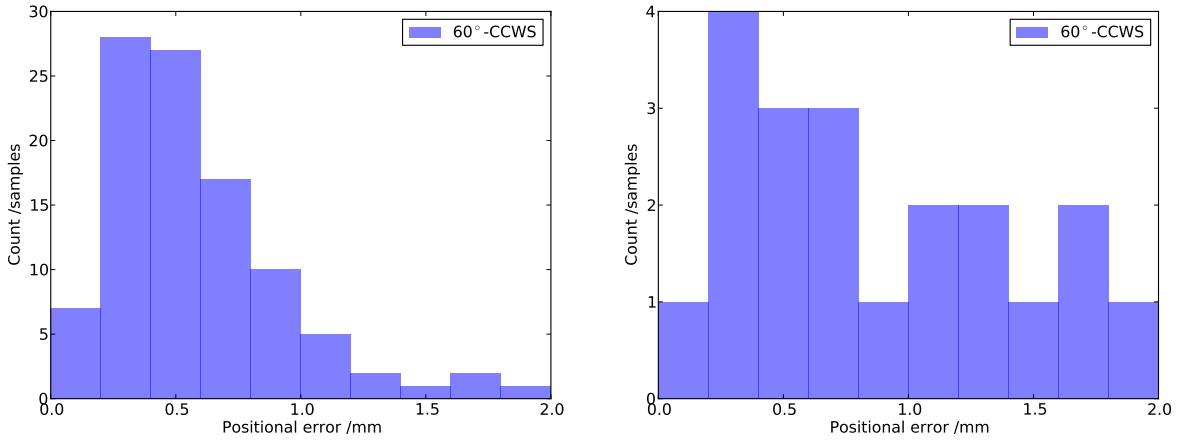


Figure 5.8: Histograms of the test location overall positional error in the registration of two copies of one channel. In (a) all the identified *Pareto* points contribute, whilst in (b) only the parameters associated with the *Pareto* point of the lowest residual error metric value are used in the mapping of the test points. As in Fig. 5.4, (b) features most of the larger errors shown in (a) indicating that the mappings associated with the other *Pareto* points have lower associated errors.

changes that may occur and then need to be compensated by registration. Therefore the significance of possible uncompensated effects in data (for example relating to probe skew or complex, localised temperature effects - see Section 4.2) has not been examined.

For such an examination, the second file introduced into the registration would need to be one from a separate scan of the same part of the component as shown in the first file, rather than an exact copy of the first file. While such files from repeat scans are available, the lack of significant geometric features that could be used as registration features in almost all files means that the registration shown currently relies on noise-like features, which is inadequate for effective registration of differing files. Undoubtedly, the feature extraction and pairing algorithms applied (see Chapter 4) could be improved to provide a greater degree of the necessary robustness to permit at least a demonstration. However, this would not be sufficient to permit practical use, as the robustness would still not be comparable to that achievable with significant geometric features in the data, as described in Section 4.3.

5.3 Single Acquisition, Multi-Modal Registration

This section concerns the alignment of several channels of different types acquired during the same acquisition. As explained in Chapter 4, the involved multi-modal comparisons are in practice tackled by focussing on data from the calibration scans. The hope is that the challenge of different inspection modalities, giving rise to differences in the appearance of signals from component features, can be overcome by examining the reflections from the simple side-drilled holes of the calibration piece. Fortunately, the correcting adjustments the registration is anticipated to require are small, as all the probes used in the rotor bore inspection are fixed in the probe pans and the relative probe positions may be readily measured. In fact, it is conceivable that no adjustment is required.

The data used here is from all four ultrasonic probes, acquired in a calibration scan before the inspection of the rotor bore itself. For testing purposes we really require knowledge of the correct alignment between inputs, so that, as in Section 5.2, the effectiveness of the registration can be quantified in terms of the positional error of test points. The alignment of inputs is however only exactly known if these inputs are simulated, as in cross-channel case, where one input is a distorted copy of the other. Such simulation is not possible in the case of the multi-modal data considered here, at least not without resorting to tools such as finite-element (F.E.) modelling [54], as even for a simple reflector like a side-drilled hole, the signals in one channel will not just be a copy of another channel's signals with modified acquisition parameters. Additionally, the objective here is really to test the extent to which the applied model provides an adequate representation of distortions in real data, rather than the model self-consistency. Therefore introducing simulated data, from F.E. for example, risks defeating the object of this testing.

Relying on real data as the test input means the reference point problem remains - it is not obvious which test locations should coincide exactly in the event of perfect alignment. The approach adopted provides a reasonable if imperfect solution: given the well-defined calibration hole features in all the data channels, the mid-points of the faces that are normal to the holes (and hence the bore) are assumed to coincide exactly when registered correctly - see Fig. 5.9. While this assumption could hold exactly in theory, in practice it suffers from several problems.

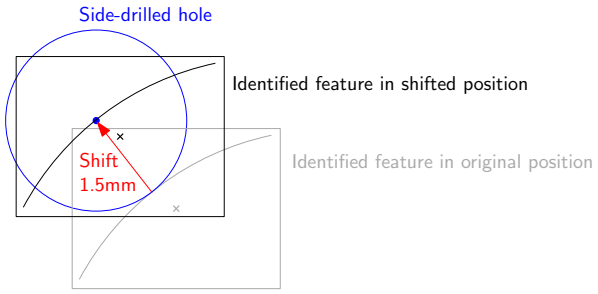


Figure 5.9: Schematic of a side-drilled hole and the associated registration feature identified in the data, in an axial projection. The feature here is a rectangle around the arc-like reflection of the side-drilled hole seen in the data as depicted. For the purposes of registration, this feature is shifted by the radius of the side-drilled hole, so that reflections from different directions coincide at the centre of the hole. The mid-point of this shifted feature in the plane shown, marked by a cross, is used here as a test location to assess the effectiveness of registration. Note the discrepancy in the position of this point and the centre of the hole, illustrating the sort of error that can occur in the test locations.

The most significant problems concern the axial direction, related to the axial sample density. This sample density is the reciprocal of the scan pitch (nominally 5mm) and, as indicated by Fig. 5.10, is by far the coarsest compared with the other directions. So if during the feature extraction stage (see Section 4.3) an inaccuracy of just a single index location occurs, this immediately gives a 5mm spatial error. The other problem is that the ends of the calibration holes are rather sub-optimal targets for determining axial (as opposed to circumferential or radial) alignment given their small diameter compared with the circumference of the test piece: at the axial position at which the main beam of the probe is axially exactly aligned with the end of a target hole the probe is unlikely to be in the required rotational position. For the assessment of axial alignment of the probes, a target that extends around the full circumference of the test piece is highly desirable. While the axial ends (at the faces) of the disk section of the test piece (containing the side-drilled holes, see Fig. 5.11) might be thought to provide such a target, the reflections in the region are unclear and complicated by the joining material, linking the disk section to the test piece guiding tube. Therefore a modified calibration piece design, for example enhanced by the addition of a set of circumferentially distributed, part-thickness side-drilled holes of a constant size, depth and radial position, is one recommendation from this project.

A further objection to the use of the calibration hole end points as reference locations for

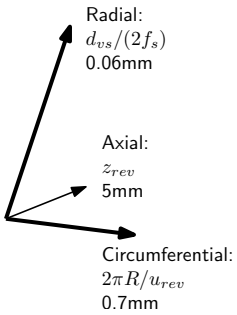


Figure 5.10: The 3D sampling intervals, which are the reciprocal of the sampling densities. The values quoted are approximate typical values, for a shear wave channel and a position close to the bore surface. The vectors shown are not to scale.

Figure 5.11: A photograph of a rotor bore calibration test piece. The disk section containing the target reflectors (side-drilled holes) and the guide tube are clearly identifiable. The guide tube also extends out of the far-side of the disk section. Note the joining material between the two sections of the test piece. *Photo taken by author at RWE npower's Ferrybridge site.*

the registration is that they are not well-spaced, especially in a circumferential direction, so provide no information about the alignment achieved in other regions of the component. Note that the calibration hole end points also require pairing across scans to enable comparisons, a further possible cause of errors.

The positional errors of these calibration hole end-point test locations in counter-clockwise shear, short-focus compression and far-focus compression channels relative to the channel designated as the global *base* in testing, the clock-wise shear channel, are shown in Fig. 5.12 prior to registration. Note that due to the described problems associated with using the calibration hole end-points as reference locations, the best achievable positional error

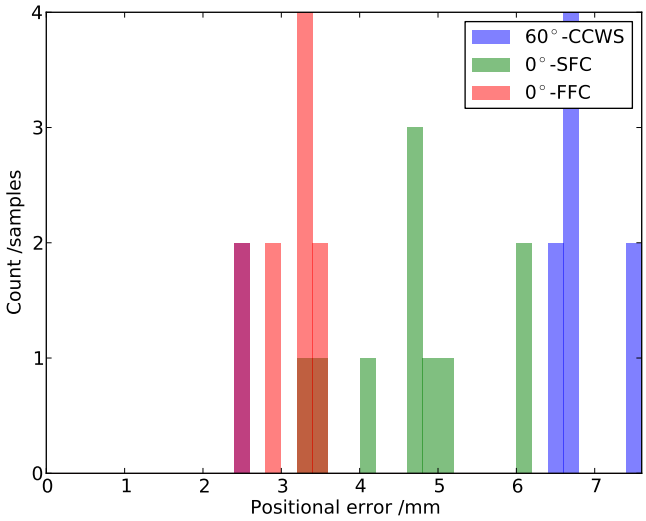


Figure 5.12: Histogram of the positional error of the multi-channel test locations relative to their positions in the global *base* channel prior to registration. In practice these plotted errors are related to errors in the set-up of the probes on the probe pans, but the error values shown include an indeterminate contribution from the described reference point problem. The legend abbreviations relate to the counter-clockwise shear (CCWS), short-focus compression (SFC) and far-focus compression (FFC) channels.

will never be zero, and will differ for different locations. Therefore the best that can be hoped for in effective registration is that the positional errors are seen to generally decrease - not just some locations at the expense of others. While this is not hugely satisfactory, it should however be borne in mind that in view of the described challenges in determining suitable reference points, the error values plotted represent a worst case for subsequent fusion processing.

5.3.1 Low dimensional parameter space

An initial optimisation was completed using a low dimensional parameter space, focussing on the ten variables considered to be most significant (from Table 5.1 and the logic of Section 4.6.2), and a two-dimensional objective function space. The output in the latter space is shown in Fig. 5.13. This differs somewhat from the objective function space plots seen so far as there is no plateau visible in the residual-based metric. This is because with a greater number of paired features, spread across several channels, these are very unlikely to all have no overlap at once. Note also that the theoretical optimum is not known for this optimisation, and given the differing data inputs, it is certain that the origin will

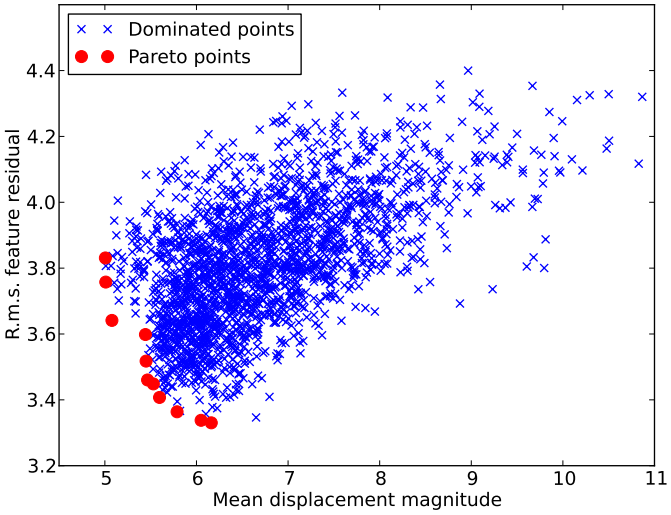


Figure 5.13: The 2-dimensional objective function space for the low dimensional parameter space multi-modal registration of 4 different channels of the same acquisition. Several *Pareto* points have been identified.

never be reachable.

Applying parameter combinations underlying the identified *Pareto* points gives rise to the test location positional error histograms shown in Figs. 5.14a & 5.14b. As before, only the parameter combination associated with the *Pareto* point of the lowest residual contributes to the second plot. A marginal improvement over the starting state (Fig. 5.12) is detectable in this. The former does show that significantly lower errors can be achieved by individual points, but that some improvements are at the expense of other points - especially in the short-focus compression channel, where many points are seen to have a worse error than in the starting state.

5.3.2 Higher dimensional parameter space

To examine whether the registration could be improved by allowing the full set of 16 model variables (from Table 5.1 and the logic of Section 4.6.2), to be optimised over, the testing was repeated with this higher dimensional parameter space, leaving all other control parameters unchanged. The resultant objective function plot is shown in Fig. 5.15, which is rather similar to that of the lower dimensional parameter space optimisation (Fig. 5.13). The positional errors derived from the identified *Pareto* points are illustrated in Figs. 5.16a & 5.16b. The former plot is rather similar to the equivalent plot in the lower

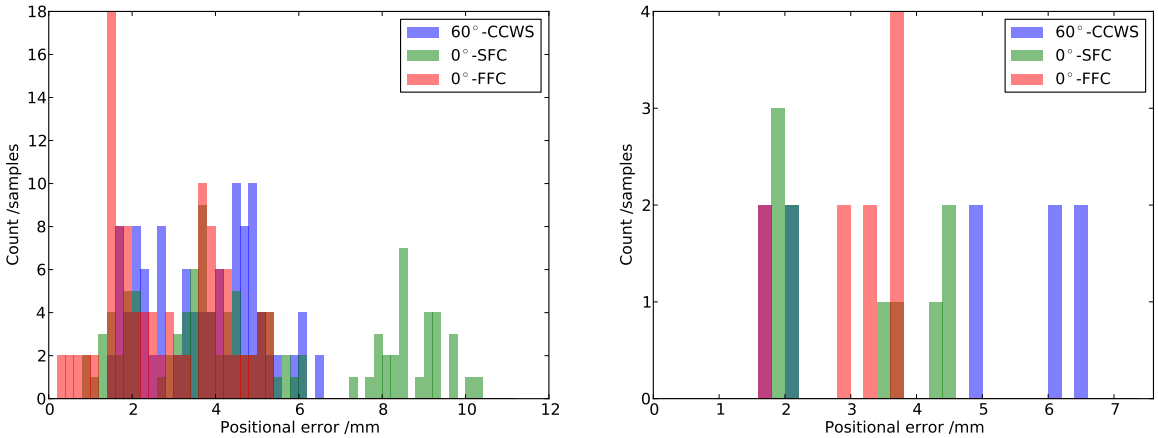


Figure 5.14: Histograms of the test location overall positional error in the low dimensional parameter space registration of all 4 channels of 1 acquisition. In (a) all the identified *Pareto* points contribute, whilst in (b) only the parameters associated with the *Pareto* point of the lowest residual error metric value are used in the mapping of the test points.

dimensional parameter space case, Fig. 5.14a - again, low errors have been reached in some locations, but, given the larger errors seen in the short focus compression channel, some of these improvements appear to be at the expense of other test point locations. The errors for the lowest residual metric *Pareto* point (Figs. 5.14b & 5.16b) are also quite similar though the latter does provide a slight improvement. Based on this it would seem the benefits of using a more comprehensive set of parameters and hence more flexible model in the registration is of limited benefit especially given the computational cost of having to explore a higher-dimensional variable space.

5.3.3 Higher dimensional objective function space

To complement the increased dimensionality of the parameter space considered above, the objective function space dimensionality was also increased in the hope this would improve overall performance, perhaps by reducing the extent to which improvements in some locations appear to be at the expense of others. Therefore the previous registration test was repeated using a 16-dimensional parameter space (from Table 5.1 and the logic of Section 4.6.2) and 6-dimensional objective function space (from de-activating metric merging, see Section 4.6.1, across three comparisons).

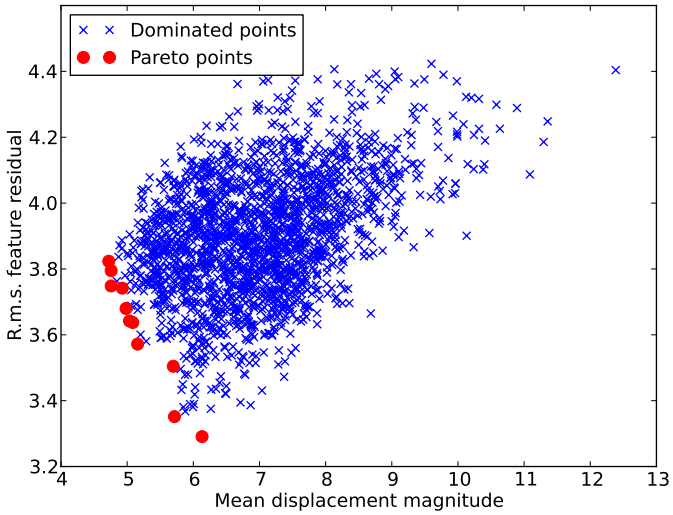


Figure 5.15: The 2-dimensional objective function space for the case of the higher dimensional parameter space multi-modal registration of 4 different channels of the same acquisition, which is quite similar to the plot for the previous lower dimensional parameter space (Fig. 5.13).

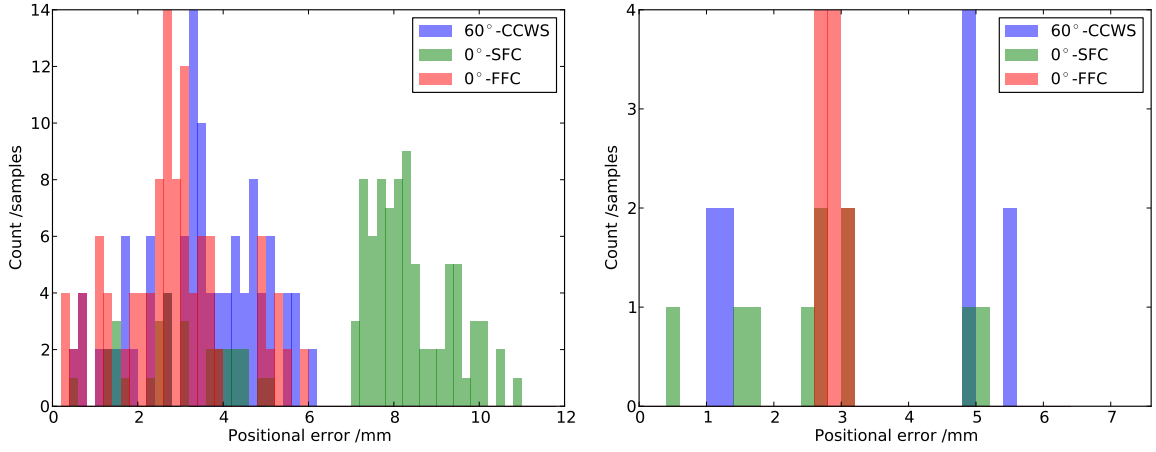


Figure 5.16: Histograms of the test location overall positional error in the higher dimensional parameter space registration of all 4 channels of 1 acquisition. In (a) all the identified *Pareto* points contribute, whilst in (b) only the parameters associated with the *Pareto* point of the lowest residual error metric value are used in the mapping of the test points.

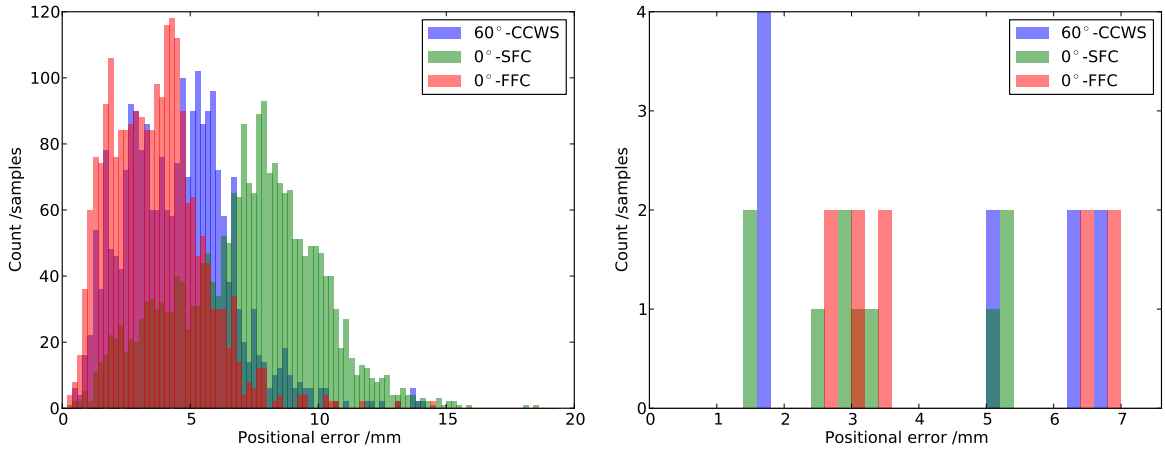


Figure 5.17: Histograms of the test location overall positional error in the registration with higher dimensional parameter and objective function spaces of all 4 channels of 1 acquisition. In (a) all the identified *Pareto* points contribute, whilst in (b) only the parameters associated with one of the *Pareto* points of a lowest residual error metric value are used in the mapping of the test points.

In such a case the objective function space cannot be visualised effectively so no attempt to illustrate it is made here. The errors associated with the identified *Pareto* points are shown in Figs. 5.17a & 5.17b, for all *Pareto* points and one of the the *Pareto* points with the lowest r.m.s. residual metric value (given that three of the six metrics are of the r.m.s. residual type, of Section 4.5.1, there is likely to be more than one *Pareto* point with the lowest r.m.s. residual metric value). The first plot shows a great spread of errors, related to the high number of identified *Pareto* points underlying the distributions. As in the previous registration trials, the errors for the short focus compression channel are seen to lag behind those for the other channels. While the second plot indicates something of an improvement over the starting state (Fig. 5.12) it cannot be described as an improvement over the comparable plots in the preceding registration tests. Therefore the benefits of using a higher dimensional parameter space in the optimisation appear to be limited at this stage.

5.3.4 Artificial distortion

The tests in Section 5.3 have so far only attempted to improve on the positional error associated with the nominal parameter values. As explained previously, the required

adjustments are expected to be small anyway, to the extent that it is conceivable that no registration adjustment is necessary or even achievable. Furthermore, the described reference point problem (see top of Section 5.3) compromises the extent to which the effectiveness of the registration can be demonstrated. To provide a more convincing illustration of the action of the cross-channel registration, in this section we introduce an additional parameter distortion, of the sort the registration aims to protect against, that the algorithm then attempts to recover from. Such a distortion in practice corresponds to the probe pan set-up being recorded incorrectly, maybe due to operator error in the measurement. We also now revert to the initially used optimisation settings, varying ten parameters and assessing performance in terms of two objective functions, bearing in mind that the applied distortion can readily be restricted to the varied parameters. It should be noted though that most of the test is now concerned again with the self-consistency of the model rather than the extent to which it is an adequate representation of physical phenomena (see Section 5.2.3).

The chosen distortion corresponds to all the probes on the pan (bar the probe providing the *base* channel, the clockwise shear) being rotationally and axially shifted, by a couple of degrees and millimetres, respectively, and a temperature change affecting the speed of sound. This is designed to reflect the sort of distortion that could occur in practice. The effect of this distortion on the initial test location positional error is shown in Fig. 5.18, indicating a significant error increase over the initial state without an additional distortion seen in Fig. 5.12. The objective function space output of the registration optimisation is presented in Fig. 5.19. While this plot identifies several *Pareto* points as in previously seen objective function spaces, the shown split of the point cloud reveals one of the special features of this optimisation. This bi-modal (potentially multi-modal) behaviour is due to the feature pairing snapping from one possible set of (binary) correspondences to another as the (continuous) parameters being optimised over change. The error histograms of the test location mappings from the parameter combinations associated with the *Pareto* points are shown in Figs. 5.20a & 5.20b, for all *Pareto* points and only the *Pareto* point with the lowest residual metric value, respectively. These are quite similar to the corresponding plots seen for the initial multi-modal registration considered, Figs. 5.14a & 5.14b, indicating that the additional distortion has had little effect on the output

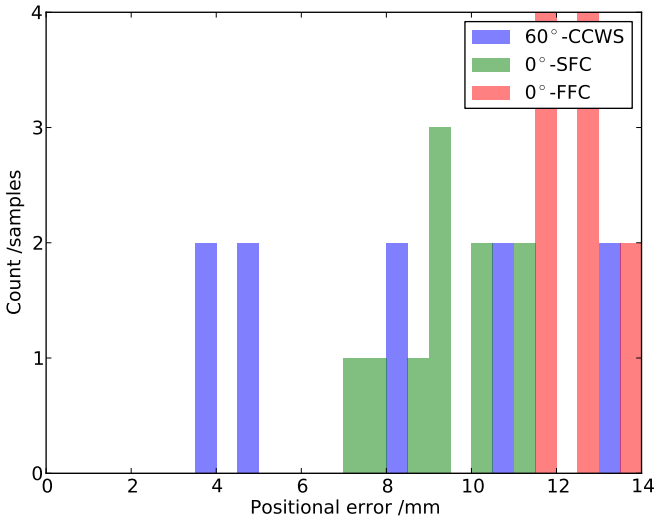


Figure 5.18: Histogram of the positional error of the multi-channel test locations relative to their positions in the global *base* channel prior to registration after the introduction of an artificial distortion in addition to the error resulting from the nominal parameter values.

obtained. Perhaps surprisingly, the errors seen in the plots for the single *Pareto* point are in fact lower here, for the initial state with an additional distortion. Certainly the results demonstrate the ability of the registration to recover from a substantially wrong initial state.

5.3.5 Discussion

The results presented here have attempted to demonstrate the ability of the software to correctly register several different channels, in which data was acquired in a single acquisition. In practice this corresponds to correcting errors in the set-up of the probe pans. The algorithm makes extensive use of the available calibration scan data to help overcome the challenges of different modalities, building on the assumption that the physical probe settings will be exactly the same during the calibration and the subsequent main scan. While the deployed data acquisition model has been shown to be self-consistent and broadly consistent with physical effects, the lack of knowledge of the exactly correct registration of the available test inputs and the fact the required adjustments to the nominal parameter values are expected to be small, mean that it is difficult to provide more convincing performance illustrations. This situation could be significantly improved with a superior calibration test piece design, as outlined at the start of Section 5.3. Moreover, if

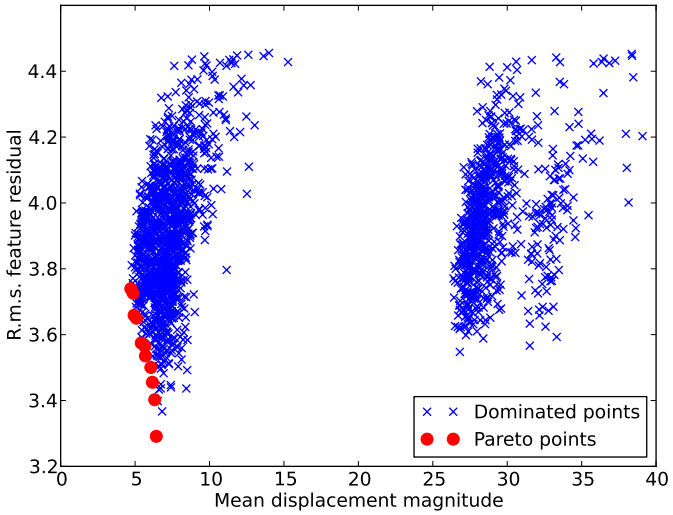


Figure 5.19: The 2-dimensional objective function space for the case of the registration of 4 different channels of the same acquisition following the introduction of an additional, artificial distortion. As before, several *Pareto* points have been identified. However, this plot also illustrates one of the special features of this optimisation: the multi-modal behaviour indicated by the splitting of the point cloud is due to the feature pairing snapping from one possible set of correspondences to another as the parameters being optimised over change.

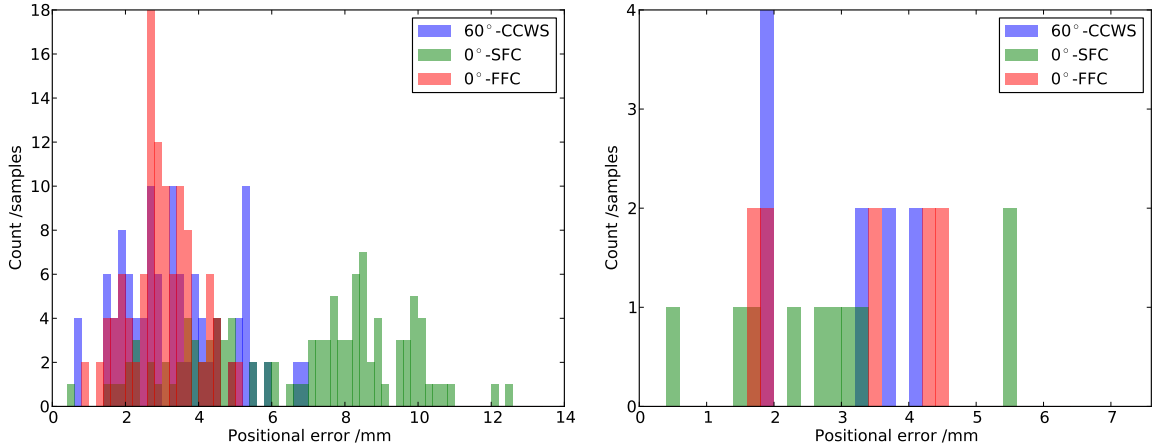


Figure 5.20: Histograms of the test location overall positional error in the registration of all 4 channels of 1 acquisition after the introduction of an additional, artificial distortion. In (a) all the identified *Pareto* points contribute, whilst in (b) only the parameters associated with the *Pareto* point of the lowest residual error metric value are used in the mapping of the test points.

the probe set-up in the scanner could be assured to the highest levels of confidence during the inspection, the single acquisition, multi-modal registration described here could be made redundant.

5.4 Registration of Two Full Acquisitions

Having in the previous two major sections considered first cross-acquisition, mono-modal registration then single acquisition, multi-modal registration, we now consider the case of both simultaneously, during the full registration of two multi-channel acquisitions - a total of eight inputs. Given the analysis presented so far, the performance is assessed in exactly the same manner, separating errors in the two types of comparison. Not only does this approach make the results here more directly comparable with those preceding, but also means difficulties in the assessment of multi-modal registration need not undermine analysis of within-modality registration, and the consequences of errors in subsequent data comparison processing are clearer.

The registration calculations in this section were completed considering only the five variables in each channel seen as the most significant in practice: z_{rev}, z_0, ϕ_0, d_{rev}, d_0 - see Section 4.6.2. Given the different classes of parameter identified, this meant there were a total of 25 variable parameters. As in most previous registration computations, the objective function space was two-dimensional. However, because of the very high dimensionality of the parameter space, the number of evaluations was doubled to 4000, corresponding to 200 generations of 20 points. Even so, the sample density in parameter space is low compared with the previous calculations.

The test input consists of the same 4 channel input files as in Section 5.3, but duplicated and distorted for the second acquisition (much like a single channel was in Section 5.2). The distortion corresponds to the probe pan of the second acquisition being set-up inaccurately, the scan then being started at a different point and the temperature being different (altering the wave speeds), compared with the first acquisition. The error histograms for the chosen combination of distorting parameters are shown in Figs. 5.21a & 5.21b, for single-modality and cross-modality comparisons, respectively. In line with previous results, the former uses a set of random, exact test locations uniformly spread across

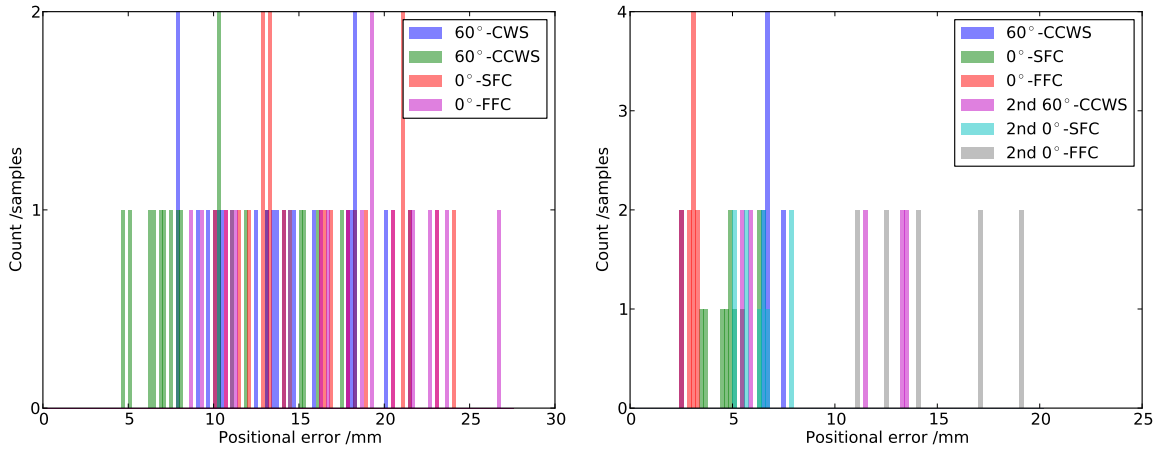


Figure 5.21: Histograms of the positional error of the test locations prior to registration following the introduction of small distortions to the parameters of all channels of the second acquisition. In (a) single modality, cross-acquisition errors (in practice from equipment for a second acquisition not being set-up the same as for the first) are shown, using a set of random test locations uniformly spread across each channel domain. In (b) the single acquisition, multi-modal errors (in practice owing to probe misalignments on the probe pans) are presented for both acquisitions present in each case using the positions of the calibration hole end-point locations relative to those of the *base* channel (the clockwise shear channel) of that acquisition.

each channel domain, whereas the latter relies on the side-drilled hole end mid-points as reference locations, with all the associated difficulties. It should be noted that as no additional distortion was applied to the channels of the first acquisition, the low dimensional parameter space case examined in Section 5.3 is actually a subset of the situation here.

5.4.1 Registration scheme 1

The computation was completed using the registration scheme 1 of the previous chapter (see Fig. 4.15). This scheme relies on three cross-modality comparisons (using calibration scan data) for the first acquisition and a mono-modal, cross-acquisition comparison (using main scan data) for each of the four channel modalities. The objective function space plot for the 4000 evaluations completed is shown in Fig. 5.22, with identified *Pareto* points. The error histograms for these are presented in Figs. 5.23a & 5.23b, for cross-acquisition and cross-modality comparisons respectively. These show a significant, general improvement over the initial state (Figs.5.21a & 5.21b), though individual components have not all improved. For example, the points for the first short focus compression

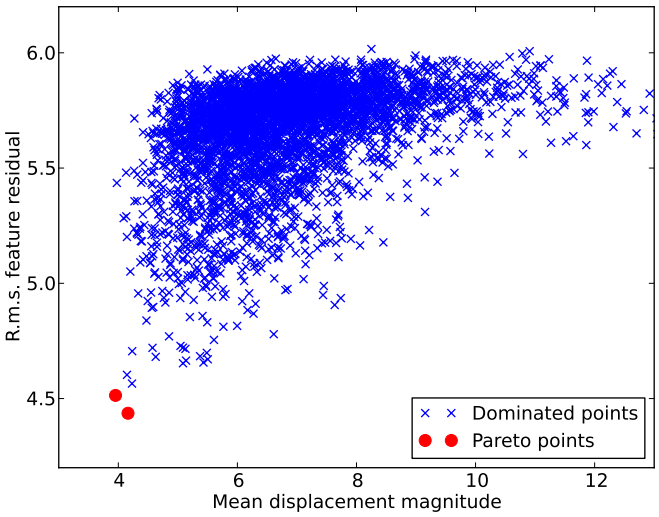


Figure 5.22: The 2-dimensional objective function space for the full registration of 2 multi-channel acquisitions using registration scheme 1, with 4000 evaluations. The two identified *Pareto* points are highlighted.

channel in the cross-modality all appear to have a higher error than before (similar to what was seen previously). It is also noteworthy that the results cannot match those obtained for the previous calculations in which components of this registration were considered in isolation (e.g. Figs. 5.4 & 5.14a). The most likely explanation for this is simply the high dimensionality of the parameter space and the relatively limited exploration of this over the course of the optimisation - completing more evaluations may well improve this.

5.4.2 Registration scheme 2

The computation was repeated, but using the second registration scheme (see Fig. 4.16). This relies on three multi-modal comparisons within each of the two acquisitions (using the calibration data for both acquisitions to check for probe set-up errors in each) and a single mono-modal, cross-acquisition comparison (using the main scan data to fix the alignment of the probe pans for the two acquisitions relative to each other). The objective function space plot for the 4000 evaluations completed is shown in Fig. 5.24, with identified *Pareto* points. The error histograms for these are presented in Figs. 5.25a & 5.25b, for cross-acquisition and cross-modality comparisons respectively. While both show an improvement over the initial, pre-registration state (Figs. 5.21a & 5.21b), they are both (especially the first) worse than their equivalents under registration scheme 1 (Figs. 5.23a

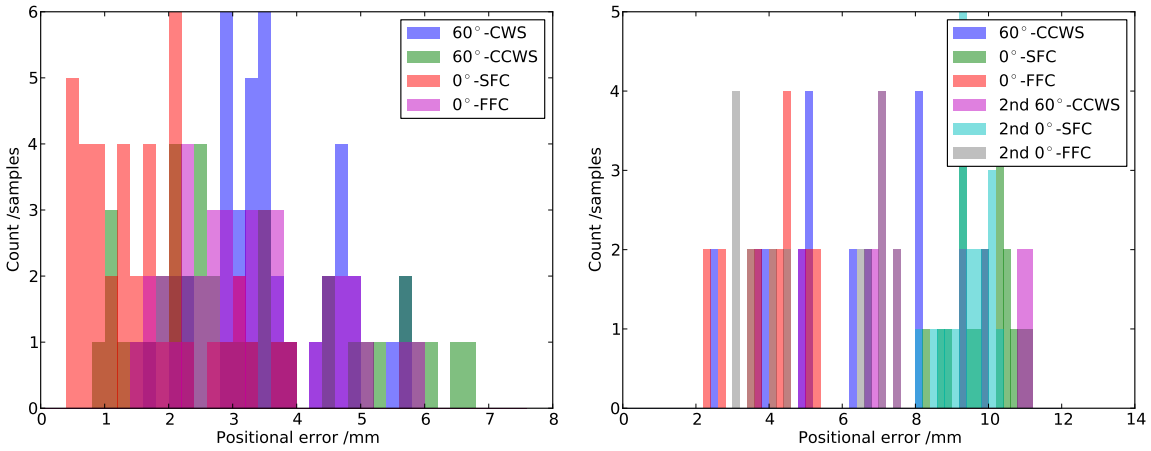


Figure 5.23: Histograms of the test location overall positional error for the registration scheme 1 alignment of 2 full acquisitions of 4 channels each. In (a) single modality, cross-acquisition errors (in practice from equipment for a second acquisition not being set-up the same as for the first) are shown, mapping the set of random test locations using the parameters underlying the *Pareto* points. In (b) the single acquisition, multi-modal errors (corresponding in practice to probe misalignments on the scanner head) are presented, for both acquisitions present, mapping the calibration hole end-point locations using the parameters associated with the *Pareto* points.

& 5.23b). Additionally, the first plot shows that there are significant inter-channel discrepancies in the mono-modal errors. Specifically, the clockwise shear channel has very low errors, while the short focus compression channel consistently (across all *Pareto* points and test locations - i.e. for all registration outputs and all over the component) exhibits very high errors. This would appear to be an extreme case of one channel improving at the expense of another. The likelihood of this occurring is increased by the combination (in the current implementation, by taking the mean) of the objective functions from all the channel comparisons, reducing the total number of objective functions to two (see Section 4.6.1). As the number of comparisons increases, the ability of any one to steer the direction of the optimiser is diminished. This situation could potentially be improved by using a more advanced combination operator than the mean, or by not combining the metrics at all and optimising over a higher dimensional objective function space. As in the registration scheme 1 case, the relatively low number of evaluations (given the dimensionality of the parameter space) means the output is likely to improve with more evaluations, and the mono-modal channel discrepancy is likely to be reduced. The relatively poor performance seen here compared with the alternative scheme, especially in the mono-modal

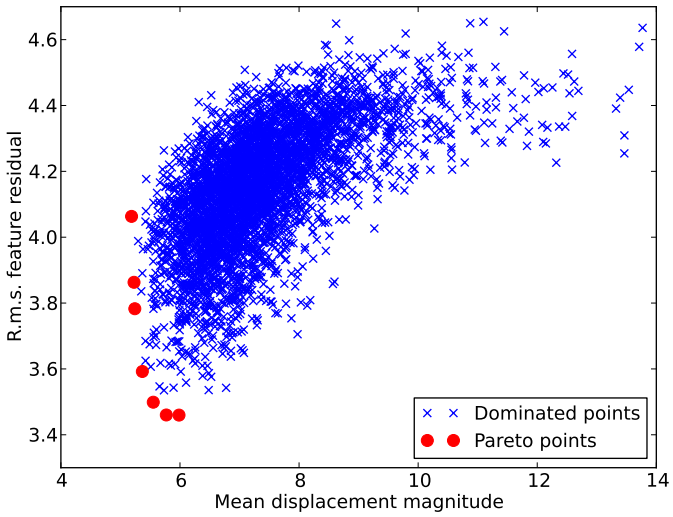


Figure 5.24: The 2-dimensional objective function space for the full registration of 2 multi-channel acquisitions using registration scheme 2, with 4000 evaluations. Several *Pareto* points have been identified.

errors, is explained by the way that the mono-modal comparisons for which the errors are plotted (with the exception of the *base* channel, the clockwise shear) depend on both the mono-modal registration computation for the *base* channel and the multi-modal registration comparisons. There can be considered to be extra degrees of separation between the channels for which the errors are plotted as these are only indirectly aligned (see Fig.4.16), which in itself is liable to lead to a lag effect in the registration, but because this extra degree of separation adds a dependence on the problematic multi-modal registration (see Section 5.3), results are particularly compromised.

5.4.3 Discussion

The results in Section 5.4 have shown that progress in full multi-channel, dual-acquisition registration has been made - the results presented under registration scheme 1 are certainly promising. However, the high dimensionality of the parameter space does at this point present a significant challenge if computations are not to take unreasonably long: the result of Sections 5.4.1 and 5.4.2 took around 11.0hrs and 8.0hrs to compute, respectively (the disparity is understood to be due to differences in file access). It might

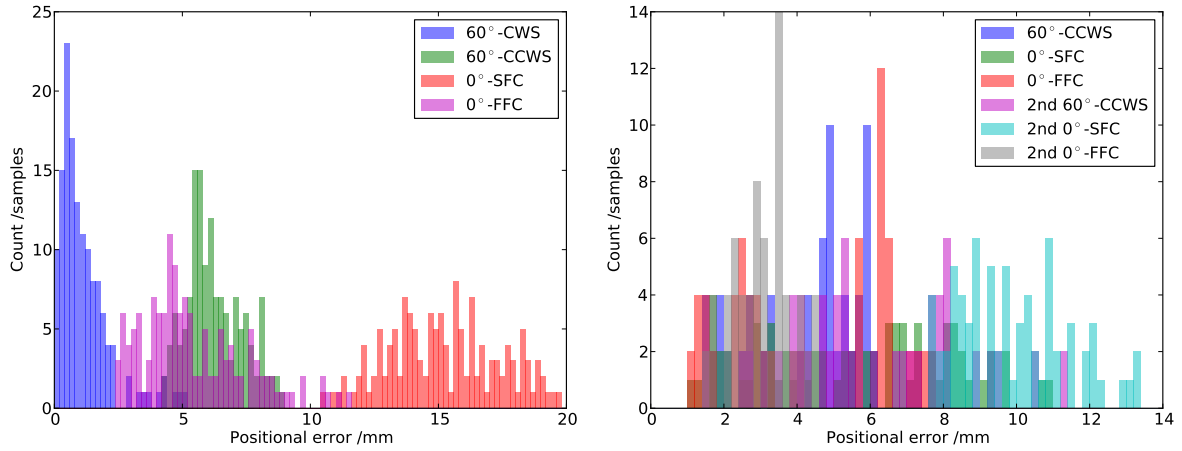


Figure 5.25: Histograms of the test location overall positional error for the registration scheme 2 alignment of 2 full acquisitions of 4 channels each. In (a) single modality, cross-acquisition errors (in practice from equipment for a second acquisition not being set-up the same as for the first) are shown, mapping the set of random test locations using the parameters underlying the *Pareto* points. In (b) the single acquisition, multi-modal errors (corresponding in practice to probe misalignments on the probe pans) are presented, for both acquisitions present, mapping the calibration hole end-point locations using the parameters associated with the *Pareto* points. Note the significant discrepancies in the performance of different channels seen in (a).

be possible to reduce that dimensionality by making carefully justified assumptions. For example, as temperature changes are likely to be the cause of all variations in the speed of sound, a suitably accurate analytic model of the effect of temperature on steel would allow the shear and compression sound speeds to be related so that only one needs to be varied in the optimisation. Of course the code as it stands could also certainly be further speed-optimised and also run on a newer, more powerful computer. While not reducing the number of objective functions to just two is appealing as an approach for reducing the extent to which improvements in one channel occur at the expense of another, this does further increase the computational cost. Specifically, the author did experiment with not combining contributions from different channels at all, resulting in a 14-dimensional objective function space, but found the associated hypervolume integral used by the optimiser to be prohibitively slow to compute. A compromise solution worth investigating as further work would be to combine metrics of one of the two types, but not of the other, giving an 8-dimensional objective function space, which should be manageable.

While the second registration scheme does have some advantages, based on the perform-

ance seen, without further changes and testing, registration scheme 1 should be used wherever possible.

5.5 Summary

The chapter has shown and described the results of testing the developed registration system for the rotor bore data. The results for simple, mono-modal, cross-acquisition testing are very convincing, those for multi-modal and full, multi-channel, dual-acquisition registration are less clear. Part of the problem is the suitability of the available data. Improvements could certainly be made, both to the inspection procedure and the registration algorithm, and several recommendations have been made to this effect. In particular, changes to ensure the presence of meaningful signals from geometric features in the main scan files, and a revision of the calibration piece design are very desirable. It is also worth emphasising that if the set-up of probes could be assured the need to perform multi-modal registrations could practically be eliminated, to the extent that the registration computations would only be deployed to enable the comparison with another acquisition, in principle then only using mono-modal comparisons.

The consequences of the registration (or lack thereof) for subsequent processing must also be considered. The data fusion processing described in the next chapter for combining signals from several channels acquired during a single acquisition is directly affected by the within-acquisition, multi-modal registration, but has a certain tolerance for positional error or uncertainty. However, if the error is too large (especially in regions of heightened sensitivity) this can mean that a defect signal seen in two channels is not considered to be from the same source, such that the two do not reinforce each other, thereby increasing the possibility of the defect being missed (see Detection Results, Chapter 7). On the other hand, the data subtraction-based comparison of channels from two acquisitions, that may be applied as an effective pre-stage in the data fusion, hinges critically on the mono-modal alignment of inputs, relying entirely on the fusion data / noise model to handle errors that arise. Fortunately, mono-modal registration can in principle be achieved to a high degree of accuracy.

It is the author's belief that the demonstration and discussion of the complex features

of the registration in the rotor bore application (such as multi-modal behaviour in the objective function space and the comparison of registration schemes 1 & 2) constitutes a contribution to knowledge. While these features are not unique to the application considered, and are likely to have appeared in the wider optimisation literature [73–80], the author is not aware of any similar work in the field of NDT.

Chapter 6

Data Fusion

6.1 Introduction

This chapter develops a general approach for the probabilistic combination of signals in an identified region of the component, containing spatially coincident sections of arbitrary amplitude fields, from different data channels whose domains intersect that region.

The intention is for such a probabilistic evaluation system to be used as a Data Fusion Detector (DFD). In view of the project objectives described in Chapter 1, this is not designed to be fully automatic or offer any kind of defect classification capability, in contrast to several detection approaches reviewed below. Rather the DFD is expected to provide a skilled human operator with a novel sequential analysis interface: the software will suggest component regions for further manual review ranked by a displayed indication severity metric computed based on the data fusion. All the conventional data displays and analysis tools, such A-scan cursors, will be available to the operator, relating to the vicinity of the suggested regions. In principle the software will continue returning component regions until the set of returned regions makes up the entire component. However, in practice the human operator will review the first few by traditional means, and having satisfied him-/ herself that the most serious indications are benign, will terminate the analysis. Therefore this interface leaves all sentencing decisions to the operator but provides an advanced analysis aid to focus the operator's attentions on the component regions where his or her skills are best applied. In addition to significant time (and

hence cost) savings, such a system offers improved inspection reliability, both because the data fusion is more rigorous and comprehensive than any simplistic analysis scheme and because there is less scope for fatigue and human error.

In simple practical terms, the data fusion system must distinguish between noise and flaw signals (Section 6.1.1), and must do so across all channels (Section 6.1.2). There are multiple possible approaches for combining signals (Section 6.1.3), but for a probabilistic approach independent samples must be obtained from the recorded amplitudes (Section 6.3). Furthermore, because data in different channels may for example have been collected with different gain settings, some form of channel-wise normalisation is required. Converting amplitudes to probabilities within each channel is a convenient means of achieving that (Section 6.5). These probabilities can be merged across channels using statistical methods to then assign a single probability to any location in the data acquisition domain (Sections 6.6 & 6.7). Comparing these fused probabilities calculated for many candidate locations allows those locations to be ranked (Section 6.8), giving the expected output of the DFD.

6.1.1 Signal mixture model

The widely-used signal mixture model offers a simple (single-channel) representation of the problem that we seek to address and is illustrated in Fig. 6.1. The basic premise, in the terminology adopted here (which differs from that used in communications theory), is that the recorded signal is described by a distribution that is the weighted combination of distributions for the noise and for a flaw of a specified type. These component distributions, expressed in terms of conditional probabilities (e.g. $p(\text{signal} \mid \text{flaw})$ is the probability distribution of signal given that it is caused by a flaw), are not available in practical data analysis. The objective of the analysis is to determine whether a particular recorded signal is the result of noise, or more importantly, a flaw. The traditional approach is to threshold the signal amplitude, declaring signals that fall above to be due to a flaw and below due to noise. As shown in the figure, the probability of detection (POD) and probability of false alarm (PFA) may be computed for a given threshold, and sweeping this threshold also allows the receiver operating characteristic (ROC) to be built up. [86–88]

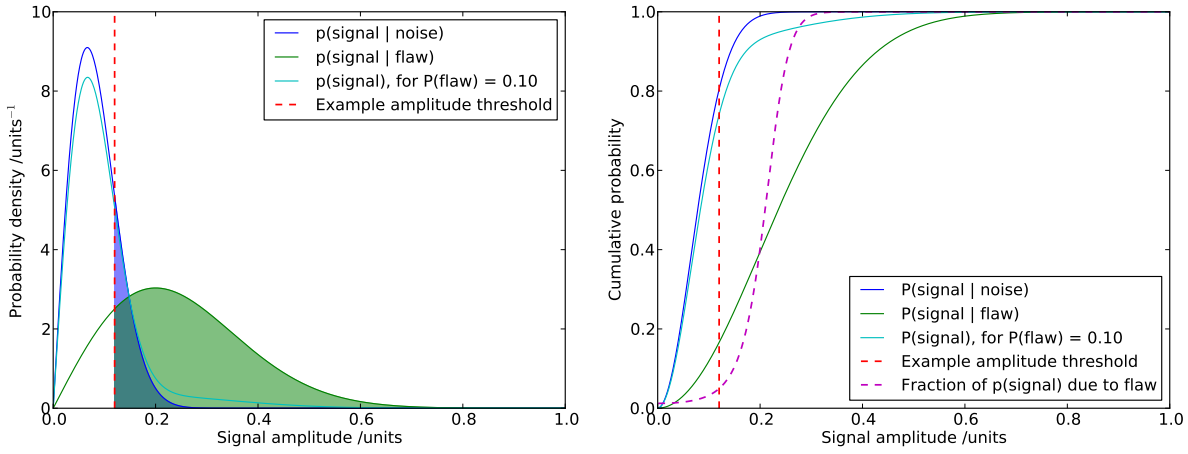


Figure 6.1: A signal mixture model. In (a) probability density functions (p.d.f.s) are shown for the noise distribution, the flaw distribution and the overall signal distribution, formed as a weighted sum from the two other distributions given the flaw prior. Plot (b) shows the corresponding cumulative distribution functions (c.d.f.s), and the fraction of the signal explained by the flaw distribution. Both plots also feature an example of an amplitude threshold, by which signals are traditionally taken to originate from a flaw if they fall above the threshold amplitude. For a given amplitude threshold the probability of detection (POD) and probability of false alarm (PFA) may be computed by integrating the p.d.f. $p(\text{signal} \mid \text{flaw})$ and $p(\text{signal} \mid \text{noise})$, respectively, from the threshold up as shaded in (a) - and so these metrics may be read off (b).

The approach described for identifying flaw signals is in many ways unsatisfactory. Firstly, even though there are algorithms for the choice of threshold value [57], this choice is rather arbitrary, and in general the selection is not probabilistic, with little regard for the consequences in terms of POD and PFA. Secondly, there is the fact that this a form of binary classification, when a fuzzy class membership [89, 90] would allow the uncertainty to be taken into account and be more helpful for ranking signals - consider for example replacing the step-function-like threshold with a smoothly varying curve such as the fraction of the signal explained by the flaw distribution shown in the cumulative distribution mixture model plot Fig. 6.1b. Additionally, focussing on raw, i.e. as-recorded, amplitudes neglects a lot of information about the data acquisition that has the potential to improve the chances of a correct classification (see also next subsection).

We therefore desire a classifier system that effectively expresses signal membership of the noise class as a probability, in the [0,1] interval. In view of the desired output from the detector, and the wish to minimise the number of assumptions required, the classifier

to be designed is of the *one-versus-rest* form, where no attempt is made to distinguish between the signals from different types of flaw. [82]

6.1.2 Multi-channel data

In addition to the single-sample evaluation considerations described so far, we wish to exploit the fact that in automated inspection each part of the component is typically interrogated several times. The benefits of the data fusion this implies are both intuitively obvious and well-documented. For example, Horn and Mayo [91] consider the improvement in NDE reliability by combining different types of inspections, and the results show that significant reliability gains can be achieved by even relatively simple means of combining two independent sets of test data.

Therefore the spatially coincident sections of diverse amplitude fields, from different data channels whose domains intersect a region of interest, should be combined in a logical and general way to aid the region's classification. Different types of combination, depending on field modality, need to be considered. Firstly, there is the combination of field sample values within a single channel, especially relevant as any real defect is expected to be visible in more than one probe position even within a single channel, given the dimensions of any physical defect and the probe's beam spread, for example. Then there is the combination of channels of differing modality (e.g. using different types of ultrasonic probe, or even using a completely different technology, such as eddy-current), acquired during one and the same inspection. Finally, there is the incorporation of field samples from a procedurally identical, but previous, inspection, which then moves the processing into the realm of baseline subtraction [51, 52] and change detection [92].

Data fusion literature defines different levels at which fusion may take place: broadly, data-level, feature-level and decision-level [93,94]. The first refers to combining raw amplitudes, the second to extracting compatible features and combining these, and the last, the highest level, relates to combining the decisions reached by analysing different data channels independently. While all three levels have applications, in general, higher levels result in coarser results than lower level processing, but permit savings in data handling volumes. There is also one critical restriction to bear in mind in deciding at which level

to undertake fusion: data-level processing requires amplitude values to be compatible - for example subtracting amplitudes generated by sensor systems based on different technologies is largely meaningless. Mono-modal datasets, that is to say of identical type (for our purposes, identical ultrasonic probe specification - so not all ultrasonic channels are considered to be one modality), are compatible after registration and may therefore be processed at the data-level, in raw amplitude form. Conversely, multi-modal datasets, for example containing data from shear and near focus compression ultrasonic probes, are not compatible and must be fused at a higher level, preferably the feature-level. What the project ultimately implements is a hybrid scheme, utilising data-level processing where possible and moving to the feature-level to overcome differences in channel modality.

6.1.3 Possible approaches

There is a vast array of data fusion algorithms, especially in the context of military applications [95]. However, a more limited range has so far been trialled on NDE data, in part because some of the most complex considerations in military uses are not relevant - surveys are provided by [94,96]. Here we mention some of the more prominent possibilities, disregarding those based on artificial intelligence, in line with the project objectives (see Chapter 1).

The simplest approaches rely directly on fundamental arithmetic operators such as addition, or taking the mean or maximum, operating on raw amplitudes or extracted features [97–99]. While these techniques can offer some benefits for some inputs, they lack theoretical grounding and are not probabilistic. Moreover, their performance is easily compromised by a single poor data value (perhaps due to that acquisition having been incorrectly started) in the inputs. Voting methods for decision fusion [100], and related fuzzy combination possibilities [90], are similar, with the additional problem of implicitly relying on the specification of thresholds.

Bayesian inference is based on using *Bayes' theorem* to compute the posterior probability $P(\text{flaw} \mid \text{signal})$ [98,99]:

$$P(\text{flaw} \mid \text{signal}) = \frac{P(\text{signal} \mid \text{flaw}) \times P(\text{flaw})}{P(\text{signal})} \tag{6.1}$$

While the output quantity is a useful probability, given that in practice only $P(\text{signal})$ is available and $P(\text{signal} \mid \text{flaw})$ is specific to every type and size of possible defect, these flaw distributions and priors, $P(\text{flaw})$, would need to be assumed based on external information [98, 99], a situation we wish to avoid in line with the project objectives.

Another approach offering attractive theoretical capabilities but suffering from practical obstacles is *Dempster-Shafer* fusion [91, 98, 99, 101]. This is a generalisation of *Bayesian* inference, for instance allowing several propositions to be considered simultaneously, and accounting for lack of knowledge (ignorance). At the heart of the approach is *Dempster's rule of combination*, operating on probability masses, which are related to probabilities but also applicable to sets of events. Unfortunately, the technique suffers as there is no universal way to compute the mass probabilities from the available sensor inputs with typical approaches relying on numerous assumptions, and even a small change in the values can have a dramatic effect on the output [94, 96].

Classical inference can be considered a simplification of *Bayesian* inference, based on hypothesis testing. A null hypothesis H_0 is defined alongside the antithetical alternative hypothesis H_1. A probability is assigned to a random variable according to how likely the value observed, x_a, or one more extreme (i.e. higher), is to occur under the null hypothesis H_0 data model for the random variable. This is the *p-value* of the random variable under the distribution, equal to the integral of the p.d.f. over the interval (x_a, ∞), i.e. $P(x_a \mid H_0) = 1 - P_{cdf}(x_a)$, as illustrated in Fig. 6.2. The *p-value* may also be expressed in terms of the survival function, $P_{sf}(x_a)$, where $P_{sf}(x) = 1 - P_{cdf}(x)$, which then gives the *p-value* directly as a function of the random variable. The definition of a *p-value* means that this statistic can serve as a measure of the extent to which the observed value can be explained by the data model, which for our purposes will essentially be a noise model. Usually, a significance level, typically $\alpha = 5\%$, is used to threshold the *p-value*: if the computed value lies above the threshold, this is deemed not to refute the null hypothesis, H_0, whereas a value below is interpreted as a rejection of H_0 (e.g. "noise") and the alternative hypothesis H_1 (e.g. "non-noise", i.e. "flaw") is deemed to apply. Note that the significance level in effect specifies the expected false-call rate. [95, 96, 102]

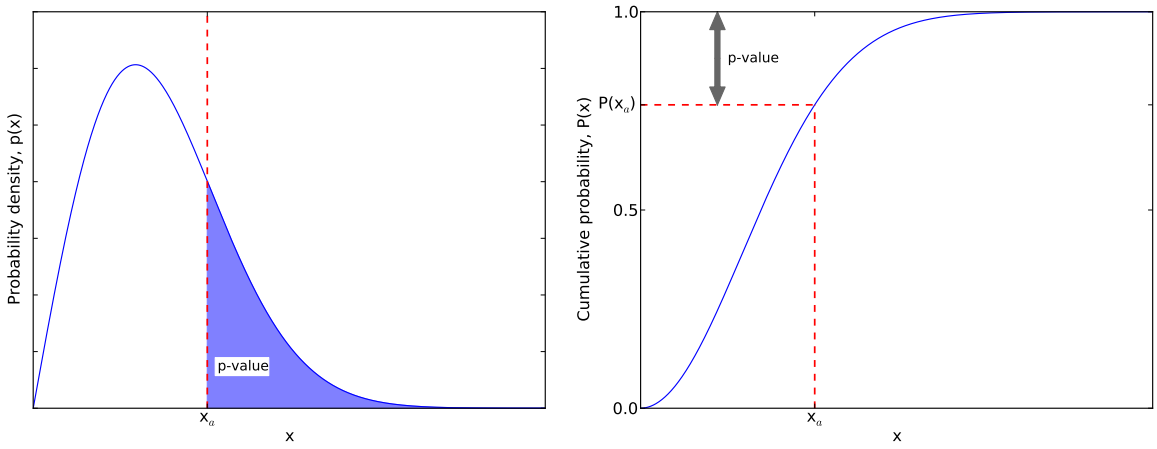


Figure 6.2: The *p-value* of an observed random variable x_a under the distribution of the variable x, in (a) as an area under the p.d.f., and (b) as seen on the corresponding c.d.f..

6.2 Channel-Specific Processing

As explained in the chapter introduction, for data from a given channel modality, data-level processing is possible, before the move to the feature-level to overcome differences in channel modality. Therefore, within each individual channel useful processing steps specific to each modality, for instance taking into account different probe effects, may be applied prior to cross-channel fusion. The processing may take the form of image processing algorithms, such as noise filtering or the Synthetic Aperture Focussing Technique (SAFT). An implementation of the SAFT algorithm for the sort of single probe data considered in this project is described in Appendix B, providing one possible way of incorporating a beam model in each channel. If there is data from an equivalent channel in a previous acquisition available, registered to the data in the current channel, then this within-modality processing allows the historic data to be incorporated through baseline subtraction [51,52] or image change detection [92] approaches. Note that subtraction-type processing should take place after techniques such as SAFT are applied to the individual inputs, as defining an appropriate beam model may not be possible for the subtracted field, and in case the image processing is non-linear. The currently implemented scheme is simplistic differencing of aligned (by registration) mono-modal data channels, which should at least provide a baseline output given that more complex algorithms have the

potential to give superior performance.

Having completed all channel-specific processing on the data channels independently, to combine fields from different probe types feature-level fusion becomes necessary. Given that the sought output of the data fusion processing is a probability, as will be explained, the feature chosen for this is a set of probabilities.

6.3 De-correlation

6.3.1 Motivation

The main source of inspiration how to process amplitude fields in terms of probabilities are the statistics used in functional Magnetic Resonance brain Imaging (fMRI), as described for example by the edited book [102]. In this area of medical research, multi-dimensional fields of MRI data are analysed to identify "activated" regions of the brain under various experimental conditions, to make deductions about brain function. The statistics developed and applied in this field also permit for example complex models to be hypothesis tested in the presence of confounding effects. While much of the work is inevitably entirely specific to the analysis of brain activity, some of the more basic processing is relevant, as the input has similarities to this project's data.

One of the initial problems that the fMRI literature encounters in attempting to apply statistics to amplitude fields is that (with exceptions) statistics requires data samples to be independent, identically distributed (i.i.d.) [103]. This is absolutely not the case for the various amplitude field samples that are the starting point for the analysis in the project, as within each channel there will be significant correlations in samples at proximate locations, which mean the samples can not be considered independent. It should be stated however, that samples drawn from different data channels are considered statistically independent for the purposes of this analysis, so what follows in this section refers to samples from a single channel.

The described correlations between proximate locations vary in magnitude and causes with direction. Thompson et al [104] showed that in the probe displacement direction(s), the extent of overlap of the insonified volume for adjacent probe locations and the sample

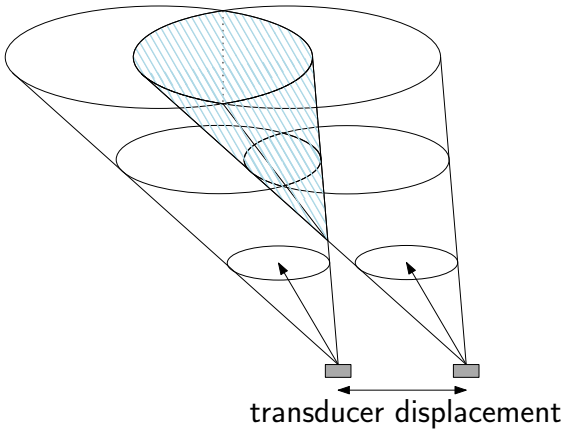


Figure 6.3: Schematic of the overlap of two conical ultrasonic beams, emitted at a slight displacement from each other.

microstructure are the main parameters controlling spatial correlations. The former varies with depth, controlled by the beam spread of the probe and the probe step size between A-scan acquisitions (see Fig.6.3), while the scale of the latter compared with the inspection wavelength determines the nature of the interaction of the sound with the medium and specifies how much averaging of grain contributions occurs over an insonified region. On the other hand, in the time-axis or depth direction the pulse length (or equivalently, bandwidth) together with the A-scan sampling frequency and material microstructure govern the extent of sample correlation.

In addition to the mentioned factors contributing to sample correlations in the data acquisition, subsequent signal processing can introduce further sample dependencies. For example, envelope detection will necessarily add considerable correlations to samples along the time-axis. The effect of processing by techniques such as SAFT (see Appendix B) is less obvious, but SAFT is recognised to exploit multiple A-scans over an aperture to enhance the axial resolution, hence potentially reducing some sample dependencies, but in an aperture-dependent way.

These correlation considerations also have direct applications, in the specification of optimal inspection sampling procedures, to ensure coverage is assured and independent data maximised but redundant data minimised [105]. In practice, for confidence in the detection of small reflectors that fall exactly between two sample locations and so as to preserve the capability of averaging out random noise without the loss of independent information, slight oversampling is desirable.

6.3.2 Method

Having established that the available data consists of varyingly correlated samples, due to a number of contributing factors, for conventional statistics to be applicable, steps must be taken to convert the samples to what might be considered independent, identically distributed (i.i.d.) samples. A rather empirical approach is taken that avoids the need to untangle the various contributions and compensate for them individually. First, the extent of sample dependence is estimated from the data by taking the autocorrelation of the samples in different directions in its index space, taking care to exclude high amplitudes from prominent reflectors, as one of the defining features of signals from reflectors of interest is that they have a correlation differing from the background. The index space arrangement of sample values is used rather than the spatial arrangement, not only as this is computationally attractive, but because the index space axes are closely aligned with the axes along which correlations are to be compensated for. For each autocorrelation a characteristic correlation length is defined from the width of the peak around the zero relative shift position. Then, borrowing from fMRI data analysis, the sampled field of voxels (individual data values, each assigned to a small region of 3D space), which is itself a discrete representation of a continuous physical field, is further discretised. The scale of discretisation in each dimension is given by the maximum of the correlation length in that dimension and the estimated registration error, such that a poor registration accuracy (see Chapter 5) will compromise the resolution of the subsequent fusion processing. Note that advice from fracture mechanics about the grouping of possible defects could potentially also be incorporated into the calculation. The formed multi-voxel chunks of space are referred to as resolution elements, or *resels*, representing a single, approximately independent sample [102, 106, Part 4]. Figure 6.4 illustrates the difference between *resels* and the *subsets* referred to in Section 3.5.5.

6.3.3 Resolution elements (resels)

The exact definition of the *resels* is complicated by the, in practice non-trivial, determination of the correlation lengths to use. In this work the correlation length was defined as the full-width at half-maximum (F.W.H.M.) of the average autocorrelation function of

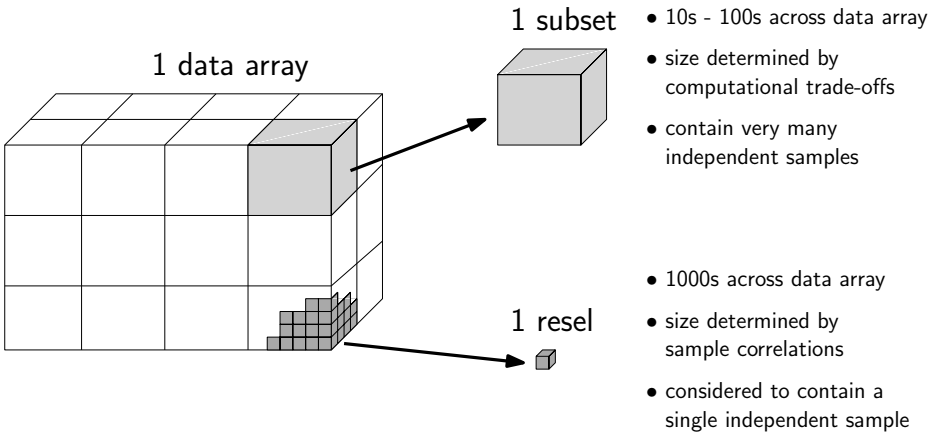


Figure 6.4: An index-space illustration comparing how the data array may be broken down into either *subsets* or resolution elements (*resels*). Both subdivisions are constrained by integer arithmetic so the sizes may vary slightly over the array to fit a whole number of subdivisions across the number of samples available in each array dimension. Note not all *resels* on the visible faces of the array have been drawn.

the data in each index-space dimension, as illustrated in Fig. 6.5 [102].

This is a crude approximation as a result of the difficulty of computing meaningful correlation coefficients on envelope detected, and hence entirely non-negative, signals, where normalisation introduces artefacts. The adopted procedure could probably be improved by incorporating alternative means of assessing signal independence, such as those used in some data reduction techniques [83]. The constraints of integer arithmetic, as in practice an integer number of elements have to be fitted across an integer number of samples in each dimension of the data array, also impact upon the application of the chosen correlation length, although requiring the edge of the dataset to be a *resel* edge is physically an arbitrary choice. Furthermore, in the current implementation the *resel* size in each dimension is fixed across the index space dataset. Even though this corresponds to radially increasing *resel* spatial sizes (and hence decreasing *resel* density see Fig. 6.6), it is accepted that this will lead to rather imperfect de-correlation. An implementation using spatially varying index space *resel* sizes is suggested as further work. Nonetheless, the current implementation appears sufficient for present purposes, given that the correlation length is in effect only a lower bound on the size of *resels* to provide adequate de-correlation, though using larger *resels* incurs a spatial resolution penalty and reduces the number of samples available for subsequent statistical calculations.

The value assigned to each *resel* is computed from the contained voxel values, reducing

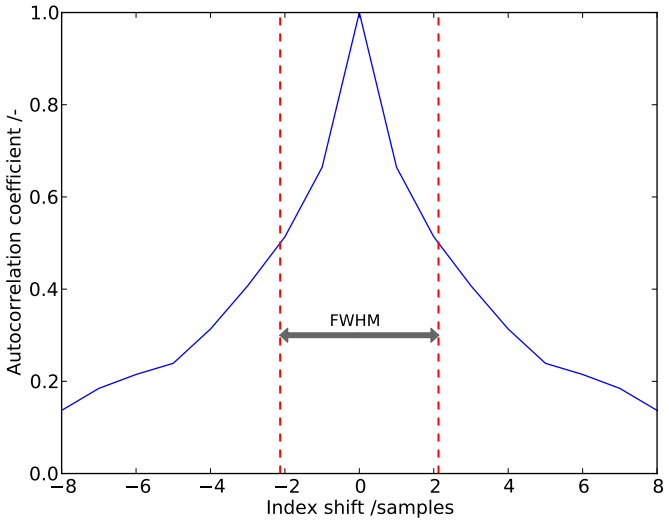


Figure 6.5: Deducing the correlation length from an autocorrelation function, using the definition provided by the full-width at half-maximum (F.W.H.M.). Note the function is necessarily symmetric about the zero shift position.

these to a single representative value - here the maximum is used. This quantity foregoes the reduction in random noise that might be possible by averaging, but it has the advantages of being conservative and compatible with traditional amplitude gate-based perspectives. At first the reduction of multiple voxel values within a *resel* to a single number may seem like gratuitous discarding of data, but in theory this procedure will merely discard sample values that provide only duplicate information.

The density (the reciprocal of volume) of *resels* across the component is of interest for examining the distribution of statistically independent samples and potentially using this to optimise inspection procedure in terms of the density of raw samples acquired. Moreover, considering this density across multiple channels, taking into account each coverage area, indicates where the scope for data fusion is greatest, and indeed, least. Figure 6.6 provides graphs of these variations within- and cross-channel for the rotor bore inspection.

Having reduced the amplitude field of each channel to a large number of approximately i.i.d. amplitude samples, the next step is to combine those *resels* within- and cross-channel falling inside a defined spatial region of interest to ultimately yield a single, useful probability for that region.

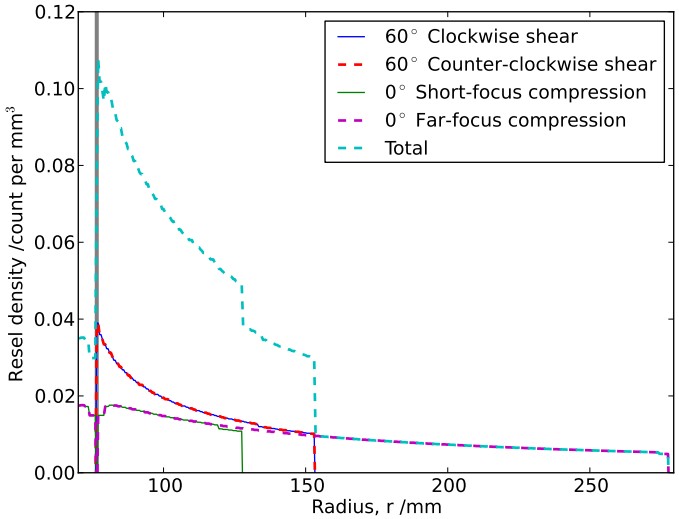


Figure 6.6: The resolution element density along a radial line from the centre of the component, based on the currently implemented constant (index-space) *resel* size. The density is shown separately for each of the available ultrasonic channels (aligned radially according to their nominal positions), together with the effective cross-channel total. Density variations observed in the separate channels are the result of the *resel* volume varying due to the complex scan geometry. Between channel variations are due to differences in probe type and channel spatial domain or coverage. The vertical line in the plot demarcates the radius of the component, such that the signal before that is before the front wall, in the ultrasonic wedge, and can safely be ignored.

6.4 Combination Considerations

For (feature-level) fusion of *resel* values identified as falling within a chosen spatial region of interest across all data channels, the *resels* must be expressed in terms of a compatible feature. The sought output of the data fusion processing is a probability, so probabilities are the feature of choice, such that the feature-level fusion may be performed using relatively conventional statistics, operating on nominally i.i.d. inputs. The probabilities used must represent the significance of each *resel* within its channel, so that the feature is invariant to differences in amplitude scale between channels, for example.

6.4.1 Classical inference & hypothesis testing

The approach adopted is based on classical inference and hypothesis testing, as outlined in Section 6.1. While classical inference has significant limitations, for instance by not being able to directly provide evidence for the null hypothesis, it makes fewer assumptions and has greater intuitive appeal than for example *Dempster-Shafer* approaches (see Section

6.1.3), and should prove adequate for our *one-versus-rest* classification scenario [95, 96, 102]. The simplicity of classical inference is especially appealing as [91] suggests that the simple fusion methods already offer good performance, and that while complex techniques may give further gains, these have to be weighed up against increased computational effort and the need to make further assumptions.

The basic classical inference introduced in Section 6.1.3 may be extended to multivariate data. In the presence of k independent tests, all the random variables recorded may be converted to *p-values*, $q_1, \ldots, q_k$ on the basis of null hypotheses $H_{01}, \ldots, H_{0k}$, and examined from the perspective of a universal, or family-wide, null hypothesis H_0, which states that all of $H_{01}, \ldots, H_{0k}$ are true. The corresponding universal alternative hypothesis H_1 is that at least one of the component alternative hypotheses is true [107]. Defining a universal null hypothesis is common practice in biological meta-studies [108] and fMRI [102]. However, there are circumstances under which doing so before applying a significance level threshold may either not be desirable or strictly valid, for example interesting results might be disregarded as deemed to have arisen "by chance" [109]. For the purpose of the data analysis here this procedure provides the justification for converting the *resel* values to *p-values* and then evaluating these together - and because no significance level is specified *a priori*, such criticisms of these procedures are irrelevant.

The data model to apply in the conversion to *p-values* is developed in Section 6.5. Having incorporated all channel-specific effects into the calculations of these probabilities, these *p-values*, regardless of channel of origin, are now compatible features, and may be combined together. This collection of *p-values* is then finally reduced to a single output probability, as discussed at length in Section 6.6.

6.5 Data Model

6.5.1 Null hypothesis

As explained previously, the objective of the data model is to facilitate the conversion of *resel* amplitudes to *p-values*. The critical characteristic of *p-values* is that under the null hypothesis, these values are uniformly distributed over the interval [0,1]. This property

follows from the definition of the cumulative distribution function, c.d.f. (and hence survival function, s.f., equal to 1-c.d.f.): samples drawn from an arbitrary p.d.f. will be mapped to U(0,1) by the c.d.f. (or s.f.) of the distribution (see graphical definition of *p-value* in Fig. 6.2). An alternative explanation is that to generate random variable samples from an arbitrary p.d.f., samples from U(0,1) may be passed through the distribution's inverse c.d.f., mapping the U(0,1) samples to the arbitrary distribution's random variable space - so the reverse transformation is provided by the c.d.f.. The null hypothesis assumed for each *resel* is that it is a random sample from the data model distribution, which must be defined.

6.5.2 Local data modelling

Initially, an attempt was made to define a global model for the data of each channel. For instance, one might conceivably define a "characteristic noise region" (by picking a region free from features and asserting that the signals are "typical" noise), and use an associated distribution to assess amplitudes throughout the component in terms of their compatibility with that noise model. However, in practice this is entirely unsuitable for many data channels due to systematic variations of noise behaviour across a scan. The problem is illustrated in Fig. 6.7, where over the length of every A-scan there is a transition from a low noise region to a very noisy region. The complexity of this data cannot meaningfully be represented by a single distribution that implicitly assumes some degree of uniformity. A global distribution based on some "characteristic noise region" would additionally fail to provide uniformly distributed *p-values* even under the null hypothesis (i.e. drawing random *resel* samples).

Therefore the decision was made to use local data modelling, and exploit symmetries of the scan to define an environment of similar *resels* for each *resel* location. In the two inspection types studied, rotational symmetry features prominently and it is reasonable to assume that *resels* in the same ring about the axis are in the same environment. Importantly, large circumferential (spanning a substantial fraction of the circumference) defects are unheard of in these applications, so the reduced sensitivity to signals of such defects is not a problem.

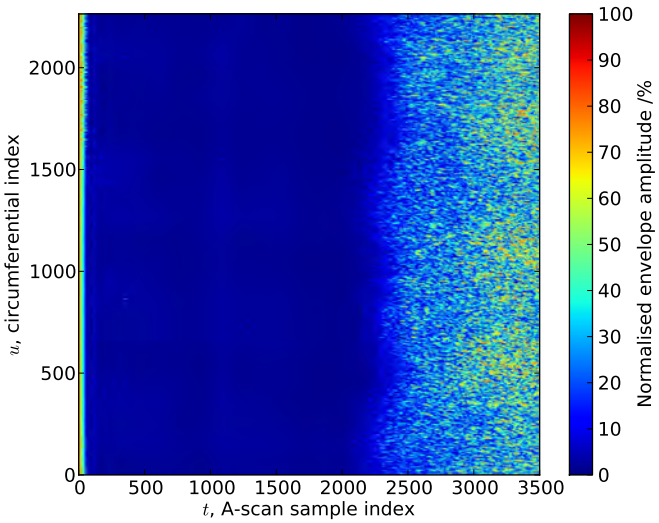


Figure 6.7: Distance-amplitude corrected B-scan image corresponding to data recorded during one rotation of a turbine disk from the aerospace application introduced in Section 2.3, with the probe scanning radially inwards. On the left side the front wall echo is visible - the water path is not shown. Towards the right side, at higher A-scan sample indices, corresponding to longer times of flight and greater depths, there is a marked increase in coherent noise.

Assessing each *resel* in terms of its local environment in this way can be thought of as normalising these amplitude samples circumferentially, with the chosen distribution within each hoop mapping the amplitudes to the interval [0,1]. Within each local environment, for our purposes each *resel* ring, there will in general be wide range of *p-values*. Note that this scheme is extremely general, and can be adapted to suit a range of component geometries and inspection types. Even in the case of highly complex geometries lacking obvious symmetries, such as aerospace fan blades, the same principles can be applied by considering each *resel* in terms of the set of *resels* at an equivalent position across the large number of practically identical parts available.

6.5.3 Discrete data

In considering how to construct a data model distribution from a collection of local *resel* amplitudes, it is necessary to bear in mind that this collection of amplitude samples is discrete, both by having been quantised during the digital recording process by an analogue-to-digital converter, and by virtue of the fact that only a finite, integer number of samples is available. Both these contributions are almost universally applicable to

the processing of digital data, though in some cases the level quantisation may be lost by subsequent data processing, for example performing envelope detection on digitally recorded full-waveform A-scans. It will be shown that these forms of discretisation can invalidate the application of statistics that at least implicitly assume continuous inputs.

As has been established, the uniform distribution (of *p-values*) over the interval [0,1] is central to hypothesis testing. The continuous version of this distribution is shown in Fig. 6.8a, and this is the only possible probability density function (p.d.f.). However, considering discrete versions of U(0,1), it soon becomes apparent that there exists an infinite number of distinct possibilities, defined by the discretisation levels available. For instance, [0,1] may be split up into any integer number of levels of regular spacing, as exemplified in Fig. 6.8b, where these is still a strong similarity to the continuous version. However, the interval may alternatively be split into an irregularly spaced set of levels, demonstrated in Fig. 6.8c, where the visual similarity to the continuous distribution is practically lost. However, the form of the cumulative distribution function (c.d.f.) shown in Fig. 6.8d, with points lying on a straight-line between the origin and (1,1), is common to all possible versions of U(0,1), continuous or discrete, and so this will be used as the defining property of any *p-value* distribution considered.

One further property of discrete distributions of note is that, whereas in continuous distributions the probability of occurrence of any specific choice of random variable is infinitesimal, this is not true for discrete distributions. There, the probability of occurrence of any of the available random variable discretisation levels x_i is specified by the corresponding probability mass function (p.m.f.) value $p(x = x_i)$, and the probability is identically zero elsewhere. This means great care must be taken in defining the c.d.f., as there is a difference between a definition based on $p(x < x_a)$ and $p(x \leq x_a)$, where x_a is the specific choice of random variable at which the c.d.f. is being evaluated. Moreover, when defining the survival function from the c.d.f., it must be noted that for discrete x, $p(x \geq x_a) \neq 1 - p(x \leq x_a)$, as $p(x = x_a) \neq 0$.

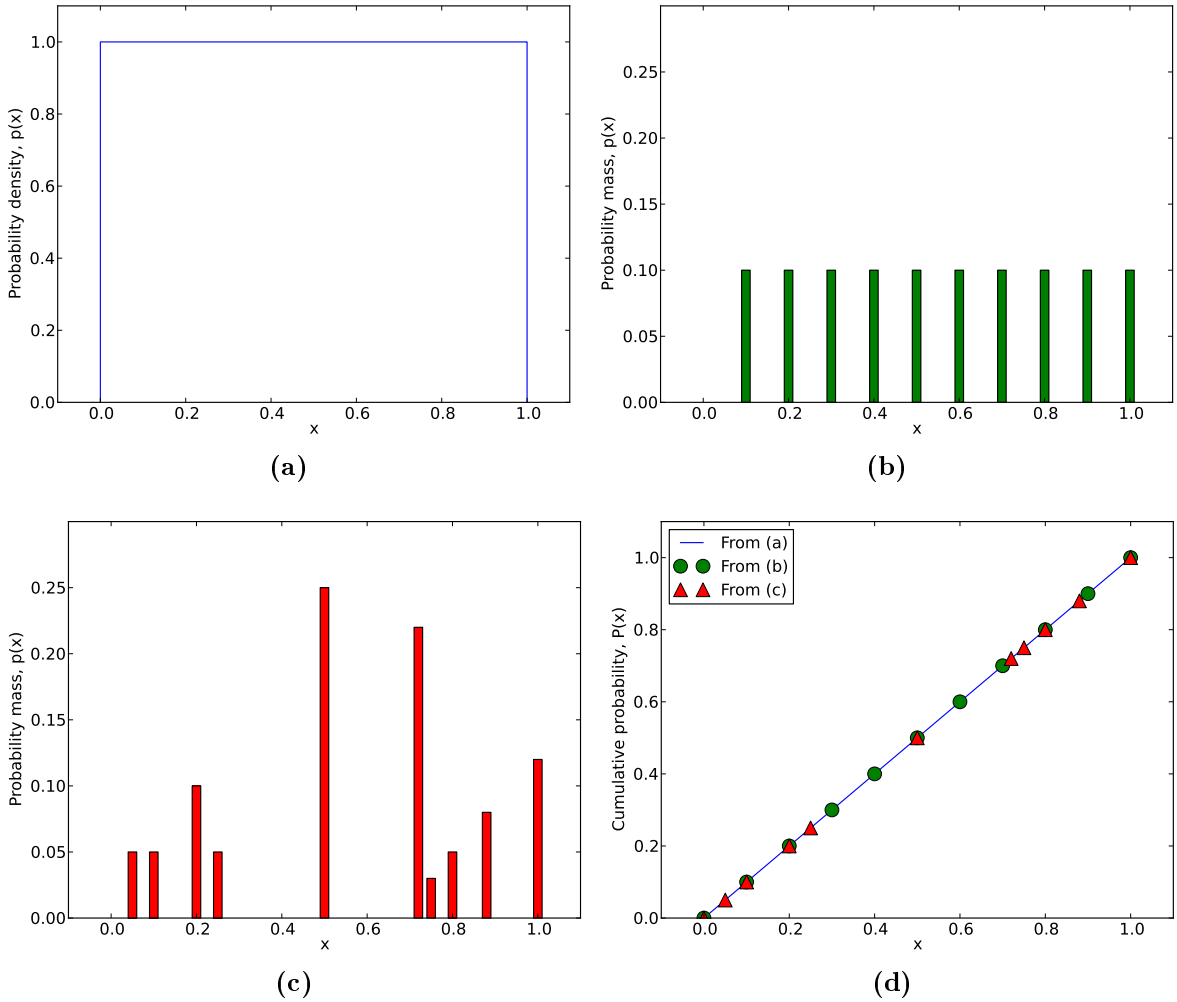


Figure 6.8: Uniform distributions over [0,1] interval, U(0,1), as a function of a random variable x. In (a) the continuous probability density function (p.d.f.) is shown, (b) shows the probability mass function (p.m.f.) of a possible discrete version of the distribution with ten quantisation levels, and (c) presents an alternative p.m.f. for a different set of ten discretisation levels. Plot (d) provides the cumulative distribution function (c.d.f.) for the previous three distributions: this straight-line c.d.f., between the origin and (1,1) is common to all versions of the uniform distribution over the interval considered, as indicated by the way the points derived from (b) and (c) map onto the line derived from (a).

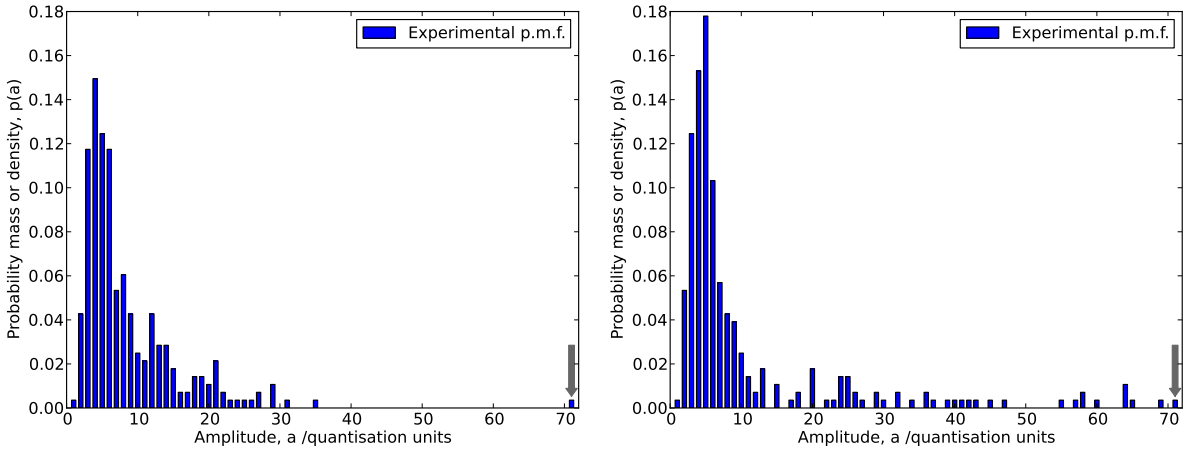


Figure 6.9: The amplitude distributions (p.m.f.s) found in two circumferential rings of rotor bore *resel* data, taken from the same data channel and acquisition. Note that the amplitude values are quantised due to the digital signal recording used and the probability mass values are quantised by virtue of the fact that these are experimental distributions, built up from the limited number of samples available in a ring of *resels*. Whilst the sample at the highest amplitude (indicated by arrow) is both identical in amplitude and of the same frequency across the two plots, differences between the remaining distributions should be reflected by the data model's assessment of those samples. Intuitively, the likelihood that the highest amplitude sample can be explained by the remainder of the distribution (which may be considered noise, using that term in a broad sense) is significantly higher for (b) than (a).

6.5.4 Semi-parametric model

The data model is required to yield the expected null hypothesis behaviour of a uniform U(0,1) output when drawing random samples. The model should also retain information about relative amplitudes, and hence the effective signal-to-noise ratio. The significance of this property can be explained using Fig. 6.9: the model should reflect the differences between the two distributions, even when assessing the highest amplitude sample in each, and despite the amplitudes (and probability mass) of those samples being equal. Intuitively, the probability that the highest sample can be explained by the remainder of the distribution is lower for Fig. 6.9a than Fig. 6.9b, as in the former there is significant separation between that sample and the remaining distribution, unlike in the latter.

Satisfying both requirements simultaneously is difficult, and as described in Appendix C, several possibilities for modelling the data were found to be wanting. The model described here and ultimately adopted offers a very good compromise between objectives,

exploiting the fact that the sensitivity to relative amplitudes creates the greatest concern at the highest amplitudes where the signal most likely reflects a cause other than noise, however the vast majority of samples will lie well below such extreme values. This majority of samples, up to a (high) amplitude threshold, can be fitted to a discrete distribution to assure conformity with the expected output under the null hypothesis, and the above-threshold samples can be represented separately, fitting a suitable analytic distribution to this tail. This sort of tail fitting can be facilitated by a generalised *Pareto* distribution [110,111]. The c.d.f. of such a distribution is specified by:

$$F_{\xi,\sigma}(x) \quad = \quad \begin{cases} 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi} & \text{for } \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\sigma}\right) & \text{for } \xi = 0 \end{cases} \tag{6.2}$$

where ξ and σ are parameters of the distribution, and $\sigma > 0$, $x \geq 0$ if $\xi \geq 0$, but $0 \leq x \leq -\sigma/\xi$ if $\xi < 0$.

An example of this semi-parametric model is shown in Fig. 6.10. The determination of the threshold (indicated by the dotted transition line in figure) at which to transition from one part of the model to the other can be critical, as setting the threshold too low invalidates the tail fit, but an excessively high threshold leaves very few points against which to map the continuous distribution. In the current implementation the threshold is broadly set to capture the top 5% of the samples in the tail (so the transition line in Fig. 6.10 is seen at a cumulative probability of 0.95), though this could almost certainly improved through a more advanced scheme [111]. Imposing a continuity constraint on the fitted curve at the transition point from the discrete distribution reduces the number of parameters to solve for in a maximum-likelihood optimisation.

In principle such a distribution provides the previously explained sought sensitivity to relative amplitudes in the domain of interest (the very highest amplitudes) at minimal cost to the validity of a uniform U(0,1) expected output under the null hypothesis. The two distributions of Fig. 6.9 allow the sensitivity to relative amplitudes to be demonstrated. Applying a purely discrete model (see Appendix C, Fig. C.1) to both to compute the *p-value* of the highest amplitude yields 3.6×10^{-3} in each case. Switching to the semi-parametric model gives the *p-values* 5.5×10^{-4} and 9.3×10^{-3}, for the distributions of Figs.

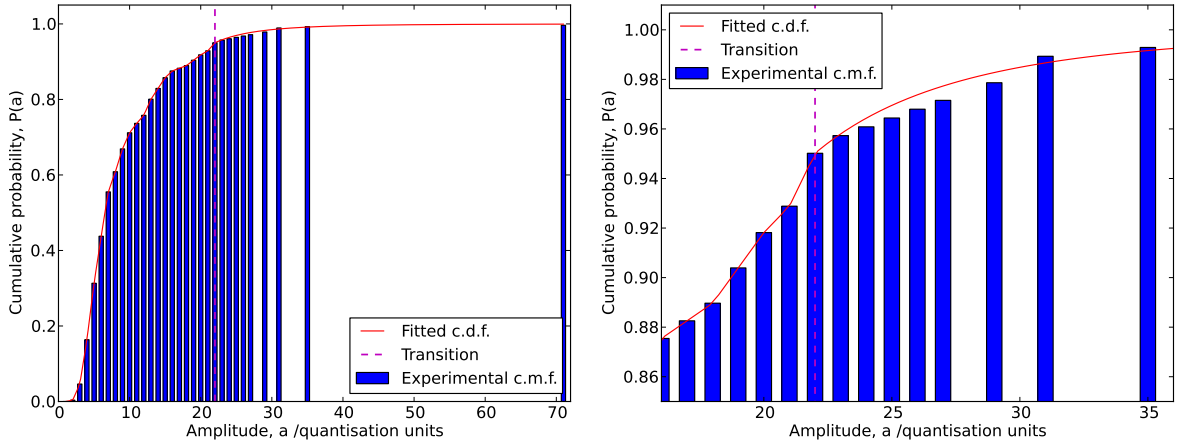
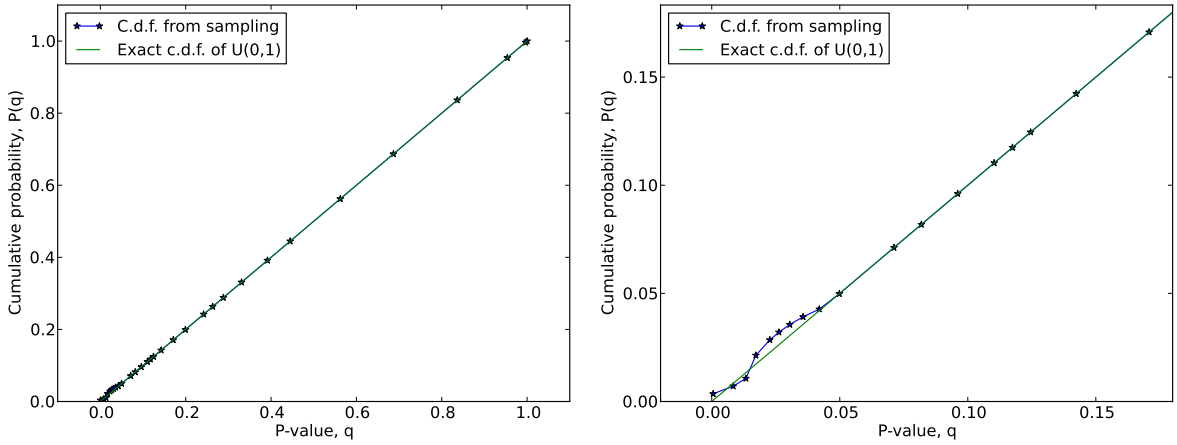


Figure 6.10: A semi-parametric model applied to the amplitude distribution found in one circumferential ring of rotor bore *resel* data (the cumulative version of the distribution presented in Fig. 6.9a). In (a), the distribution is practically indistinguishable from a purely discrete distribution (see C.1), whereas the zoomed version (b) reveals the key difference: up to the transition threshold the distribution exactly matches the discrete, piecewise-linear fit, but above a *Pareto* distribution is fitted to the upper tail of c.m.f.. This is seen as a smooth curve that no longer exactly matches the tops of the bars. The *p-value* at the very highest amplitude recorded is nearly an order of magnitude lower than in the equivalent purely discrete model (see Appendix C, Fig. C.1). Note that such differences in *p-value* are too small to be evident graphically, but matter greatly in the subsequent fusion statistics.

6.9a (see also Fig. 6.10) and 6.9b, respectively. Thus, the fact the amplitude in question is better separated from the remainder of the distribution in the former than the latter distribution is reflected in the *p-values* - the value for the former is more than an order of magnitude lower than for the latter. The result of testing the behaviour of the model under H_0 is shown in Fig. 6.11, where the match to the expected output distribution is perfect apart from a slight deviation at the very lowest *p-values* (corresponding to especially high amplitudes). This deviation is considered negligible (confirmed later, in Section 6.7.6) as it for example only means that a *p-value* of 0.02 has an associated cumulative probability of about 0.025 rather than exactly 0.02 - a difference easily buried in sampling effects. This semi-parametric model has therefore been demonstrated to provide the desired sensitivity to effective signal-to-noise ratio (at the highest amplitudes, where this matters, as in Fig. 6.10) at negligible cost to the validity of the null hypothesis expected output.

Note that this semi-parametric, local model makes no attempt to explicitly separate noise and flaw contributions in the received signal. The model is based simply on drawing



(a) **(b)**

Figure 6.11: Cumulative distribution function (c.d.f.s) for the output of the semi-parametric model shown in Fig.6.10. The full set of amplitudes that were used to generate the model were converted to *p-values* using the model, and the c.d.f. was determined for these. Plot (a) shows the full domain, while (b) a zoomed section at the origin. As should be expected for a discrete model, the samples follow the continuous U(0,1) c.d.f. exactly, except at the very lowest *p-values* (corresponding to especially high amplitudes) where there is a slight deviation, the result of the *Pareto* tail-fit.

attention to the most extreme amplitudes, which are the most likely to be the result of something other than noise, and exploits the fact that indications will only occupy a tiny fraction of the data volume, and hence *resels*. This data model will from now on be termed the Local Empirical Noise (LEN) model. It is used to convert all channels' *resel* amplitudes to *p-values*. How these *p-values* should be combined is the subject of the next section.

6.6 Combing p-values: Consensus Test

6.6.1 Introduction

To combine *p-values* relating to different independent tests there are so-called consensus, or family-wide, *p-value* tests available, all based on the fact that under the null hypothesis the *p-values* should be random samples from the uniform distribution U(0,1) distribution. As explained previously (see Section 6.4.1), this approach relies on the definition of a universal null hypothesis H_0, and a corresponding alternative hypothesis H_1. For our pur-

poses the former corresponds to the *p-values* of *resels* considered together in an identified spatial region of interest being random samples from U(0,1), and the alternative implies that they are not. Note that the approach hinges on detecting correlations between *resels* that are spatially proximate, typically due to different data channels interrogating the same spatial domain.

The general procedure underlying these consensus, family-wide, *p-value* tests is to form a test statistic from the *p-value* inputs, and then map that test statistic value to a (single) output *p-value* by invoking the c.d.f., or more exactly survival function (s.f.), of the test statistic. Conceptually, there can be multiple different consensus tests, all of which under H_0 receive as input multiple random samples from U(0,1) and map these onto a single U(0,1) output, because this requirement still leaves freedom of choice in the weighting of inputs. For example, in some tests *p-values* are able to "cancel symmetrically" about the 0.5 distribution mid-point (so that e.g. combining *p-values* 0.3 and 0.7 would yield 0.5), whereas in others low *p-values* are permitted more influence than the corresponding high values. For the data processing to be conservative, the test chosen must be of the latter sort, and not allow a single low *p-value* to be cancelled out in effect by an "opposite but equal" high *p-value* in the consensus test input. Note that AND and OR logical combinations bound the range of possible combination methods (see also Results, Section 7.2.1). [107]

Fisher test

One such test, widely used elsewhere and implemented as the primary consensus test is known as *Fisher's combined probability test* or *Fisher's method* [112]. While this test has been criticised as unsuitable for some applications [108], based on the analysis presented in [107] it is largely to be preferred over alternatives for the purposes of the project. The test forms a test statistic from the sum of the natural logs of the *p-values* in the collection, equivalent to taking the natural logarithm of the product of all the inputs [113]:

$$s_k = -2\sum_{i=1}^{k} \ln(q_i) \tag{6.3}$$

where s_k is the test statistic for k *p-value* inputs q_i. Under the null hypothesis, this quantity follows a chi-squared distribution with $2k$ degrees of freedom, $s_k \sim \chi^2_{2k}$, so the survival function appropriate to that χ^2 distribution may be used to convert the test statistic value to an output *p-value* q. The explanation for this is that twice the negative logarithm of a U(0,1) input follows a chi-squared distribution, and the sum of independent chi-squared variables is itself chi-squared, with a number of degrees of freedom equal to the total of the inputs' degrees of freedom. [114]

Tippett test

An alternative test examined is *Tippett's combined probability test* (Tippett is unrelated to Tippetts, the author's colleague [13]), which is based on taking the minimum of the provided *p-value* samples [113, 115]. The probability of no lower value existing under the null hypothesis is then computed, applying the definition of mutual statistical independence $P(\cap_{i=1}^{n}(1-q_i)) = \prod_{i=1}^{n}(1-q_i)$, so:

$$q = 1 - (1 - \min(q_1, \ldots, q_k))^k \tag{6.4}$$

using the same notation as before.

Comparison

The behaviour of these consensus tests may be illustrated graphically for an input of 2 *p-values*, by plotting the so-called rejection region, where the output falls below the specified significance level α. Such a plot comparing the two tests considered here is shown in Fig.6.12, illustrating some of the important properties of the tests, such as input commutativity. [107]

For the remainder of this chapter we focus on the *Fisher* test, as this is the test currently used, but the *Tippett* test is considered in detail in Appendix D. Of the two, the *Fisher* test is largely the test of choice as the output does not only vary with the minimum input value, thus providing greater distinguishing power.

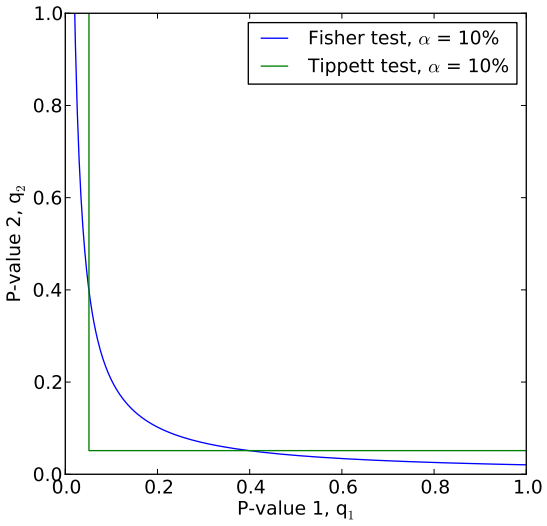


Figure 6.12: The rejection region at a 10% significance level for 2 possible consensus *p-value* tests with 2 input *p-values* - the null hypothesis is rejected in the region beneath and to the left of the curves. Note that both thus enclose 10% of the total plotting area, both are symmetric about the ascending diagonal and neither touches the axes so a high input is never able to eliminate the effect of a low *p-value* input.

6.6.2 Discrete data and the p-value discretisation problem

The discussion of consensus testing has, at least implicitly, assumed that the *p-value* inputs are samples from the continuous U(0,1) distribution under the null hypothesis, rather than from any discrete variant (discussed in Section 6.5.3). However, the *p-values* that result from the application of the local, semi-parametric model (developed in Section 6.5) to convert *resel* amplitudes to probabilities will be discrete. Not only that, but different *p-values* originating from different local distributions will be discretised differently, in general, with different numbers of irregularly spaced levels available.

Naïvely applying the consensus tests as described to discrete, rather than continuous, *p-value* inputs causes some remarkable, problematic distortions, illustrated in Figure 6.13. Here the output of the *Fisher* consensus test is examined, presented as a c.d.f. of consensus *p-values*, for inputs drawn from the different U(0,1) distributions presented previously, in Fig. 6.8. Whilst the continuous input case closely matches the expected null-hypothesis behaviour of a U(0,1) output, all the plots for discrete *p-value* inputs show severe deviations from this.

In practice such distortions render the test unusable for discrete inputs, as even under

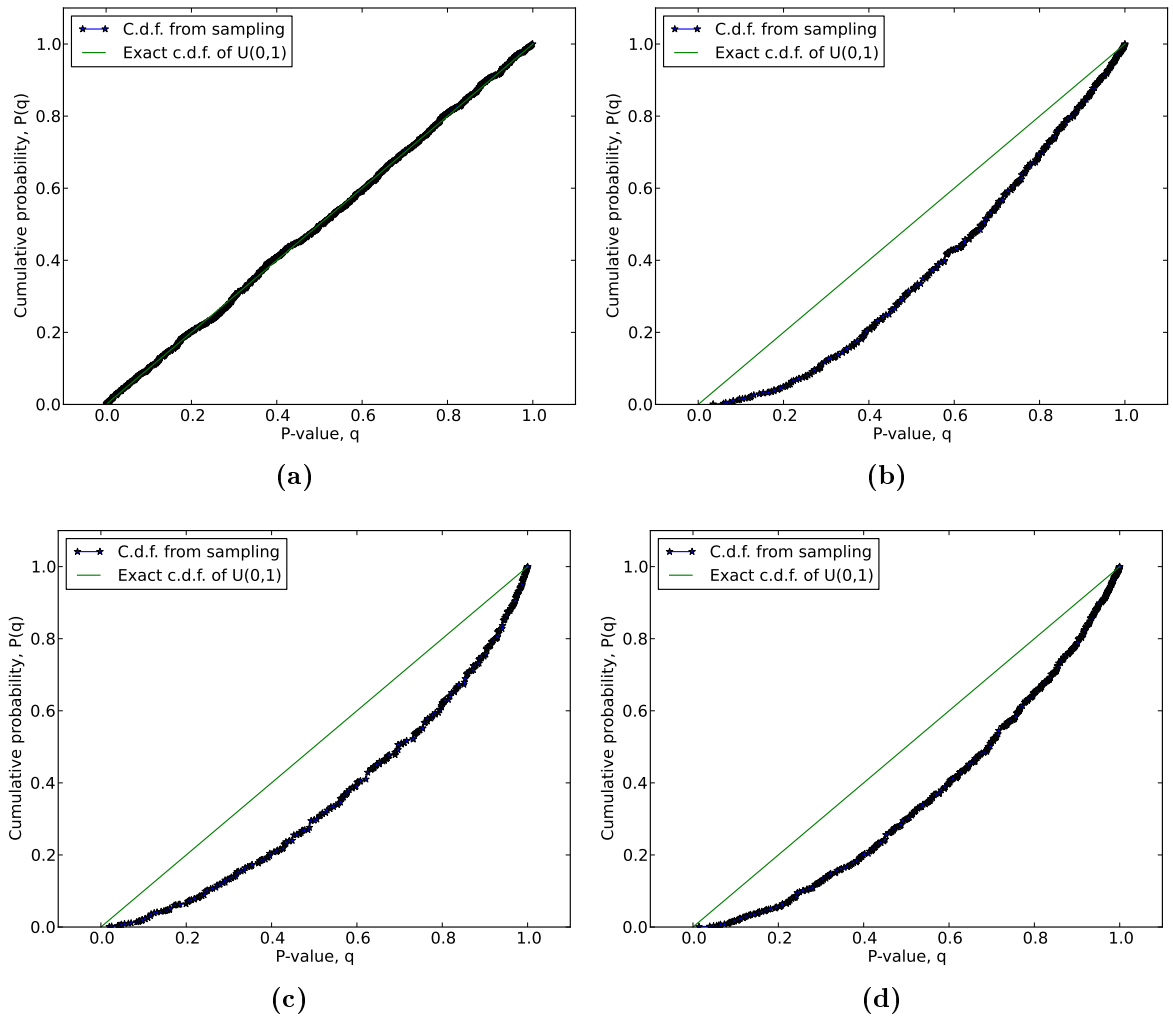


Figure 6.13: Cumulative distribution functions (c.d.f.s) for the output *p-value* of the *Fisher* consensus test for different simulated inputs, compared against the theoretically expected. In (a) 5 *p-values* drawn (by *Monte Carlo* sampling, simulating null hypothesis conditions) from the continuous U(0,1) distribution (see Fig. 6.8a) are combined, and the experimental distribution closely matches the theoretical. However, in plots (b) - (d) large deviations from the continuous U(0,1) c.d.f. are observed: plots (b) and (c) are based on drawing 5 inputs from the discrete distributions shown in Figs. 6.8b and 6.8c, respectively, while (d) was produced by drawing 2 samples from the former and 3 from the latter. Note the differences between plots (b) - (d) are limited as they are all based on the same number (i.e. 5) of input *p-values*.

the null hypothesis the probability of a significant result substantially differs from the significance level. For example, in Fig. 6.13b a *p-value* of 0.2 only has a cumulative probability of about 0.05 of occurring - not 0.2 as should be expected. While it is possible that the effects observed for discrete inputs are not a concern in the sort of meta-studies in which these consensus tests are typically applied as the input *p-values* are always suitably continuous, in principle this is a problem relevant to a wide range of situations, bearing in mind that all digital data is quantised.

It is noteworthy that this problem has parallels to the so-called familywise error rate (FWE) in conventional multiple testing. There, because of the large number of individual hypothesis tests being considered together, there is a high probability of falsely rejecting the universal H_0, that none of the tests are significant. Adjustment procedures exist, the most common of which is the *Bonferroni* correction that scales down the significance level to be applied to individual tests by the number of such tests considered together [102,116]. Alternatively, the individual *p-values* themselves may be adjusted [117].

6.7 Compensation Schemes - Fisher Test

6.7.1 Principles

The observed deviations from the required null hypothesis behaviour (as in Fig. 6.13) were investigated further by studying the intermediate stage of the consensus testing process, the test statistic and its survival function (s.f.), for the problematic, discrete input cases. The discrete-input specific discrepancy underlying the departures from the required null hypothesis behaviour is shown in Fig. 6.14. In each of the plots a large mismatch is seen between the continuous, χ^2 distribution s.f. that the input samples are expected to follow from continuous statistics, and the discrete s.f. that the samples actually follow, computed by *Monte Carlo* (M.C.) sampling the *p-value* inputs. Naïvely applying the continuous, original s..f. to the discrete inputs can be seen to give rise to excessively high survival probabilities, equal to the output consensus *p-values*.

The s.f. computed by M.C. sampling provides the first of three possible compensation schemes identified and this possibility is studied further in the next subsection. The

graphs in Fig. 6.14 also include the s.f. relevant to the other two compensation schemes to be introduced. The first involves breaking down the input discretisation using additive random noise, in a dither compensation scheme described further in Section 6.7.3, which also demonstrates that the input discretisation is to blame for the problems experienced. The second additional scheme is an analytic approach, explained in Section 6.7.4, and only shown for the first of the plots in the figure.

6.7.2 Monte Carlo

This compensation scheme is simply based on applying the actual, discrete s.f. to the computed test statistic value, rather than the s.f. expected from continuous statistics theory. Doing so retains the properties of the chosen consensus test and associated test statistic, as such a process does not affect how the multiple inputs are merged into a single value. The s.f. to use can readily be computed *a priori* for any discrete inputs and choice of test statistic, although while the continuous *Fisher* test s.f. only relies on the number of *p-values* in the input, the M.C. scheme requires knowledge of all possible discretisation levels that each input *p-value* could take. Under the null hypothesis these inputs each follow a U(0,1) distribution, so the discretisation fully specifies each input distribution as the probability of each level can be computed from the c.d.f. of a U(0,1) distribution. To then obtain the required s.f., the test statistic is computed many times using M.C. samples drawn from the discrete input distributions to build up a distribution - in effect simulating the expected null hypothesis behaviour.

The effectiveness of the scheme is demonstrated by Fig. 6.15, where the plots showing the poor null hypothesis behaviour (Fig. 6.13) are reproduced but with the application of the M.C. s.f.s: the distortions previously observed have been removed. It is worth emphasising that there are two types of *Monte Carlo* sampling underlying these plots: first the M.C. simulation of the s.f. under the null hypothesis given the input discretisation that makes up the compensation scheme, then the testing by M.C. sampling to build up consensus output distributions under the null hypothesis and hence demonstrate the validity of the compensation.

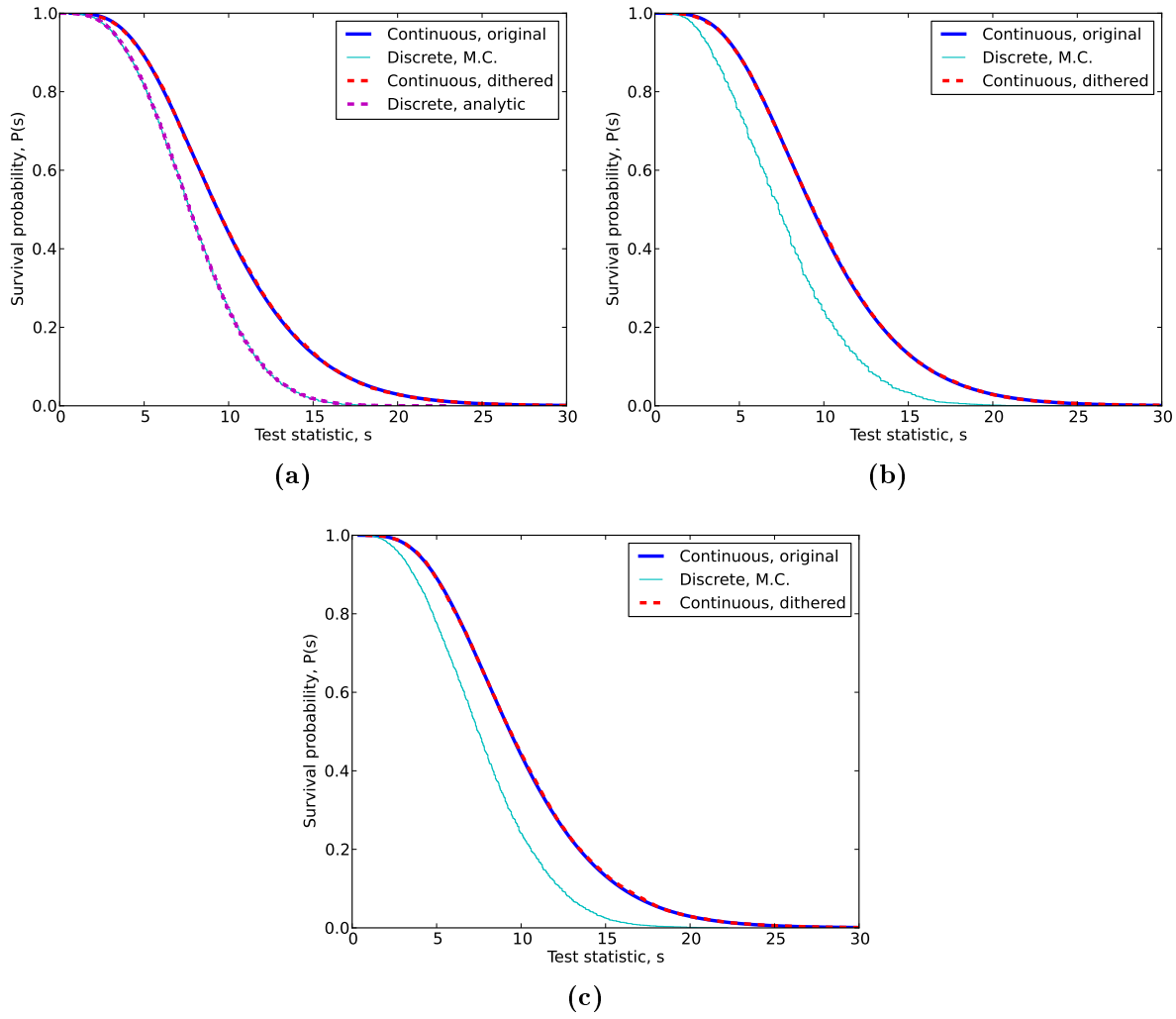


Figure 6.14: *Fisher* consensus test statistic survival functions (s.f., equal to 1-c.d.f.) corresponding to (a) 5 samples from the distribution in Fig. 6.8b, (b) 5 samples from the distribution in Fig. 6.8c, (c) 2 samples from the former and 3 from the latter. In each plot there is a significant discrepancy between the continuous s.f. and the M.C. version. In each case the continuous s.f., produced using the dither compensation scheme to break down the input discretisation closely matches the continuous s.f., demonstrating that the unexpected effects identified are indeed caused by discretisation effects. Plot (a) additionally includes the s.f. computed analytically for the discrete inputs, and this distribution is seen to coincide with the M.C. curve. Note the discrete s.f.s are steeper than the continuous curves, and that the three sets of plots only differ slightly as they are all based on the same number (i.e. 5) of input *p-values*, so that the original, continuous curves shown are in fact identical.

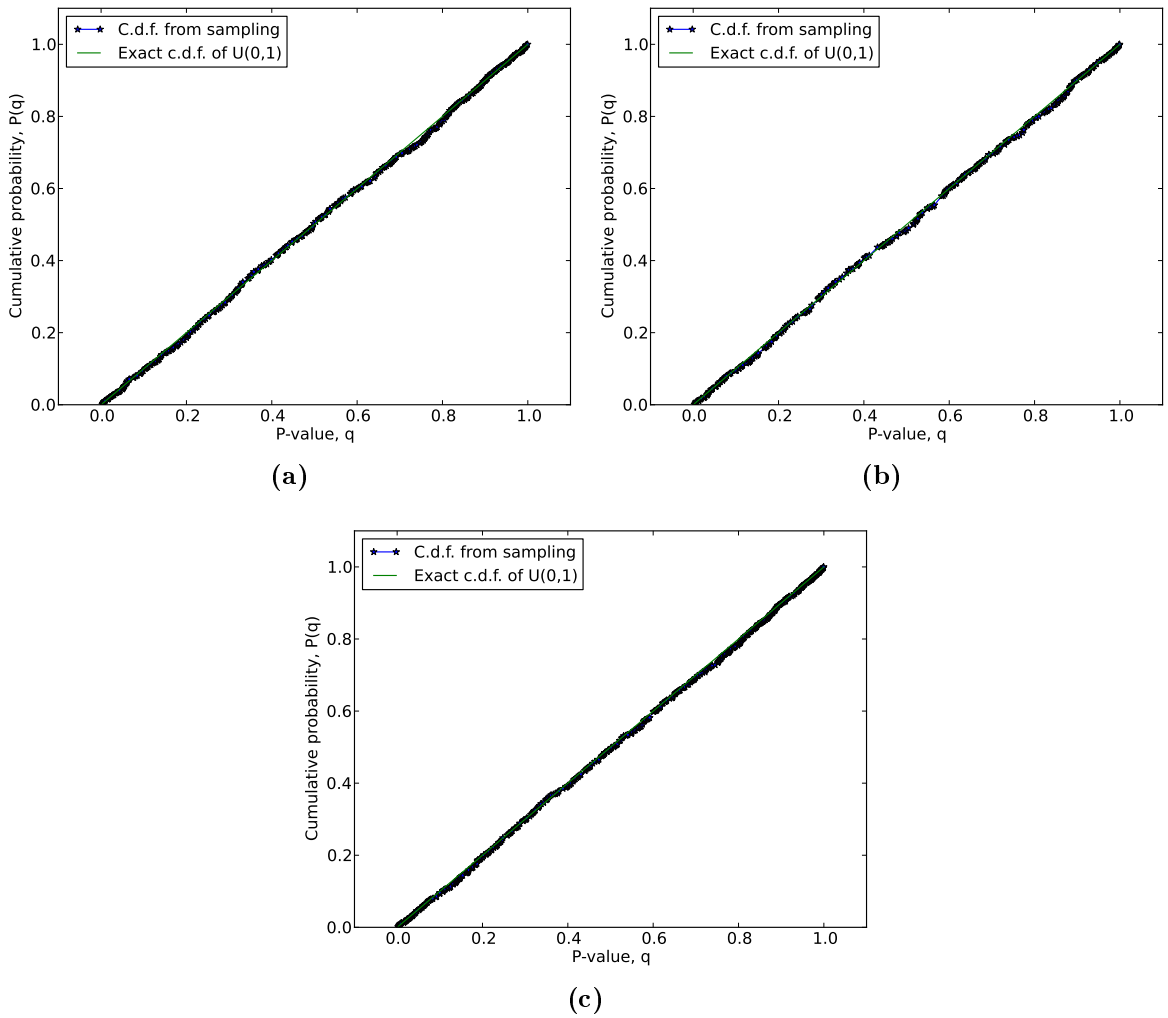


Figure 6.15: Output c.d.f.s for *Fisher* consensus test for different simulated inputs with *Monte-Carlo*-based discretisation compensation. Figures (a) - (c) correspond to the uncompensated plots seen in Fig.6.13 (b) - (d), respectively, but use the *Monte Carlo* s.f.s shown Figs. 6.14 (a) - (c), respectively. In each case the experimental distribution is seen to closely correspond to the continuous U(0,1) distribution, demonstrating the effectiveness of the compensation.

6.7.3 Dithering

This compensation scheme was initially developed to demonstrate that the problematic effects experienced are purely the result of the input discretisation. The underlying principle is to break down the discretisation of the input *p-values* by dithering, adding random noise to each appropriate to the effective width of the discretisation level of the inputs. Therefore this scheme, like the *Monte Carlo* option, requires full knowledge of the inputs' discretisation levels - in general it is not adequate to add uniform noise drawn from some fixed interval such as $[\frac{-1}{2n}, \frac{1}{2n}]$, where n is the number of discretisation levels, and $\frac{1}{n}$ would be the width of ever level in the case of evenly spaced levels. Again like the M.C. option, this scheme is applicable to all discretisations and test statistic choices, with this scheme essentially serving as a preliminary calculation step to make the continuous statistics applicable to a discrete input.

Having blurred the discrete *p-value* inputs by dithering it is then again correct to apply the s.f. expected from continuous statistics to the test statistic formed. This is confirmed by the curves from dithering shown in Fig. 6.14 very closely matching the original, continuous s.f.s, as expected. The performance of this compensation scheme is demonstrated by Fig. 6.16, where again the previously identified problem under H_0 (Fig. 6.13) is shown to have been eliminated.

Note that this dithering of *p-values* is equivalent to uniformly dithering the original amplitudes across each quantisation level when a discrete data model is used to convert between the two quantities, and implicitly makes an assumption about the nature of the quantisation error. While the assumed uniformity may be reasonable in some situations, especially if there are significant gradients in the recorded data distribution, assuming that the recorded samples are drawn from a smooth, continuous underlying function is more appropriate. This then calls for a smooth data model such as one based on Kernel Density Estimation (KDE, see Appendix Section C.3), with associated problems. The other problematic feature of the dithering scheme in practical use is that the consensus output is stochastic and rather than a single output value for a given input of *p-values* a distribution would have to built up, making comparisons of consensus outputs difficult.

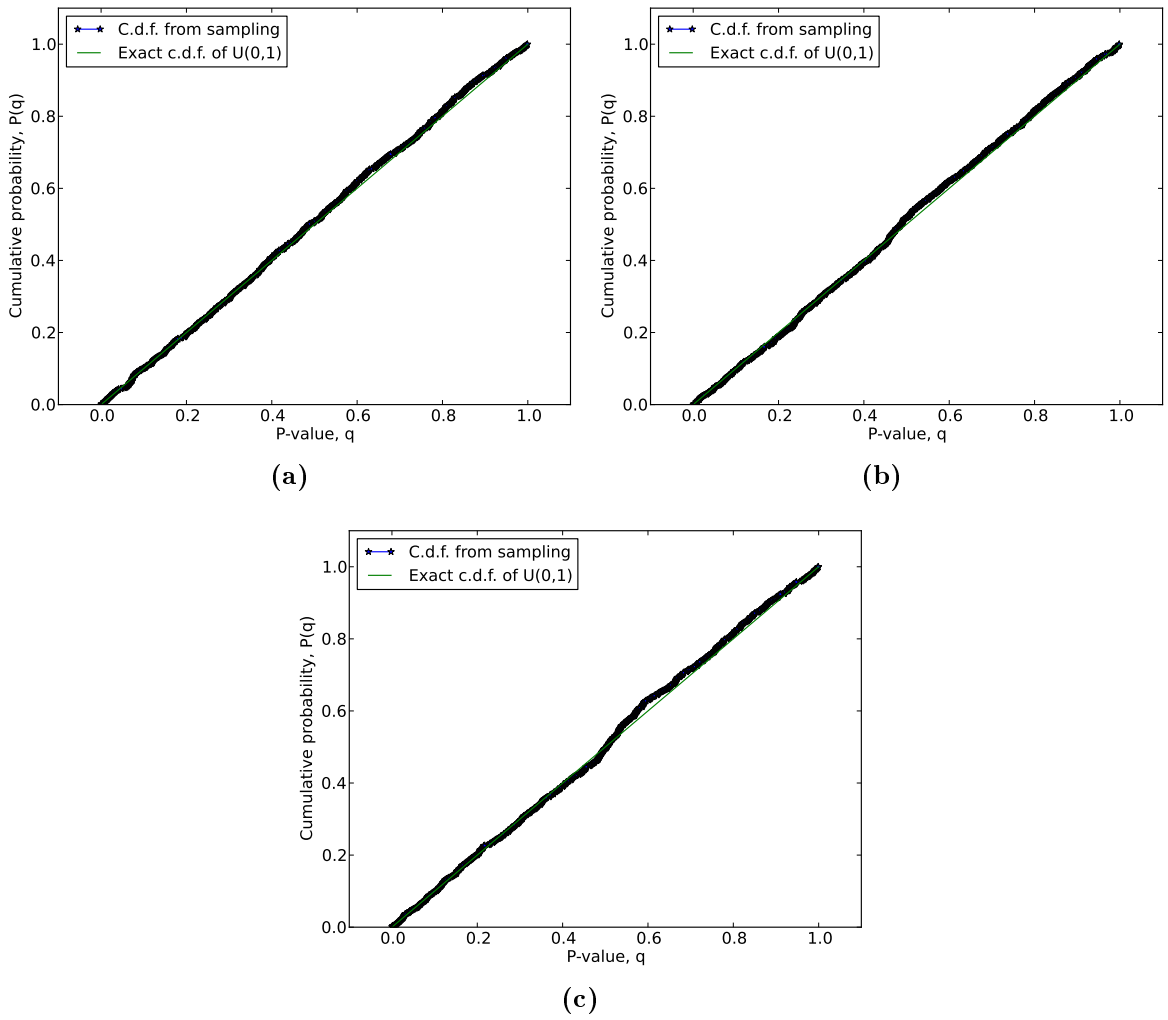


Figure 6.16: Output c.d.f.s for *Fisher* consensus test for different simulated inputs with dithering discretisation compensation, effectively turning the discrete inputs into quasi-continuous values. Figures (a) - (c) correspond to the uncompensated plots seen in Fig.6.13 (b) - (d), respectively. The computation exploits the coincidence between the quasi-continuous and continuous s.f.s seen in Fig. 6.14 (a) - (c). In each plot the experimental distribution is seen to closely correspond to the continuous U(0,1) distribution, demonstrating the effectiveness of the compensation.

6.7.4 Analytic

This compensation scheme was a by-product of an effort to gain a theoretical understanding of the processes giving rise to the unexpected behaviour of the consensus tests for discrete *p-value* inputs. The starting point of this analysis is the realisation that any discrete distribution can be thought of as a sequence of weighted *Dirac* delta functions δ:

$$p(x) \quad = \quad \sum_{a=1}^{n} w_a \delta\left(x - x_a\right) \tag{6.5}$$

gives the probability mass function, p.m.f., $p(x)$ corresponding to n discretisation levels at positions x_a with weight w_a. Note that given normalisation, $\sum_{a=1}^{n} w_a = 1$.

The *Fisher* consensus test statistic involves the summation of multiple samples drawn from different such distributions. The distributions in question are not just the U(0,1) distributions determined from the available discretisation levels, but these rescaled, taking the natural logarithm of the input *p-values* and multiplying by -2 (without affecting the probability associated with each level). The probability distribution of the sum of independent random variables is given by the convolution of the contributing probability distributions [103]. Convolutions meanwhile are typically completed in the *Fourier* domain, where the operation simply corresponds to a multiplication:

$$\mathcal{F}\left(f * g\right) \quad = \quad \mathcal{F}\left(f\right) \cdot \mathcal{F}\left(g\right) \tag{6.6}$$

where $\mathcal{F}$ represents the *Fourier* transform, and f, g are arbitrary functions.

Noting that

$$\mathcal{F}\left(\delta\left(x - x_a\right)\right)\left(\omega\right) \quad = \quad \exp\left(-2\pi i \omega x_a\right) \tag{6.7}$$

where $i = \sqrt{-1}$ and ω is the *Fourier* space variable, it follows that

$$\mathcal{F}\left(p(x)\right)\left(\omega\right) \quad = \quad \sum_{a=1}^{n} w_a \exp\left(-2\pi i \omega x_a\right) \tag{6.8}$$

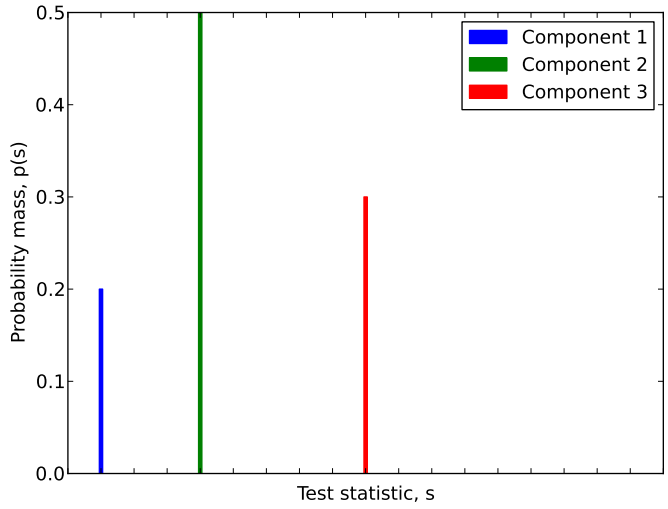
and hence that the sought distribution $p_{n_1,n_2\ldots n_m}(x)$ is

$$p_{n_1,n_2\ldots n_m}(x) \;=\; \mathcal{F}^{-1}\left[\prod_{b=1}^{m}\left(\sum_{a=1}^{n_b} w_{a,b}\exp\left(-2\pi i\omega x_{a,b}\right)\right)\right] \tag{6.9}$$

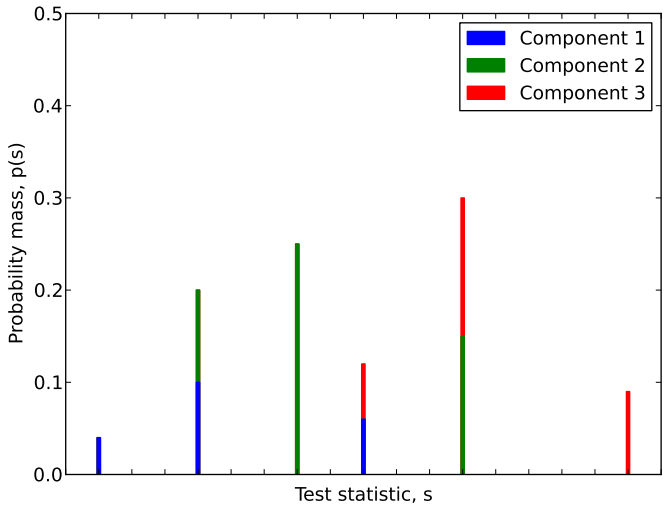
for the general case of m distributions being convolved (from m *p-values* being combined in the consensus test), each with n_b δ-function components, where $b = 1, 2\ldots m$. Expanding the right-hand expression gives $\prod_{b=1}^{m} n_b$ terms, each the product of m elements $w_{a,b}\exp\left(-2\pi i\omega x_{a,b}\right)$. Each one of these terms may then be reduced to a single exponential of the form $\left(\prod w_{a,b}\right)\exp\left(-2\pi i\omega\left(\sum x_{a,b}\right)\right)$. From Equation 6.7 taking the inverse *Fourier* transform gives a δ-function at position $\sum x_{a,b}$, weighted by $\prod w_{a,b}$. In practice the number of terms can be reduced as some of the *Fourier*-space exponents, and hence δ-function positions, will be the same, so the associated weights can be summed for that position. The principle of this calculation is illustrated graphically in Fig. 6.17, where a short sequence of δ-functions is convolved once with itself. Having computed the analytic p.m.f. in this manner, the s.f. can readily be determined from a cumulative sum.

The computational implementation of the calculation to be able to evaluate the exact survival function to use for any collection of possible *p-value* discretisation levels is not straightforward. A brute-force approach, without for instance exploiting symmetries, necessitates the evaluation of two m-dimensional arrays with edges $n_{1,2\ldots m}$, one containing all possible combinations of $x_{a,b}$ summed and one for all possible products of $w_{a,b}$. For all but the simplest cases this risks quickly becoming infeasible given the poor scaling with input size. Therefore only a limited version was implemented, restricted to small and relatively simple inputs - enough to validate the analysis and compare it with the other compensation schemes described but less useful for practical use. The author is confident that the analysis would hold for more complex inputs, too, and that comparable analyses could be developed for other forms of test statistic.

A s.f. computed in this manner was already included in Fig. 6.14a, and an enlarged section of that plot is reproduced in Fig. 6.18, comparing this analytic s.f. to that from the *Monte Carlo* approach. Note also the very close correspondence between the two lines, where almost every deviation in the analytic distribution is mirrored in the *Monte Carlo* version, with occasional minor deviations due to the stochastic nature of *Monte*



(a)



(b)

Figure 6.17: Illustration of a discrete p.m.f., corresponding to a sequence of weighted *Dirac* δ-functions in (a), and its convolution with itself in (b). The colours in the second plot indicate which of the δ-functions in the first that part of the distribution originates from: there are 9 different contributions making up (b), corresponding to 3, the number of components of (a), squared. Note that the distributions are functions of the test statistic, not *p-value* as in some similar previous plots.

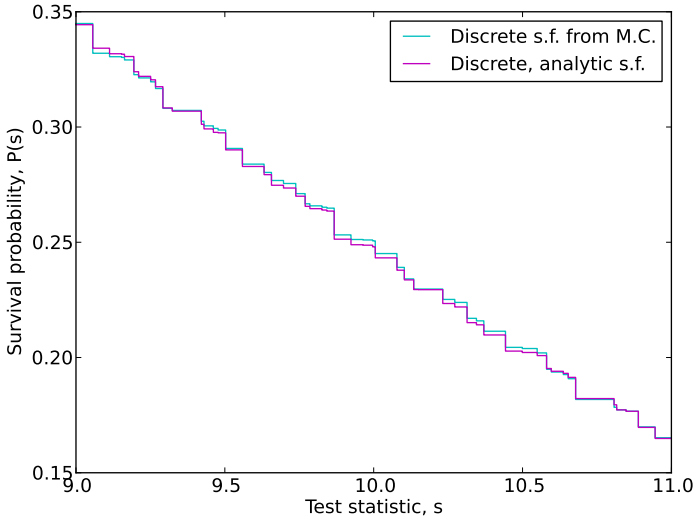


Figure 6.18: A close-up view of part of the survival functions (s.f.s) seen in Fig. 6.14a for the *Fisher* test statistic, comparing *Monte Carlo* (computed using 10^4 samples drawn from U(0,1)) and analytic distributions. The stair-step structure of these discrete distributions is evident - note that the "risers" mark transitions from one discrete level to another, with the upper level being valid up to and including the value of the test statistic at the "riser".

Carlo, typically maintained over a few "steps" given the underlying cumulative sum. This provides a convincing cross-validation of the two approaches.

The analysis presented here in fact describes *Fisher's exact test*, typically used in the context of so-called contingency tables [118]. Moreover, this is based on the hypergeometric distribution, which is known to be hard to evaluate and a range of tools have been used to tackle it, including *Monte Carlo* methods [117, 119].

6.7.5 Practical considerations

In practice the compensation scheme of choice is the *Monte Carlo* approach. Given the limited applicability of both the computation and implementation of the analytic approach, that option can rapidly be dismissed. The dithering approach in turn can be dismissed as, by using the continuous test rather than *Fisher's exact test*, it sacrifices distinguishing power, critical for the sort of multiple testing, examining many *p-values* simultaneously, used here [117]. Additionally, the dithering approach for instance requires assumptions to be made about the nature of the inherent quantisation error of the data. Meanwhile, the M.C. approach is relatively easy to implement in a fully general manner,

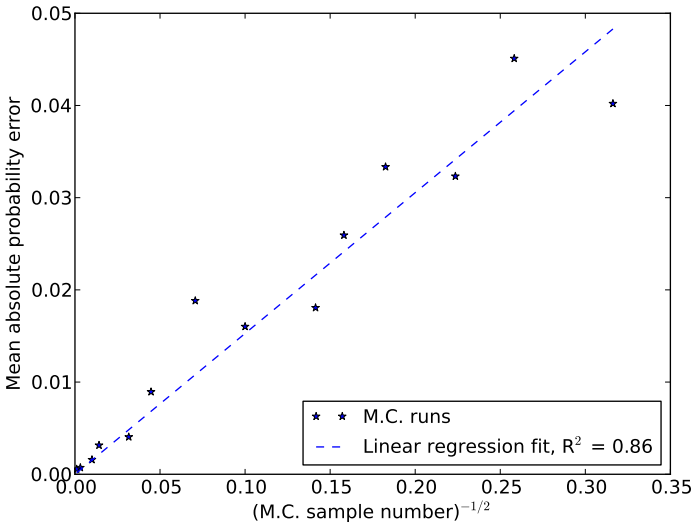


Figure 6.19: *Monte Carlo* convergence to analytic s.f. for varying numbers of M.C. samples (drawn from U(0,1)). For each sample size the s.f. was computed, and the deviation from the analytic was quantified by the mean absolute difference across the length of the curve. This quantity is shown plotted as a function of the reciprocal square root of the M.C. sample number (so large sample numbers appear closer to the origin), as according to M.C. theory the points should then lie on a straight line through the origin. Such a line has been fitted to the plot, and the relatively high R-squared value means that this is indeed a reasonable model of the M.C. convergence behaviour.

and yields good results quickly. As a indicative example, the evaluation of the 10^4 points used for the generation of Fig. 6.18 took about 20 times less computation time than the analytic approach, yet the outputs are very similar, to the extent that the differences are likely to be immaterial for practical purposes.

The quantity that directly links the accuracy of the M.C. evaluation and the computational effort required to achieve it is the number of samples used. Typical M.C. theory states that the former improves as a function of the reciprocal of the square root of this sample number [120]. The extent to which this convergence relationship holds has been tested, comparing the s.f. deviation from the exact, analytic distribution for different numbers of M.C. evaluations. The result is shown in Fig. 6.19, confirming that the general theory provides a reasonable model in the application here.

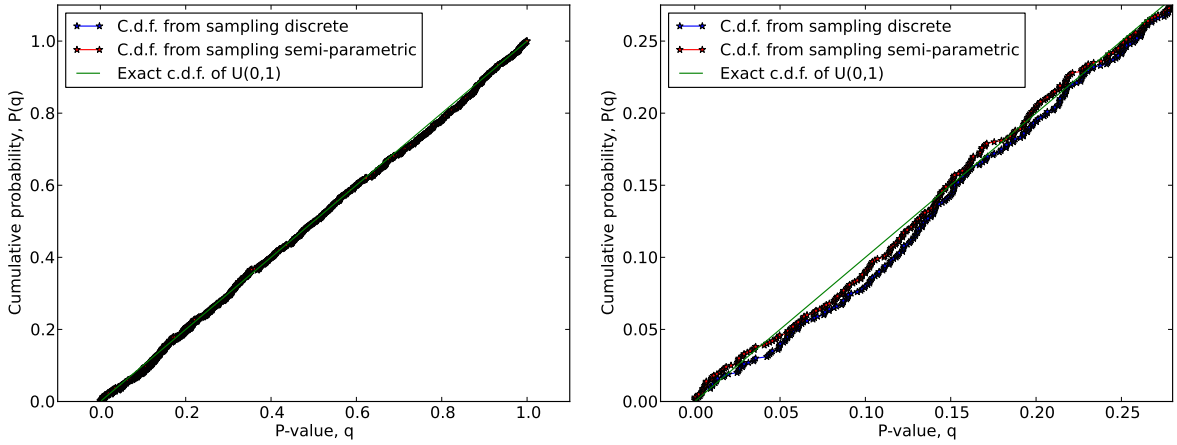
There are a number of computational approaches to reduce the computing time required to give a certain accuracy. A simple option is to perform the evaluations in parallel, using several CPU cores simultaneously [121]. A numerical technique to give consistent coverage of the space being explored by random sampling, and hence potentially reduce

the number of samples needed for a given accuracy, is to draw in the inputs from a low-discrepancy sequence spanning the [0,1] interval. These pseudo-random sequences avoid random clustering of samples that would occur with a more conventional random number generator, and the *Sobol* sequence is a commonly used example [120,122]. A more complex technique is importance sampling: this involves biasing the used samples to the part of the [0,1] interval where they provide greatest value in the output, and then compensating for the biasing in the collation of the output [120]. Because in practice the consensus *p-values* of indication region *resels* will be usually very low, the resolution of the s.f. for that consensus *p-value* range is critical to be able to distinguish between regions at all. Achieving the required accuracy by unbiased M.C. sampling is very wasteful, but this problem can readily be addressed by the application of importance sampling, as implemented and described further by Tippetts [13].

6.7.6 Combining consensus test and semi-parametric model

As a final test of the compensated consensus test, the behaviour of the consensus output when operating on inputs generated by the semi-parametric data model (see Section 6.5.4) applied to experimental data was studied. The resultant c.d.f.s for such trials are shown in Fig. 6.20. This both demonstrates the effectiveness of the consensus test compensation and provides evidence in support of the earlier assertion that the deviation from the null hypothesis behaviour introduced by the semi-parametric model is negligible.

However, Fig. 6.21 indicates that the induced deviation increases with the number of *p-values* combined in the consensus test, and will also increase with the magnitude of the semi-parametric tail fit error. This does mean that some caution should be applied in interpreting consensus outputs in terms of their likelihood under the null hypothesis in cases where this cumulative effect is likely to occur, although the 25 input case considered is a rather extreme case for practical purposes. If this is not acceptable, the fully discrete data model should be relied upon instead of the semi-parametric model proposed. Nonetheless, given the benefits of the semi-parametric model, this cause of order uncertainty must be considered relative to others, such as the M.C. discretisation compensation and the quantisation error of the original data.

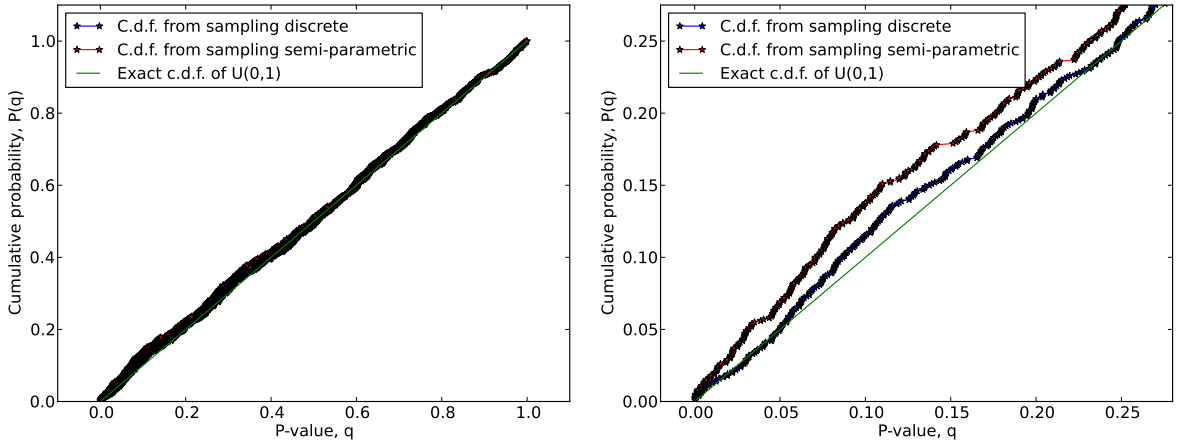


(a) **(b)**

Figure 6.20: The output c.d.f.s for combining the semi-parametric data model with the *Monte Carlo* compensated *Fisher* consensus test, applied to experimental data. The plots are based on the amplitude distribution found in one circumferential ring of rotor bore *resel* data used previously - see Fig. 6.9a in Section 6.5.4. Sets of 5 *p-values* were obtained using both the (exact) discrete and the semi-parametric models, then these were passed through the *Monte Carlo* compensated *Fisher* consensus test to build up distributions for the output consensus *p-values*. The only underlying difference between the curves shown is the choice of data model, even the random numbers used in the M.C. were identical. In (a) no difference is discernible between the two curves, only in the zoomed plot (b) is the effect of the semi-parametric model's tail fit identifiable. This provides good evidence that the data model and the compensation scheme work well together, and the deviation from the null hypothesis behaviour due to the semi-parametric model is indeed negligible.

6.7.7 Comparing and interpreting consensus test outputs

Having used the described probabilistic data fusion system to assign a consensus *p-value* to many if not all *resels* in the multichannel dataset, we wish to compare these values. Specifically, we wish to rank the component regions as described in the chapter introduction. Importantly, all the fused *p-values*, while strictly discrete, are samples of the continuous uniform U(0,1) distribution to a very close approximation. This is related to the fact that the number of possible unique input combinations from the contributing distributions increases approximately exponentially with the number of signals being combined, as exemplified by the many steps in the plot of Fig. 6.18, for example. This essentially continuous behaviour is demonstrated in the results of the next chapter, and it allows the *p-values* originating from different underlying distributions to be directly compared.



(a) **(b)**

Figure 6.21: A variant of Fig. 6.20, based on combining 25 (rather than 5) *p-values* drawn from both the (exact) discrete and the semi-parametric models in a consensus test. Again, in (a) the differences between the curves are hardly visible, but in the zoomed plot (b) differences are quite apparent. The deviation exhibited by the semi-parametric-based curve is more noticeably greater than in the preceding figure, illustrating that combining more *p-values* allows the error to accumulate.

In principle, significant discretisation may still exist in the consensus *p-values*, either due to the underlying distributions containing extremely few unique entries or the number of input channels being limited to unity - such that the consensus *p-value* is merely a LEN *p-value* (see Section 6.5.4). Under such circumstances, to maintain the appropriate null hypothesis behaviour in the multiple testing that is implied by ranking, the *p-values* should be adjusted for the discretisation, in a further layer of compensation. The appropriate process in this case is essentially the application of the discrete *Tippett* test, described in Appendix D [117].

In the applications considered, only the second of the possible causes of this problematic behaviour can occur: in the rotor bore application part of the component is only interrogated by a single channel. Rather than worrying about violating the null hypothesis behaviour or applying a further compensation scheme, the *resels* for which no fusion is possible are eliminated from the results, as these are of little interest for fusion performance assessment anyway.

6.8 Computation of Probabilities for Indications

6.8.1 Indication Severity Probability

Using the methods described in the previous sections, a consensus *p-value* may be assigned to any spatial region of interest in the data volume: the data field samples are de-correlated and turned into *resels*, the values of which are then converted to *p-values* using a local, semi-parametric data model, and these probabilities are then fused using a consensus test, with appropriate discretisation compensation. Subtracting this value from unity gives a probability of greater intuitive appeal, increasing for less noise-like regions. This is termed the indication severity probability.

While this value is expected to be displayed for an operator to see when he or she is reviewing the sequence of suggested indications (see Section 6.1), it should be emphasised that this should only be used as a rough guide. The decision of when to terminate the analysis must be made by the operator based only on the viewed signals, and therefore for instance no hard threshold (equivalent to a significance level for the *p-value*) should be set *a priori*.

6.8.2 POD, PFA and ROCs

For the purposes of the results chapter that follows it is necessary to explain how the probability of detection (POD), probability of false alarm (PFA), and hence also the receiver operating characteristic (ROC) can be computed for the data fusion output given some test data with known indications. Recall from the start of the chapter (Fig. 6.1) that typical calculations of POD and PFA rely on computing integrals of distributions p(signal | flaw) and p(signal | noise) for a given amplitude threshold, which may be varied to build up an ROC curve. Here the process is similar, but completed in terms of indication severity probability rather than signal amplitude, with a chosen significance level taking the place of the amplitude threshold.

Noise *p-value* distribution

Dividing the entire component volume into candidate indication regions and computing indication severity probabilities for each yields the probability density function p(severity | noise). The implementation of this full evaluation is non-trivial, and not only because of the sheer computational load (itself a significant concern previously [85,123], overcome using, with the help of the software techniques outlined in Chapter 3) and the fact that care should be taken to omit the regions containing the known indications. Rather, the major challenge concerns the decision of which *p-values* to fuse, given that the originating data channels are described in terms of aligned coordinate axes (post-registration) but different coordinates. Essentially, in each channel physical space is discretised differently, and in a somewhat arbitrary manner (see definition of *resel*, Section 6.3.3), such that determining associations between *resels* requires some approximation. The author's colleague Tippetts chose one approach, the present author a second:

- Define a global set of *resels*, of a size roughly compatible with the sizes determined in the individual channels. Interpolate the data for each channel to the *resel* coordinate system. Evaluate the values of the *resels* from the interpolated amplitudes. Complete all subsequent fusion processing on the global *resels*, where association between points from different channels is now clear.

- In each channel define *resels* locally, associating each with its centre location. Compute which *resel* of every channel every *resel* of every channel lies in. Eliminate duplicate *resel* combinations. Complete fusion processing for all remaining combinations.

The first (Tippetts') implementation has the advantage of relative simplicity. However, the need to interpolate introduces further noise and the global *resels* are rather unphysical, and incapable of dealing with, for example, beam-spread effects unique to individual channels. The second (the author's) approach is more rigorous but also more complicated. No matter what approach is used, *resel* edge effects can arise in practical use, where essentially a data point falls the wrong side of a *resel* boundary and so is not associated correctly with its counterparts in the other channels. This then has the potential to

lead to an unduly low indication severity probability. Given that the location of each data value is somewhat uncertain (see chapter Chapter 4) and may be considered to be described by a spatial probability density function (which may be visualised as a point cloud), it is inevitable that sometimes these distributions are split by *resel* boundaries, no matter what spatial discretisation is chosen. However, to avoid the worst of these effects, the *resels* should of course be of a scale larger than the spatial uncertainty of the data - emphasising the importance of careful *resel* choice, taking into account the accuracy of the achievable registration.

For a highly conservative system, one could consider adaptive association, searching the local environment, on a scale of the positional uncertainty, for the lowest set of *p-values* to fuse. Such an approach would not only be demanding to implement but would undermine (at least without modified compensation) the expected null hypothesis behaviour. Nonetheless, this is a suggestion for further work.

Flaw *p-value* distribution

Having computed p(severity | noise), the corresponding p(severity | flaw) distribution ideally requires a collection of indication severity probabilities to be available for a number of indication regions considered to contain equivalent defects. However, often only a single such indication severity probability is available, which is then used to create a distribution empirically. This is done here by making the assumption that the amplitudes from the defect vary like the local noise distribution, creating for each contributing *resel* a duplicate of the local noise distribution with the mean shifted to the value of the *resel*. This process is illustrated in Fig. 6.22. Drawing samples (by further *Monte Carlo* sampling) from these empirical amplitude distributions and recomputing the severity probability of the indication gives the required p(severity | flaw) probability density function. This approach is considered conservative, for example given that summing maxima from two sets of values gives the upper bound on the maximum of the summed underlying values. Moreover, the model assumptions will similarly affect different detectors if several are being compared, so should be irrelevant in such relative analyses.

As will become clear in the results that follow, moving from simple amplitude distributions to the consensus (indication severity) probability space that incorporates knowledge of

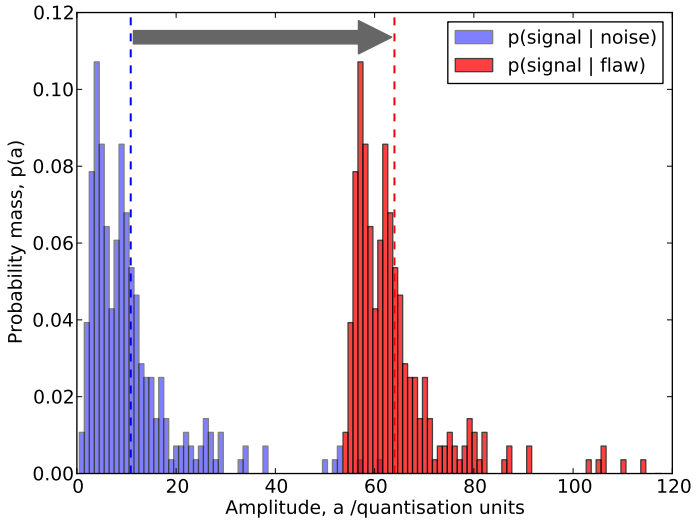


Figure 6.22: The process used to build an empirical distribution p(signal | flaw) from a single *resel* sample. The local distribution of *resels*, excluding the one containing the indication, forms p(signal | flaw). This distribution is duplicated and has its mean shifted to the value of the *resel* containing the indication to serve as p(signal | flaw). The distribution means are shown as dotted lines.

local data distributions and correlations between independent samples of the same spatial region allows substantially improved separation of noise and defect signals to be achieved. This is borne out by the improved ROC plots of the next chapter.

6.9 Summary

The chapter has described in detail a novel yet general process by which samples from diverse amplitude fields may be fused to form a single probability. The basic processing stages are data de-correlation, local data modelling to convert amplitudes to probabilities taking into account local data statistics, then fusion of the probabilities for a spatial region of interest using a consensus test. The processing scheme is essentially free of arbitrary thresholds and makes very few assumptions. Novel elements include the semi-parametric data model and the application of *Fisher's exact test*. The author is not aware of a comparable data fusion system having been developed elsewhere, in any field, and the work described here may have applications beyond the confines of Non-Destructive Evaluation. Next we consider the performance of the data fusion detector, compared with more traditional systems, when operating under a range of circumstances.

Chapter 7

Detection Results

7.1 Introduction

In this chapter we examine the performance of the data fusion detection system developed in the preceding chapter. Suitable registration (see Chapters 4 & 5) of data inputs is assumed to have taken place, so that the data inputs are all in the same global coordinate system. Detection results of indications in an industrial rotor bore dataset are presented, first looking at several scenarios in a single acquisition before considering the case of a second, baseline acquisition being available. Finally, due to the lack of suitable known real indications in the available rotor bore data, a selection of detection results obtained for the disk inspection application is reproduced from Tippetts' thesis [13] to demonstrate performance on real defects.

Unless otherwise stated, all results presented in this chapter are based on the data for the inspection of a low pressure stage 2 turbine rotor at RWE npower's Didcot A coal-fired power plant, Forging C001 993 Y21657, inspected in January 2007 as job number GBF4201. The files used in the analysis here all relate to the section of the bore 2.4 - 2.9m axially from the alternator end of the rotor bore, given that the 6m long bore is inspected in 50cm sections (with a little overlap). The inspection report concludes that "No recordable responses were observed during the Ultrasonic Inspection of the bore." [20], so the focus here is on seeing what sort of possible defect indications were missed. Additionally, artificial defect-type signals are inserted in the data for testing purposes.

By default, each rotor bore data analysis computation used 200 *Monte Carlo* samples per set of *p-values* to fuse. Each such computation took around 3.0hrs to complete, on the computer of Section 3.2.1, running the *Monte Carlo* evaluation in parallel on three cores of the CPU. Figure A.4 provides an example of the sort of *Operator* (see Section 3.4) configuration relevant to the calculations of this chapter.

7.2 Simulated Test Inputs

In this section we test the detector with entirely artificial, simulated inputs, to confirm that the system behaves as expected and gain a qualitative understanding of its performance. There is some overlap here with the extensive analysis presented by Tippetts [13]: Section 7.2.1 offers re-worked versions of plots found in Chapter 5 of the cited publication.

7.2.1 Qualitative performance

Here we compare the Data Fusion Detector (DFD) with the standard logical AND and OR operators, that effectively provide bounding fusion possibilities. These two operators are typically used with pairs of binary / boolean inputs, but to enable a comparison with the DFD we need to be able to operate on multiple float probability inputs, taking unity minus a *p-value* to correspond to a "probability of success". We apply binomial theory to compute the probability of obtaining "all successes" (for AND) and "at least one success" (for OR). The AND consensus test output q_{AND} for N input *p-values* q_i is then:

$$q_{AND} = 1 - \prod_{i=1}^{N}(1 - q_i) \tag{7.1}$$

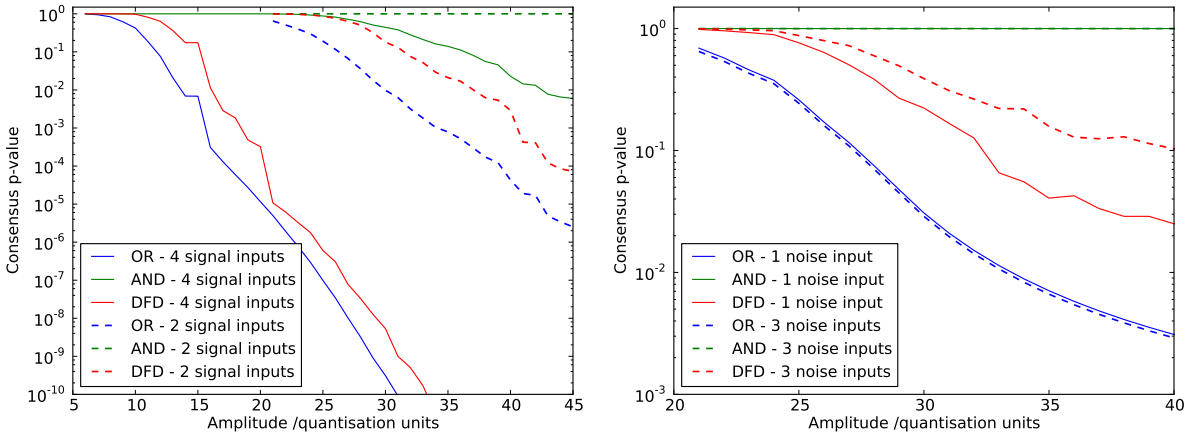
The corresponding OR consensus test output q_{OR} is:

$$q_{OR} = \prod_{i=1}^{N} q_i \tag{7.2}$$

Figure 7.1a shows how the three detectors vary when provided with four or two inputs of the same amplitude (but differing *p-values*, as drawn from different *resel* distributions

of the industrial dataset). Considering the four input plots first, as should be expected, all three detectors give a decreasing output consensus *p-value* as the amplitude increases, and the OR consensus test output falls off at a lower amplitude and more rapidly than the AND consensus test. The DFD is seen to transition from behaving like an AND test to behaving like an OR test as the amplitude increases. This is (qualitatively) ideal behaviour for a detector, as the AND-like behaviour is desirable at low amplitude to suppress the false-call rate, yet OR-like behaviour is required at higher amplitudes so as to maintain sensitivity to amplitude spikes in a single input. The exact balance and transition between these two states is governed by the specific implementation of the data fusion, for example in the tail fit of the semi-parametric data model (see preceding chapter), and is open for debate and adjustment. If only two rather than four inputs are provided, similar trends are observed, but for all tests the fall in consensus *p-value* output is less pronounced and delayed to higher amplitudes. This can rationalised as four defect signals provide greater confidence of the presence of a defect than just two defect signals at the same amplitude.

Figure 7.1b compares the three detectors in a different test scenario: a single input *p-value* resulting from the varied amplitude is effectively corrupted by either 1 or 3 noise-like inputs. Again, the DFD is seen to behave as an intermediate between AND and OR consensus tests, and a transition from AND-like behaviour to more OR-like behaviour is visible as the amplitude increases. Note that the AND consensus test output does not vary as that test is undermined by any noise inputs - potentially dangerous in practice, as if any channel does not contain a significant defect response (for instance, due to the defect reflectivity being highly directional, or maybe a break-down in probe coupling) the defect would be missed. As the number of noise inputs is increased the DFD performance is reduced but not completely undermined, whilst the other two tests are essentially unaffected. This shows that the DFD has some resilience to corrupting noise inputs, but the reduction in sensitivity does mean that if there is prior knowledge that a defect response will be very poor in a particular channel, for example due to the directionality of the defect or probe near-field effects, that channel should be manually excluded from the fusion process for improved detection performance. A practical example of such a scenario is considered in Section 7.3.3.



(a) **(b)**

Figure 7.1: A comparison the Data Fusion Detector (DFD) compared with AND and OR operators, acting on the same *p-value* inputs. In (a) the detectors receive 4 or 2 inputs of the same amplitude (but differing *p-values*, as drawn from different distributions). Qualitatively, the DFD is seen to transition from behaving like an AND test to behaving like an OR test as the amplitude increases. The fall in consensus *p-value* output is less pronounced and delayed to higher amplitudes if only 2 rather than 4 inputs are provided. In (b) a single input *p-value* resulting from the varied amplitude is effectively corrupted by either 1 or 3 noise-like inputs. The DFD is seen to not reach as low outputs when more noise is added, whereas AND and OR combinations are almost insensitive to the increase plotted.

7.2.2 Output under the null hypothesis

In this section we consider the output of the Data Fusion Detector (DFD) under null hypothesis conditions to confirm the effectiveness of the *Monte-Carlo*-based evaluation described in the previous chapter when operating on real industrial data. To impose null hypothesis conditions the *resels* of each channel were permuted randomly circumferentially (leaving the *resel* amplitude distributions unchanged). Additionally, to avoid distorting effects the importance sampling used elsewhere in the *Monte Carlo* evaluation of the fusion was disabled [13, 120], and the number of samples taken instead increased to 500 from 200 used elsewhere.

Complete sets of fused *p-values* were computed, both using the DFD with the default semi-parametric data model and then switching to the purely discrete model (see previous chapter). In Fig. 7.2 the cumulative distributions of the outputs are shown and compared against a continuous uniform distribution over the [0,1] interval, as in many of the plots of the preceding chapter - but here for a larger, complete industrial input. A slight oscillation

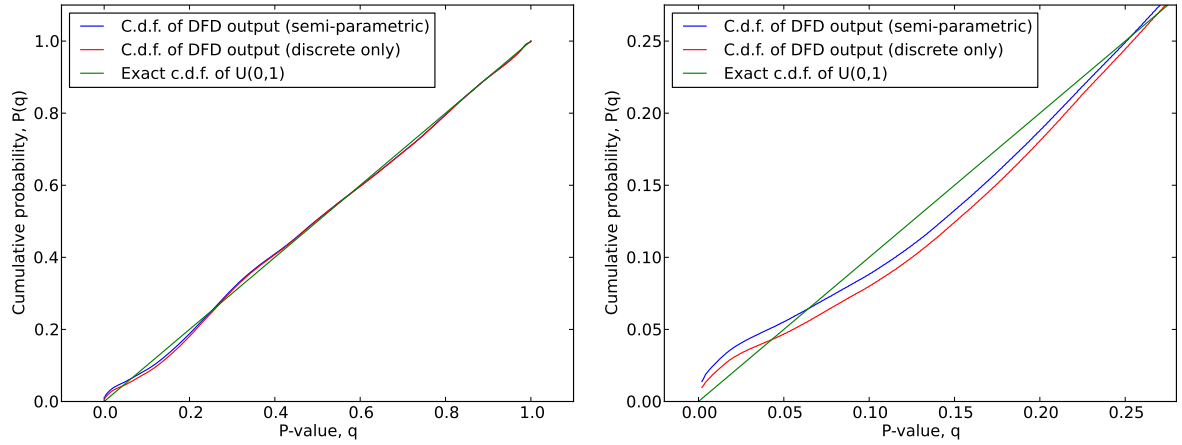
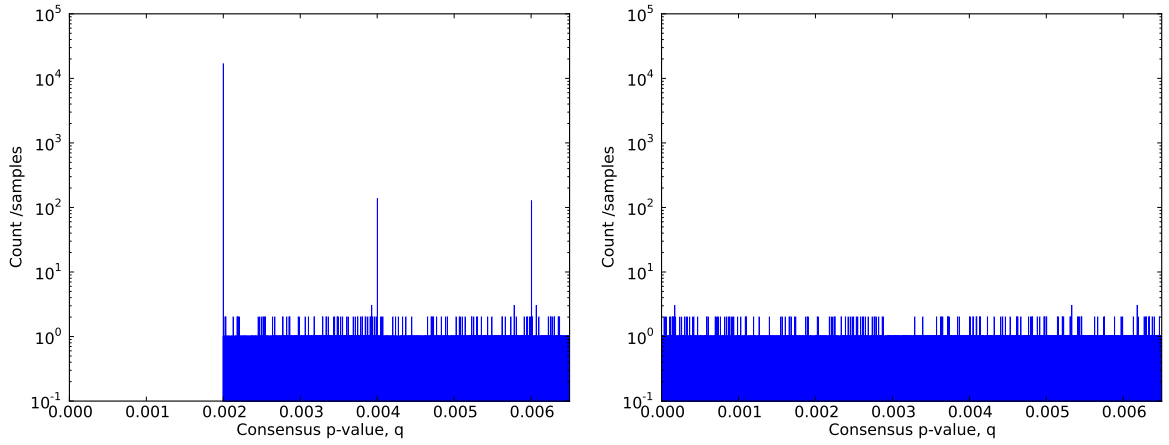


Figure 7.2: The behaviour of the Data Fusion Detector (DFD) under imposed null hypothesis conditions. The c.d.f. of the output *p-values* is compared against that of a continuous uniform distribution over [0,1] interval, for both the semi-parametric data model and the purely discrete alternative model. A slight oscillation of the DFD output curves around the theoretical line is detectable in (a), and although this is more apparent in the zoomed version of the plot shown in (b), the correspondence is very good, for both data models. This provides a practical confirmation of the results of the previous chapter.

of the DFD curves around the theoretical line is detectable in (a), and although this is more apparent in the zoomed version of the plot shown in (b), the correspondence is very good, for both data models. This confirms that the compensation for *p-value* quantisation is effective in practice and the distortion introduced by the semi-parametric data model is very limited.

The plots of Fig. 7.3 illustrate the structure of the consensus *p-values* resulting from the evaluation of the *Monte Carlo* (M.C.) generated survival functions, and the benefits of applying importance sampling. The data are presented as bar charts - not histograms - so that the strictly discrete data structure is exposed. In (a) several interesting features are identifiable: the extremely large spikes at multiples of 0.002, the 0.002 cut-off, above which the values are continuous to a very close approximation - with only the duplicate appearance of a few values revealing that the output is technically discrete (the probability of even a single duplicate for continuous values, taking into account computational numerical precision and the number of samples, is infinitesimal). The spikes are the result of most samples falling on the M.C. sampled locations (by definition of M.C.), and these have an associated probability that is an integer multiple of the reciprocal of the number



(a) **(b)**

Figure 7.3: Bar charts - not histograms - of the DFD output over a small section of the range to illustrate the data structure resulting from the *Monte Carlo* (M.C.) processing. For (a) importance sampling was disabled, and the number of M.C. samples set to 500. Several interesting features are identifiable: the extremely large spikes at multiples of 0.002, the 0.002 cut-off, above which the values are continuous to a very close approximation - with only the duplicate appearance of a few values revealing that the output is technically discrete. In (b) importance sampling was re-enabled, and the M.C. sample number reduced to the default of 200. The previously seen cut-off and spikes are no longer present, despite the reduced sample number.

of M.C. samples used - 500 here. The outputs between the spikes result from interpolating the M.C. survival function, and the extrapolation of this explains the cut-off - without importance sampling it is not possible to reach a probability below the reciprocal of the M.C. sample number. To generate plot (b) importance sampling was re-enabled, and the M.C. sample number reduced to the default of 200. Even with this smaller number of samples (and hence reduced processing time) the plot shows that extremely low output *p-values* have been reached, and the spikes have also disappeared such that the output is more uniform than without importance sampling. Thus the practical significance of this technique has been demonstrated.

7.3 Rotor Bore Inspection

In this section we examine several possible indications in a complete rotor bore dataset, and use these to assess the performance of the novel Data Fusion Detector (DFD). Specifically, this is compared with a basic Amplitude Threshold Detector (ATD), as used

by Rolls-Royce in their disk inspection (see also results from that in Section 7.4), that considers the highest raw amplitude detected across all channels at all locations. Additionally, for each individual channel we consider the Local Empirical Noise (LEN) detector that relies on the local data model underlying the DFD to normalise amplitudes to their local environment - an early step in the DFD evaluation (see Section 6.5).

All results presented are for the fusion of all *resels* of all four available ultrasonic channels. Note that the *resels* in the far-focus compression wave channel that lie beyond the domain of the other channels, such that no fusion is possible (see Fig. 6.6 of previous chapter) have been excluded from the analysis. Although in principle the fusion system is compatible with fusion-free *p-values*, these are of limited interest for demonstrating the benefits of the DFD. Moreover, care should be exercised as in principle further quantisation compensation may be required to make these comparable with fused outputs (see previous Chapter), especially given the large number of fusion-free *p-values* in this case.

7.3.1 A first possible indication

In the absence of known indications in the dataset we could examine, we can instead identify several positions in the dataset from the DFD output and analyse these as possible indications. We first search for the very lowest consensus *p-value* (out of 1.2×10^6 values) - corresponding to the highest indication severity probability. The spatial location of this *resel* is marked in Fig. 7.4, alongside the locations of the *resels* with the next nine lowest *p-values* and the location of the *resel* of the very highest amplitude across all channels. Some of these points will be considered further in the later sections.

The A-scans associated with the *resels* contributing to the identified very lowest DFD output are presented in Fig. 7.5. Note that because of its depth the location of interest lies beyond the domain of the short-focus compression channel, so only the remaining three channels provide inputs. In each A-scan the location of the *resel* of interest is highlighted, critical given the complex geometry of the inspection. Even so, looking at just the A-scans it is hard to understand why this location has been selected by the detector, especially as each A-scan contains other points that are similar or higher in amplitude than the *resel* of interest. However, looking at just these A-scans can be misleading, as there is no way

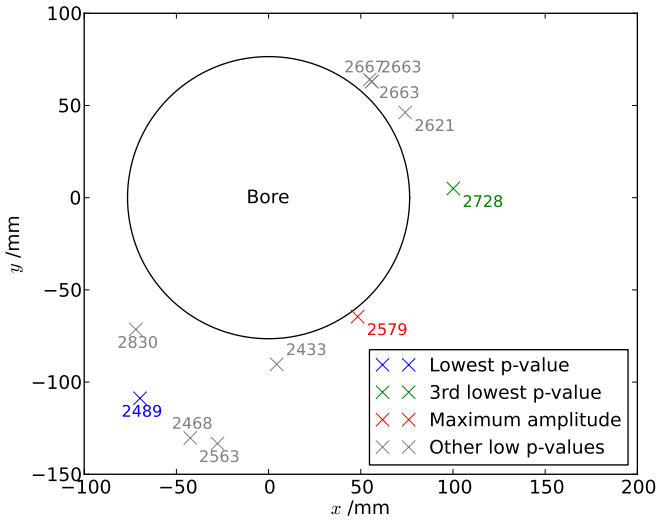


Figure 7.4: The positions of some possible indications, as seen looking down the length of the bore, with the axial locations indicated as annotations, in millimetres. The locations of the ten lowest *p-values* are plotted, and amongst these the very lowest and the 3rd lowest are highlighted as they will be considered further in these results. Also shown is the location of the maximum amplitude recorded across all channels, as this location will be considered in Section 7.3.3.

to judge depth-dependent effects.

More useful for our purposes are the *resel* amplitude distributions that the contributing *resels* lie in. These are shown in Fig. 7.6 for clockwise shear, counter-clockwise shear and far-focus compression channels, in (a), (b) and (c), respectively. In each the amplitude at the possible indication is marked. The indication amplitude is the highest amplitude in the distribution in two of the input distributions, and second highest in the third. This provides a initial explanation of why this location should be of interest to the DFD.

To further understand the operation of the fusion mechanics, we consider how each amplitude distribution is used to convert the amplitude at the candidate indication to a *p-value*. The conversion is illustrated in Fig. 7.7, where for the distribution of each channel the Local Empirical Noise (LEN) *p-values* are plotted, highlighting the *p-value* in each channel contributing to the possible indication with a circle. Note that the contributing *p-value* in the clockwise shear channel is lowest, as should be expected, given that the underlying amplitude is most clearly separated from the remainder of the distribution in Fig. 7.6. This is largely the result of the *Pareto* tail-fit of the semi-parametric data model (see preceding chapter), which also gives rise to the straight-line segments seen at higher amplitudes in Fig. 7.7.

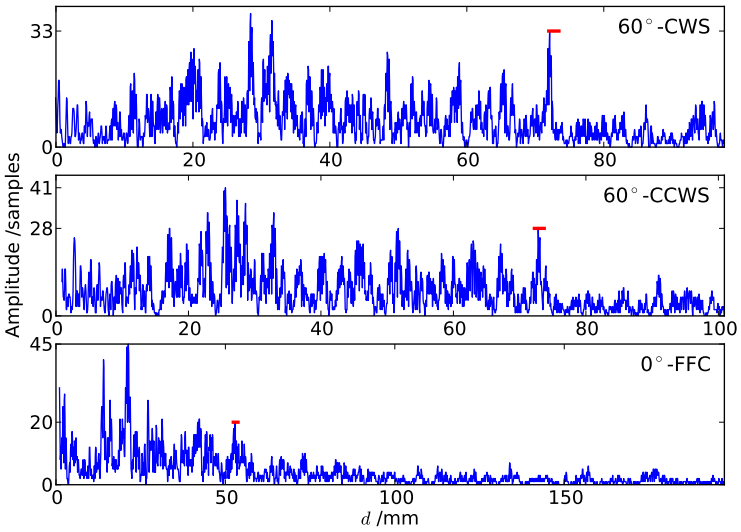


Figure 7.5: The A-scans for the location of the very lowest consensus *p-value*. Only three channels contribute to this particular location as it lies at a depth beyond the domain of the short-focus compression channel. The red lines mark the extent and amplitude of the *resel* in question. Note the differences in depth along the A-scan of that *resel*, resulting from the complex geometry of the inspection.

Having concluded that the location identified by the DFD as having the lowest overall *p-value* is not an unreasonable candidate for a defect, we now treat that location as a known indication. This then permits the computation of the receiver operating characteristic (ROC), by the means described at the end of the preceding chapter, and hence a rigorous comparison of the DFD with a more conventional detector such as the ATD. The ROC curves thus calculated are shown in Fig. 7.8. As ROCs may be considered to be the *Pareto* points of a multi-objective optimisation (see Section 4.7), only the convex hull of the originally computed ROC points is shown, where in fact every point on the hull plotted may be reached by stochastic interpolation between the end points of that line segment [82]. The plot features a logarithmic scale for the probability of false alarm (PFA) to accommodate the wide range of values observed, but is limited to high probability of detection (POD) values, as only these are of interest. Unless stated otherwise, all subsequent ROC plots are presented in this manner.

Studying the plot more closely shows that the LEN detectors, using the individual channels, are seen to each already outperform the ATD, by offering substantially lower PFA for an equal POD. Combining these LEN inputs into the DFD provides a substantial further improvement, lowering the PFA by about two orders of magnitude at the same

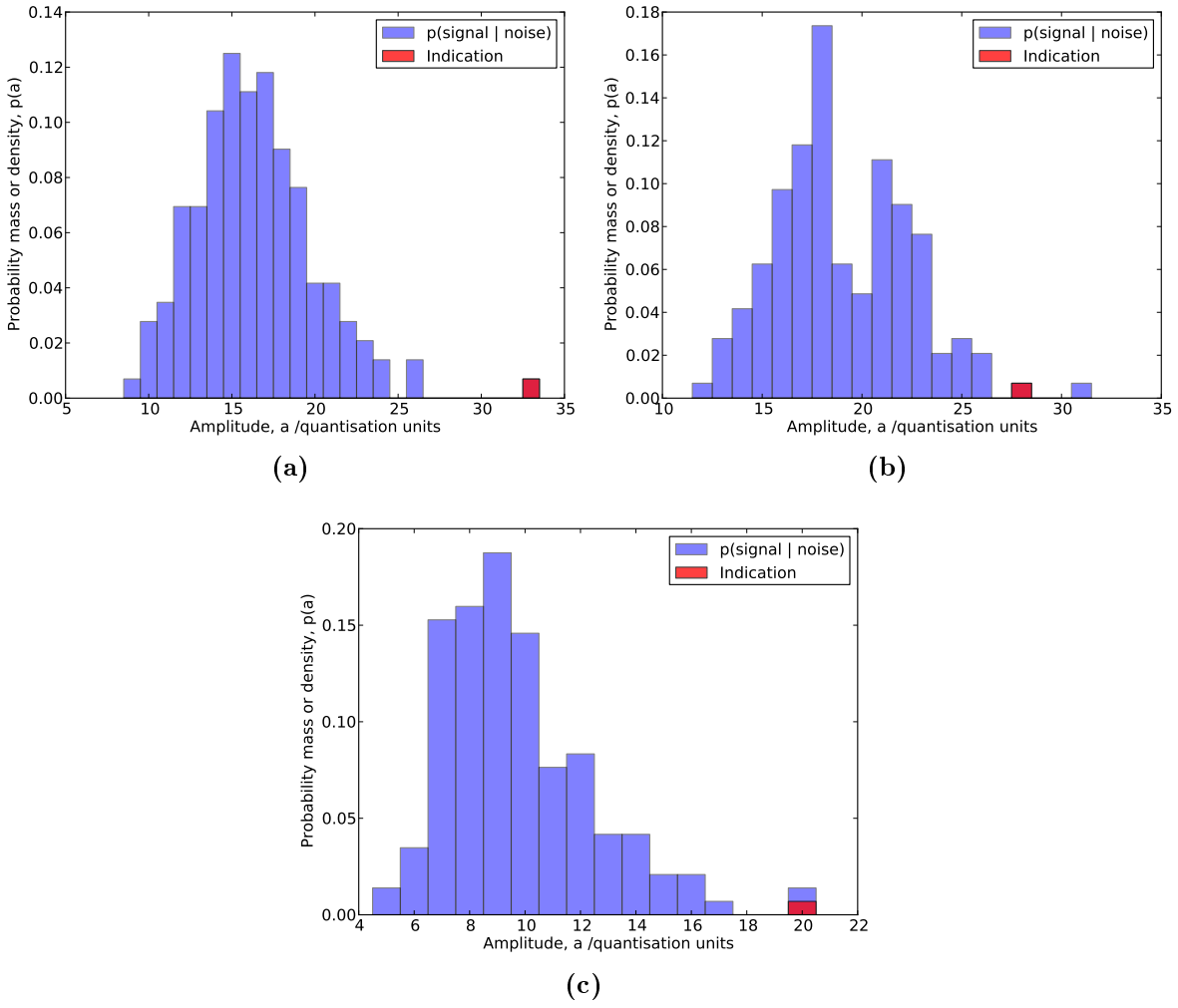


Figure 7.6: The *resel* amplitude bar charts for the *resels* contributing to the very first data fusion detector output. Plots (a), (b) and (c) relate to the clockwise shear, counter-clockwise shear and far-focus compression channels, respectively. In each the amplitude at the possible indication is marked. The indication amplitude is the highest amplitude in the distribution in two of the input distributions, and second highest in the third.

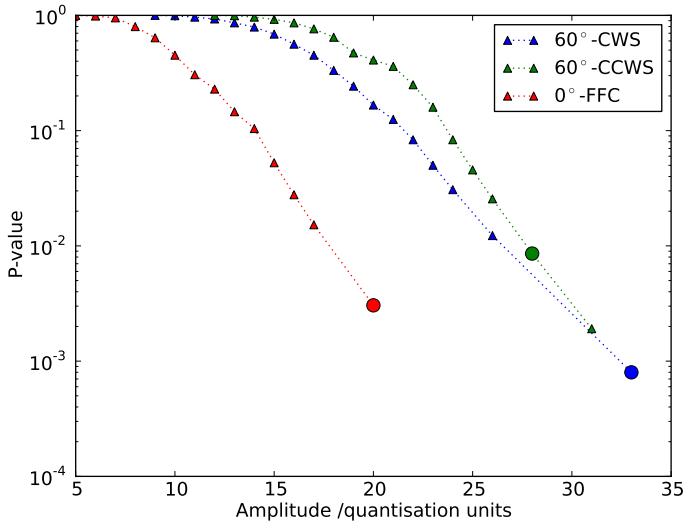


Figure 7.7: Converting the *resel* amplitude distributions contributing to the location with lowest consensus *p-value* to Local Empirical Noise (LEN) *p-values*. The *p-value* in each channel at the *resel* of the possible indication is marked by a circle. The straight-line segments seen at higher amplitudes are a consequence of the *Pareto* tail-fit (see previous chapter).

high POD compared with the best of the LEN detectors (for the clockwise shear channel). This provides an impressive demonstration both of the data model (embodied in the LEN detectors) and the benefits of subsequent fusion processing.

7.3.2 Another possible indication

We now consider an alternative possible indication in the dataset, this time selecting the location identified by the DFD as having the third lowest associated consensus *p-value*, or equivalently third highest indication severity probability. The physical location of this point was already marked in Fig. 7.4. The analysis here exactly matches that of Section 7.3.1, starting with a presentation of the relevant A-scans in Fig. 7.9, where red lines mark the extent and amplitude of the *resel* in question. Unlike in the previous case, all four channels contribute here.

The *resel* amplitude distributions are shown in Fig. 7.10, for clockwise shear, counter-clockwise shear, short-focus compression and far-focus compression channels, in (a), (b), (c) & (d), respectively. Examining the marked amplitude at the possible indication in each plot reveals that the recorded amplitude is higher in the shear channels than compression channels, but the indication amplitude is the highest amplitude in both compression chan-

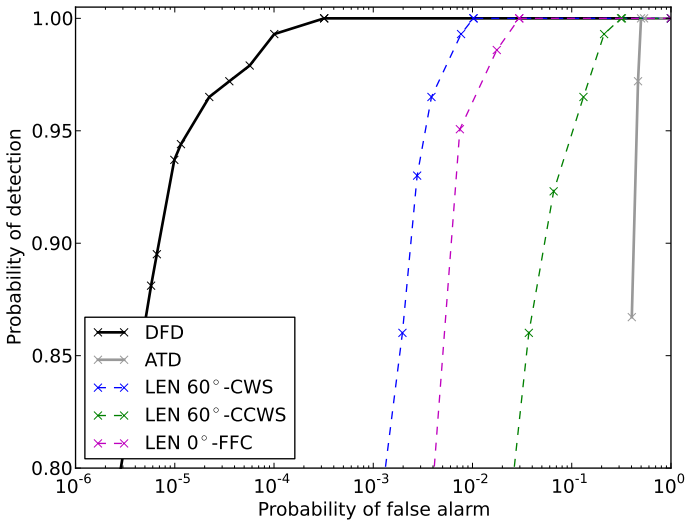


Figure 7.8: The ROC curves relating to the location identified as having the lowest overall consensus *p-value*. The Local Empirical Noise (LEN) detectors, using the individual channels, are seen to each already outperform the Amplitude Threshold Detector (ATD). Combining these inputs into the Data Fusion Detector (DFD) provides a substantial further improvement, lowering the probability of false alarm (PFA) by about two orders of magnitude at the same, high, probability of detection (POD).

nel distributions, but considerably lower by rank in the shear channel inputs. Therefore in this case the local amplitude normalisation of the DFD is particularly significant.

Figure 7.11 shows this conversion of amplitudes to LEN *p-values* for the contributing distributions, highlighting the *p-values* that are then fused for the location of interest.

As the location suggested by the DFD as having the third highest indication severity probability does not appear to be unreasonable, we now treat this as a known indication to enable the computation of ROCs. Note that no other known indications are included in this calculation here, so that the location labelled a known indication in Section 7.3.1 here contributes to the false calls. The resulting curves for DFD, ATD and the LEN detectors of the four individual channels are plotted in 7.12. The two compression channel LEN detectors outperform the ATD, but the two shear channel LEN detectors are worse, as might have been expected from Fig. 7.10. Combining the LEN *p-values* in the DFD provides a further improvement over any of the inputs, even though two of the channels are individually poor, thereby significantly outperforming the ATD. This provides a further convincing demonstration of the DFD, and specifically illustrates the robustness of the detector to noise-like inputs.

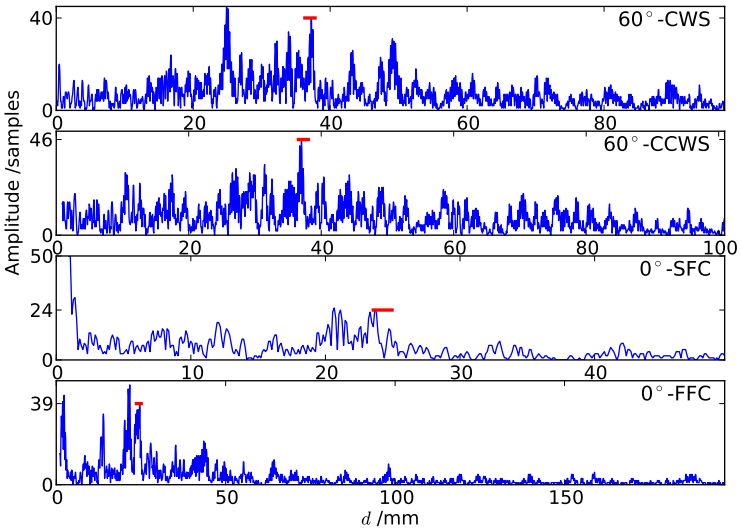


Figure 7.9: The A-scans for the location of the third lowest consensus *p-value*. Here all four channels contribute. The red lines mark the extent and amplitude of the *resel* in question. Note the differences in depth along the A-scan of that *resel*, resulting from the complex geometry of the inspection.

7.3.3 The location of the highest amplitude

Next, to provide a strongly contrasting example, we consider the location of the very highest amplitude across all channels. By definition this is the first output of the ATD, whereas this location is 7191 (out of 1.2×10^6) in the ranked output of the DFD. The physical location of this point was shown earlier in Fig. 7.4.

The relevant A-scans are presented in Fig. 7.13. The very highest amplitude is 110 quantisation samples, as seen in the short-focus compression channel. Note that the amplitudes of the *resels* of the other channels are very much lower, both in absolute terms and relative to the remainder of the A-scan. The shallow depth indicated for the *resels* of interest will feature in the discussion later.

Figure 7.14 shows the *resel* amplitude distributions relating to the location of the highest overall amplitude. The indication amplitude is by far the highest in the short-focus compression channel, not only in absolute terms (as already seen in Fig. 7.13), but also relative to the rest of the distribution. Especially the shear channels only provide very low amplitudes at the location of interest, at least compared with the remainder of the distributions.

The conversion of these distributions to LEN *p-values* is illustrated in Fig. 7.15. As

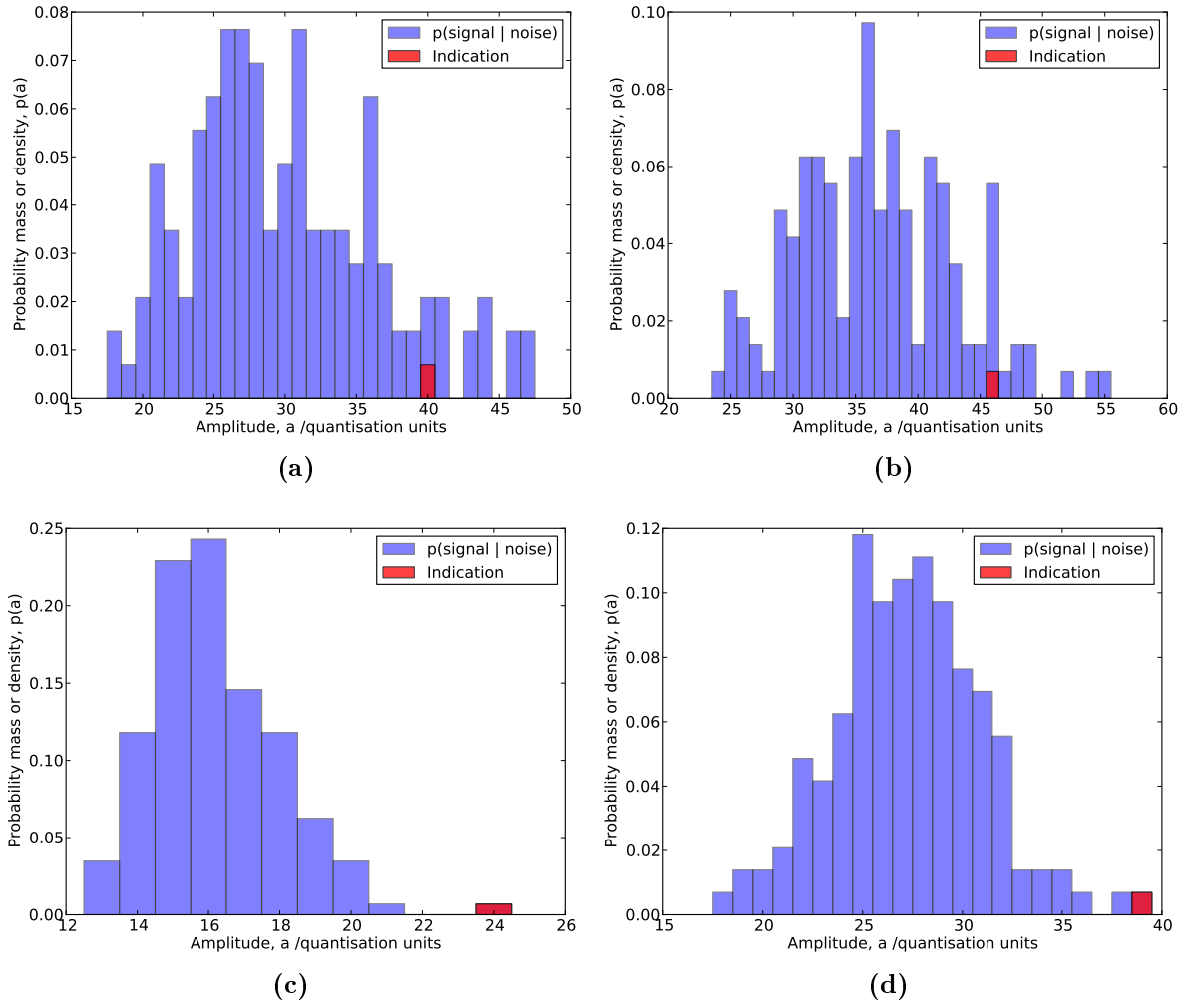


Figure 7.10: The *resel* amplitude bar charts for the *resels* contributing to the third data fusion detector output. Plots (a) - (d) relate to the clockwise shear, counter-clockwise shear, short-focus compression and far-focus compression channels, respectively. In each the amplitude at the possible indication is marked. The indication amplitude is the highest amplitude in the compression channel distributions, but considerably lower by rank in the shear channel inputs.

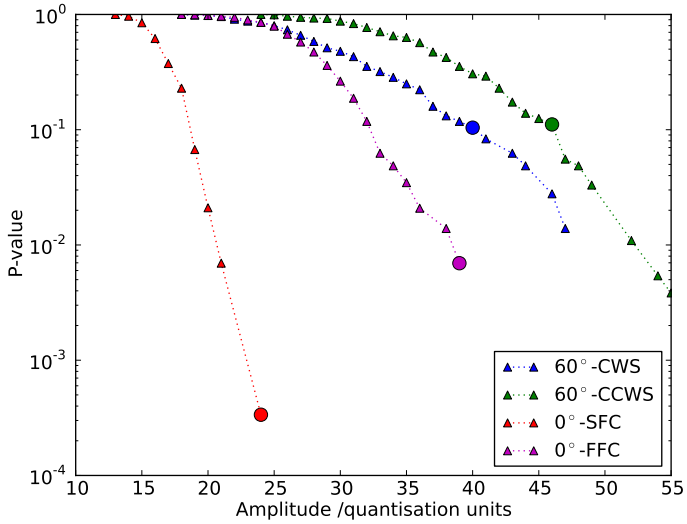


Figure 7.11: Converting the *resel* amplitude distributions contributing to the location with third lowest consensus *p-value* to Local Empirical Noise *p-values*. The *p-value* in each channel at the *resel* of the possible indication is marked by a circle.

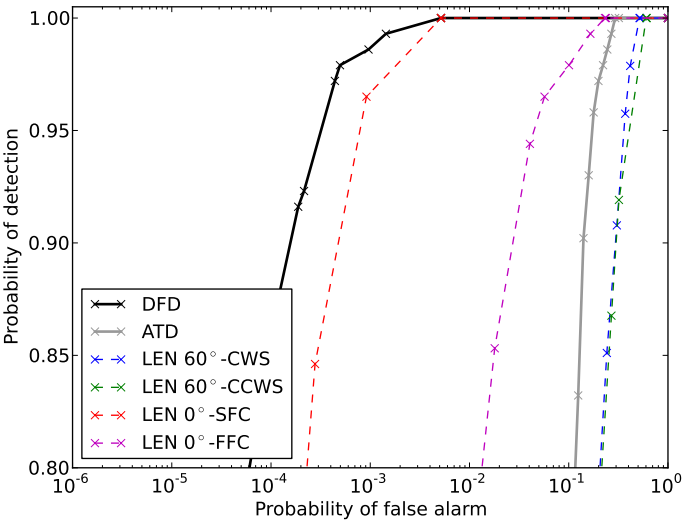


Figure 7.12: The ROC curves relating to the location identified as having the third lowest overall consensus *p-value*. Two of the LEN detectors outperform the ATD, but the two for the shear channels are worse. Combining the LEN *p-values* in the DFD provides a further improvement over any of the inputs, even though two of the channels are individually poor, convincingly outperforming the ATD.

might have been expected given the amplitude distributions, the *p-values* at the location of interest for the shear channels are so high that they are hard to identify. The very high amplitude in the short focus compression channel responsible for this location being selected in the first place does give a very low *p-value* as should have been expected. However, it is perhaps surprising that despite the extremely high amplitude the *p-value* reached is in fact rather less low than some of those seen previously, for example in the short-focus channel in Fig. 7.11. This is the result of the tail-fit in the semi-parametric data model (see previous chapter) and hence the shape of the distribution being fitted to. Comparing the underlying amplitude bar charts allows the observed differences to be rationalised. Inevitably though, some computed fits will be more appropriate than others and more generally there is scope for debate about how exactly the tail-fit should behave.

As for the previous possible indications, we compute the ROCs, having declared the location of the highest overall amplitude to be a known indication. The output curves are shown in Fig. 7.16. In line with the distributions observed, here the short-focus compression LEN detector and the ATD outperform the DFD. This is caused by the three other LEN detectors performing poorly, so inputs from these to the DFD undermine its performance. Therefore this is effectively an example of a single high amplitude, defect-type input to the DFD being corrupted by three noise-type inputs, much as in Section 7.2.1. This however does not necessarily mean that the DFD is performing incorrectly - it is conceivable that the large amplitude signal in the short-focus compression channel has a benign cause, and the indication severity probability of that location should indeed be down-weighted given that no comparable signal is seen in the other channels. On the other hand it is perhaps more likely, given the shallow depth of the identified location, that only the short-focus compression channel is really able to get a significant reflection from a flaw at that position. If this were the case, this would suggest that at least the shear channels should not be included in the fusion in this spatial region as it is known *a priori* that these will only corrupt the output. Such a change would substantially improve the DFD performance (in line with Fig. 7.1b).

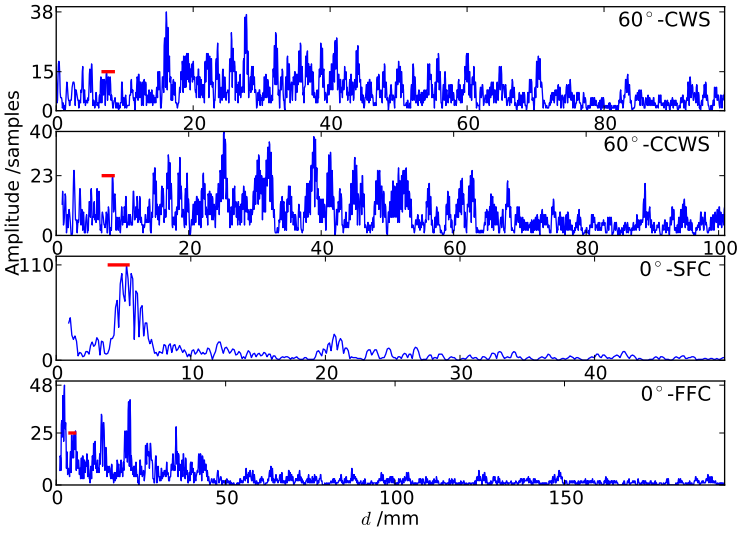


Figure 7.13: The A-scans for the location of the very highest amplitude observed over all input channels. The red lines mark the extent and amplitude of the *resel* in question. The very highest amplitude is 110 quantisation samples, seen in the short-focus compression channel. Note that the amplitudes of the *resels* of the other channels are very much lower, both in absolute terms and relative to the remainder of the A-scan.

7.3.4 Artificial indication insertion - 1

Especially in the absence of real known indications, it can be informative to insert an artificial target signal in the data. A target signal was therefore inserted into the raw data and this examined as a known indication. The position of this was chosen to lie in the same *resel* rings examined in Section 7.3.3 for the overall highest amplitude observed, at approximately 90° to the previously considered location. Therefore the distributions seen in Fig. 7.14 remain relevant. The amplitude of the signal inserted was 32 quantisation samples across all channels, a reasonable but not extremely high value given these original amplitude distributions. Note the parallels between this scenario and that considered in Fig. 7.1a.

The A-scans containing the artificial target signal are shown in 7.17, where in each the amplitude of the *resel* at the considered location is 32 quantisation units, as should have been expected. The fact the inserted signal is an unphysical *Dirac* delta function as shown does not detract from the analysis given the discretisation provided by the *resels*. The corresponding *resel* amplitude distributions are provided by Fig. 7.18. These plots only differ from those in Fig. 7.14 by the way one of the samples has been replaced by one

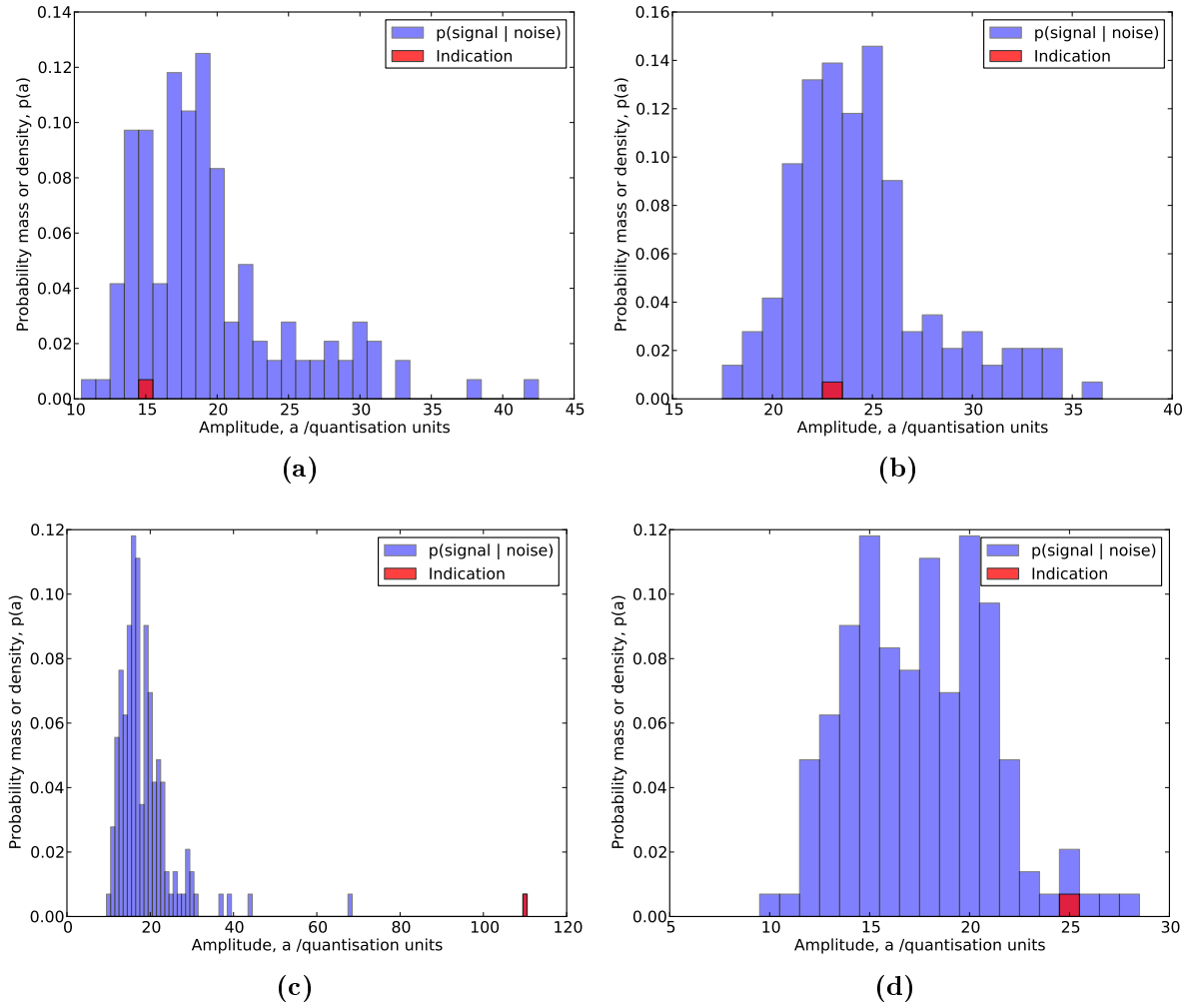


Figure 7.14: The *resel* amplitude bar charts for the *resels* relating to the location of the highest amplitude across all channels. Plots (a) - (d) relate to the clockwise shear, counter-clockwise shear, short-focus compression and far-focus compression channels, respectively. In each the amplitude at the possible indication is marked. The indication amplitude is by far the highest in the short-focus compression channel, but much lower in the other inputs, especially the shear channels.

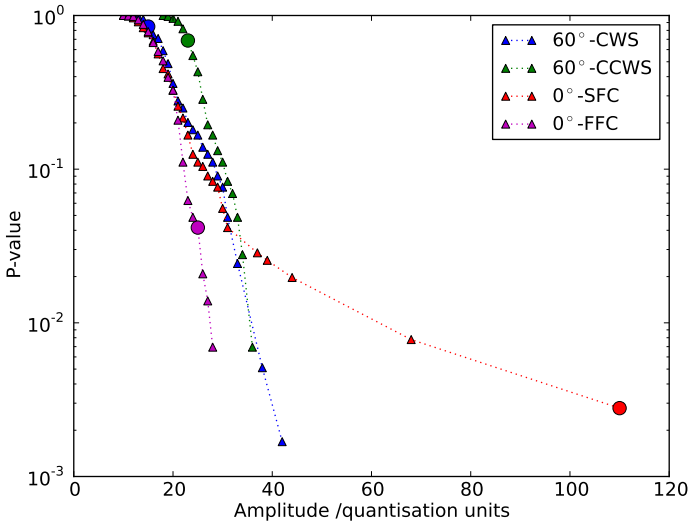


Figure 7.15: Converting the *resel* amplitude distributions associated with the location of the very highest amplitude seen across all channels to Local Empirical Noise *p-values*. The *p-value* in each channel at the *resel* of the possible indication is marked by a circle.

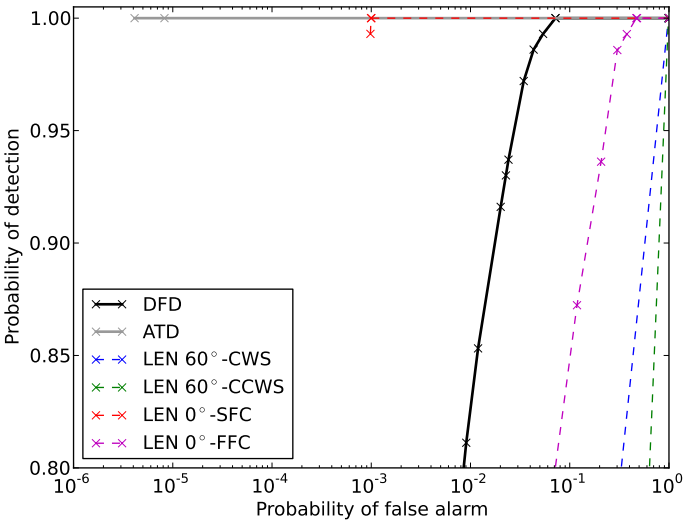


Figure 7.16: The ROC curves for the location of the very highest amplitude seen across all channels. In this case the short-focus compression LEN detector and the ATD outperform the DFD. This is caused by the three other LEN detectors performing poorly, so inputs from these to the DFD undermine its performance.

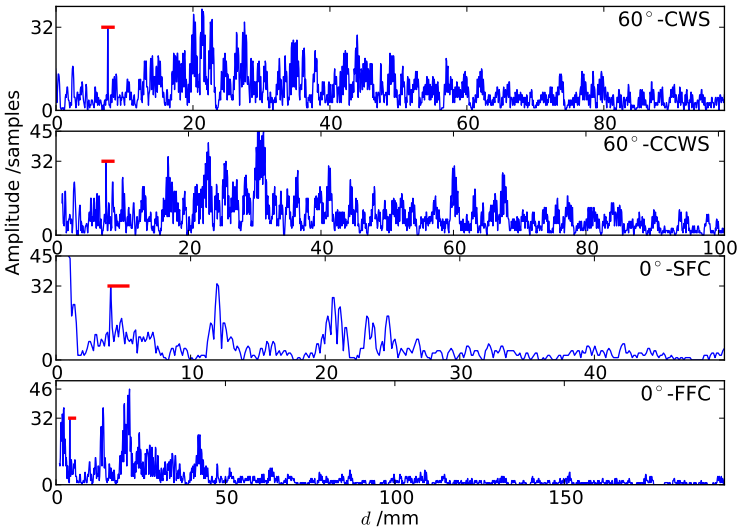


Figure 7.17: The A-scans showing the cross-channel artificial insertion of a target signal. The red lines mark the extent and amplitude of the *resel* of interest - all are at the amplitude of the inserted signal, 32 quantisation samples.

at an amplitude of 32 quantisation samples, and by the fact that amplitude is now the amplitude at the indication. Figure 7.19 illustrates the conversion of these distributions to *p-values*, and although similar to Fig. 7.15, it emphasises how the identical amplitudes at the indication are converted to quite different *p-values* depending on the surrounding distribution.

The ROCs for the artificial target indication are plotted in Fig. 7.20. Much as in Fig. 7.8 for the very first possible indication considered, the LEN detectors all already outperform the ATD (especially in the far-focus compression channel), but the DFD provides a further significant improvement over any of the alternatives, again demonstrating the benefits of fusion.

7.3.5 Artificial indication insertion - 2

In the previous chapter the possible problem of *resel* edge effects was mentioned (Section 6.8.2). We now change the position of the target reflector by just a couple of millimetres from that seen in Section 7.3.4, leaving everything else unchanged, to demonstrate this effect.

The corresponding A-scans are plotted in Fig. 7.21. These do bear a resemblance to those

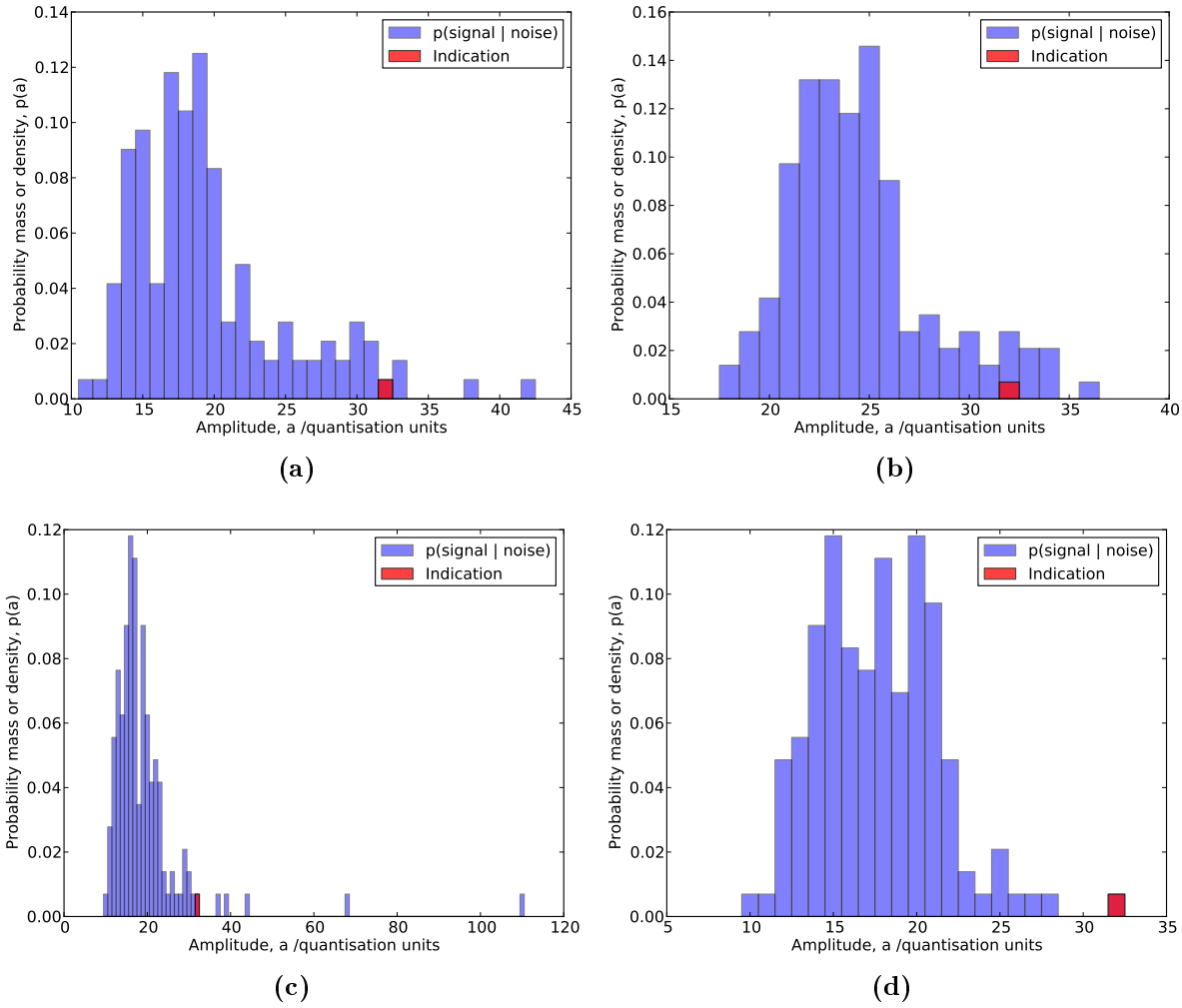


Figure 7.18: The *resel* amplitude bar charts for the *resel* containing the artificially inserted target signal. Plots (a) - (d) relate to the clockwise shear, counter-clockwise shear, short-focus compression and far-focus compression channels, respectively. In each the amplitude at the indication is marked, and essentially only this element distinguishes these plots from their counterparts in Fig. 7.14. Only in the far-focus compression distribution is the inserted amplitude of 32 quantisation units particularly high.

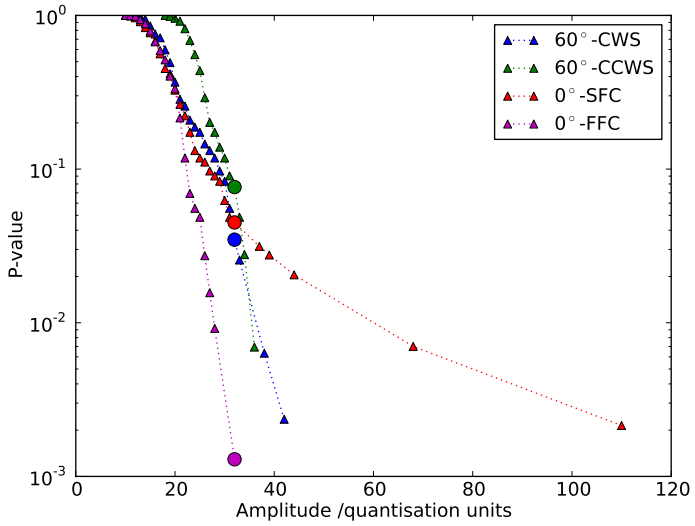


Figure 7.19: Converting the *resel* amplitude distributions associated with the location of the inserted artificial target signal to Local Empirical Noise *p-values*. The *p-value* in each channel at the *resel* of the possible indication is marked by a circle - note that in this case these circles all lie at an amplitude of 32 quantisation units, but have quite different *p-values*.

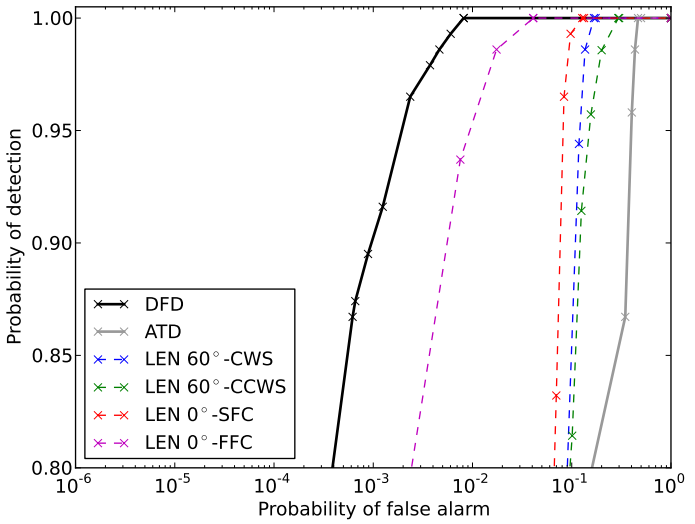


Figure 7.20: The ROC curves for the location at which an artificial target signal at an amplitude of 32 quantisation units was inserted across all channels. Much like the ROC computed for the first detector output (Fig. 7.8), the LEN detectors all already outperform the ATD (especially in the far-focus compression channel), but the DFD provides a further significant improvement over any of the alternatives.

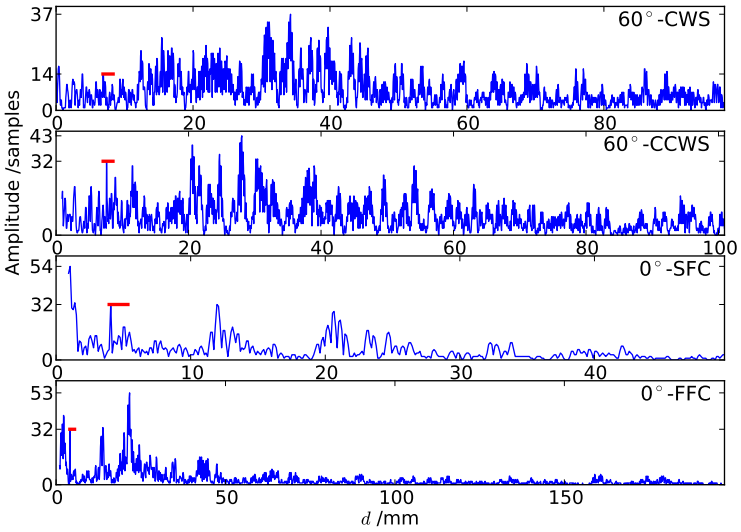


Figure 7.21: The A-scans showing the cross-channel artificial insertion of a target signal where that signal in one channel has been associated with an adjacent *resel*, rather than the one seen here. The red lines mark the extent and amplitude of the *resel* of interest. Note that in this case the inserted amplitude of 32 quantisation samples is not seen in the uppermost, clockwise shear channel.

seen for the original target signal location, in Fig. 7.17. The most significant difference, symptomatic of the *resel* edge problem, is that the inserted amplitude of 32 quantisation samples is not seen in the clockwise shear channel, as in that channel the inserted signal has fallen in the circumferential neighbour of the *resel* considered for the fusion here. As the *resel* amplitude distributions and the subsequent conversion to LEN *p-values* is barely different from the previous case, those plots are omitted here. Proceeding on to the ROCs shown in Fig. 7.22 and comparing these to those in Fig. 7.20, the LEN detector curves for the three unaffected channels are unchanged, but the LEN for the compromised clockwise shear channel is worse, such that the DFD in this case only just outperforms the best of the LEN detectors, in the far-focus compression channel. While the benefits of fusion are still evident, especially compared with the ATD, these results demonstrate how *resel* edge effects can significantly compromise DFD performance, as suggested in the preceding chapter.

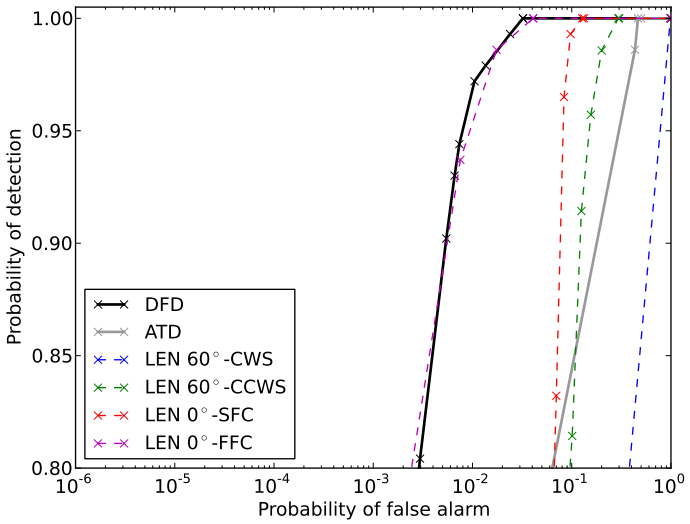


Figure 7.22: The ROC curves for the location at which an artificial target signal at an amplitude of 32 quantisation units was inserted across all channels, for the case where in one of the channels (clockwise shear) the target signal falls into an adjacent *resel* rather than the one for which the ROC is shown. Compared with Fig. 7.20, the LEN detector curves for the three remaining channels are unchanged, but the LEN for the compromised clockwise shear channel is worse, such that the DFD in this case barely outperforms the best of the LEN detectors, in the far-focus compression channel.

7.3.6 Subtracting a previous acquisition

Baseline subtraction [51, 52] has been mentioned previously as a means to improve detection sensitivity if a baseline dataset is available. As the rotor bore inspection is a repeated, in-service inspection, such a dataset may frequently be found, and it is therefore appealing to exploit this. The basic data subtraction system implemented is therefore applied here in an attempt to improve the detection of the artificial target signal seen in Section 7.3.5, which was compromised significantly by the inserted signal falling in an adjacent *resel* in one of the channels (the clockwise shear).

The baseline dataset used is a complete copy of the original dataset, prior to the insertion of the artificial signal, but distorted to a slightly incorrect set of registration parameters. This is designed to mimic an imperfect registration (due to inaccurate parameter choice, rather than the registration model being inadequate). As will be evident, the performance of the dataset comparison hinges critically on the quality of the registration. If this is poor, artefacts will be created by baseline subtraction that potentially swamp any signals of interest [53, 54]. The magnitude of the introduced distortions was chosen to be comparable

to those in the best registration results obtained in Chapter 5, for the alignment of two copies of the same channel across two acquisitions. It should therefore be noted that this represents an extremely good alignment, which would be very demanding to achieve in practice, and is beyond the current full dataset registration capability.

The relevant A-scans in the dataset post-subtraction are shown in Fig. 7.23. These are significantly different from those seen without subtraction, in Fig. 7.21, and the effect of the subtraction is seen to vary considerably from channel to channel: the counter-clockwise shear channel provides an example of essentially perfect subtraction, but the others include significant artefacts of the subtraction due to less perfect registration in these channels.

Figure 7.24 provides the ROCs computed, to be compared with the equivalent case without a subtraction of a secondary acquisition shown in Fig. 7.22. Now, the LEN detectors for clockwise shear and short focus compression channels are worse than before, but the other two have improved significantly - allowing the DFD to ultimately provide a substantial improvement over the non-baseline-subtracted scenario. The fact this is possible despite the two very poor LEN inputs provides further evidence of the robustness of the DFD to unhelpful inputs. These results also provide an insight into the potential capabilities of a baseline subtraction system in this context, and may motivate the implementation of a more advanced dataset comparison technique than subtraction [92] and the further development of the registration.

7.4 Disk Inspection

The otherwise extensive examination of the performance of the Data Fusion Detector (DFD) for the rotor bore application in Section 7.3 was lacking an analysis using known real indications, as no such suitable targets are available for that inspection. Therefore a selection of detection results obtained for the disk inspection application described in Chapter 2 is reproduced from Tippetts' thesis [13] to demonstrate detector performance on real defects.

The results found in this section are based on data acquired from scanning Rolls-Royce's seeded defect disk 5, which is one of several test-pieces containing realistic defects after

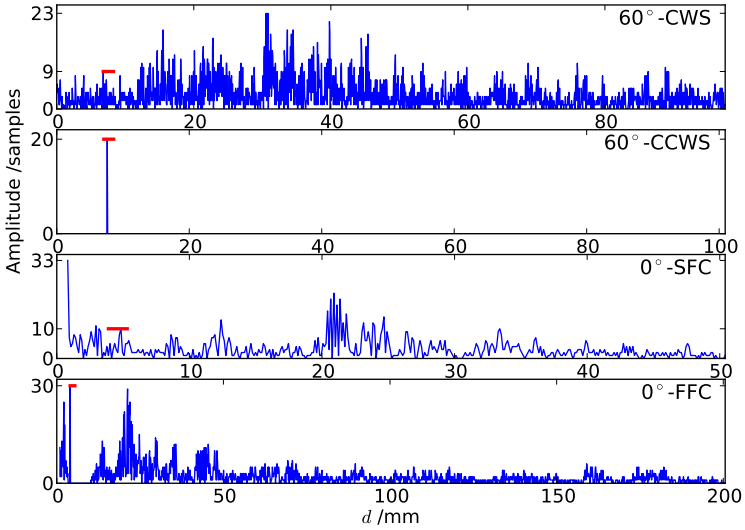


Figure 7.23: The same A-scans seen in Fig. 7.21, complete with artificial target signal, after subtracting out a second, baseline acquisition. The effect of the subtraction is seen to vary considerably from channel to channel here, with the counter-clockwise shear channel providing an example of essentially perfect subtraction, but the others including artefacts of the subtraction due to imperfect registration.

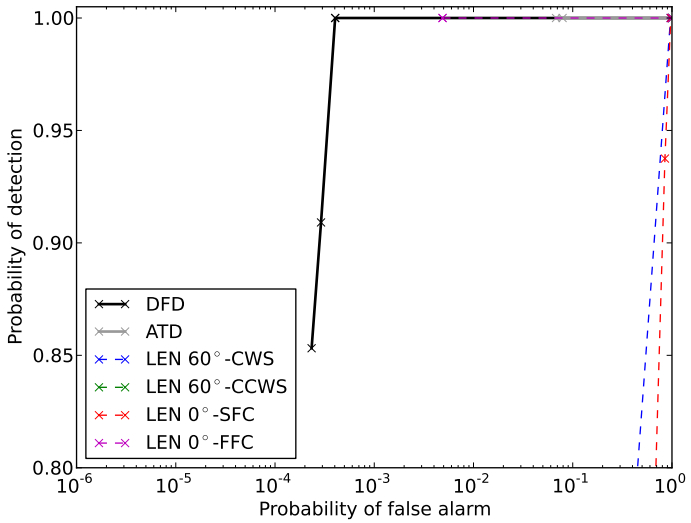


Figure 7.24: The ROC curves for the location at which an artificial target signal at an amplitude of 32 quantisation units was inserted across all channels, for the case where in one of the channels (clockwise shear) the target signal falls into an adjacent *resel* rather than the one for which the ROC is shown, but the data fusion includes a baseline subtraction pre-stage. Several of the lines are obscured by their coincidence with the DFD curve. Compared with the equivalent case without a subtraction of a secondary acquisition shown in Fig. 7.22, the LEN detectors for the clockwise shear and the short focus compression channels are worse than before, but the other two have improved significantly. This allows the DFD to ultimately provide a substantial improvement over the scenario without baseline-subtraction.

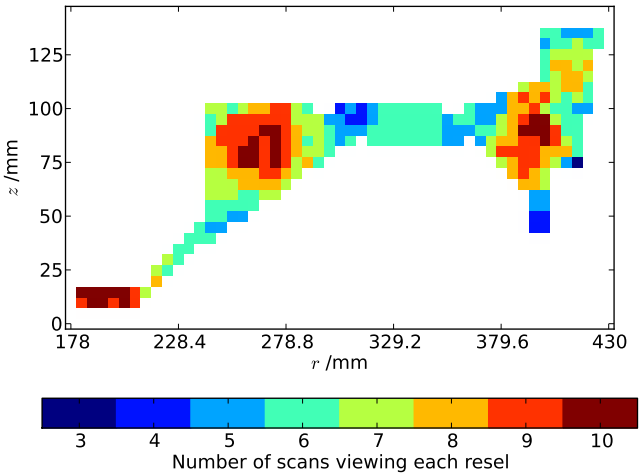


Figure 7.25: The coverage map for the disk inspection, counting the number of times each *resel* in a radial cross-section of the component is insonified across all scans collected. The disk axis of rotational symmetry is the z-axis, and the radius r is measured from this. Note that every *resel* in the cross-section is scanned at least 3 times, and some are viewed up to 10 times. This means there are significant opportunities for data fusion.

having been forged from contaminated billet. The disk in question has five known point-like defects from contaminant inclusions, identified by traditional ultrasonic and radiographic methods, including an ultrasonic Amplitude Threshold Detector (ATD), with a threshold set in line with the currently deployed Rolls-Royce inspection procedure (at 64 amplitude quantisation samples). A schematic diagram, including approximate locations of the known indications considered here, was provided in Fig. 2.5b. The disk is considered to be an outstanding test piece, providing highly realistic targets. Further details of the test component are given in Tippetts' thesis.

The coverage map for the disk inspection, counting the number of times each *resel* in a radial cross-section of the component is insonified across all scans collected, is presented in Fig. 7.25. The plot shows that every *resel* in the cross-section is scanned at least three times, and some are viewed up to ten times. This implies there are significant opportunities for data fusion, but does also mean that the detection sensitivity will almost certainly be quite position dependent. Compared with the coverage for the rotor bore, as presented as a *resel* density in Fig. 6.6, it is clear that for the disk inspection the fusion opportunities are higher. On the other hand, the microstructure of the disk titanium compared with the rotor steel means that the former is generally harder to inspect than the latter due to the likelihood of many false calls from grain noise.

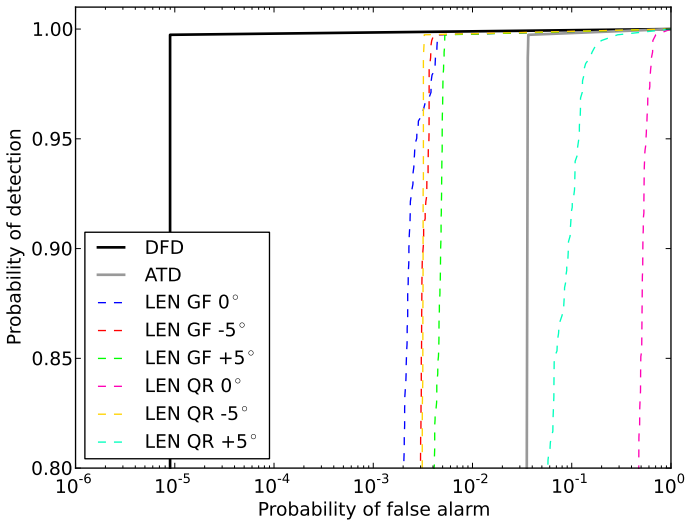


Figure 7.26: The ROC curves for the first known indication of the disk. The channels of the LEN detectors are specified in terms of the scan surface and the probe angle to the surface normal. This indication is seen in 6 scans, 4 of which provide a LEN detector superior to the conventional ATD. But the DFD comprehensively outperforms all of the inputs - and in fact provides an essentially perfect detector, as the probability of false alarm value reached is as low as can be numerically estimated from the available data.

The data for the ROC plots that follow are taken directly from Tippetts. This means that while most of the underlying mechanics (and indeed lines of code) are identical to those for the rotor bore application seen previously, there are some implementation differences, due to the authors taking slightly different approaches. For example, these calculations use the alternative scheme for associating *resels* from different channels with each other, as mentioned in Section 6.8.2 of the previous chapter, and no attempt is made to restrict the ROCs to the convex hull. Therefore some caution should be exercised in comparing the plots of these results with those from the preceding sections.

First, we consider the ROCs computed for the first known indication, considered to be the easiest to find. These curves are plotted in Fig. 7.26. This indication is seen in six scans, four of which provide a LEN detector already superior to the conventional ATD. But the DFD comprehensively outperforms all of the inputs. In fact it provides an essentially perfect detector, as the probability of false alarm (PFA) reached at optimal probability of detection (POD) is as low as can be numerically estimated from the available data. This is a highly convincing demonstration of the power of the DFD, operating on a realistic known indication.

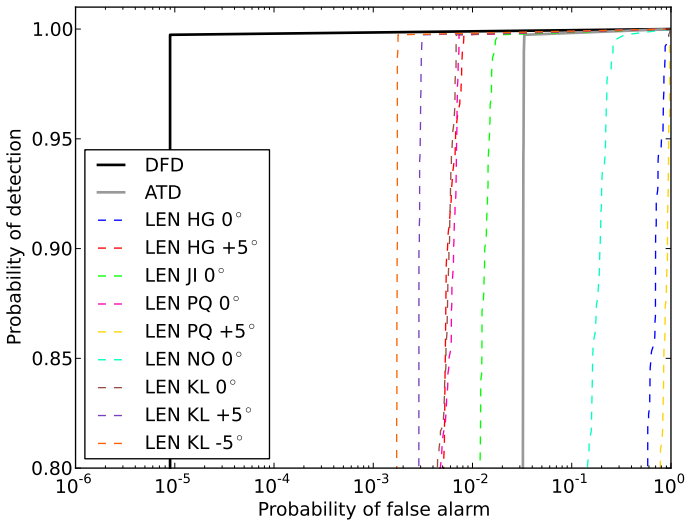


Figure 7.27: The ROC curves associated with the fourth known indication of the disk. The channels of the LEN detectors are specified in terms of the scan surface and the probe angle to the surface normal. This location is inspected by 9 scans, 5 of which provide a LEN detector that is better than the ATD. Combining all the LEN inputs in the DFD again gives a very substantial further improvement.

Now we consider the ROCs for the fourth known indications, one that is harder to find than the first examined. The results are shown in Fig. 7.27, revealing that in this location an impressive nine scans contribute. Five of these provide a LEN detector that is better than the ATD. Combining all the LEN inputs in the DFD again gives a very substantial further improvement, lowering the PFA by over two orders of magnitude compared with even the best of the LEN detectors (the one for the KL $-5°$ scan). The DFD performance has thus further been verified.

It is also noteworthy that the five *resels* of the lowest consensus *p-value* output, and hence highest indication severity probability, also correspond exactly to the five known indications in the disk. Therefore, if the DFD were deployed in this inspection as proposed, with a human inspector reviewing *resel* locations sequentially, as ranked by consensus *p-value*, the known indications would be immediately identified. But in such a scenario, the human inspector may not stop at the fifth *resel*. Given that the contaminated billet used in this disk may have also given rise to further, as-yet undetected defects, we now continue looking through the sequence of ranked *resels*. Ignoring *resels* that are immediately adjacent to those already identified as containing the known indications, caused by some of the signal from these defects spilling into neighbouring *resels* (essentially a

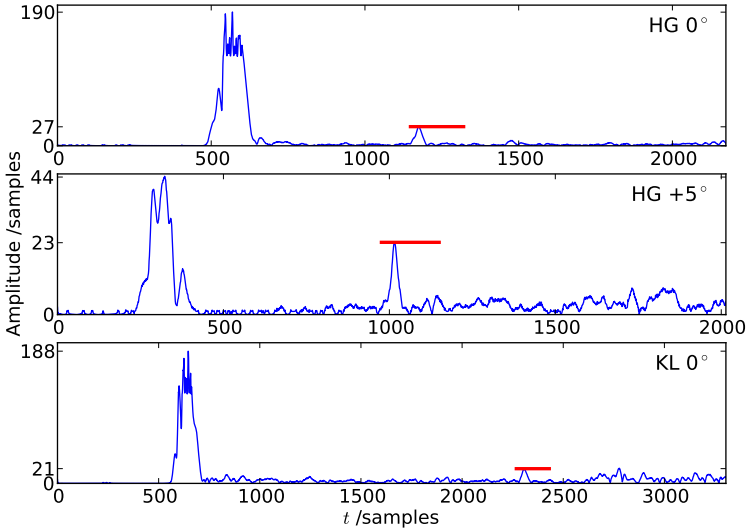


Figure 7.28: The A-scans for the signals that have been identified as originating from a possible sixth, and previously undetected, indication in the disk. The red lines indicate the extent and amplitude of the *resel* in question. Note how low the amplitudes are, especially compared with the industrially conventionally applied global threshold at an amplitude of 64 samples. Note also that the A-scans in the three other channels that view the location are not shown as they appear to only contain noise.

resel edge effect), a possible sixth indication is located. The A-scans associated with this are presented in Fig. 7.28, ignoring those channels that only appear to contain noise. The amplitudes at the indicated *resel* locations are very low, especially compared with the industrially conventionally applied global threshold at an amplitude of 64 samples - the currently applied detection system would have no chance of detecting this possible indication.

The ROC curves for this possible sixth, and previously unknown, indication are shown in Fig. 7.29. This location is seen in a total of six scans. Three of the LEN detectors are worse than the ATD, three better - corresponding to the channels for which the A-scans were plotted in Fig. 7.28. Yet again, the DFD outperforms all the alternative detectors convincingly.

It has so far not been possible to confirm the presence of this candidate indication by independent means, after the original radiographs for the disk were found to be of inadequate quality (partially due to degradation with age). Nonetheless, the A-scans do strongly suggest the presence of a defect at the position identified. This therefore is a striking further demonstration of the capabilities of the Data Fusion Detector.

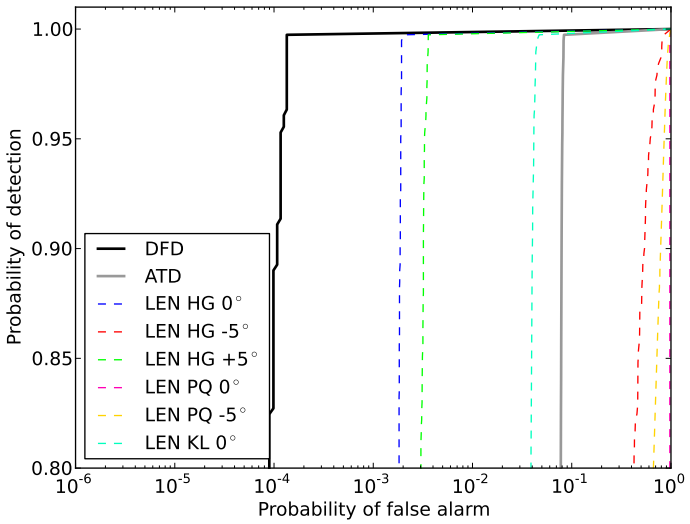


Figure 7.29: The ROC curves for a possible sixth, and previously unknown, indication identified in the disk. This location is seen in 6 scans. Three of the LEN detectors are worse than the ATD, three better - and yet again, the DFD outperforms all the alternatives convincingly.

7.5 Summary

In this chapter the performance of the developed Data Fusion Detector (DFD) has been tested extensively. Initially, the detector was tested with specific artificial inputs to assess the behaviour of the system in such specific, hypothetical circumstances. Then DFD was used to analyse several possible targets in a complete rotor bore dataset, including the location of the highest indication severity probability, the location of the highest overall amplitude, the locations at which artificial signals were inserted, first without then with baseline subtraction. Finally, key results obtained by Tippetts using the DFD operating on a disk inspection dataset with known indications were presented.

Overall, the performance of the DFD compared with the alternatives detailed is dramatic, providing a convincing validation of the development work of the previous chapter. However, we have also explored some of the limitations and complex features of the DFD, finding circumstances under which performance is compromised. While the final results (Section 7.4) are reproduced from Tippetts' thesis, and there is some inevitable overlap in his results and those for the rotor bore presented in Section 7.3, the author does claim novelty for the results and discussion unique to the rotor bore application, specifically relating to the insertion of artificial target signals and the use of a baseline dataset.

This page intentionally left blank

Chapter 8

Conclusions

8.1 Thesis Review

Chapter 1 provided an introduction to the field, an explanation of the industrial context of this thesis and an overview of the project objectives (see Section 1.4). The following chapter, Chapter 2, gave further background material, explaining in detail the two automated NDE procedures that are the primary applications of the data analysis system developed: the RWE npower rotor bore scan and the Rolls-Royce disk inspection. We identified shortcomings in the current procedures that constitute the motivation for this project. Chapter 3 highlighted some of the software engineering and computational challenges of the project, stated the division of responsibilities between the author and his colleague Tippetts [13], and explained some of the tools adopted to overcome the identified challenges.

Chapter 4 explained the system devised for the registration of multiple channels of rotor bore inspection data, potentially from two different data acquisitions. That system is specific to the rotor bore inspection, with a number of features differentiating it from that presented by Tippetts for the case of the disk inspection application [13], but much of the underlying logic and framework is applicable to a great range of multi-channel registration situations. Chapter 5 then presented results obtained from testing the registration algorithm. These results were recognised to be somewhat compromised by the lack of suitable testing data, but the chapter yielded several suggestions for procedural

improvements to the inspection.

Chapter 6 developed the data fusion (and detection) system for probabilistic evaluation of spatially overlapping amplitude fields. The basic processing stages are data de-correlation, local data modelling to convert amplitudes to probabilities taking into account local data statistics, then fusion of the probabilities for a spatial region of interest using a consensus test. Chapter 7 then presented the results of extensive tests on the Data Fusion Detector (DFD), convincingly demonstrating the superiority of that system over conventional alternatives.

In this final chapter we now highlight the main findings identified, state the author's claims to knowledge contribution and make suggestions for further work.

8.2 Main findings

The key findings are organised according to whether they relate to the registration processing, the data fusion work, or constitute suggestions for a revised data acquisition procedure in the context of a semi-automatic data analysis system, such as the one described in this thesis.

8.2.1 Registration

On the registration side the use of a multi-objective optimisation (relying heavily on the work done by Tippetts [13]) was found to be advantageous. This was both because it allows the properties of different objective functions (Section 4.5) to be exploited and provides a set of *Pareto* optimal outputs, rather than a single optimum, thereby allowing uncertainty in the registration to be assessed (Section 4.7). The devised approach includes several features unique to the rotor bore inspection. The calibration scan was identified to offer the best opportunity to address the challenge of sensors of different modalities recording dissimilar signals from the same reflector (Section 4.3). Uncertainty in the feature matching (Section 4.4) which therefore allowed the feature correspondences to vary during the registration was seen to give rise to multi-modality in the objective function space exploration (see, for example, Fig. 5.19). The wrapping period, the number of A-

scans per revolution, of the helical geometry in the rotor bore inspection was determined to be a problematic parameter in the registration optimisation. However, this problem was resolved by the innovative pre-computation of this parameter, allowing the parameter to be eliminated from the subsequent optimisation (Section 4.8.1). Additionally, the author developed a systematic approach for the effective alignment of multiple data channels of differing types and potentially split across two separate acquisitions, relevant to the case of data from a historic inspection being available for comparison. This included the definition of the objective functions to use (Section 4.6.1), choosing one of two identified schemes (see Figs. 4.15 & 4.16), as well as the choice of the parameters to vary in the optimisation (Section 4.6.2). While each of the schemes for defining the objective functions was found to have some advantages over the other, the results of Section 5.4 showed that at least in the current implementation the first scheme (involving cross-acquisition comparisons for all channel modalities) is to be preferred. The registration results also showed that the system performance and the assessment thereof is currently compromised by the suboptimal nature of the available data (see also Section 8.2.3), and that the very high dimensionality of the parameter space in some multi-channel registration calculations remains a computational challenge.

8.2.2 Data Fusion

On the data fusion and detection side of the software, Chapter 6 developed a novel data fusion framework appropriate to the evaluation of large, systematically sampled amplitude fields that overlap spatially, as is seen in automated NDE. The need to de-correlate the field data was established (by defining *resels*, Section 6.3.3), as was the need to use local data modelling. The semi-parametric model of Section 6.5.4 was found to provide the desired sensitivity to effective signal-to-noise ratio (at the highest amplitudes, where this is important) at negligible cost to the validity of the null hypothesis expected output. *Fisher's exact test*, in practice evaluated using an importance-sampling-enhanced *Monte Carlo* technique [13], was established to be a good means of combining spatially coincident *p-values* computed from the data model in different data channels. In the tests of Section 7.2.1, this system was shown to provide qualitatively ideal behaviour for a fusion system.

The further results of Chapter 7 convincingly demonstrate the power of the developed Data Fusion Detector (DFD) over conventional detection approaches, such as those currently used in industry. The detector performed well in a range of inspection scenarios using full rotor bore inspection datasets, including with the insertion of artificial target reflectors. Moreover, the potential of effectively differencing with a previous inspection dataset was demonstrated. Finally, the results obtained by Tippetts using the DFD on data from the inspection of a disk with known defects, reproduced from [13], show how impressively well the system performs in this second application.

8.2.3 Revised Procedure

There are four suggestions for an improved rotor bore inspection procedure to enable the effective operation of a software data analysis system of the sort described in these pages.

Signals from geometric features in main scan data

Section 4.3 discussed how the registration of data channels from different acquisitions is currently undermined by the lack of meaningful geometric reflections in the main scan data. Therefore the author recommends that the procedure be changed to ensure the presence of such features in (at the very least one channel of) every section of the bore inspection. Probably the simplest way to achieve this would be to increase the length of the collected A-scans, such that elements of the rotor's external features are captured, though this would be subject to the directivity of potential reflectors relative to the probes. It is also accepted that this would both increase the data volumes and potentially reduce the data collection speed, but cross-acquisition registration is critical for the comparison of acquisitions, and cannot reliably proceed without significant features in the data to compare.

Improved calibration test piece

As explained in Section 5.3, an improved calibration test piece design would be highly desirable to facilitate the effective use of the calibration scans to check the alignment of the probes on the scanner head. The through-thickness side-drilled hole target reflector of the

current design are unsuitable for precision registration of the different probes, especially in the axial direction. Therefore it is suggested that a new design, featuring, for example, a set of circumferentially distributed, part-thickness side-drilled holes of a constant size, depth and radial position, be introduced.

Consistency in acquisition settings

As emphasised in Chapter 5, if the set-up of the probes could be assured to very high accuracy, multi-modal registration to check the alignment of the probes in the scanner head could be made practically redundant. This would have significant computational advantages, but would require inspectors to take the utmost care in the equipment set-up and recording thereof. In practice it is likely that, even with some slight modifications to the procedure, at least a quick check of the probe alignment would still be desirable, to minimise the scope for human error.

Full wave signal capture

In Chapter 2 it was mentioned that the rotor bore scanner equipment, unlike the disk immersion scanner, currently only saves envelope detected A-scans, rather than full RF (Radio Frequency) signals. This is potentially very problematic for later signal processing. Therefore the author urges that this be changed. Saving full RF signals does not preclude inspectors from still viewing envelope detected signals, and the 1-bit cost of doubling the effective dynamic range to include negative values should not be an excuse given the current size of the datasets.

8.3 Contribution

The author's primary original contribution to knowledge is the data fusion framework of Chapter 6. The basic processing stages are data de-correlation, local data modelling to convert amplitudes to probabilities taking into account local data statistics, then fusion of the probabilities for a spatial region of interest using a consensus test. The processing scheme is essential free of arbitrary thresholds and makes very few assumptions. Novel

elements include the semi-parametric data model and the application of *Fisher's exact test*. The author is not aware of a comparable data fusion system having been developed elsewhere, in any field, and the work described here may have applications beyond the confines of Non-Destructive Evaluation. The performance of this system was impressively demonstrated in Chapter 7.

Secondary claims to academic merit arising from this thesis relate to elements of the registration unique to the rotor bore application that is the focus of this work. Specifically, the approach adopted for addressing the challenge of the wrapping period in Section 4.8.1, the framework for the alignment of multiple data channels in Section 4.6, and the assessment of the two registration schemes used to define the objective functions in Section 5.4 all constitute contributions to knowledge.

8.4 Further work

Reviewing the project objectives of Section 1.4, it should be clear that this project has achieved significant success in addressing most of these. Nonetheless, there are of course many possibilities for further work, some of which build on the recommendations of Section 8.2.3. Arguably the most important general task is to ensure that the software system developed is taken on and integrated into a live industrial project, to ensure the knowledge embodied in this thesis and the underlying code base is harnessed whilst fresh and the transfer into industry completed. Here some more specific suggestions for further work are presented, grouped as pertaining to the implementation, the registration, or the data fusion.

8.4.1 Implementation

As was mentioned in Section 3.6 for the software to be usable in an industrial context, it would need to be provided with an effective and user-friendly graphical user interface (GUI). Additional speed optimisation work would also be important in practice. It is likely for instance that the *Monte Carlo* evaluation of *Fisher's exact test* (see Chapter 6) could be performed in parallel, not on cores of the CPU as currently, but rather a GPU

(Graphics Processing Unit) with significant speed improvements. Unfortunately, given the complexity and memory usage of the registration computation, it is unlikely this would be suitable for a GPU implementation. It might be desirable to also implement a set of rapid, basic checks to be completed on the uploaded data prior to more advanced calculations. This would give an inspector immediate feedback on the quality of the data acquired, ensuring that no re-scan, for instance due to a loss of coupling, is required. In the presence of such initial feedback it is likely that the time the computations could be permitted to take in a practical, industrial setting would be increased.

Section 3.4 did establish that the developed software is not as general as perhaps originally hoped, with significant code that is highly application specific. Therefore the extension to other forms of automatic inspection would require considerable further work, but given the results presented, is likely to be worthwhile. Similarly, the focus on bulk wave ultrasonic testing in this work means that the incorporation of alternative testing techniques, including other forms of ultrasound such as Time-OF-Flight-Diffraction (TOFD), would necessitate more research but seems worth pursuing. The software framework both for the registration and the data fusion is readily extensible both to new inspection types and new data modalities. The specific example of the data from the rotor bore scanner's eddy current array is discussed further in the following sections.

8.4.2 Registration

If meaningful signals from geometric features became available in the main rotor bore scan files, in line with the recommendations of Section 8.2.3, the feature extraction algorithm used for the main scan data (see Section 4.3.2) would need to be revised appropriate to the reflections present. Moreover, as indicated in Section 4.4 the feature pairing algorithm could be enhanced by incorporating the *principle of similarity*, requiring the development and integration of suitable feature descriptors. The encoder data that the rotor bore scanner collects alongside the main data channels could also be integrated into the registration, for instance in the initialisation of the optimisation (see Section 4.9). Meanwhile, in line with Section 4.8.2 the calibration scan could potentially be used to extract the parameters for a complex beam model, and that beam model could then be exploited in

later data fusion processing. A more distinct suggestion is the integration of the eddy current data from the rotor bore inspection (see Section 2.2) into the analysis that has so far used only the four ultrasonic channels. This would require the additional development of the multi-modal registration capability, probably exploiting the calibration scan further.

8.4.3 Data Fusion

There are many opportunities for further work on the data fusion system. Amongst the options that might be considered refinements of the current implementation described in Chapter 6 is an improvement of the definition of the *resels* to use, to make the definition more robust, accommodate significant size variations along the length of A-scans and perhaps incorporate an operator's choice of preferred size. Another suggestion is the mentioned (Section 6.7.7) application of a further discretisation compensation, based on the *Tippett* test (see Appendix D) to avoid the null hypothesis behaviour being undermined in certain circumstances. The detection results from the rotor bore inspection (see Section 7.3) also implied that down-weighting some channels in some parts of the data domain, as they are unlikely to be able to contribute anything useful there, could be a topic of further research. Similar down-weighting to handle data channels for which the implicitly assumed channel independence does not hold could also be studied. Section 6.8.2 mentioned the possibility of a highly conservative solution to the problem of associating *resels* of different data channels, and this may well be worth investigating, and may be of particular interest to potential industrial users.

Suggestions for future work not mentioned elsewhere include the refinement of the selection of the *resels* that make up the distribution against which each *resel* is compared. In the current implementation this is based on the arrangement of the data in the 3D index space representation, but for the helical wrapping of the rotor bore data this is affected by arbitrary choice of the cut used in the data unwrapping (see Section 3.5.4), so a selection more appropriate to the helical geometry could be defined. Also, the normalisation process that distribution is used in could be modified to attempt to compensate for the fact that in the rotor bore inspection, the gravity-induced flow of the couplant (see Section

2.4) does slightly break the rotational symmetry of the data that the current local data modelling relies upon.

An additional, but related, extension would be the modification of the data fusion implementation to allow the similar processing of datasets from components that are not axi-symmetric as the rotors and disks focussed on in the project. For example, disks that have been machined to contain a sequence of slots around the outer edge, to ultimately hold blades in, are more complex in shape than the rectilinear form inspected currently. Such slotted disks also lack the full rotational symmetry of the currently inspected form, instead only having rotational symmetry in integer steps of the angle between edge slots. However, this discrete rotational symmetry should be enough to permit an analysis using the data fusion system of Chapter 6, if the *resels* are very carefully defined to match the disk structure and the normalising distributions of *resels* are defined to only include *resels* from equivalent positions, separated by that angular slot-step. An extension to other component symmetries should also be possible, as long as the symmetry is reflected in the microstructure. A further step be would be the extension to components devoid of significant symmetries, but where the high throughput of nominally identical inspected parts in principle permits a population-wide comparison of equivalent *resels*.

Last, but not least, the integration of the eddy current data from the rotor bore inspection (see Section 2.2) into the ultrasonic analysis so far would certainly be interesting and improve the overall near-surface sensitivity. In principle the only complication in the data fusion framework should be the definition of the depth of the amplitude field, computed from the skin-depth of the eddy currents.

This page intentionally left blank

Appendix A

Operator Network Diagrams

As described in Section 3.4, the program developed consists of a number of discrete *Operator* units, each with a well-defined and restricted remit, and its own set of parameters. These *Operators* interact through a directed network, where each exposes its available outputs to the downstream *Operators* for them to request as an input to their own calculations. As explained previously, Tippetts designed and implemented the tools to enable this flexible and extensible software structure [13], and while there is some overlap in the *Operators* used, the network configuration is largely unique to each application. Here we present some illustrations of a network appropriate for the rotor bore inspection to provide a greater understanding of the implementation underlying the calculations, described in Chapters 5 & 7, relating to registration and data fusion respectively.

The diagrams that follow were all automatically created by the software, using the *pydot* library [124], but were subsequently manually edited for the reader's benefit. The diagrams all relate to a single code run of registration and data fusion, for which the network shown was automatically configured from a few key inputs. However, in contrast to the processing described elsewhere in the thesis, to avoid excessively complex diagrams here the code only operates on two input channels (of the four available in the rotor bore inspection) per acquisition. Two acquisitions are considered together. Therefore the diagrams for instance apply to the case of two sets of both shear channels being available for registration and fusion.

The diagrams display sections of a single network. Each is presented as a directed graph,

where arrows represent the data flow between *Operators* plotted as rectangles at the nodes of the graph. The diagrams are related by nesting, giving rise to multiple levels in the network. Such nesting is facilitated by so-called *Compound Operators*, which in fact each contain a sub-network. These network building blocks enable highly complex linkages to be created, and importantly also provide the mechanics for linking parameters of the contained *Operators*, critical for the implementation of Section 4.6.2. A further complication arises due to the possibility of partial nesting of sub-networks, giving rise to structures that are hard to visualise.

Figure A.1 shows the network of *Operators* at the highest level in the network structure. The presence of levels below is indicated by two nodes being marked with "Compound" and some red text alluding to a separate diagram. This network shows how the bulk of the code for registration and data fusion are related including the registration optimisation.

Going a level deeper to examine the sub-network nested in the top-most "Registration" *Operator*, we reach the network structure shown in Fig. A.2. Close inspection reveals that this sub-network corresponds to registration scheme 1 from Section 4.6, as there is a single "Calibration Feature Pair" *Operator*, acting on inputs from two channels from the same acquisition - indicative of a multi-modal comparison. Note also how each "Feature Pair" *Operator* feeds data into two associated *Operators* for the calculation of registration error metrics. The label "Compound" and the red text for each of the inputs to the sub-network, along the top edge, reveals that there is a further, lower level of the network to explore.

That lowest level is shown in Fig. A.3, for any one of the *Compound Operators* of the previous diagram. This shows how the operations of importing data, creating the data acquisition model and extracting features for registration, for both calibration and main scans of the channel, are related. Note that the outputs of the containing *Compound Operator* need not be drawn only from the last *Operator* of the sub-network.

Finally, if we return to Fig. A.1 and now probe the "Data Fusion" *Operator*, the contained sub-network shown in Fig. A.4 is revealed. This is the software structure used in the implementation of Chapter 6.

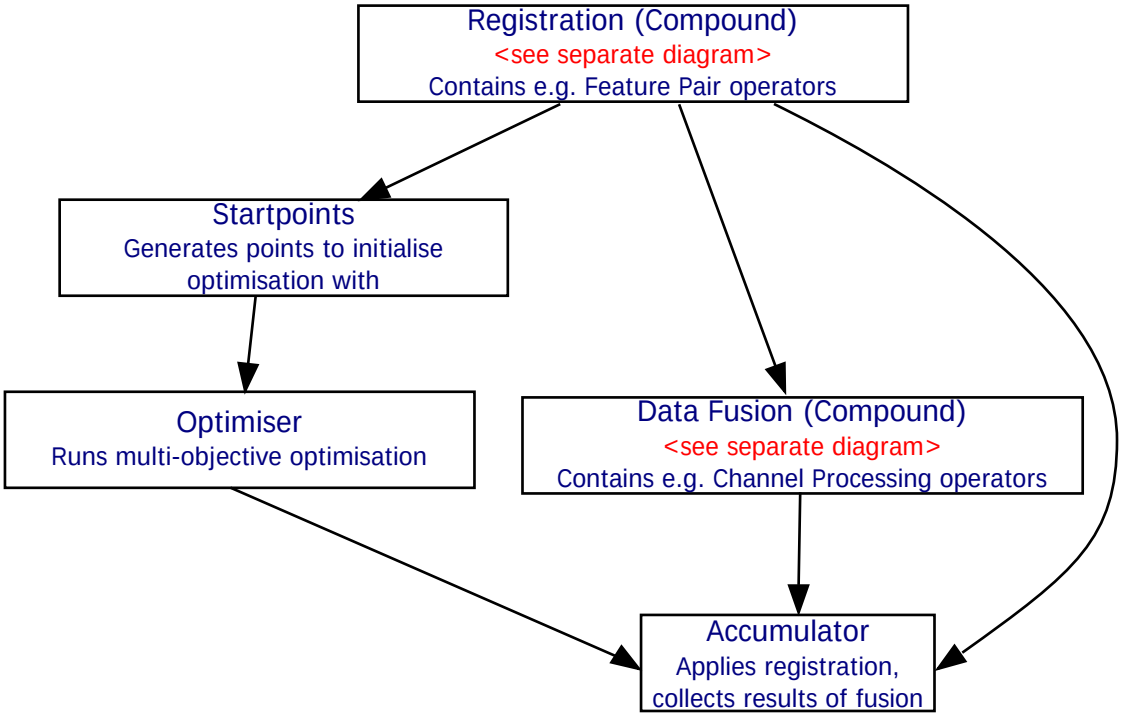


Figure A.1: A directed graph representation of the network of *Operators* at the highest level in the network structure. Arrows represent the data flow between *Operators* plotted as rectangles at the nodes of the graph. The presence of levels below in the network is indicated by two nodes being marked with "Compound". The separate diagrams alluded to are shown in Figs. A.2 & A.4, for "Registration" and "Data Fusion" *Operators*, respectively. This network shows how the bulk of the code for registration and data fusion are related - note the resemblance to Fig. 3.1.

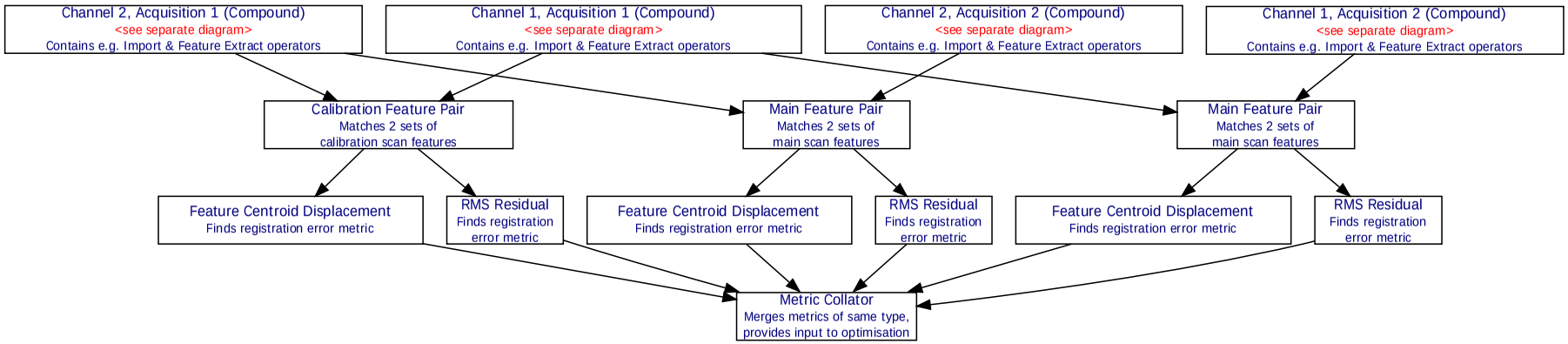


Figure A.2: The *Operator* sub-network used in the registration, based on registration scheme 1 of Section 4.6. The left-most "Feature Pair" *Operator* is part of a single-acquisition, mono-modal comparison, the other two relate to cross-acquisition, mono-modal comparisons. The *Operators* along the top edge all contain further sub-networks, of the sort shown in Fig. A.3.

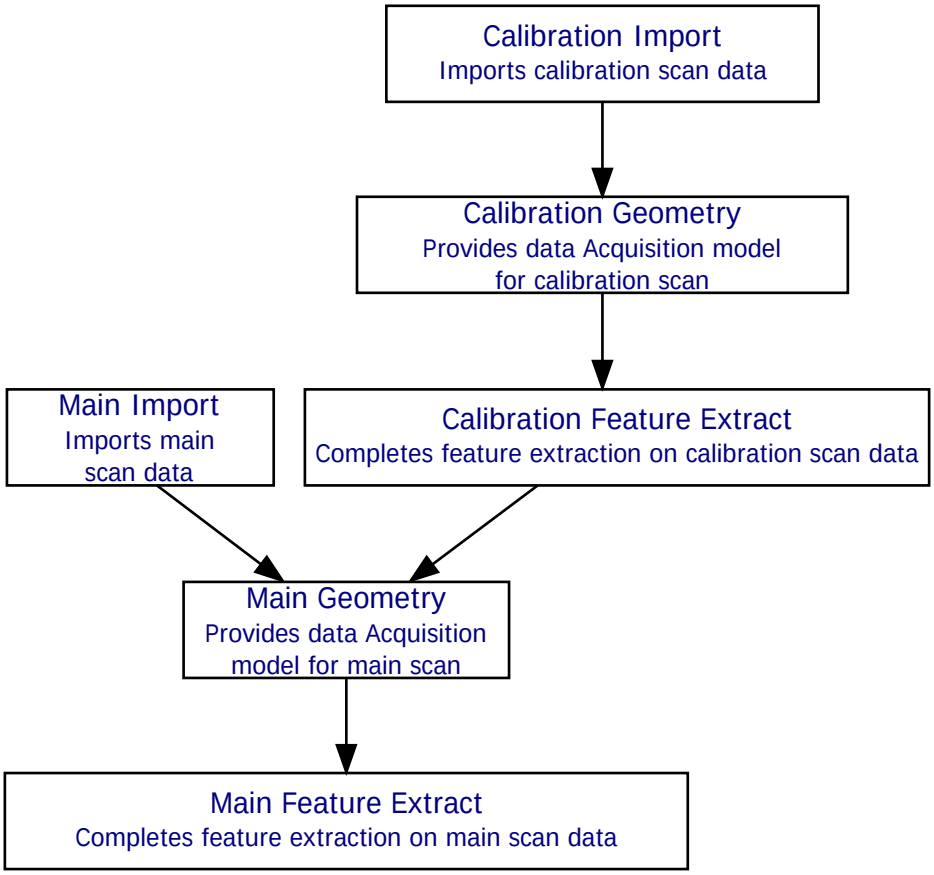


Figure A.3: An example of the sub-network at the lowest level of the *Operator* network. An identical structure exists for every channel contributing to the data analysis. The structure shows how the operations of importing data, creating the data acquisition model and extracting features for registration, for both calibration and main scans of a given channel, are related.

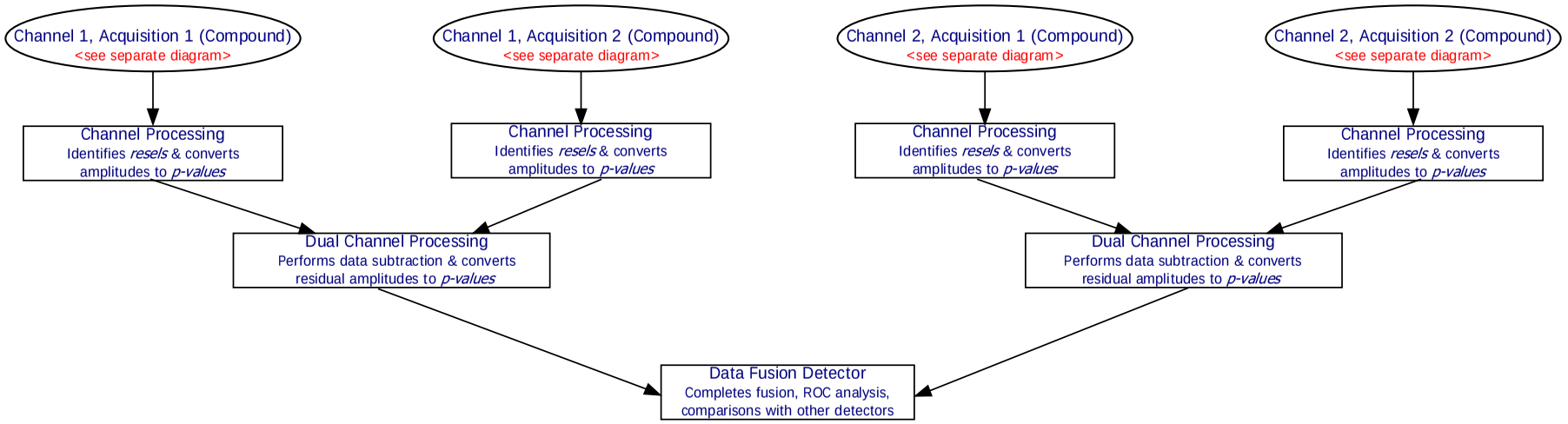


Figure A.4: The *Operator* sub-network for the completion of the data fusion. The fact the *Operators* along the top-edge are shown as ellipses rather than rectangles indicates that they are only partially in the sub-network shown. Individually, these *Operators* again contain the sub-network shown in Fig. A.3. Note that the seen "Dual Channel Processing" *Operators* act on inputs of the same channel from two acquisitions and only exist for the case of two acquisitions being compared. If all provided channels are from a single acquisition, the "Channel Processing" *Operators* link directly with the "Data Fusion Detector".

Appendix B

Synthetic Aperture Focussing Technique (SAFT)

A version of the Synthetic Aperture Focussing Technique (SAFT, itself a version of the Total Focussing Method, TFM, specific to pulse-echo-only array data) that is a standard ultrasonic array imaging technique can be applied and has been implemented in the program. Based on the simple equation for the TFM imaging algorithm in 2D is given in [125], the SAFT intensity I at a location (x, z) in the imaging plane is given by :

$$I(x,z) \quad = \quad \| \sum h_{trx} \left(\frac{2 \cdot \sqrt{(x_{trx} - x)^2 + z^2}}{c} \right) \| \tag{B.1}$$

where $h_{trx}(t)$ is the time domain A-scan signal for transducer at location $(x_{trx}, 0)$, and c is the speed of sound. This equation is clearly readily extensible to 3D.

The SAFT implementation effectively provides an advanced alternative to single channel amplitude interpolation (nearest neighbour or spline), but can also serve directly as a physically meaningful way of combining ultrasonic signals, either within a single channel, or indeed across several data channels. In contrast to typical applications to array data the input data here is based on transducer locations that form a coarse, non-Cartesian grid in a complex 3D geometry. Moreover, for the case of data from RWE npower, the signals are also envelope detected - eliminating the possibility of useful signal cancellations from destructive interference. Therefore, from the outset, not only is the algorithm more

difficult to apply, but given how sub-optimal the inputs to the algorithm are, one cannot hope to match the outstanding performance of the algorithm seen in array data processing papers. Fortunately, our focus is on SAFT as a means of incorporating a beam model and potentially suppressing noise independently within each channel, prior to subsequent cross-channel processing, rather than direct imaging. Additionally, some rudimentary trials on entirely simulated data suggested that even with, for example, envelope detected data the algorithm can deliver useful gains over simple signal interpolation.

While the theory of SAFT imaging is intuitive and simple, its application here is significantly complicated by the input data not being from a simple linear array. Given the complex geometry, and the non-point transducers, local beam models must be considered to decide which A-scans are able to contribute to a particular imaging target location. This then also means that different points in the image include contributions from different numbers of A-scans, so the need for standardisation such as averaging must be examined.

Researching possible beam models for the square-shaped 60 degree shear wave transducers, it became apparent that there a significant number of models, of varying complexity. However, no one analytic model could be found that was directly applicable to a rectangular, angled beam probe acting on a curved sample surface - each one of these attributes gives rise to significant distortions compared with the basic field from a “standard” circular normal probe [6,126]. Briefly, the possibility of using a modelling package such as CIVA [127] to generate a numerical beam model was contemplated, but then discarded for relying on an unjustified number of assumptions and being excessively complicated for the purposes outlined. It was reasoned that all the models would have severe limitations in practice and as only a coarse one was required, it was decided to use the simplest roughly applicable model - keeping the implicit assumptions to a minimum and simplifying implementation, in return for accepting the model limitations. Given how small a component of the project and the software this beam modelling is, this approach would seem entirely rational.

Hence, a beam model, centred around the probe main beam angle and based on the directivity of a rectangular oscillator expressed in terms of sinc functions, from [6, Section 4.5], was applied:

$$\begin{aligned} p &= p_0 \left(\frac{\sin X_1}{X_1}\right)\left(\frac{\sin X_2}{X_2}\right) \\ X_1 &= \pi D_1 \left(\frac{\sin \gamma_1}{\lambda}\right) \\ X_2 &= \pi D_2 \left(\frac{\sin \gamma_2}{\lambda}\right) \end{aligned} \tag{B.2}$$

where p is the pressure, and p_0 that at the transducer face, $D_{1,2}$ are the two transducer edge lengths, $\gamma_{1,2}$ the corresponding angles from the beam centre line and λ is the wavelength of the ultrasound. In the current algorithm implementation, the ring of the most central zero crossing of the directivity function is used as the angular cut-off for the contribution of A-scans. The other condition that controls which A-scans are allowed to contribute to a particular target location is what lines of "sight" are possible between the target location and the collection of A-scan sample locations, taking account the potential obstacle of the bore surface and length of the A-scans.

The related questions of weighting amplitudes (including distance-amplitude decay) and standardising for different numbers of A-scans contributing at different points, have not been resolved, though different possibilities have been tried out. Certainly without weighting or standardisation, performance is impeded by step changes in the number of A-scans used at certain positions in the geometry.

Example results obtained using the SAFT code are promising - Fig. B.1 compares a 2D image evaluated by spline interpolation of the amplitude field against the equivalent obtained by a particular version of the SAFT function. The former certainly has more background scatter than the latter. Whilst the improvement in image quality is not overwhelming, it is worth bearing in mind that A-scans combined at every point is small compared with the application of SAFT to traditional array data, and that number varies with depth as a function of the particular beam model employed in the algorithm. Adding more A-scans acquired at a significantly different angle, from a second data channel, would give greater improvements, but this thesis focusses on combining data channels at the feature level with SAFT merely as a supporting preliminary step, so these possibilities have not been explored further.

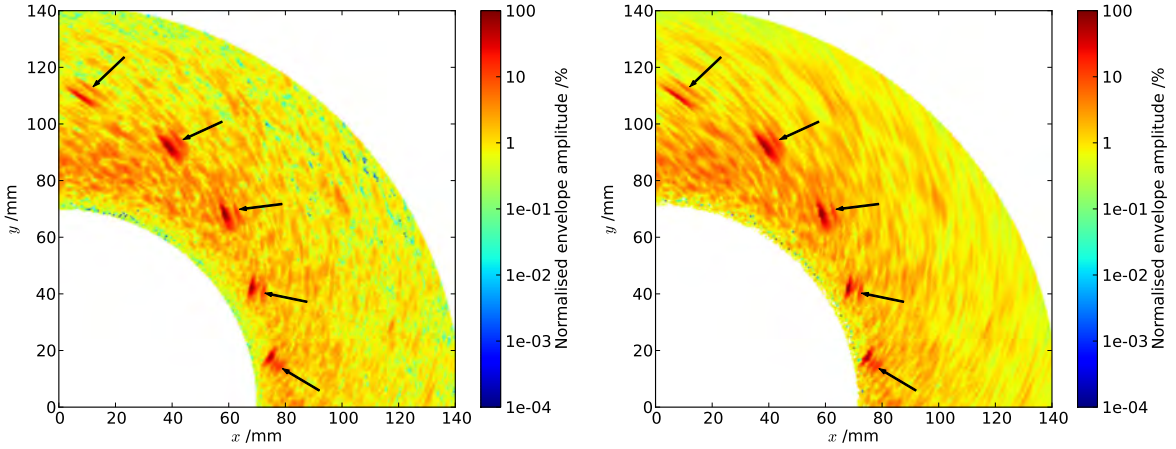


Figure B.1: Demonstration of the Synthetic Aperture Focussing Technique (SAFT) capability, treating a single data channel as having been generated by a sparse array. In (a) a relatively standard B-scan-type image of part of the rotor bore calibration piece is shown, (b) provides the SAFT version of the same cross-section. The images are each normalised to the same logarithmic colour scale, and arrows mark the creeping wave signals that follow each side-drilled hole reflection. Note that SAFT has smoothed out some of the noise along arcs corresponding to the beam spread, and thus helped to draw out the side-drilled hole reflection and creeping wave signals.

The most interesting feature of the SAFT output example shown are the "secondary hole reflections", small reflections apparently lying in the shadow of the side-drilled holes and highlighted with arrows in Fig. B.1. These "secondary hole reflections" are also visible in the original, reference image, but there they are not particularly striking due to the comparable level of surrounding scatter.

As an aside, these "secondary hole reflections" were investigated further, given that their regular pairing with the reflections from the side-drilled hole reflectors implied a physical significance. The author was prompted to examine possible creeping wave effects. The paper [128] by Nagy et al describes an ultrasonic inspection technique based on examining the creeping wave signals that can propagate around circular holes in a sample. According to the presented theory, the delay of the creeping wave signal with respect to the main reflection from the hole is given by:

$$\tau \quad = \quad (2+\pi)\,\frac{R}{c} \tag{B.3}$$

where R is the hole radius and c the creeping wave velocity. The creeping wave, a

leaky Rayleigh mode, travels at a velocity slightly below the shear wave velocity. The consistency of this theory with the observed "secondary hole reflections" was checked by comparing the "half-skip" travel times. Taking measurements from a SAFT image and averaging the experimental value was 1.30μs, whilst the theoretical value, given a hole diameter of 3mm and approximating the creeping wave velocity to be 90% of the shear wave velocity, was 1.33μs. Especially considering the approximations made in this analysis, this is very good agreement.

This page intentionally left blank

Appendix C

Data Model Development

C.1 Discrete model

The simplest type of distribution that can be formed from a collection of discrete samples (of amplitude, here) is a discrete distribution, where the p.d.f. is simply given by the frequency of each discretisation level's occurrence divided by the total number of samples. Such a p.d.f. is illustrated in Fig. C.1. The corresponding survival function will, by definition, given that the model distribution matches the data precisely, produce exactly the expected null hypothesis behaviour of a uniform distribution of *p-values* for random samples of the input amplitudes. Note the available output *p-value* levels may be non-uniformly spaced, depending on the relative frequencies of the different levels in the input discrete samples, but will be separated by integer multiples of $\frac{1}{n}$, where n is the finite number of samples making up the distribution.

While the simplicity of this distribution and exact U(0,1) *p-value* output behaviour under the null hypothesis make the distribution an attractive choice, it does suffer one significant drawback in the proposed application. This is that the distribution output has no sensitivity to relative amplitudes, related to signal-to-noise, and is entirely determined by the different frequencies of the input amplitudes, with no regard for differences in these amplitude levels (beyond their order). Expressed more practically, given that our interest for detection will be focussed on the highest amplitudes, the *p-value* of the highest amplitude is unaffected by how much higher that amplitude is than the remainder of the

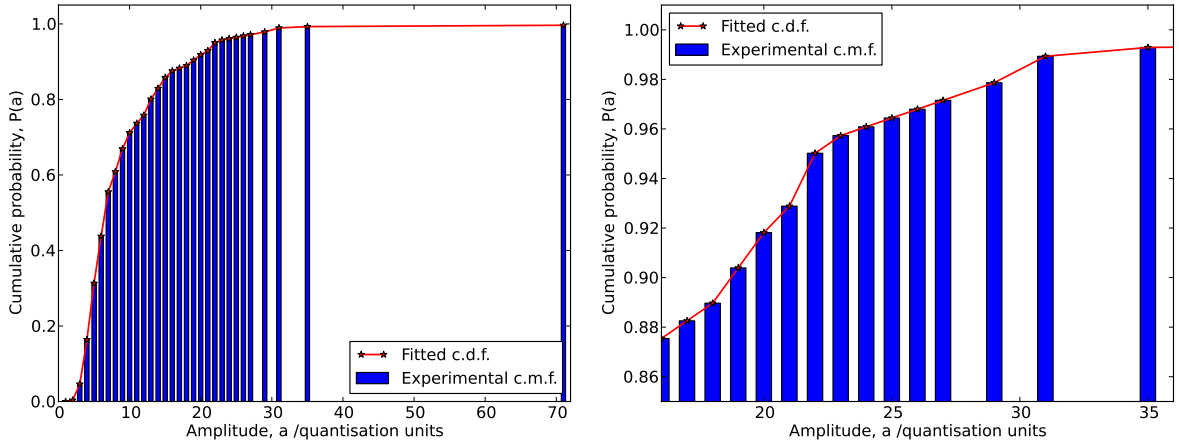


Figure C.1: A purely discrete data model, demonstrated on the amplitude distribution found in one circumferential ring of rotor bore *resel* data. Plot (a) shows the cumulative distribution function implied by the raw data (red line and blue bars respectively). In (b) an enlarged section of (a) is shown to illustrate the piece-wise linear nature of the "fitted" distribution - straight-line segments link the tops of the bars. The *p-value* at the very highest amplitude recorded is 3.6×10^{-3}, which is exactly equal to the reciprocal of the number of samples making up the distribution, as is to be expected given the single occurrence of this amplitude.

distribution (see Fig. 6.9). In this sense, using a discrete distribution involves discarding valuable information about the input amplitudes, a waste to be avoided.

C.2 Analytic model

An alternative to the discrete distribution considered in the previous subsection is to fit an analytic distribution [83]. The choice of analytic form to use in this case should be determined based on the theoretically applicable distribution, and the determination of the distribution's parameters can then be completed using an optimisation of a metric such as maximum likelihood. This parametric technique is at the other extreme of the range of possible data models to the discrete distribution in the sense that it is only minimally data-driven and the experimental samples (typically hundreds) are reduced to just a handful of parameters. The model does provide the sought sensitivity to the relative amplitude magnitudes.

Figure C.2 illustrates the sort of problem that may be encountered adopting this approach - an experimental *resel* amplitude distribution from a turbine disk scan is shown. Theory

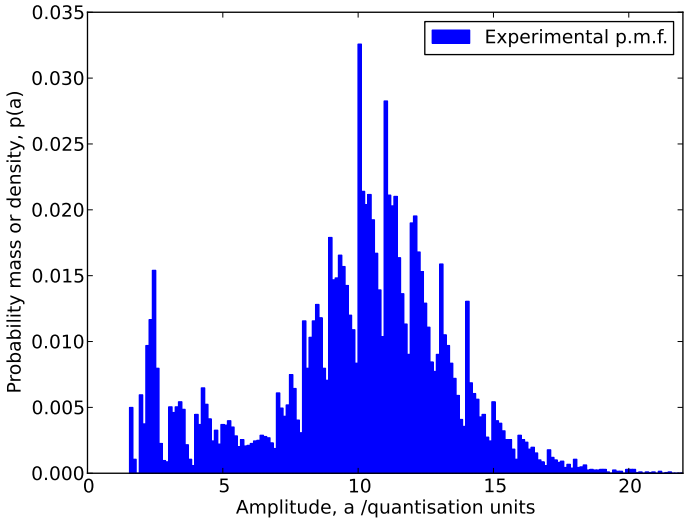


Figure C.2: The distribution of envelope-detected amplitudes recorded in one of the turbine disk scans. The plot illustrates some of the complex distribution features that may arise in practice that would be impossible to adequately represent with a simple parameterisation from fitting an analytic distribution. Note that the experimental amplitudes are in this case effectively continuous as full waveform data was captured and envelope detection only applied in post-processing.

predicts the Rice distribution to be a good analytic, parametric model of envelope-detected amplitudes such as those making up the experimental distribution [129]. However, it is clear from the plot that there are a number of complex features (note the periodic structure over parts of the p.m.f.) that no such simple parametric distribution would ever be able to match with an acceptable fitting error. This fitting error is important because it undermines the ability to match the expected null hypothesis behaviour for random input samples.

C.3 Kernel Density Estimation (KDE) model

Kernel Density Estimation (KDE) is a data-driven, non-parametric data model that was examined but then discarded. It may be used to estimate a distribution from a collection of samples by superposing instances of a chosen kernel function centred on each sample. The kernel widths in this processing are free parameters, and can be adjusted to give different levels of smoothing, possibly in an adaptive fashion, as a function of amplitude [130]. The choice of the kernel widths provides a means to adjust the trade-off between how

closely the output under the null hypothesis matches U(0,1) and the desired sensitivity to relative amplitudes. The fact that no generally applicable, satisfactory compromise could be found in initial testing, and concerns about the free choice of kernel shape and width giving scope for significant arbitrary decisions, meant that this approach was dismissed.

Appendix D

Compensation Schemes - Tippett Test

D.1 Motivation

This alternative consensus test was implemented and studied for comparison with the *Fisher* test focussed on (Section 6.7), not only with a view to providing the operator with a choice but also for academic interest. The *Tippett* (not to be confused with Tippetts [13]) test has great intuitive appeal - recall from Section 6.6 that the consensus *p-value* for k inputs q_i, $i = 1, 2 \ldots k$, under this test is given by:

$$q \;=\; 1 - (1 - \min(q_1, \ldots, q_k))^k \tag{D.1}$$

Its simplicity allows a comprehensive analysis to be hoped for. Additionally, the test provides an explicit link to combinatorics, and the "multiple comparison problem", when a specified significance level is adjusted according to the number of independent tests simultaneous examined [102, 108, 116].

Given that the function used to reduce the inputs to a single value is the minimum (also described as the first order statistic [103]), it would seem that the scope for the input discretisation to have an effect is limited, and so this is an interesting situation to investigate. It is however clear that for a discrete input the possible values that the output can take is limited to the total number of distinct values amongst the possible discretisation levels of the inputs. This may mean that the test in practice is of limited

use - when the lowest *p-value* ever reached corresponds to let us say 30% of being a false-call the test can hardly be considered helpful. These considerations are borne out by the c.d.f.s produced testing the *Tippett* test under the null hypothesis conditions in Fig. D.1. The graphs show that the need for a discretisation compensation scheme is not as great as for the *Fisher* test, as the continuous test at least behaves as expected for many inputs. However, Fig. D.1d indicates that there is still a need for compensation in the case of discrete inputs drawn from multiple distributions. Furthermore, the unhelpfully limited output range is obvious. Yet there exists the intriguing possibility of maybe leveraging multiple occurrences of the observed minimum *p-value* (related to using higher order statistics than first order), an eventuality unique to discrete inputs, to extend the output range to lower consensus *p-values*. Note that the *Monte Carlo* (Sec. 6.7.2) and dithering (Sec. 6.7.3) schemes for discretisation compensation previously described as applied to the *Fisher* test are still applicable, but as will be shown, we can do better still and deepen our understanding by studying the *Tippett* test in detail.

D.2 Principles

The approach taken is to consider the minimum value of all possible input combinations, across the discretisation levels of all inputs, as well as the frequency of occurrence, and compute the relevant survival probability - the probability of such a combination, or a more extreme one, occurring. Figure D.2 illustrates the process, setting $r = \min(q_1, \ldots, q_k)$ and letting j represent rank of r in the ascending l distinct values amongst the possible discretisation levels of the inputs, $j = 0, 1 \ldots l-1$. The basic *Tippett* test relies purely on the first, left-most column of the diagram, corresponding to one or more instances of the recorded minimum *p-value*. For a given frequency f the overall survival function probability decreases monotonically as j decreases, given for example that $(1 \times r_1)$ is a subset of $(1 \times r_0)$ in the sense that the former is less extreme than the latter. Conversely, for a given j, higher multiples of r_j are guaranteed to have a lower overall survival probability P_{sf} than lower multiples, as lower multiples are subsets of higher multiples. However, there is no such guarantee (for all possible distributions of r_j) considering different frequencies of different r_j. For example, $(3 \times r_1)$ is not a subset of

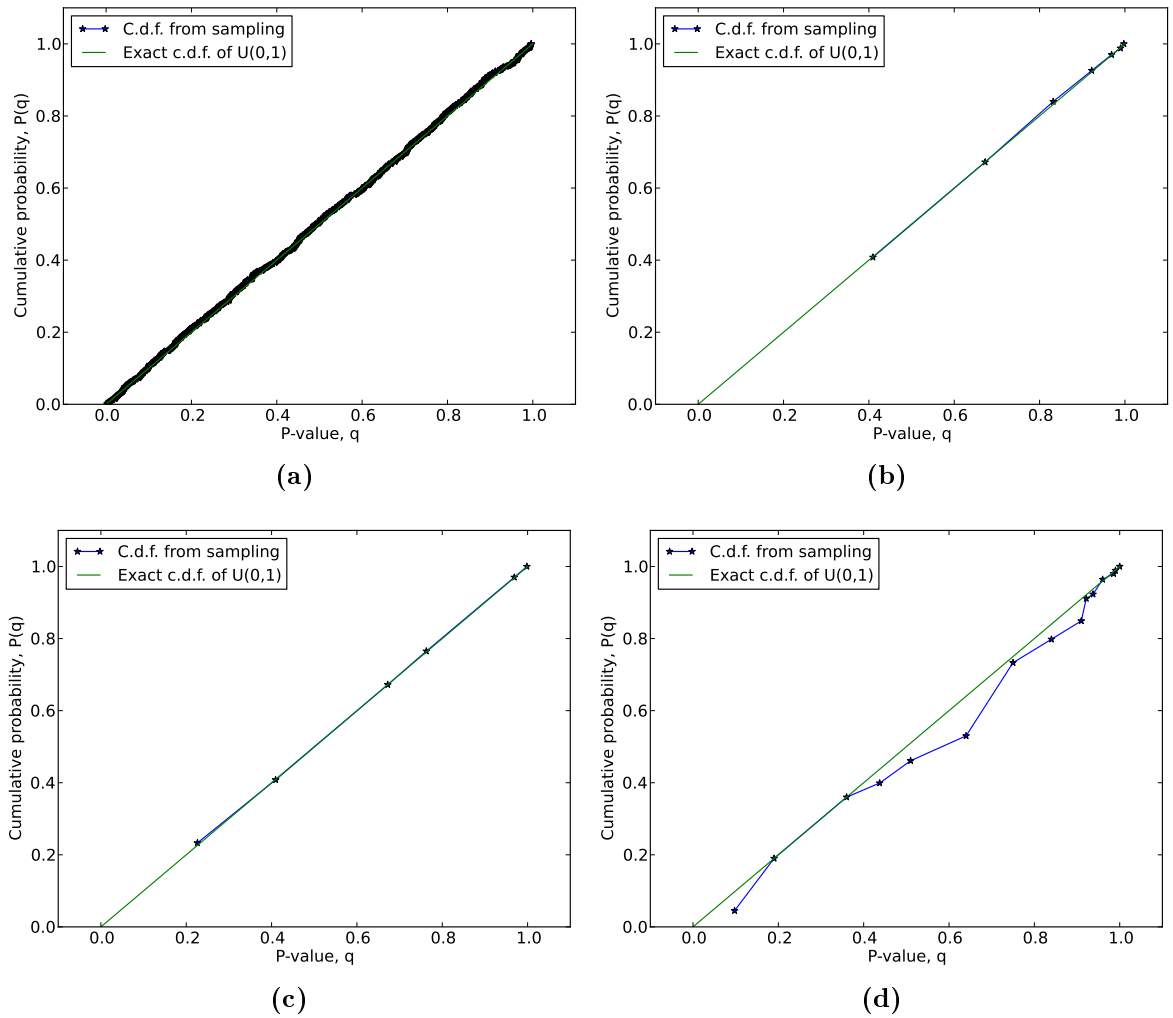


Figure D.1: Cumulative distribution functions (c.d.f.s) for the output of the *Tippett* consensus test for different simulated inputs, compared against the theoretically expected. In (a) 5 *p-values* drawn from the continuous U(0,1) distribution (see Fig. 6.8a) are combined, and the experimental distribution closely matches the theoretical. Plots (b) and (c) are based on drawing 5 inputs from the discrete distributions shown in Figs. 6.8b and 6.8c, respectively. Plot (d) was produced by drawing 1 sample from the former and 1 from the latter (rather than 2 & 3, respectively, in comparable plots). In (b) and (c) the sampled points lie very close to the locations expected from the continuous distribution, while (d) displays significant disparities between the plotted lines. Additionally, it is clear in all these discrete input cases that only a limited number of points of the c.d.f. are achievable, consistent with the form of the test statistic, but that also the lowest *p-value* obtainable is rather high, probably too high to find sensible use.

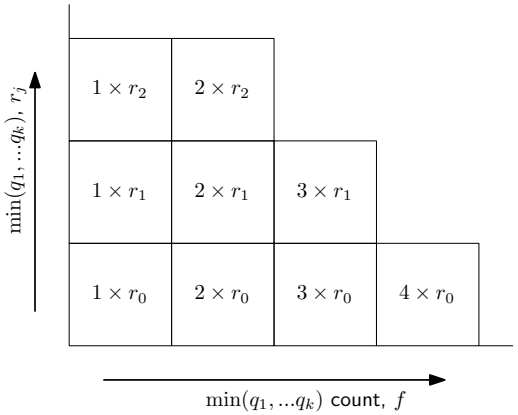


Figure D.2: Diagram to illustrate multiple occurrences of the lowest observed (discrete) *p-value*, in relation to other, similarly unlikely possibilities. The different *p-value* discretisation levels that could be minimal amongst the inputs are denoted r_j, for $j = 0, 1 \ldots l-1$, where l is the total number of distinct values amongst the possible discretisation levels of the inputs, and these are ordered vertically with $\min(r_{0,1 \ldots l-1}) = r_0$. The horizontal axis meanwhile shows the number of occurrences f of the recorded minimum *p-value*, with the row length of each r_j determined by the maximum number of times that value could occur across the input distributions.

$(2 \times r_0)$, nor is the reverse true.

Using the higher multiple possibilities to sub-divide the single instance possibilities that are the only ones available under the basic *Tippett* test should permit a greater number and range of possible output *p-values*. However, it must be possible to arrange the permitted combinations in an order such that each combination is a subset of all preceding ones, given the cumulative nature of the survival function. This imposes a restriction on which combinations can be considered, and in general (for all possible distributions of r_j), means those entries lying away from the axes in the diagram cannot be contemplated, but those in the bottom-most row of the diagram can be. Hence the *Tippett* test is extended to consider frequencies greater than unity of the lowest possible *p-value* amongst all the input distributions, which will extend the output range of the test towards zero, in the range where more possible outputs are most needed.

As an aside, there is one practical complication that needs some consideration for the multi-occurrence extension to be viable under the greatest possible range of circumstances. It may be that the very lowest possible *p-value* amongst all the input distributions r_0 can only occur once, or simply has a minuscule probability of occurring even once compared with the next lowest possibility r_1. In these cases the extension by considering multiple occurrences of *p-values* is effectively blocked by the existence of that r_0. The problem

may be overcome by effectively folding that r_0 into r_1 and eliminating r_0 as a possible value from the input discretisations. The apparent loss in consensus output resolution is more than compensated for by the way that higher multiples of r_1 can then be considered. Such a system has been implemented but is not discussed further here.

To compute the required survival probabilities, and correctly even for multi-distribution input cases where the basic *Tippett* test fails (see Fig. D.1d), it is necessary to recognise that this combinatorial problem is governed by the binomial distribution, the fundamental probability mass function of which is given by [103]:

$$p(g,k,r) = {}^{k}C_g\, r^g\, (1-r)^{k-g} \tag{D.2}$$

$${}^{k}C_g = \frac{k!}{g!\,(k-g)!} \tag{D.3}$$

for k, $g \in \mathbb{N}_0$, where in the application considered $r = \min(q_1, \ldots, q_k)$. This corresponds to the probability of picking g "successes" from k "attempts" and r is the probability of a "success" in a single "attempt". The sought survival probability is then:

$$P_{sf}(g,k,r) = 1 - \sum_{i=0}^{g} {}^{k}C_i r^i\, (1-r)^{k-i} \tag{D.4}$$

Note that for $g = 0$ the *Tippett* test Equation D.1 is recovered, and $g = f - 1$, where f is frequency of occurrence of r. The equation shown is applicable to samples drawn from a single distribution, but if there are multiple input discretisations contributing, the survival probability form can become considerably more complicated. Two evaluation implementations are discussed in the sections that follow.

D.3 Analytic

The analytic approach is firmly based on Equation D.4. The current implementation is restricted to cases where there are no probability contributions from multiple input discretisations, even when the inputs are drawn from multiple input discretisations. This will be the case when levels less than or equal to r are only found in a single input discretisation, or there is only a single occurrence of r to consider, either because only a

single occurrence featured in the input *p-values* or because $r > r_0$ and higher multiples are therefore not considered, as explained in the previous section. In the former case, the survival probability is given by Equation D.4 applied to the single contributing distribution. In the latter case, probability contributions from different input discretisations may be simply combined as

$$P_{sf}(g=0) \;=\; 1 - \prod_{b=1}^{m} \left(1 - P_{sf}(g=0, k_b, r_b)\right) \tag{D.5}$$

for m input distributions, each providing k_b samples and r_b being the largest *p-value* in the discretisation for which $r_b \leq r$.

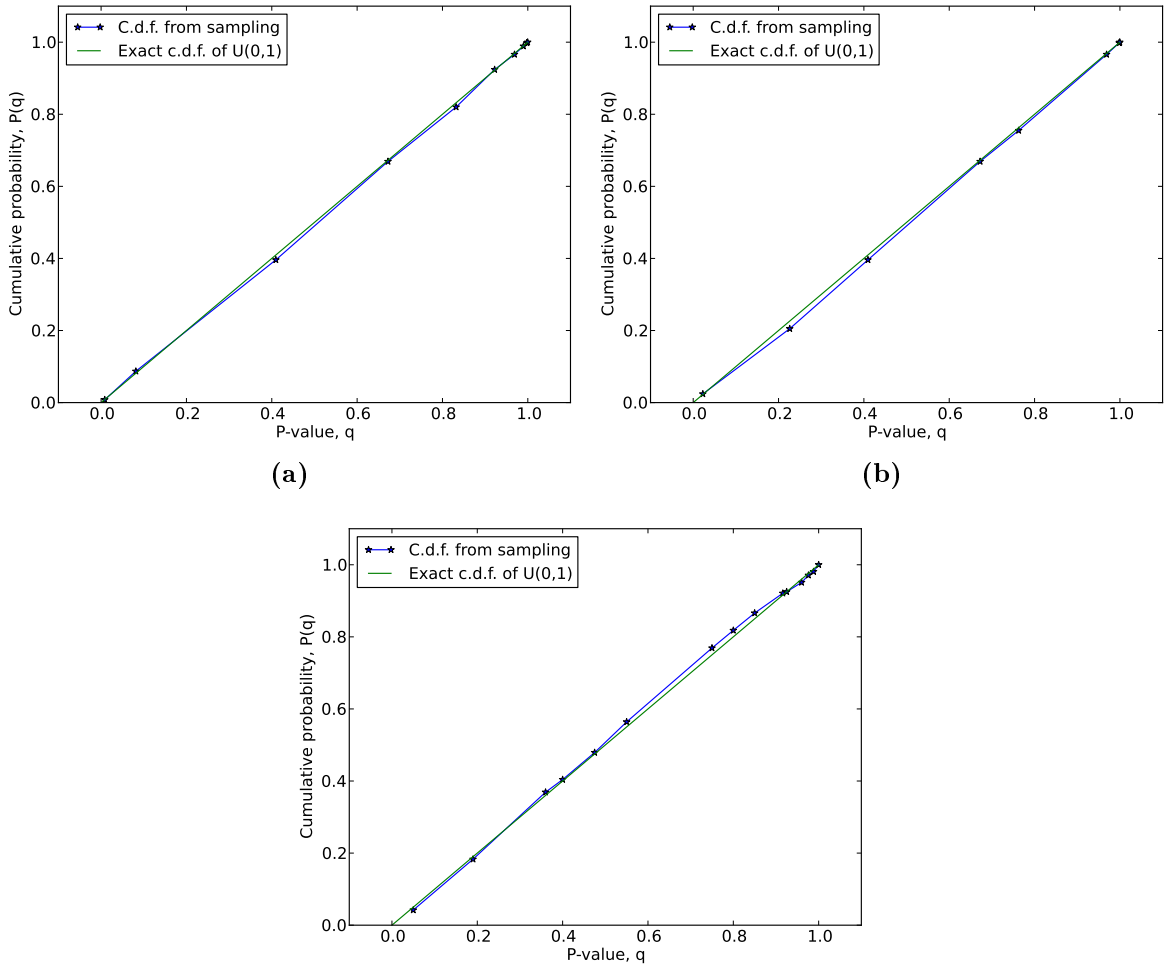
The effectiveness of the scheme is demonstrated in Fig. D.3, where the plots not only show the required null hypothesis behaviour, but also span a greater number and range of *p-values* than before, such that the test now holds greater promise of being useful.

D.4 Monte Carlo

This scheme for the correct computation of the required survival probability for an identified (frequency x lowest *p-value*) case is based on *Monte Carlo* (M.C.) sampling. The scheme here differs from the general compensation scheme described in Section 6.7.2. Unlike the analytic scheme described in the last section, this scheme is not restricted to any special, simple cases.

The approach is to draw random samples from the binomial distributions corresponding to each input discretisation, counting the number of such samples across the combined set that at least match the observed minimum *p-value* occurrence frequency f. Repeating this M.C. sampling many times allows the fraction of trials that give an output at least as extreme as the one observed, identical to the sought survival probability, to be computed. Note that here the number of M.C. samples immediately affects the output consensus *p-value* by being the denominator of the fraction computed.

The power of the *Monte Carlo* compensation scheme is demonstrated in Fig. D.4, where again the null hypothesis behaviour is as expected and a greater number and range of outputs are reached than using just the basic *Tippett* test. An example of the scheme



(a) (b)

(c)

Figure D.3: Output c.d.f.s for *Tippett* consensus test for different simulated inputs with analytic discretisation compensation. Plots (a) - (c) correspond to the uncompensated plots seen in Fig.D.1 (b) - (d), respectively, but compute outputs analytically. In each case the experimental distribution is seen to closely correspond to the continuous U(0,1) distribution, including for the case of discrete samples drawn from two distributions - unlike in Fig. D.1d - thereby demonstrating the effectiveness of the compensation. Note the greater number and range of *p-values* than before in all plots.

operating on a case that the restricted analytic scheme could not handle is not included, but the performance in such cases is as expected.

The broader applicability of the *Monte Carlo* scheme means that this is the scheme most used in practice, though it is possible to use automatically the exact, analytic approach for simple cases where that implementation will work.

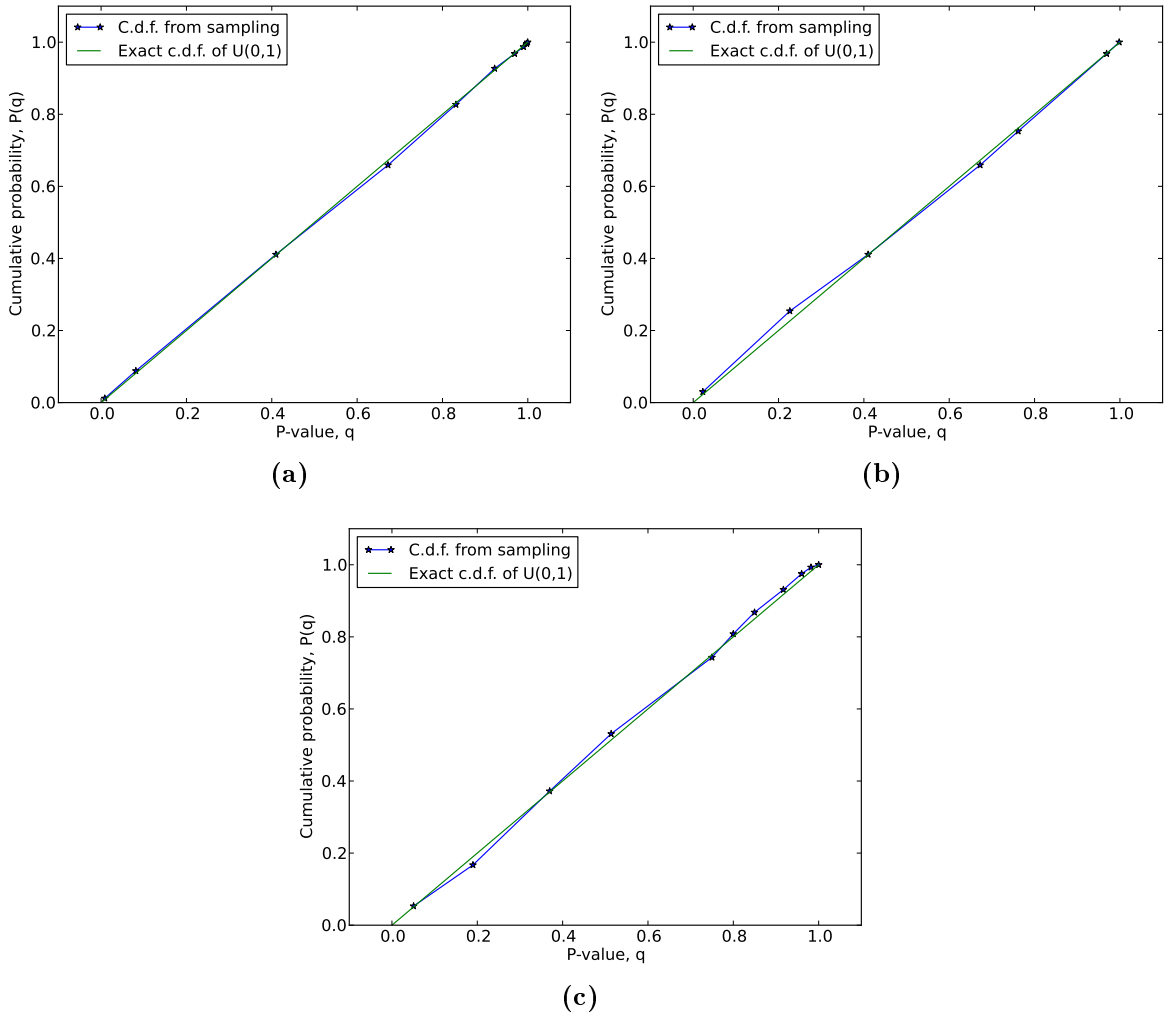


Figure D.4: Output c.d.f.s for *Tippett* consensus test for different simulated inputs with *Monte-Carlo*-based discretisation compensation. Plots (a) - (c) correspond to the uncompensated plots seen in Fig.D.1 (b) - (d), respectively, but compute outputs by taking M.C. samples from binomial distributions. As for the analytic compensation scheme compensation scheme (Fig. D.3), the experimental distribution now closely matches the continuous U(0,1) distribution, and a greater number and range of *p-values* is available than without compensation. Note that K-means clustering was used to suppress M.C. sample plotting artefacts [131].

This page intentionally left blank

References

[1] R. Halmshaw, Non-Destructive Testing, Edward Arnold, London, 1987.

[2] H. Masoom, R. S. Adve, R. S. C. Cobbold, Target detection in diagnostic ultrasound: Evaluation of a method based on the CLEAN algorithm, Ultrasonics 53 (2) (2013) 335–344.

[3] C. G. Windsor, L. Capineri, Automated object positioning from ground penetrating radar images, Insight 40 (7) (1998) 482–488.

[4] D. C. Wright, A. Miltreyger, M. Bron, S. Rabinovich, 1000 Gates - a novel approach and method for ultrasonic inspection of aero-engine disc forgings, Presentation at British Institure of NDT Conference, Telford, UK (2013).

[5] A. Rogerson, A personal perspective on the early developments in inspection qualification and reliability assessment in the UK nuclear industry, in: Proceedings of 18th World Conference on Non-Destructive Testing, Durban, South Africa, 2012.

[6] J. Krautkrämer, H. Krautkrämer, Ultrasonic Testing of Materials, 4th Edition, Springer, Berlin, 1990.

[7] W. Rummel, Nondestructive inspection reliability - history, status and future path, in: Proceedings of 18th World Conference on Non-Destructive Testing, Durban, South Africa, 2012.

[8] C. Müller, M. Bertovic, M. Pavlovic, D. Kanzler, U. Ewert, J. Pitkänen, U. Ronneteg, Holistically evaluating the reliability of NDE systems - paradigm shift, in: Proceedings of 18th World Conference on Non-Destructive Testing, Durban, South Africa, 2012.

[9] M. Bertovic, B. Fahlbruch, C. Müller, J. Pitkänen, U. Ronneteg, M. Gaal, D. Kanzler, U. Ewert, D. Schombach, Human factors approach to the acquisition and evaluation of NDT data, in: Proceedings of 18th World Conference on Non-Destructive Testing, Durban, South Africa, 2012.

[10] W. E. Gardner, Improving the Effectiveness and Reliability of Non-Destructive Testing, International Series on Materials Evaluation and Non-Destructive Testing, Pergamon Press, 1992.

[11] RWE npower, accessed: 23-01-14.
URL http://www.rwe.com/web/cms/en/97682/rwe-npower/about-us/our-businesses/

[12] C. Ward, The use of eddy current arrays as a replacement for dye-penetrant inspection of generator and exciter end-rings, Insight 47 (9) (2005) 543–546.

[13] T. Tippetts, Improved reliability of automated non-destructive evaluation, Ph.D. thesis, Imperial College London, Department of Mechanical Engineering (2013).

[14] A. McNab, I. Dunlop, A review of artificial-intelligence applied to ultrasonic defect evaluation, Insight 37 (1) (1995) 11–16.

[15] S. Yella, M. S. Dougherty, N. K. Gupta, Artificial intelligence techniques for the automatic interpretation of data from non-destructive testing, Insight 48 (1) (2006) 10–20.

[16] C.-Y. Lee, J.-D. Kwon, Y. Chai, K.-S. Jang, Stress and fracture analyses of nuclear power plant LP turbine rotor discs, KSME International Journal 14 (2) (2000) 207–214.

[17] R. Lyon, Confirming plant integrity by specialised NDE inspections, Insight 47 (9) (2005) 551–558.

[18] L. Coughtrey, K. Stamps, C. Ward, R. Spencer, NDT work instruction: Procedure for the ultrasonic and eddy current surface inspections of rotor shaft bores and adjacent bodies, Tech. rep., RWE npower plc (2008).

[19] P. I. S. Ltd., Turbine shaft inspection system - TSIS.
URL http://www.phoenixisl.com/tsis-mdh-prism

[20] L. Coughtrey, P. Crowther, Engineering report: Didcot 'A' LP2 rotor bore inspection, job no: GBF4201, Tech. rep., RWE npower plc (2007).

[21] R. Herblot, Eddy current C-scan display for array probe bore tube inspection, Tech. rep., R/D Tech Europe (2002).

[22] Zetec UltraVision.
URL http://www.zetec.com/2010/06/ultravision/

[23] L. Coughtrey, P. McKay, P. Crowther, Engineering report: Didcot 'A' LP2 rotor bore inspection, job no: GBF4229, Tech. rep., RWE npower plc (2008).

[24] A. Chamberlain, R. Edwards, RWE npower plc reference document for ultrasonic inspection methods, Tech. rep., RWE npower plc (2008).

[25] ScanMaster.
URL http://www.scanmaster-irt.com/

[26] I. Yalda, F. Margetan, K. Han, R. Thompson, Survey of ultrasonic grain noise characteristics in jet engine titanium, in: D. O. Thompson, D. E. Chimenti (Eds.), Review of Progress in Quantitative Nondestructive Evaluation, Vol. 15, Springer US, 1996, pp. 1487–1494.

[27] F. J. Margetan, I. Yalda, R. B. Thompson, J. Umbach, U. Suh, P. J. Howard, D. C. Copley, R. Gilmore, Ultrasonic grain noise modeling: Recent applications to engine titanium inspections, in: D. O. Thompson, D. E. Chimenti (Eds.), Review of Progress in Quantitative Nondestructive Evaluation, Vol. 16, 1997, pp. 1555–1562.

[28] D. Kahneman, Thinking, Fast and Slow, Penguin, London, 2011.

[29] Python language reference, version 2.7.
URL http://www.python.org

[30] G. van Rossum, J. de Boer, Interactively testing remote servers using the python programming language, CWI Quarterly 4 (4) (1991) 283–303.

[31] T. Rentsch, Object oriented programming, SIGPLAN Notices 17 (9) (1982) 51–57.

[32] Enthought Scientific Computing Solutions.
URL https://www.enthought.com/

[33] T. E. Oliphant, Python for scientific computing, Computing in Science Engineering 9 (3) (2007) 10–20.

[34] E. Jones, T. Oliphant, P. Peterson, et al., SciPy: Open source scientific tools for Python (2001-).
URL http://www.scipy.org/

[35] F. Alted, I. Vilata, et al., PyTables: Hierarchical datasets in Python.
URL http://www.pytables.org/

[36] J. D. Hunter, Matplotlib: A 2D graphics environment, Computing In Science & Engineering 9 (3) (2007) 90–95.

[37] P. Ramachandran, G. Varoquaux, Mayavi: 3D Visualization of Scientific Data, Computing in Science & Engineering 13 (2) (2011) 40–51.

[38] J. Daintith, E. Wright, Lazy evaluation, in: A Dictionary of Computing, Oxford University Press, 2008.

[39] M. Cornwell, Anatomy of a solid-state drive, Queue 10 (10) (2012) 30–36.

[40] F. Goichon, G. Salagnac, S. Frénot, Swap fairness for thrashing mitigation, in: K. Drira (Ed.), Software Architecture, Vol. 7957 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2013, pp. 311–315.

[41] Q. Koziol, HDF5, in: D. Padua (Ed.), Encyclopedia of Parallel Computing, Springer, 2011, pp. 827–833.

[42] J. Cupitt, K. Martinez, VIPS: An image processing system for large images, Proceedings of SPIE 2663 (1996) 19–28.

[43] C. Wei, R. Liu, M. Zwicker, H. Pfister, Hardware-accelerated adaptive EWA volume splatting, in: Visualization, 2004. IEEE, 2004, pp. 67–74.

[44] P. Hastreiter, T. Ertl, Integrated registration and visualization of medical image data, in: Computer Graphics International 1998 Proceedings, 1998, pp. 78–85.

[45] B. Zitova, J. Flusser, Image registration methods: a survey, Image and Vision Computing 21 (2003) 977–1000.

[46] J. B. A. Maintz, M. A. Viergever, A survey of medical image registration, Medical Image Analysis 2 (1) (1998) 1–36.

[47] O. E. Mealha, M. J. Ribeiro, C. Rocha, J. J. Pedroso de Lima, A. S. Pereira, B. S. Santos, M. F. Botelho, 3D registration and integrated visualization of multimodal clinical data, in: Proceedings of the 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Engineering Advances: New Opportunities for Biomedical Engineers, Vol. 1, 1994, pp. 600–1.

[48] J. Yang, J. P. Williams, Y. Sun, R. S. Blum, C. Xu, A robust hybrid method for nonrigid image registration, Pattern Recogn. 44 (4) (2011) 764–776.

[49] J. Kybic, M. Unser, Fast parametric elastic image registration, Image Processing, IEEE Transactions on 12 (11) (2003) 1427–1442.

[50] R. Szeliski, Computer vision: algorithms and applications, Springer, 2010.

[51] A. J. Croxford, P. D. Wilcox, G. Konstantinidis, B. W. Drinkwater, Strategies for overcoming the effect of temperature on guided wave structural health monitoring, in: Proceedings of SPIE, Vol. 6532, 2007, pp. 65321T–1–10.

[52] P. D. Wilcox, A. J. Croxford, J. E. Michaels, Y. Lu, B. W. Drinkwater, A comparison of temperature compensation methods for guided wave structural health monitoring, in: D. O. Thompson, D. E. Chimenti (Eds.), Review of Quantitative Nondestructive Evaluation, Vol. 975, American Institute of Physics, 2008, pp. 1453–1460.

[53] T. Clarke, Guided wave health monitoring of complex structures, Ph.D. thesis, Imperial College London, Department of Mechanical Engineering (2009).

[54] A. Galvagni, Pipeline health monitoring, Ph.D. thesis, Imperial College London, Department of Mechanical Engineering (2013).

[55] N. Brierley, EngD 18-month progress report: Improving the reliability of automated non-destructive inspection, Tech. rep., Imperial College London (2011).

[56] A. Collignon, F. Maes, D. D., D. Vandermeulen, P. Suetens, G. Marchal, Automated multi-modality image registration based on information theory, Information processing in medical imaging 3 (1995) 263–274.

[57] M. Sezgin, B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation, Journal of electronic imaging 13 (1) (2004) 146–168.

[58] T. Tippetts, N. Brierley, P. Cawley, Data registration for automated non-destructive inspection with multiple data sets, in: D. O. Thompson, D. E. Chimenti (Eds.), Review of Progress in Quantitative Nondestructive Evaluation, Vol. 1511, American Institute of Physics, 2013, pp. 1773–1780.

[59] M. Torki, A. Elgammal, One-shot multi-set non-rigid feature-spatial matching, in: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010, pp. 3058–3065.

[60] L. Dianchao, S. Cheng, A brief introduction of feature matching 2008 IEEE Region 5 Conference, Vol. ISSU, 2008, pAGE.

[61] M. Pilu, A direct method for stereo correspondence based on singular value decomposition, in: Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, 1997, pp. 261–266.

[62] G. L. Scott, H. C. Longuet-Higgins, An algorithm for associating the features of two images, Proceedings: Biological Sciences 244 (1309) (1991) 21–26.

[63] L. S. Shapiro, J. Michael Brady, Feature-based correspondence: an eigenvector approach, Image and Vision Computing 10 (5) (1992) 283–288.

[64] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Conference on Computer Vision and Pattern Recognition.

[65] L. Nan, W. Han, Feature extraction using evolutionary weighted principal component analysis, in: Systems, Man and Cybernetics, 2005 IEEE International Conference on, Vol. 1, 2005, pp. 346–350 Vol. 1.

[66] K. Yan, R. Sukthankar, PCA-SIFT: a more distinctive representation for local image descriptors, in: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, Vol. 2, 2004, pp. II–506–II–513 Vol.2.

[67] A. E. Johnson, M. Hebert, Using spin images for efficient object recognition in cluttered 3D scenes, Pattern Analysis and Machine Intelligence, IEEE Transactions on 21 (5) (1999) 433–449.

[68] C. H. Papadimitriou, Combinatorial optimization: algorithms and complexity, 2nd Edition, Dover, 1998.

[69] P. Viola, W. M. Wells III, Alignment by maximization of mutual information, International Journal of Computer Vision 24 (2) (1997) 137–154.

[70] P. Xu, D. Yao, A study on medical image registration by mutual information with pyramid data structure, Computers in Biology and Medicine 37 (3) (2007) 320–327.

[71] N. Brierley, EngD 12-month progress / transfer report: Improving the reliability of automated non-destructive inspection, Tech. rep., Imperial College London (2010).

[72] N. Brierley, T. Tippetts, P. Cawley, Improving the reliability of automated non-destructive inspection, in: D. O. Thompson, D. E. Chimenti (Eds.), Review of Progress in Quantitative Nondestructive Evaluation, Vol. 1430, American Institute of Physics, 2012, pp. 1824–1831.

[73] C. Voglis, I. E. Lagaris, Towards "ideal multistart". A stochastic approach for locating the minima of a continuous function inside a bounded domain, Applied Mathematics and Computation 213 (1) (2009) 216–229.

[74] C. Audet, J. J. E. Dennis, Analysis of generalized pattern searches, SIAM Journal on Optimization 13 (3) (2002) 889–903.

[75] K. O. Jones, Comparison of genetic algorithm and particle swarm optimisation, in: International Conference on Computer Systems and Technologies - CompSysTech, ACM, Varna, Bulgaria, 2005, pp. IIIA.1.1–IIIA.1.6.

[76] D. R. Jones, M. Schonlau, W. J. Welch, Efficient global optimization of expensive black-box functions, Journal of Global Optimization 13 (1998) 455–492.

[77] D. R. Jones, A taxonomy of global optimization methods based on response surfaces, Journal of Global Optimization 21 (4) (2001) 345–383.

[78] R. Poli, Particle swarm optimization, Swarm intelligence 1 (1) (2007) 33–57.

[79] J. A. Vrugt, B. A. Robinson, Improved evolutionary optimization from genetically adaptive multimethod search, Proceedings of the National Academy of Sciences of the United States of America 104 (3) (2007) 708–11.

[80] J. A. Vrugt, B. A. Robinson, J. M. Hyman, Self-adaptive multimethod search for global optimization in real-parameter spaces, IEEE Transactions on Evolutionary Computation 13 (2) (2009) 243–59.

[81] J. E. Fieldsend, R. M. Everson, Multi-objective optimisation in the presence of uncertainty, in: The 2005 IEEE Congress on Evolutionary Computation, Edinburgh, 2005, pp. 243–250.

[82] P. Flach, ROC analysis, in: C. Sammut, G. Webb (Eds.), Encyclopedia of Machine Learning, Springer US, 2010, pp. 869–875.

[83] A. Connor, Automated analysis of non destructive evaluation data, Ph.D. thesis, Imperial College London, Department of Mechanical Engineering (2011).

[84] S. Leary, A. Bhaskar, A. Keane, Optimal orthogonal-array-based latin hypercubes, Journal of Applied Statistics 30 (5) (2003) 585–598.

[85] N. Brierley, EngD 30-month progress report: Improving the reliability of automated non-destructive inspection, Tech. rep., Imperial College London (2012).

[86] B. Olin, W. Meeker, Applications of statistical methods to nondestructive evaluation, Technometrics 38 (2) (1996) 95–111.

[87] M. Wall, Modelling of inspection reliability, Engineering Science and Education Journal 6 (2) (1997) 63–72.

[88] J. Ogilvy, Model for predicting ultrasonic pulse-echo probability of detection, NDT&E International 26 (1) (1993) 19–29.

[89] I. Cornwell, A. McNab, Towards automated interpretation of ultrasonic NDT data, NDT&E International 32 (1999) 101–107.

[90] I. D. Hall, A. McNab, G. Hayward, Improved ultrasonic image generation through tomographic image fusion, Ultrasonics 37 (1999) 433–443.

[91] D. Horn, W. R. Mayo, NDE reliability gains from combining eddy-current and ultrasonic testing, NDT&E International 33 (2000) 351–362.

[92] R. J. Radke, S. Andra, O. Al-Kofahi, Image change detection algorithms: A systematic survey, IEEE Transactions on Image Processing 14 (3) (2005) 294–307.

[93] D. L. Hall, J. Llinas, An introduction to multisensor data fusion, Proceedings of the IEEE 85 (1) (1997) 6–23.

[94] L. Zheng, D. S. Forsyth, J. P. Komorowski, K. Hanasaki, T. Kirubarajan, Survey: State of the art in NDE data fusion techniques, Instrumentation and Measurement, IEEE Transactions on 56 (6) (2007) 2435–2451.

[95] D. L. Hall, S. A. H. McMullen, Mathematical Techniques in Multisensor Data Fusion, 2nd Edition, Artech House Inc, Boston, Massachusetts, 2004.

[96] X. E. Gros, A review of data fusion processes and applications, Insight 39 (9) (1997) 652–657.

[97] C. Kohl, M. Krause, C. Maierhofer, J. Wöstmann, 2D- and 3D-visualisation of NDT-data using data fusion technique, Materials and Structures 38 (9) (2005) 817–826.

[98] X. E. Gros, J. Bousigue, K. Takahashi, NDT data fusion at pixel level, NDT&E International 32 (1999) 283–292.

[99] X. E. Gros, P. Strachan, D. W. Lowden, Fusion of multiprobe NDT data for ROV inspection, in: Oceans MTS/IEEE, Vol. 3, 1995, pp. 2046–2050.

[100] Y. H. Lu, J. E. Michaels, Feature extraction and sensor fusion for ultrasonic structural health monitoring under changing environmental conditions, IEEE Sensors Journal 9 (11) (2009) 1462–1471.

[101] X. E. Gros, P. Strachan, D. W. Lowden, Theory and implementation of data fusion, Research in Nondestructive Evaluation 6 (1995) 227–236.

[102] K. Friston, J. Ashburner, S. Kiebel, T. Nichols, W. Penny, Statistical Parametric Mapping, Academic Press, 2007.

[103] R. V. Hogg, J. W. McKean, A. T. Craig, Introduction to Mathematical Statistics, 7th Edition, Pearson Education, Boston, MA, 2013.

[104] R. B. Thompson, F. J. Margetan, L. Yu, A formal theory for the spatial correlation of backscattered ultrasonic gain noise, in: D. O. Thompson, D. E. Chimenti (Eds.), Review of Quantitative Nondestructive Evaluation, Vol. 24, American Institute of Physics, 2005.

[105] M. O'Donnell, S. Silverstein, Optimum displacement for compound image generation in medical ultrasound, IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control 35 (4) (1988) 470–476.

[106] C. Schneider, Application of extreme value analysis to corrosion mapping data, in: 4th European-American Workshop on Reliability of NDE, 2009.

[107] T. M. Loughin, A systematic comparison of methods for combining p-values from independent tests, Computational Statistics and Data Analysis 47 (3) (2004) 467–485.

[108] W. R. Rice, A consensus combined p-value test and family-wide significance of component tests, Biometrics 46 (1990) 303–308.

[109] K. J. Rothman, No adjustments are needed for multiple comparisons, Epidemiology 1 (1) (1990) 43–46.

[110] D. A. Clifton, S. Hugueny, L. Tarassenko, Pinning the tail on the distribution: A multivariate extension to the generalised Pareto distribution, in: Machine Learning for Signal Processing (MLSP), 2011 IEEE International Workshop on, 2011, pp. 1–6.

[111] A. Frigessi, O. Haug, H. Rue, A dynamic mixture model for unsupervised tail estimation without threshold selection, Extremes 5 (3) (2002) 219–235.

[112] R. A. Fisher, Statistical methods for research workers, Oliver and Boyd, Edinburgh, 1925.

[113] M. Westberg, Combining independent statistical tests, Journal of the Royal Statistical Society. Series D (The Statistician) 34 (3) (1985) 287–296.

[114] A. M. Winkler, The logic of the Fisher method to combine p-values, `http://brainder.org/2012/05/11/the-logic-of-the-fisher-method-to-combine-p-values/`, accessed: 2012-12-21 (2012).

[115] M. F. Brilhante, D. Pestana, F. Sequeira, Combining p-values and random p-values, in: Information Technology Interfaces (ITI), 2010 32nd International Conference on, 2010, pp. 515–520.

[116] H. Abdi, Bonferroni and Šidák corrections for multiple comparisons, in: N. Salkind (Ed.), Encyclopedia of Measurement and Statistics, Sage, 2007, pp. 103–107.

[117] P. H. Westfall, R. D. Wolfinger, Multiple tests with discrete distributions, The American Statistician 51 (1) (1997) 3–8.

[118] W. J. Conover, Practical Nonparametric Statistics, 3rd Edition, Wiley, New York, 1999.

[119] A. Agresti, A survey of exact inference for contingency tables, Statistical Science 7 (1) (1992) 131–153.

[120] P. Jäckel, Monte Carlo Methods in Finance, Wiley, 2002.

[121] I. Foster, Designing and Building Parallel Programs, Addison-Wesley, 1995.

[122] H. Niederreiter, Quasi-Monte Carlo methods, in: Encyclopedia of Quantitative Finance, John Wiley & Sons, Ltd, 2010.

[123] N. Brierley, T. Tippetts, P. Cawley, Data fusion for improving the reliability of automated non-destructive inspection, in: D. O. Thompson, D. E. Chimenti (Eds.), Review of Progress in Quantitative Nondestructive Evaluation, Vol. 1511, American Institute of Physics, 2013, pp. 1781–1788.

[124] E. Carrera, et al., pydot: Python interface to Graphviz's Dot language.
URL `https://code.google.com/p/pydot/`

[125] C. Holmes, B. W. Drinkwater, P. D. Wilcox, Post-processing of the full matrix of ultrasonic transmit-receive array data for non-destructive evaluation, NDT&E International 38 (2005) 701–711.

[126] R. B. Thompson, E. F. Lopes, The effects of focusing and refraction on gaussian ultrasonic beams, Journal of nondestructive evaluation 4 (2) (1984) 107–123.

[127] P. Calmon, S. Leberre, T. Sollier, P. Benoist, CIVA, an integration software platform for the simulation and processing of NDT data, in: Proceedings of 16th World Conference on Non-Destructive Testing, Montreal, Canada, 2004.
URL `http://www-civa.cea.fr/en/`

[128] P. B. Nagy, M. Blodgett, M. Golis, Weep hole inspection by circumferential creeping waves, NDT & E International 27 (3) (1994) 131–142.

[129] K. K. Talukdar, W. D. Lawing, Estimation of the parameters of the rice distribution, The Journal of the Acoustical Society of America 89 (3) (1991) 1193–1197.

[130] M. C. Jones, J. S. Marron, S. J. Sheather, A brief survey of bandwidth selection for density estimation, Journal of the American Statistical Association 91 (433) (1996) 401–407.

[131] X. Rui, I. Wunsch, D., Survey of clustering algorithms, Neural Networks, IEEE Transactions on 16 (3) (2005) 645–678.

List of Publications

Written publications

[P1] N. Brierley, T. Tippetts and P. Cawley. Improving the reliability of automated non-destructive inspection, in *Review of Progress in Quantitative NDE*, Burlington, Vermont, USA. American Institute of Physics Conference Proceedings, volume 1430, pages 1824-1831. Edited by D. Chimenti, B. Thompson. 2012.

[P2] N. Brierley, T. Tippetts and P. Cawley. Improving the reliability of automated non-destructive inspection, in *Proceedings of 18th World Conference on Non-Destructive Testing*, Durban, South Africa. 2012.

[P3] N. Brierley, T. Tippetts and P. Cawley. Data processing to improve the reliability of automated non-destructive inspection, in *Proceedings of British Institute of NDT Conference*, Daventry, UK. 2012.

[P4] N. Brierley, T. Tippetts and P. Cawley. Data fusion for improving the reliability of automated non-destructive inspection, in *Review of Progress in Quantitative NDE*, Denver, Colorado, USA. American Institute of Physics Conference Proceedings, volume 1511, pages 1781-1788. Edited by D. Chimenti, B. Thompson. 2013.

[P5] T. Tippetts, N. Brierley and P. Cawley. Data registration for automated non-destructive inspection with multiple data sets, in *Review of Progress in Quantitative NDE*, Denver, Colorado, USA. American Institute of Physics Conference Proceedings, volume 1511, pages 1773-1780. Edited by D. Chimenti, B. Thompson. 2013.

[P6] N. Brierley, T. Tippetts and P. Cawley. Improving the reliability of automated

non-destructive inspection, in *Proceedings of British Institute of NDT Conference*, Telford, UK. 2013.

[P7] N. Brierley, T. Tippetts and P. Cawley. Improving the reliability of automated non-destructive inspection, in *Review of Progress in Quantitative NDE*, Baltimore, Maryland, USA. American Institute of Physics Conference Proceedings, volume 1581, pages 1912-1919. Edited by D. Chimenti, B. Thompson. 2014.

[P8] N. Brierley, T. Tippetts and P. Cawley. Data fusion for automated non-destructive inspection, in *Proceedings of the Royal Society A*, London, UK. Volume 470, number 2167. 2014.

Separate posters & oral presentations

[P8] N. Brierley, T. Tippetts and P. Cawley. Capability assessment, at *British Institute of NDT Aerospace NDT Forum*, Coventry, UK. 2013.

[P9] N. Brierley, T. Tippetts and P. Cawley. Improving the reliability of automated non-destructive inspection, at *5th European-American Workshop on Reliability of NDE*, Berlin, Germany. 2013. *Won 1st prize in poster competition.*