IMPERIAL COLLEGE LONDON

IMPROVED RELIABILITY OF AUTOMATED NON-DESTRUCTIVE EVALUATION

by

Trevor Tippetts

A thesis submitted to the Imperial College London for the degree of

Doctor of Philosophy

Department of Mechanical Engineering
Imperial College London
London SW7 2AZ

May 2014

Declaration of originality

The material presented in this thesis “Improved Reliability of Automated Non-Destructive Evaluation” is entirely my own research work under the supervision of Professor Peter Cawley. All published and unpublished material referred to in this thesis has been given full acknowledgment.

Name: Trevor Tippetts

Date: May 16, 2014

Signed:

Copyright

Abstract

In recent years, Non-Destructive Evaluation (NDE) has trended toward increased automation in data acquisition. Automated scanning has the potential to greatly increase reliability of the NDE results, but it also tends to increase the volume of data that must be inspected manually by a skilled technician. This is a time-consuming task made tedious by the fact that most of the data contains no indication of a defect. There is a great need for software that can partially automate the data analysis by prioritizing regions of interest to the inspector. This thesis describes an approach to that end, laying out a framework that is general enough to fit a wide array of NDE applications. It also describes practical considerations for the specific application of ultrasound inspection of titanium turbine discs.

Acknowledgments

My deepest thanks go to my supervisor, Prof. Peter Cawley, for superb guidance and support. I am very grateful for the opportunity I have had to work in the Non-Destructive Testing group at Imperial College.

I wish to express my gratitude to my colleague, Nick Brierley, for many hours of discussion and many lines of code. This work would not have been possible without his help and input, and I could not have hoped for a more pleasant person with whom to work.

I owe a great debt of gratitude to David Wright, Tony Dunhill, Katie Milne (now of MTC), and Edwill Escobar-Ruiz of Rolls-Royce for their assistance, support, and background knowledge. Without David Wright's carefully collected data and many answers to questions, I would have had nothing to talk about and nothing to say about it.

Many thanks to Nicolas Bonadeo at Tenaris and other members of the Research Centre in Non-Destructive Evaluation for their support and valuable discussions.

I would like to thank Andrew Ward and the kind staff at NDT Services for helping me better understand ultrasound inspection in a practical setting.

I must thank Rich and Mary Mah for their tremendous support and encouragement. Their generous assistance made a challenging experience possible, and their company helped to make it enjoyable.

I thank my parents, Dean and Teddie Tippetts, for teaching me to value education and hard work, for sacrificing so much to give me a start in life, and for tolerating my serial dismantling of household appliances as a pre-engineer.

Last and most of all, I thank my long-suffering wife, Emily. Thank you for sticking with me through a crazy dream and an international downsizing. I dedicate this dissertation to you, as well as to our wonderful sons, Damian and Ezekiel. I have learned enough from you to write ten dissertations.

Contents

List of Figures

List of Tables

Nomenclature

t, u, v	Indices to 3D ultrasound amplitude data
x, y, z	Physical Cartesian coordinates
θ, r, z	Physical cylindrical coordinates
θ_0, r_0, z_0	Hypothetical data coordinates due to beam propagation effects only
p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}	Registration parameters
N_u	Number of A-scans per circumference
φ	Ultrasound beam front surface incident angle
$\hat{n}_{refract}$	Unit vector in the direction of the refracted ultrasound beam
$\hat{n}_{FS}$	Outward-pointing front surface unit normal vector
d_{metal}	Distance between the surface incidence point and the data sample location
f_s	Data acquisition sample frequency
$r_{FS}(v)$, $z_{FS}(v)$	Nominal front surface physical coordinates
R_{tilt}	Rotation matrix that applies the effect of tilt registration mapping
$p(f, g)$	Joint probability density function between random variables f and g
p_f	Marginal probability density function for random variable f
$MI[f, g]$	Mutual information between random variables f and g
$L_p[f - g]$	L_p norm of the residual vector between data sets f and g
r_{COM}, z_{COM}	Surface echo center of mass
N_d	Number of data samples in a surface echo feature
x_{min}	Parameter value at global minimum in a single-objective optimization
f	Optimization objective function
x_{dom}	Dominated parameter vector in multi-objective optimization
N_f	Number of objective functions
N_p	Number of registration parameters

Continued on next page...

– Continued from previous page

p^*	Pareto optimal point in the parameter domain
f^*	Objective function value on the Pareto front
r^*, z^*	r and z mapped location of a Pareto optimal data point
σ_{r^*}, σ_{z^*}	Standard deviation of r^* and z^* over the Pareto front
$\delta_F(f^*)$	Hypervolume Indicator expected density
$w(f^*)$	Hypervolume Indicator weight function
$\hat{e}^*$	Unit vector normal to the Pareto front in the objective function domain
$E_{Pareto}\left[g(f^*)\right]$	Integral of a function over the Pareto front in the objective domain
J	Jacobian matrix for objectives as functions of registration parameters
f_{RBF}	Radial Basis Function approximation to an objective function
N_{RBF}	Number of Radial Basis Function nodes
$k_i(p)$	Radial Basis Function kernel at parameter point p_i, evaluated at p
a_i	Radial Basis Function kernel amplitude for node at parameter point p_i
f_{max}	Maximum value of an objective function over all Pareto points
$P(s)$	Cumulative probability of signal amplitude s, conditioned on the presence of a true flaw
H_0	Null hypothesis for a hypothesis test
p_i	ith input p-value for a consensus test
p_{LEN}	LEN detector p-value
p_{OR}	Consensus p-value from the OR test
p_{AND}	Consensus p-value from the AND test
p_{DFD}	Consensus p-value from the Discrete Fisher Detector
p_{MPV}	Consensus p-value from the Minimum P-Value test
T_{Fisher}	Fisher consensus test statistic
μ	Mean of a Pareto distribution

Continued on next page...

– Continued from previous page

σ	Shape parameter for a Pareto probability distribution
A	Signal amplitude for a Pareto probability distribution
$\Gamma(\cdot)$	Gamma function
$\Gamma(\cdot,\cdot)$	Upper incomplete gamma function
$E_{p_x}[g(x)]$	Expected value of $g(x)$, under probability distribution p_x for random variable x
N_{MC}	Number of Monte Carlo sample points
h_x	Importance Sampling density for random variable x
w_j	Importance Sampling weight for sample point j
$H^{-1}(x)$	Inverse function of the Importance Sampling cumulative distribution for random variable x
$\tilde{p}_i$	Monte Carlo p-value, sampled from the Importance Sampling distribution, for the ith consensus test input
m_i	Median of the Importance Sampling distribution for the ith consensus test input
$\tilde{w}_i$	Importance Sampling marginal weight for the ith consensus test input
$\tilde{w}$	Weight for the Importance Sampling Monte Carlo point
N_{resel}	Total number of resels in the disc
A_j	Scaled-down amplitude from scan j
A_{j0}	Original amplitude from scan j
α	Scaled-down indication scale factor
μ_{noise}	Mean of the noise distribution
σ_{noise}	Standard deviation of the noise distribution

Chapter 1

Introduction

1.1 Motivation

Advances in technology have driven capability up and costs down for automated data acquisition and storage in Non-Destructive Testing (NDT). Automated acquisition hardware can greatly improve reliability of NDT results by maintaining precise relative motion of the probes, standardized instrumentation settings, and repeatable part coverage.

Digital data storage continues to increase in speed and capacity, which makes it possible to store full records of inspection data. This also creates the possibility of improved reliability through more sophisticated data processing, beyond the relatively simple amplitude thresholding or image generation often used today.

However, many NDT inspections still rely on interpretation by a skilled technician, even when the data acquisition is automated. Parts that undergo NDT tend to be expensive to produce and have a high pass rate, since the inspection costs would otherwise be greater than the loss from scrapping questionable parts without inspection. This means that inspectors are often forced to spend a large fraction of their time verifying defect-free data. Not only does the human involvement create a time bottleneck, it can decrease reliability due to operator fatigue. Both of these problems are exacerbated as hardware capacities increase the volume of data produced in each inspection. Thus the challenges and costs are likely to increase over time unless automation can alleviate the burden on inspectors.

Many NDT applications would benefit from a computational capability to screen data automatically, but a fully automated evaluation with no human intervention would face prohibitive challenges. Computers are very useful for processing large volumes of predictable data. Anomalous inspection data from high-quality parts is, by nature, rare and unpredictable. It would be extremely difficult to write software that would properly handle all of the unforeseen, unique cases that could arise, sufficient to reassure the manufacturer. This is especially true given the cost and risk that manufactures would incur if a defect were missed in a critical application. In addition, many NDT applications fall under regulatory requirements for qualified human inspectors to take responsibility for the evaluation, which the manufacturer and inspector are not at liberty to change.

It is therefore important to emphasize that the goal for this NDT automation work is not to replace the human inspector. Rather, the goal is to divide the task between computer and human in a way that makes the best use of their characteristics. Such a system would automate the filtering out of regions that are clearly free from defect, while flagging data that are more difficult to interpret. This would assist inspectors to prioritize their time and focus their attention where it can add the most value.

A computer can apply a pre-programmed algorithm to vast quantities of data quickly, reliably, and repeatably. There are no issues with fatigue or inattention to small detail. The computer has no problem working with high-dimensional data, possibly merged from different sources. Most importantly, computational technology increases in capacity over time, in parallel with data acquisition. The two scale together, so bottlenecks are avoided both now and in the future.

Human inspectors, in contrast, are better able to interpret and respond to unexpected results. They can make decisions based on experience and context, recognize and correct an anomalous data acquisition, and know when to report a problem or ask for help. These are all functions that are very difficult or impossible to program for a computer. A tool that facilitates rather than obviates the inspector's work is more likely to be successful and accepted in practice, because it does not hide data as would a fully automated evaluation. This thesis proposes such a tool to maintain high NDT reliability while potentially reducing cost and inspection time by easing the data volume bottleneck.

What follows is a system that prioritizes inspection data for evaluation. The software selects a series of small regions from the inspection data, ordered by decreasing likelihood of containing a defect. The human inspector views the sequence of possible indications and makes all final decisions regarding the part fitness. This saves the inspector's time by immediately highlighting the most important regions to examine. The inspector is spared from combing through the rest of the data, which contains only defect-free noise.

This automated NDT system is not only faster and better-scaling, but it can also increase reliability in multiple ways, using existing data. Computational detection algorithms allow capabilities that are not possible with the current manual inspection technology. Instead of a single threshold, imposed by the inspection policy before the data are collected, automated detection can find signals relative to local noise levels. In this way it adapts itself to each individual disc and to the different regions and scan conditions within a disc. Perhaps most importantly, manual inspection considers each scan individually, but automated NDT could combine data from overlapping scans to increase reliability [1, 2].

This new automation capability is within reach, but it brings additional challenges. First, there is a need to align the data from multiple scans to make them sensible to combine. This data alignment process is known as registration, and it must be automated to achieve the benefits of efficiency and reliability [3–5]. No real system will ever achieve perfect registration, so there is also a need to quantify the residual uncertainty, using the same data that were used to perform the registration. As part of this thesis, a method was devised for estimating this uncertainty. This method could also have much wider application to other instances of multi-objective optimization.

Automated inspection also clearly requires the means to quantify defect likelihood at all points throughout the part. Efficiency is a key component in these expensive detection algorithm computations. Another contribution of this thesis is an improved numerical technique for the data fusion detector. Without the speed-up that it provides, it would not be feasible to run the detection computation on the large data sets that come from automated NDT data acquisition.

1.2 Thesis outline

The overall approach to NDT automation developed for this thesis is laid out in Chapter 2. This chapter also describes the target application for the automated system: ultrasound inspection of aircraft engine turbine discs. The work for this thesis was carried out for the turbine disc application, and the results that follow use data from those inspections. However, the approach was intended to be as generally applicable as possible. In fact, the present author worked in close collaboration with an Engineering Doctorate student, Nick Brierley, who has directed his efforts toward an NDT application in the power generation industry. Both authors wrote software and contributed to a common repository. The generality of the approach is evidenced by the extent to which this software could be shared. Chapter 2 outlines the segments of the software that were primarily written by each student.

Chapter 3 describes the concept of data registration for multiple data sets with overlapping coverage, in the context of the turbine disc application. More detail is given on the multiple stages in this computation, from feature extraction, geometry mapping, quality-of-fit metric calculation, and multi-objective optimization. Chapter 3 also contains a description of the Pareto integral estimation method. The work for this thesis developed this technique for estimating integrals over the Pareto-optimal set, based on statistics of the optimizer Pareto points, in order to quantify the uncertainty in registration results.

Results from the registration approach and software are shown in Chapter 4. Registration tests, using data from a real disc scan but altered such that an exact solution is known, establish the validity of the approach. These registration test results illustrate the capabilities of the automated registration developed for this thesis and provide guidance for the expected accuracy. Following the test cases, results are shown for registration on the full turbine disc data. To the author's knowledge, the turbine disc and Brierley's power generation application together represent an original contribution to registration between three-dimensional NDT data sets of this large scale. The particular results for this case are presented as an example for future NDT work that would need to cope with similarly large quantities of higher-dimensional data.

Once data sets are aligned with each other by registration, it is possible to use a data fusion algorithm to detect flaws and defects. Chapter 5 defines, compares, and contrasts several detection algorithms. Included is a detector that captures the behavior of the current manual inspection procedure, based on a universal amplitude threshold. This amplitude threshold detector establishes a basis for comparison. As a conceptual stepping stone, an improved single-scan detector then introduces adaptivity to local noise levels. Finally, a data fusion detector is described that can combine separate scans that view the same location from different directions. It is this data fusion detector that will provide the detection capability for the automated NDT system. Detection results for bounding test cases show the improved performance over the status quo inspection method. Chapter 5 also contains a description of an improved numerical method developed for this thesis that accelerates computation in the data fusion detector.

The detection results in Chapter 6 apply the approach outlined in Chapter 5 to the registered data. First, known indications in a disc with seeded defects serve to differentiate performance of the various detectors. Then the detectability limits for realistic defect signals are probed, by scaling down the amplitudes from known indications. Finally, Chapter 6 also shows results for several new, possible indications that had not been detected with previous inspection methods.

Chapter 7 summarizes the main findings of this thesis. Possible avenues for future work are given that could extend the present work or branch into further applications.

Chapter 2

Application and approach

This chapter lays out the objectives of the project and gives an overview of the approach. Aspects of this work were conducted in collaboration with Nick Brierley, who is pursuing an Engineering Doctorate degree with a similar project [6–8]. The development and results for this thesis necessarily targeted a particular application, to demonstrate the relevance and capabilities of the approach. The methods developed were also intended to be general and extensible. The discussion therefore also gives a brief description and references to Brierley's application, and it explains how the principles could relate to other applications.

This work was supported by the Research Centre in Non-Destructive Evaluation (RCNDE). The RCNDE is an organization that facilitates collaboration between industry and academic institutions. RCNDE coordinates NDT research to ensure research topics are relevant to medium- to long-term industrial needs. Rolls-Royce, RWE Npower, and Tenaris are a few of the member companies of RCNDE that have expressed interest in improved automation in NDT. They have all committed to sponsoring research into this area of study. Rolls-Royce, in particular, has provided data and expertise for the work in this thesis. The target application for this thesis comes from their ultrasound inspection of aircraft engine turbine discs.

2.1 Turbine disc application

The software written for this project was designed to be general and easily extended to many specific NDT applications, but it was necessary to develop and demonstrate its use in the context of a specific application. This section shows the perspective from the present author's main application.

Figure 2.1: Titanium aircraft engine turbine disc. This disc is one of 3 forgings machined to the inspection shape of a Trent 890B compressor disc from a contaminated billet. The material contamination produced five known seeded defects inside the disc. Rolls-Royce uses this disc, labeled by its test piece serial number 5, to evaluate their inspection procedures and equipment. Photo: N. Brierley

Turbine discs for aircraft engines are inspected for small internal flaws at an intermediate stage of production. The titanium forgings are machined to a shape with a rectilinear cross section, specifically to facilitate ultrasound NDT. Figure 2.1 shows a typical example of such a rectilinear disc. The disc is then secured to a rotating platform in a water-filled immersion tank, as shown in Figure 2.2. A single focused probe is positioned above the disc and records data in a pulse-echo mode. The probe scans parallel to each surface of the disc like a phonograph record or digital optical

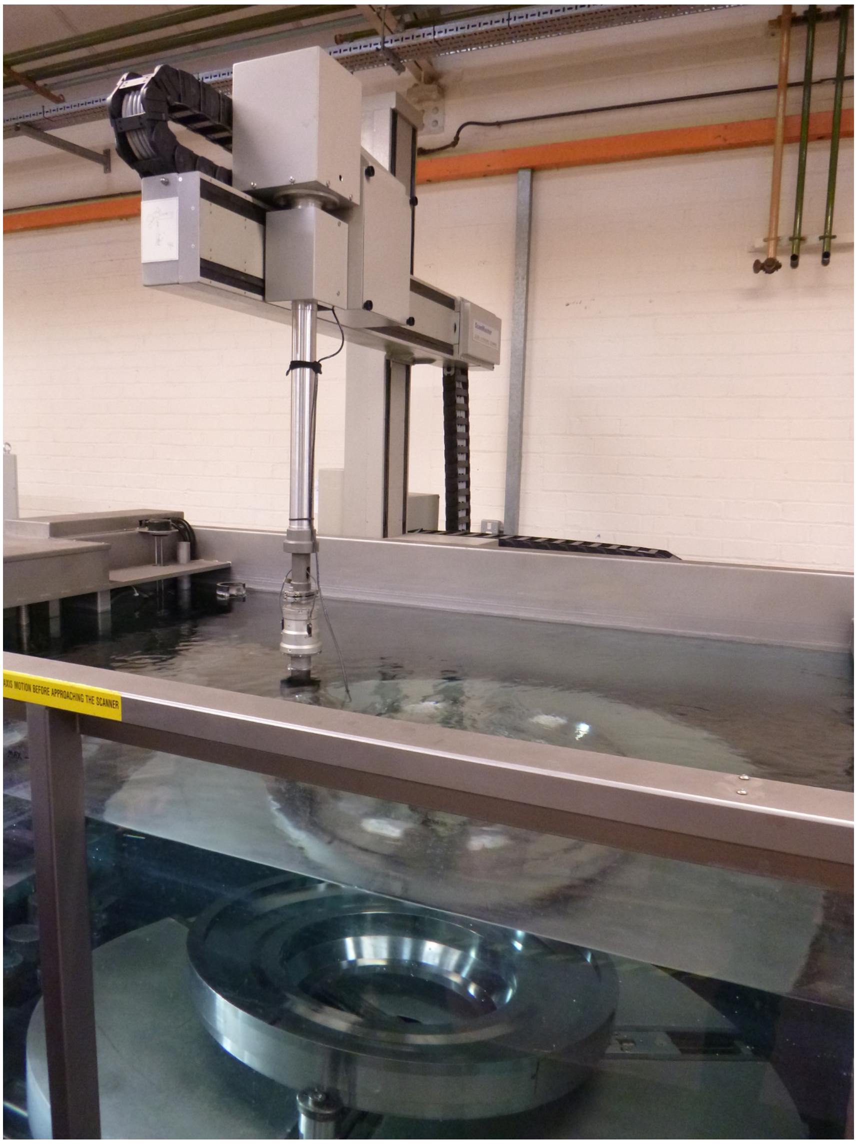


Figure 2.2: Seeded Defect Disc in the water immersion tank, ready for inspection. The disc rests on a rotating platform at the bottom of the tank. A computer-controlled, two-axis scanning frame moves the ultrasound probe radially and vertically to follow the cross-sectional profile of the disc. Photo: N. Brierley

disc player, moving in radial (r) and axial (z) steps as the platform rotates beneath. Figure 2.3 shows a closer view of the ultrasound probe as it scans the outer rim surface of the disc. The probe scans each surface three times, with three incidence angles: $0°$ (normal to surface), $+5°$, and $-5°$.

Ultrasound mode conversion is governed by the critical angles at the interface. It is possible to estimate the critical angles for transmission from the water into the titanium with the ratios of the wave speeds [9, 10]. In titanium, the longitudinal wave speed is approximately 6.17 km/s, and the transverse wave speed is 3.3 km/s.

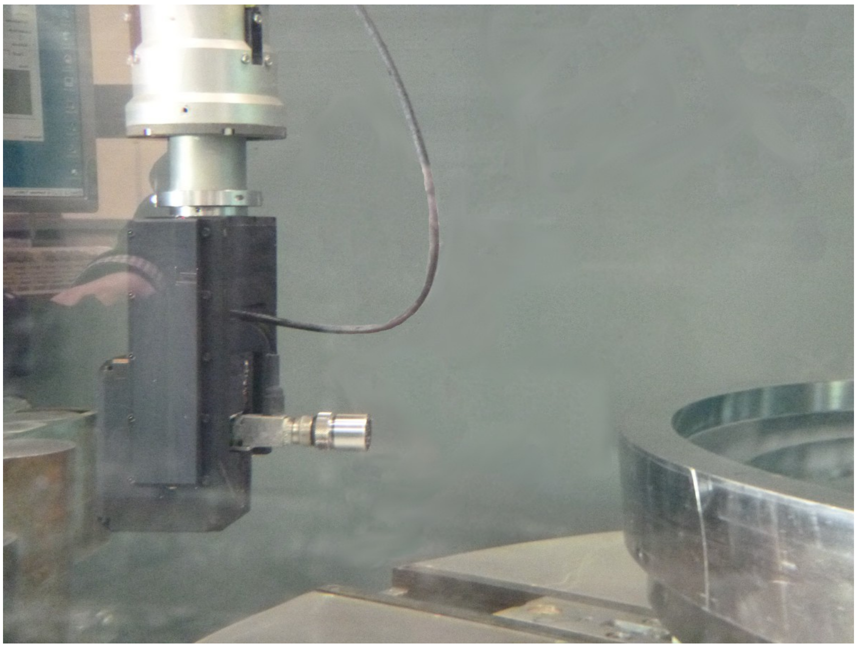


Figure 2.3: Close-up view of the turbine disc inspection. The scanner is programmed to maintain a 160 mm water path between the ultrasound transducer and the surface of the disc. Here it scans surface KJ on the rim of the disc. Photo: N. Brierley

The longitudinal wave speed in water is roughly 1.48 km/s. This makes the first and second critical angles 14° and 27°, respectively. At a 5° incidence angle, the longitudinal transmitted wave will refract at 21°.

The beam center line is well within the first critical angle at all three surface incidence angles, so most of the energy is transmitted as a refracted longitudinal wave. A smaller amplitude transverse wave will also generally propagate in the titanium. Given the variation in incidence angle across the focused beam, there is also the possibility of seeing diffracted surface waves near corners.

The author had the opportunity to observe several disc scans and discuss the procedures with the inspectors. This experience was helpful to gain an understanding of realistic conditions and practical considerations for this NDT procedure. Some of the observations related to automated inspection are reported below.

The scans use a 0.5 inch diameter, 5 MHz probe sampled at 100 MHz. There is a 0.1 inch dead zone under the front surface, due to the duration of the front surface echo. The inspected time gate extends to a maximum 4 inch depth or to the opposite dead zone, 0.1 inches inside the back wall. The probe has a beam width of approximately 1 mm at the focal point, which is typically set at just inside the part surface.

The scan resolution is typically fixed at half the width of the beam at the focal point [11]. The time-varying gain is set before the scan by calibration with a standard set of blocks with flat-bottom holes [12]. The Distance/Amplitude Correction (DAC) [13] curve derived from that calibration is applied by the sampling hardware, so the 8-bit samples use the full dynamic range and the same threshold should apply at all depths. For most scans of titanium parts, a response greater than -18 dB down from the calibration standard halts the scanning mechanism. The system then waits for the inspector to investigate manually and resume the scan.

Other focused ultrasound inspections have used multiple probes with different focal depths [14]. This multi-zone technique is designed to reduce the effect of beam spread beyond the focal point. With the single focused probe, at the standard 160 mm water gap distance, the beam can spread up to a 32.5 mm diameter at the maximum 4 inch depth in titanium. This might reduce the ability to localize a defect. It should not affect sensitivity since the DAC increases the gain to compensate for both the beam spread and other attenuative effects.

On some scanners, the Time of Flight trigger on the front surface oscillates visibly as the A-scans are plotted dynamically during the scan. This is because the rotating platform is not quite level. An annual correction removes the oscillation, but with time and usage the platforms develop a small tilt. The resulting periodic variation in water path length will require compensation by the mapping function. Inspectors also report that the wave speed in water has a noticeable dependence on temperature [15].

True positive detections are very rare, but the false call rate is roughly estimated as one out of ten discs [15]. The cause might be foreign matter on the surface, an air bubble in the water, or a part identification etched on the surface. Occasionally a small notch from irregular machining can trip the amplitude threshold. Usually the inspector can identify these harmless indications easily and continue the scan. There would be value in a system that could help the inspector find and investigate these indications without the need to interrupt and manually resume the scan.

Another phenomenon that impedes turbine disc NDT is the gramophone effect.

This happens when groove patterns from surface machining are too deep, causing a prolonged ringing in the front surface echo. Usually it only affects the front surface on the 0° scan. The gramophone effect increases the depth of the dead zone since no signal is distinguishable after the front surface until the ringing subsides. This currently requires the part to be sent back for resurfacing, since it is outside the specifications and it interferes with the simple indication threshold. It would be very worthwhile, in terms of both cost and scheduling, to adapt the dead zone to roughly 0.4 to 0.5 inches as needed, especially when this region will eventually be machined off. This is a case where, even if an automated inspection program could not accurately evaluate a region, it would still save the inspector time and effort by isolating the noisier region and salvaging the rest of the data. There is also the potential for a data fusion detector to mitigate the effect of gramophone ringing from one scan, by combining with data from the other scans that cover the same region.

Titanium tends to be a noisy material because of residual large-grain orientations. This coherent grain noise shows up as a change in the noise distribution on the A-scans [16–18]. Inspectors have the option to note the reduced sensitivity and continue the test, searching for defects greater than 4 dB above the nearby noise. The inspector moves the probe under manual control to determine if a spike is isolated or similar in amplitude to the surrounding region. These noisy scans are reported to the engine manufacturer, who can keep track of where the parts go into service and rate them for a shorter lifespan.

Variations like these from the default procedure are necessary for continuing a difficult scan, but they can be very time consuming. Fluctuations in noise levels across the part prevent the inspector from setting an automatic amplitude trigger. The scan can only proceed with manual advancement, often with a much slower rotation. The need to estimate nearby noise levels visually introduces a certain amount of subjectivity into the results. This type of inspection could be much faster and more reliable with automated screening.

2.1.1 Seeded defects

The turbine disc in Figure 2.1 has small seeded defects that Rolls-Royce uses to test their NDT hardware and inspection procedures. The Seeded Defect Disc was created by machining the disc from a titanium billet that was contaminated with foreign matter. Many NDT test pieces use small-diameter holes drilled into a block of material. The relatively smooth surfaces of the drilled hole target can have a lower directional dependance for the ultrasound reflectivity than the rough texture and irregular shapes in real defects. These drilled-hole targets are useful for calibration blocks, for example, because the signal reflected from the hole is highly repeatable. However, the material flaws in the Seeded Defect Disc are more realistic due to manner in which they were introduced into the disc. The absence of a drill hole also means they can be inspected from all sides, as with a real inspection. The flaw realism and inspection accessibility make the Seeded Defect Disc a very good test for a new inspection procedure.

Figure 2.4 shows a cross-sectional diagram of the Seeded Defect Disc. The corners of the cross section are labeled with letters A through U. Pairs of letters identify surfaces on the disc, according to the cross-sectional corners that bound the surface. The surface identifiers list the corners in the order indicating the probe motion during the scan. For example, the BA surface is the surface near the bore between the B and A corners. When this surface is scanned, the probe moves from corner B to corner A as the disc rotates underneath. The green dots in Figure 2.4 show the locations of the five known indications. The purple line shows the location of a 0.25 inch diameter hole that was drilled through the disc in order to provide a reference datum for relative locations. Figure 2.5 shows the locations for the indications and the datum hole as viewed from below the disc (i.e., viewing the TU surface).

The Seeded Defect Disc has gone through years of use as a NDT test piece. As a result, it is well characterized and known to contain five indications. The Seeded Defect Disc was re-inspected for this thesis, with data collected according to the current inspection procedures for discs of this type. All of the results in this work come from these data.

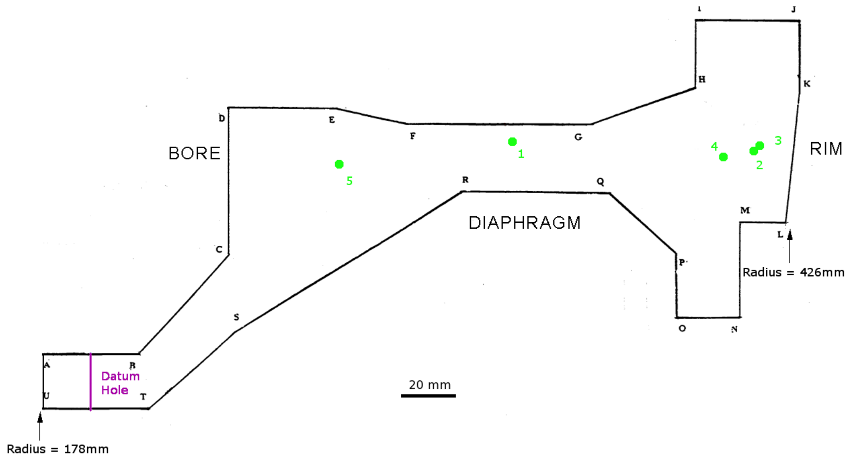


Figure 2.4: Cross-sectional diagram of the Seeded Defect Disc. Corners are labeled with letters A through U. The disc is axisymmetric, and the center line (not shown) would be parallel to the UA surface and to the left of the figure. Green dots mark the locations of the five known defects, numbered 1 through 5. A purple line shows the location of a datum hole that has been drilled through the disc to provide a location reference.

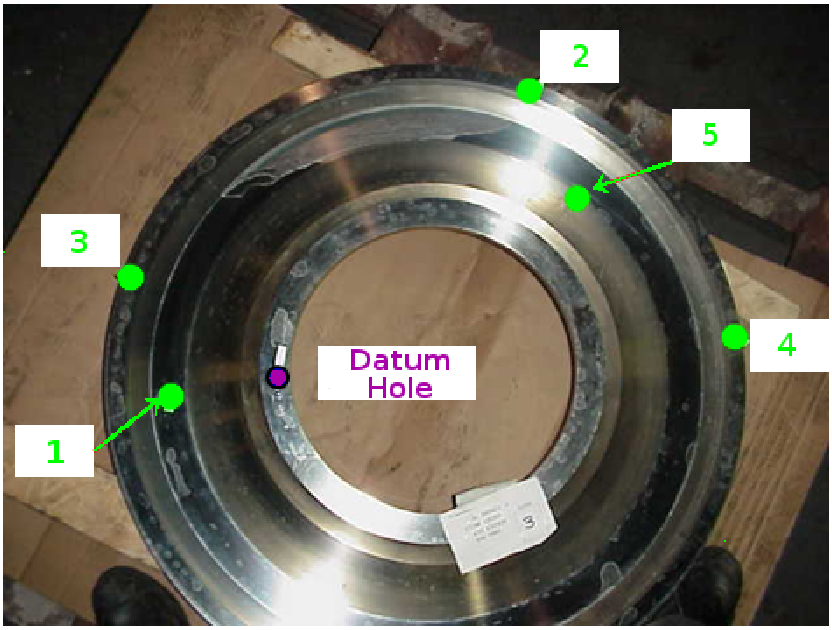


Figure 2.5: Circumferential locations of the Seeded Defect Disc indications. The cross-sectional indication locations are shown in Figure 2.4. A purple dot marks the location of a datum hole that was drilled through the disc, orthogonal to surfaces TU and BA.

2.2 Rotor bore application

Brierley's doctoral work focuses on power generator rotor bore scans [19–24]. Some power generation turbines have a rotor shaft with a bore drilled along the axis of symmetry. Power companies inspect these rotors during a periodic maintenance shutdown by pulling an array of eddy current and ultrasound transducers through the hollow bore. Figure 2.6 shows an example of the test setup. The transducer array made for this type of inspection is shown in Figure 2.7.

Figure 2.6: Rotor bore inspection. Some power generation rotors have a hollow bore. NDT technicians inspect the rotor by sliding and rotating an array of ultrasound and eddy current transducers through the bore. This photograph shows the frame of the probe manipulator mounted on the end face of the turbine rotor during such an inspection. (Image from [19].)

The rotor bore inspection is similar to the turbine disc inspection in a few important ways. First, data acquisition is automated, but evaluation is largely manual. Most of the inspectors' time is spent viewing data that show no sign of a defect. This creates a bottleneck that will only worsen as data acquisition technology improves. Second, there are multiple data sets that overlap in coverage. Current inspection capabilities do not provide any way to combine overlapping data to produce a combined result. Each data set is evaluated separately.

There are also differences between the two inspections that must be considered. The rotor bore inspection is carried out as a single scan, in that the transducer array makes a single pass over the part surface. That single scan, however, produces multiple channels of data due to the multiple transducers in the array. This is in contrast to the turbine disc, which uses a single transducer at all times but makes multiple passes over the part. The multiple sensors also make the rotor bore data multimodal: The array contains both eddy current sensors to detect surface-breaking cracks and ultrasound probes to find internal flaws. Even the multiple ultrasound probes are somewhat multimodal, in that they are made to have different focal depths and incidence angles. The monomodal data from the turbine disc application addressed by this thesis does not require a multimodal detection algorithm. However, the common detection algorithm used for both of these applications is capable of processing input data from different types of sensors and combining them sensibly into a single result.

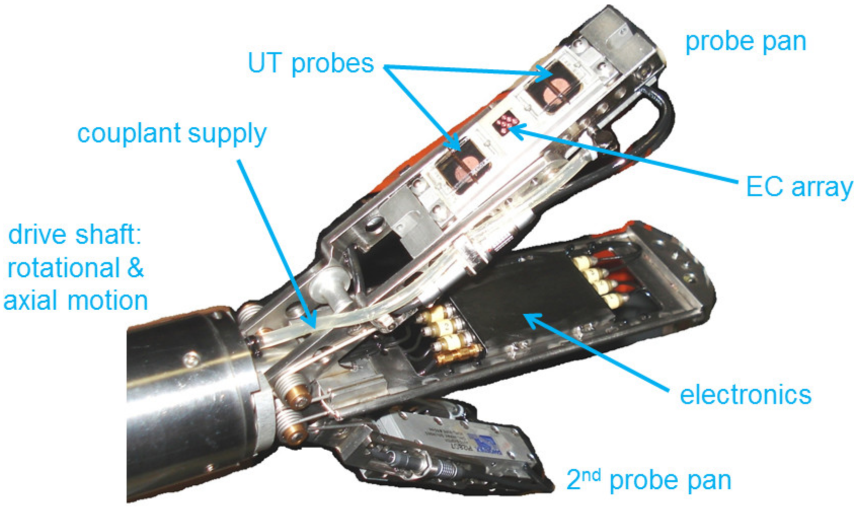


Figure 2.7: Rotor bore transducer array. Springs hold the probe pans in contact with the inner bore surface. Pulse-echo ultrasound and eddy current transducers record data as the array moves through the stationary rotor bore. The array rotates as it is pulled axially, making the transducers follow helical paths along the bore surface. (Image from [19].)

The two applications are also different in inspection history. Each turbine disc is only inspected once, during manufacturing. Each rotor has a full inspection history, with data collected at regular maintenance intervals. This opens the possibility with the rotor bore to use comparison to past inspection data, such as data subtraction processing [25]. Data subtraction could potentially detect changes over time or in comparison to a standard reference component.

2.3 Automation approach

The research objective for this thesis is an analysis software code base, with results demonstrated on real data of current interest. The target application illustrates the approach and performance of the analysis method. However, the software was intended to have the most general framework possible, and it was designed from the beginning to follow a high standard of modularity. The application-specific parts of the code are clearly identified and isolated as far as possible. The software could extend easily to many other applications by the addition of only the modules necessary for unique adaptation.

A key aspect of this work entails merging data from multiple channels or overlapping scans. Many NDT applications produce multiple data sets for each inspected part. These multiple sets might come from sensor arrays, more than one pass of a single sensor over the same scan surface, or different types of sensors producing heterogeneous data from one scan. In any of these cases, a detection algorithm could boost performance by combining data from overlapping regions of the inspected domain. If an indication is evident in two data sets, for example, an intelligent algorithm will see that each data set confirms the indication in the other. The resulting indication is reported with more confidence than if the data sets were analyzed separately.

Fusing data from multiple sets requires proper matching between the locations of their respective samples. Uncertainty in part geometry, NDT hardware position, or physical properties such as wave speed will cause deviations between real and apparent locations. In the above example, if the data sets were poorly registered with each other, what is really a single feature would appear to be two separate regions. Each one would have a possible indication in one data set but only background noise in the other. The conflicting information might actually lead to a fused-data detection with lower reliability than the single data set analysis. Accurate data registration is clearly an essential component of this work.

Much of the previous work that is most similar to this type of data fusion is in the field of image processing [26–29]. For example, Zitová and Flusser wrote a very useful survey of registration methods for visual images [30]. There are significant differences that must be addressed before attempting to apply a traditional image registration approach:

1. NDT data fusion should use knowledge of the physics behind the inspection process; it should not just be image-based.
2. Image registration usually assumes 2D data, as in the projection of a visual scene. Inspection data often come from scans over a surface, with a time history at each point, making a 3D data set. The higher dimensionality makes manipulation somewhat more complex and significantly more time-consuming.
3. Much of the previous work is focused on registration of only two images at a time [25, 31, 32], whereas NDT might use multiple sensors and/or multiple overlapping scans.

Nevertheless, Zitová and Flusser's paper outlines a framework that is a useful starting point in this discussion. They break the problem down into four basic steps: feature detection, feature matching, mapping function design, and image transformation and resampling. Analogies to these will comprise the first four steps.

First it is necessary to identify the type of feature extraction most suitable for NDT ultrasound data. Note that some previous work in data fusion refers to registration feature extraction as "detection." This thesis avoids the word "detection" in this context, since detection will refer more specifically to finding flaws in a later step.

The next step is to define mapping functions for 3D NDT data. The discussion takes the viewpoint that uncertainty in the geometry and other physical parameters represent nuisance variables. The purpose of registration is to estimate the nuisance parameters, thereby to eliminate their effect.

In this approach, feature matching corresponds to the use of registration metrics. Here registration metrics operate in a different way from image feature matching. They are analogous in that each serves to quantify the quality of fit between data sets.

Image transformation and resampling are covered in the form of a more generalized discussion on interpolation. Scalability receives particular attention, due to the potential computational cost of interpolation in large data sets.

Optimization is implicit in these steps for registration, since the registration metric only solves the forward problem of confirming a correct mapping function. After multiple data sets have been matched up by registration, a fused-data detection algorithm can use these multiple sources of information to find likely flaws.

2.4 Computational approach

Brierley and the author have dedicated considerable effort to designing a software architecture that would enable working together smoothly. It was intended from the beginning that software for the applications be as inter-operable as possible. This both avoids duplication of effort and preserves generality for future expansion.

The design allowed both researchers the freedom to modify any part of the code at any time, enabling the other to leverage the new work but without impediments from critical-path dependencies. These independent modifications allowed for the unique, individual contributions required for academic qualification. They were also necessary for bugfixes and to alleviate limitations in older versions of the code, which inevitably arise with different priorities for different users.

The solution to these requirements is a highly modular data model with consistent, well-defined interfaces throughout. Processing steps are split into black-box operators, each having specific inputs and optional outputs. The inputs and outputs of these operators are connected at a higher level as nodes in a directed graph. When output is requested from an operator, it sends requests upstream for its inputs and then performs its computation. The result is then returned to the downstream operator that made the original request.

Following is a description of the major sections of the code base shared by the present author and Brierley. These software modules are listed in order of dependence and typical data flow. The primary author is identified for each software module. Aside from the modules listed specifically below, the code base uses a considerable infrastructure which the two authors developed together.

Data file access. Any application requires software to interface with the data files produced by the inspection hardware. Code is required to read and interpret both the very large data files and the smaller files containing metadata, such as inspection parameters. This software is likely to be very specific to the application, or at least to the hardware or firmware manufacturers. The modules for each application were developed separately by each author.

Geometry. The geometry module manipulates geometric information during registration. It also uses the geometric information to perform interpolation during detection. This part of the software is also quite specific to each application. There are some aspects that are more general and might be reused, such as structured interpolation. For the most part, the geometry modules were developed separately and would need a new implementation for each new application.

Signal conditioning. The signal conditioning module is a collection of utilities that prepare the raw amplitude data for downstream processing. For example, there are routines that detect the envelope of the oscillatory signal, downsample overly dense data, and detect amplitude peaks. This part of the code could be more generally applicable and reusable for applications with similar types of data. Between the turbine disc and rotor bore applications, however, there was less in common than there might have been due to the type of data recorded. All of the turbine disc data were saved as raw ultrasound amplitudes. Much of Brierley's data consisted of ultrasound data, but they were recorded as envelope-detected rather than oscillatory signals. It was also the case that registration features were distinct in nature between the two applications. For these reasons, most of the signal conditioning was developed by each author separately.

Registration feature extraction. The registration feature extraction software operates on the conditioned signals to create data objects representing physical features. The physical features are targeted for the information they can provide to obtain the best possible alignment between multiple data sets. Like the signal conditioning module, feature extraction could in principle be shared between applications that have similar physical features. It happens that the two applications were too dissimilar in this regard. The turbine disc data has many surface echoes, while the rotor bore is a monolithic structure without surfaces that are detectible in the data. This part of the software was developed separately. However, a future application that had similar features to either of these two applications could benefit from reuse.

Registration metric calculation. Once features are extracted for registration, the software that quantifies the quality of fit is more easily generalized than the previous modules. The registration metric calculation is almost entirely shared by the two applications. The two authors both contributed to the software development for this module.

Optimization. The optimization solver that solves the inverse problem is also very general, working as a black box to the rest of the software. Practically any application of automated NDT following the overall approach outlined in this thesis could use the optimization module without modification. This module was developed by the present author. The optimizer includes the Pareto integral estimator, which is one of the claims for academic merit in this thesis.

Detection. The detection module is also very general. Other automated NDT applications should be able to use this software with minimal modification. The work for this thesis contributed a numerical technique to accelerate detector computation, which is the second main claim in this thesis. The present author also made minor contributions in the form of inferior detectors for comparison. Other than these exceptions, the detection software was developed entirely by Brierley.

This software structure is both flexible and robust. Each operator in the directed graph has a limited, clearly defined scope. The code that implements each operator makes only minimal assumptions about other operators, so that any of them can be modified or replaced without breaking another. At the highest level, functionality comes from connecting simpler building blocks so it is easy to write and maintain. Figure 2.8 showes a schematic diagram of the software modules and their primary run-time dependencies.

The software for this project is written primarily in the Python programming language [33]. Python is an Object Oriented Programming (OOP) language, and the operators use the OOP pattern of inheritance from a common base class. This facilitates consistent interfaces between different parts of the code, written by different people.

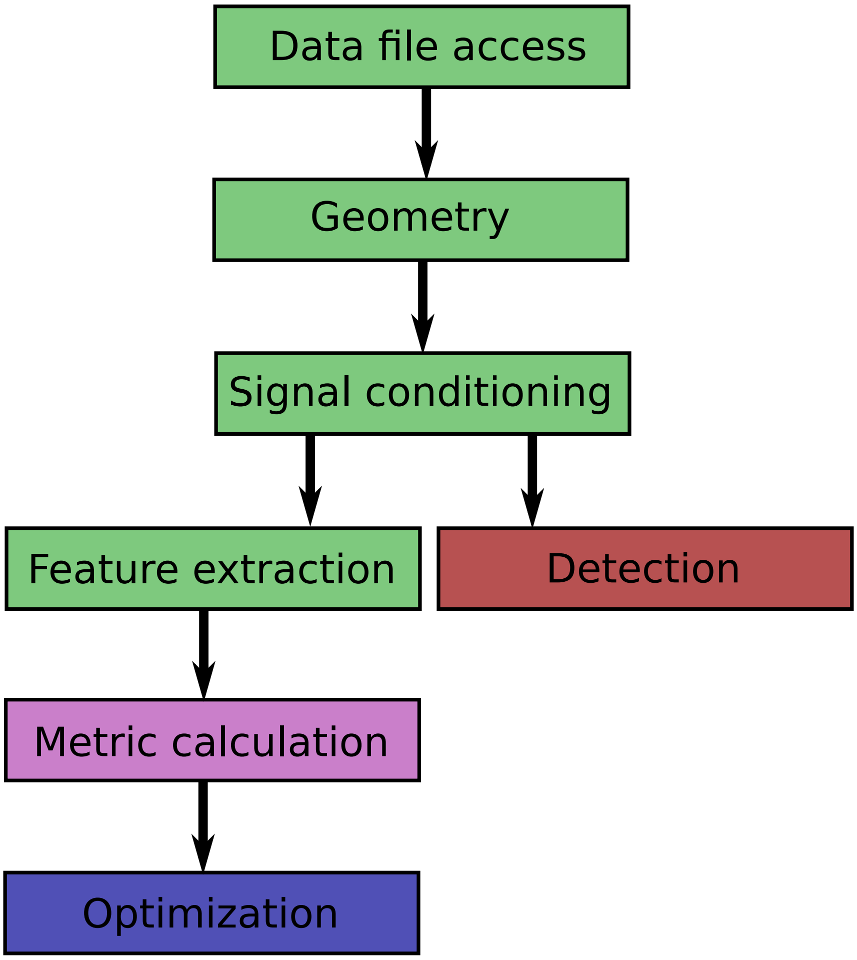


Figure 2.8: Computational approach and breakdown of responsibilities. The software implementing the automated NDT is split into modules, each illustrated by a box in the diagram. Arrows show the predominant direction of data flow during program execution. Green boxes indicate modules implemented largely independently by Brierley and the present author for their respective applications. The red Detection module was written by Brierley. The present author wrote the blue Optimization module. Metric calculation, in purple, was jointly developed by both authors.

The ability to link with code written in other languages is another key advantage of Python. This gives programmers a number of options to boost speed and memory performance in the most critical core of the program. This has allowed the code to remove performance limitations while continuing to take advantage of rapid application development in Python. The optimization algorithm used by registration is written in C++ [34]. The interface between the C++ and Python code is generated automatically with the SWIG interface generator [35]. Python's interfacing capability is also the basis for numpy, the third-party numerical library for python [36], which the NDT software written for this thesis uses extensively.

2.5 Proposed mode of operation

Under the current turbine disc inspection procedure, there is no prioritization for the data. The inspector much watch the amplitude data in real time as the disc is scanned. The vast majority of the data is only noise, but the same level of diligent attention must be applied throughout the inspection. An indication could appear at any time, and the inspector would receive no forewarning.

In the envisaged end-user operation for automated NDT, most of the data will be screened out. An indication detector uses all available data to identify and prioritize regions requiring human inspection. The software displays a zoomed view of the data, one indication at a time, for the inspector to evaluate. With each iteration of the detector, it also provides a likelihood value with the possible indication. This number quantifies the detector's estimate of the likelihood that the indication represents a real flaw.

The likelihood will decrease monotonically with each indication that the inspector checks. He or she can terminate the inspection once it is satisfactorily low, in the knowledge that all regions of the disc not yet inspected have likelihood values that are even lower. In this way the likelihood serves as not only an evaluation for the indication under investigation, but also a stopping criterion. This stopping criterion is important because the detector can otherwise continue to iterate through more indications until it has displayed the entire data set, piecemeal, to the inspector. The likelihood is intended to be informative, but it is not recommended that the stopping decision be taken away from the inspector by setting a hard lower limit.

2.6 Summary

This chapter outlines an approach to automated NDT that is followed in the Chapters to come. There is a need to speed up the inspection, but the manufacturer cannot tolerate a decrease in reliability. The safety-critical nature of the application and regulatory restrictions are major obstacles to a fully automated system, which would have no human involvement in the evaluation.

Section 2.1 gives details about the aircraft engine turbine disc application that is the focus of this thesis. The part is symmetric about the center-line axis, and the scan proceeds in the direction of that symmetry. This type of symmetry is a common feature in many NDT applications, and it can be a challenge for registration. However, it is also advantageous for detection, as later chapters will show. The general applicability of the overall approach is emphasized by Section 2.2, which gives a brief description of a second application. This rotor bore inspection application is the target of doctoral research by the present author's colleague.

Section 2.3 explains how the proposed system supports and facilitates the inspector's work. It does so without the increased risk that would come from taking away responsibility and ultimate decision-making from the human inspector. The automated inspection brings the data from the most likely indications first to the inspector's attention. It also reports a quantitative likelihood for each possible indication. These likelihoods also help the inspector to evaluate the data and decide how to report the indication. They also assist the inspector in deciding when to stop, when the likelihood drops to a level that he or she considers negligible. Section 2.4 describes how the various steps of the approach are broken down into software modules. It also clarifies the contributions made to the software by the two collaborators, and explains the areas where new applications would require additions to the code. Finally, Section 2.5 summarizes the approach as a proposed operational mode, from the NDT inspector's perspective.

Chapter 3

Registration approach

Under the current inspection procedure, each surface is scanned from three different angles as described in Section 2.1. This produces a high degree of overlapping coverage, such that every point inside the disc is interrogated from multiple directions. Fusing data from multiple surface scans by their physical location offers the possibility of significant detector performance improvement. Properly combined inspection data could detect smaller defects with smaller amplitude signals, while also rejecting noise with higher confidence.

Registration, or relative alignment between the multiple data sets, is necessary before those data sets can be usefully combined. Precision is important, since misalignment could make detection worse. Consider a case in which a defect signal is present in two scans, but the scans are misaligned. The detector would match a defect signal from one scan to a noise-only region in the other, and vice versa. The noise signal in the second data set would decrease the likelihood of detection for each of what appear to be two separate defects. It is therefore critically important to achieve sufficient accuracy in the registration stage of automated inspection.

The computer-controlled scanning hardware records signals from position encoders for locations of the scan arm and rotating platform. The encoder data give important information because they can set a starting point and bounds in which to search for an improved registration. However, it is not sufficient for two reasons. First, there are parameters that affect relative positions that are not measured by encoders.

The wave propagation speed in water and in the metal are obvious examples. It is also possible for the scanner to depart from perfect calibration, increasing with usage up to the end of its maintenance period. Any calibration deficiency is also not measured, by definition.

Second, the encoders can only measure positions of the scan hardware, not the inspected part itself. See Section 2.1 for details on the scanning geometry. The relative position between the part and the scanning platform is not known with perfect precision. This is particularly relevant to aligning scans from the upper and lower surfaces of the disc to each other. Once the scanner completes one side, the inspector must manually remove the disc from the tank. He or she then flips the disc over before repositioning it on the scanner platform. Part-to-part dimensional variability could also contribute to the uncertainty for some applications.

The first step in registration is to reduce the computational burden of registration by extracting features from the data. Section 3.1 describes the feature extraction techniques used for the turbine disc data. Once features are obtained from scans that view the same physical feature, it is necessary to adjust the position of one feature to match the location of the other. Mapping functions move the position associated with the amplitude data from the initial, assumed locations. When registration is complete, these mapping functions compensate for and cancel the initial position errors. Section 3.2 contains a description of the mapping functions and the parameters that control them.

Objective functions are metrics that quantify the lack of fit between registration features. Section 3.3 defines the objective functions applied to the turbine disc features and explains their most important properties. With objective functions defined as functions of mapping parameters, registration becomes an optimization problem. Optimization searches for parameter values that give an optimal fit, as measured by the objective functions. Section 3.4 introduces important concepts of multi-objective optimization. In particular, it explains the special properties of the Hypervolume Indicator optimization algorithm that make it the ideal choice for registration.

This chapter concludes with a new technique for evaluating integrals of any function over the set of optimal points, in Section 3.5. The registration uncertainty quantification results in Chapter 4 use this technique. It has the potential for much broader application.

The power of this novel method lies in its ability to compute the expected value for the full, theoretical Pareto set, based on statistical properties of the points sampled by the multi-objective optimization algorithm. This is significant because many multi-objective optimization algorithms have little control over the distribution of points they return, and nearly all lack any quantitative predictability. This leaves other optimization algorithms with little or no way to know how well their results approximate the theoretical Pareto front.

This technique makes it possible to compute a properly weighted mean as a single best estimate of the multi-objective optimization results. It also computes higher moments of the family of optimal points, for quantitative estimates for the registration uncertainty. These are just two examples of this new method to estimate the integral of any function on the Pareto front.

3.1 Feature extraction

The first step in data registration is extraction of features from the data to register against each other [37, 38]. Most NDT ultrasound data sets contain relatively small regions of high intensity against a background of low amplitude noise. The high intensity peaks are waves that are reflected or diffracted from geometric features. The peaks possibly include somewhat lower intensity signals from indications that require investigation and interpretation by a human inspector. The noise can come from random sources, such electronic noise, which could be reduced in principle by averaging. In some cases the noise is coherent, such as scattering from large grains, which cannot be reduced by averaging multiple acquisitions of the same data.

Successful registration will use the information from the large amplitude features to match two or more data sets to each other, so that they may be sensibly fused [39–41]. Data fusion will enable more accurate detection, simultaneously improving the inspector's confidence that a flaw will be automatically detected and that regions checked but not flagged are safely ignorable.

This framework focuses on isolating relatively small, individual features to use in registration. Multiresolution methods take an entirely different approach to feature extraction and are successful in other fields. For example, Xu and Chen registered medical images with a two-level high frequency/low frequency wavelet representation of the data [42]. Multiresolution methods work well for data in which the information content is spread more uniformly across the domain. They are also appropriate when there is no good initial guess for the mapping parameters, and the low frequency information can robustly guide optimization from a distance. However, they do not exploit the a priori knowledge of where certain scattering features of the geometry will be in the data.

The data sets collected during automatic NDT scans are often too large to register directly. Typical amplitude files total approximately 15 GB per disc. These sizes already approach or exceed commodity computing RAM capacity, leaving little to no room for calculations. Any registration computation on the full data set would take far too long, and it would spend the vast majority of that time processing noise. Therefore, the first step in registration of large data sets is a reduction operation. The goal of this reduction is to extract features that represent the necessary information, but in a much more compact form. The physical characteristics of each NDT application must guide the selection of appropriate features.

Feature extraction is necessarily application-specific, but there are important factors common to all applications. For example, features do not come from auxiliary information, such as the position encoder data. That information is accounted for prior to registration. It is precisely the relative distortions between the recorded data and the nominal or assumed positions from external sources of information that registration needs to find.

Registration features must be extracted from the NDT data that are to be positioned. They should translate and morph with the same mapping functions and parameters that will map the data for later detection. Finally, they must be sufficient in number and contain independent information sufficient to constrain the mapping parameters.

Three types of registration features are defined for the turbine disc application: datum hole features, front surface echoes, and back surface echoes. Section 3.1.1 describes the datum hole feature, which guides registration for the relative axial rotation between the upper and lower scan groups. Registration for all other mapping parameters uses the front and back surface echo features. Their extraction is described in Sections 3.1.2 and 3.1.3, respectively.

To clarify, this step identifies and quantifies characteristics of features for registration only. The selection and analysis of possible flaws is the distinct processing step of detection, discussed in Chapters 5 and 6.

3.1.1 Datum hole feature

Prior to registration, the greatest uncertainty in the disc position is the relative rotational displacement between scans of the upper and lower sides of the disc. This is because the inspector must manually flip the disc in order to access both sides.

Current procedures do not entail any registration, and so do not direct inspectors to note the angular position of the disc relative to the platform. In the absence of data fusion, any possible offset is irrelevant. Furthermore, a manual alignment would currently be very difficult. As explained below, the axisymmetric discs do not have a visible feature the inspector could use. The discs are also quite heavy and the inspector must reach over the edge of the water tank to move them, making precise manual positioning difficult even if it were required. For these reasons, the pre-registration uncertainty in the relative angle between upper and lower scans is the full 360° of the circle.

It is therefore essential to have information on the relative axial rotation between the upper and lower scans. However, both the disc itself and the scanning motion are axisymmetric. Angular registration requires some feature of the part that is not axisymmetric and is detectable in the ultrasound data.

The angular registration feature used in this work is specific to the Seeded Defect test disc. A small datum hole had been drilled through the disc, parallel to the axis of symmetry, from the BA surface through to the TU surface, as indicated in

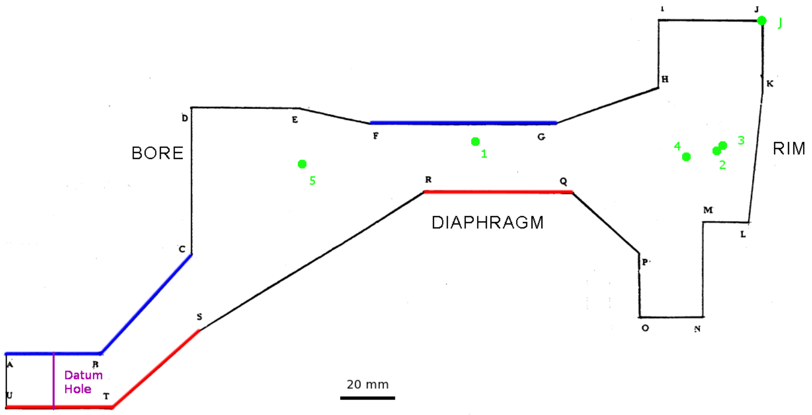


Figure 3.1: Cross section diagram of the registration approach for Seeded Defect Disc 5. The BA, CB, and GF surfaces marked in blue are the upper faces from which front and back surface echo features are extracted. The TU, ST, and QR surfaces marked in red are the lower registration faces. Green dots indicate locations to be used in Chapter 4 to evaluate registration accuracy. Corner J is located at the extreme corner of the cross section, maximal in both radius and axial height. The Datum Hole location is near the opposite corner of the cross section from J. The cross-sectional locations of Indications 1 through 5 are also labeled for reference.

Figure 3.1. The reverberations from the hole reflect more ultrasound energy back to the transducer, making a clear signal in scans of both the BA and TU surfaces.

Turbine discs in production would not have a datum hole like this, nor do they have any other kind of non-axisymmetric feature. A production-ready procedure would need some modification to the disc for the specific purpose of registration. The inspector might adhere a temporary marker to the outside. Alternatively, a small hole or divot could be drilled or stamped during manufacturing in a location that will be machined off after inspection. The key property of an angular registration feature is that the same datum must be present in the ultrasound data from both upper and lower sides of the disc.

Angular registration is a separate, preliminary optimization step before registration for the other features. Mapping in the angular direction is a simple rotation shift, applied to the lower surface scans (K through U; see Figure 3.1). The rotation mapping is a function only of the rotation shift parameter, and none of the other registration metrics is a function[1] of the rotation shift parameter. This makes the datum hole registration a decoupled problem from the rest of the registration. Not only does this allow for a simplification, it is necessary to solve the two decoupled optimizations separately. If the datum hole objective function were included in the multi-objective optimization that solves for the other parameters, the dissociated sets of objective functions would randomize each other and registration would fail.

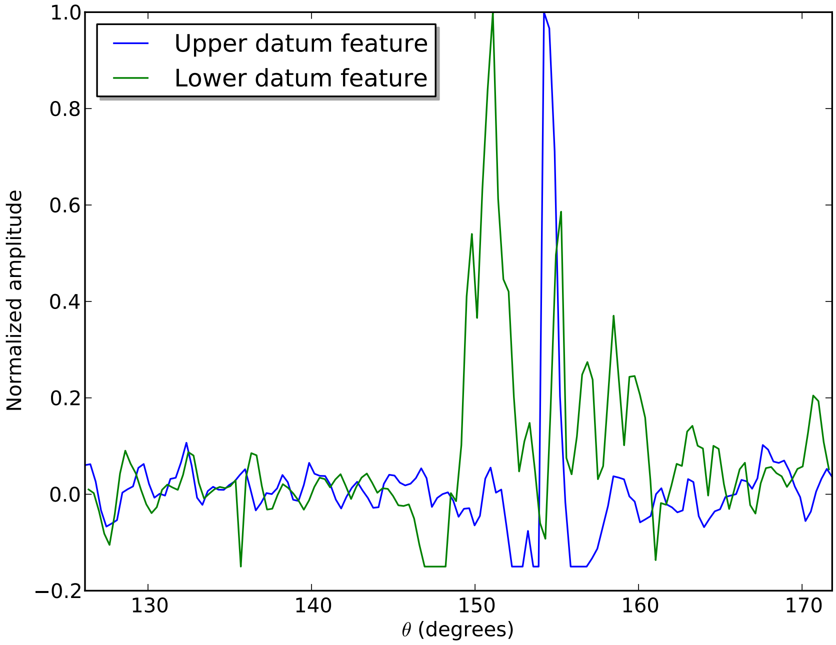


Figure 3.2: Datum hole feature from the BA 0° (upper, in blue) and TU 0° (lower, in green) scans. The higher-amplitude peak in each feature comes from the reflections off of the datum hole drilled through the disc, from the BA to the TU surface. Circumferential registration seeks a shift in θ that aligns the two feature peaks.

A datum hole feature is extracted from each of the BA 0° and TU 0° scans. First a ring of A-scans over a range of radial positions that contain the datum hole feature is extracted from the full data file. The feature extractor then takes the absolute value of the raw amplitude to obtain a rectified signal. The rectified amplitudes are summed over the time (axial) and radial directions. The summation collapses the three-dimensional data to a one-dimensional feature that is a function of angular position, as shown in Figure 3.2. The extractor scales the feature amplitudes such that the maximum values are equal to unity. The features tend to have sharp peaks at the location of the datum hole, and are noisy everywhere else.

---

[1] The other registration metrics do depend on the rotation shift parameter if the tilt angle is nonzero. In practice, the tilt angles observed are very small and this dependence is negligible. See Section 3.2.1 for more information.

It would be possible to use the Hilbert transform to obtain an envelope of the A-scan amplitude, rather than absolute value. The use of the absolute value is for computational speed, given the quantity of data to be processed.

With datum hole features extracted from both sides, it is a simple matter to find a rotational offset value that aligns the maxima of the two features. This initial rotational shift initializes a registration optimization. It is necessary to run the optimization after the initial alignment, because the datum hole features are not smooth functions and a spurious maximum could occur at a location offset from the true center of the hole. The rotational offset parameter is p_6 in Table 3.1. Section 3.3.1 describes the metric for angular registration.

3.1.2 Front surface echo

The front surface echo feature represents the reflection of the ultrasound pulse from the front surface of the disc. This echo signal is clear and unmistakeable in all of the normal-to-surface scans. In fact, because the amplifier gain is set high enough for detection sensitivity, the surface echo amplitude typically saturates the Analog-to-Digital Converter (ADC) and is clipped at the maximum value.

The inspection software records the full time history of the response (i.e., the A-scan) at each scan position. The surface echo feature extractor begins by removing the phase from the sinusoidal pulse signal with a Hilbert envelope filter. The datum hole feature in Section 3.1.1 rectified the amplitude by taking the absolute value. This simple algorithm is sufficient for the datum hole feature because the rectified amplitude is then summed along the time direction. The surface echo features must remove the phase from the raw amplitude signal, in order to avoid alternating maxima and minima that would impede comparisons during registration. The absolute value does not remove phase, but the Hilbert envelope does for band-limited signals. Figure 3.3 shows a typical example of an A-scan and its Hilbert envelope used for surface feature extraction.

In the current application area of immersion tank scanning, the ultrasound hardware uses the Time Of Flight (TOF) as a trigger during data collection. The TOF

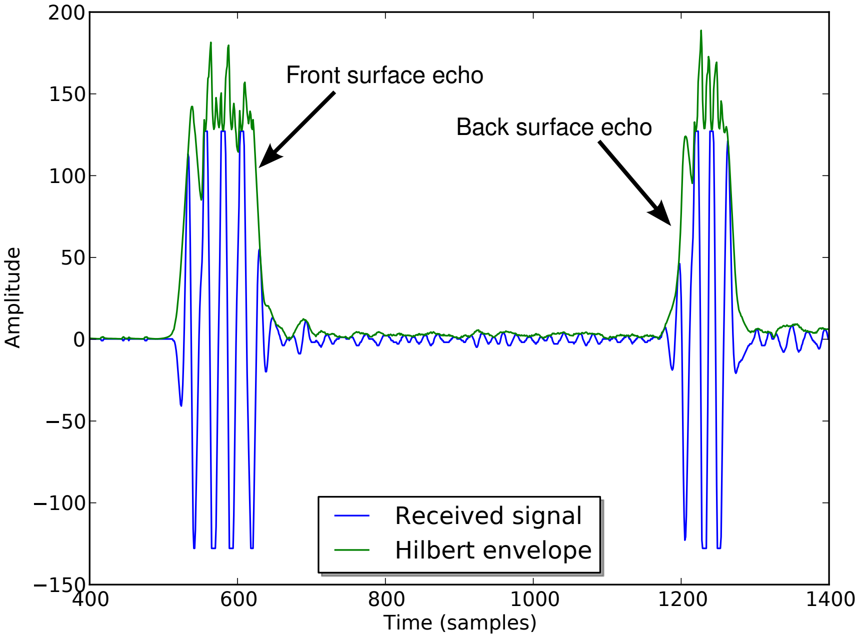


Figure 3.3: A-scan from the BA 0° scan. This A-scan was recorded at the midpoint between the A and B corners. The front and back surface echoes are easily identifiable, at $t = 590$ and 1230, respectively. The peak envelope amplitude and its location are used in the front surface feature. The back surface echo, including 256 time samples surrounding the peak, is extracted to form the back surface feature.

is usually defined as the first crossing of the envelope through a given threshold, within a given time gate. The TOF characterization of the front surface location is useful for data acquisition, in which the only requirement is to decide when to start saving amplitude samples to disk.

Registration, on the other hand, requires a different definition. The width of the front surface echo can vary over the surface, for example due to ringing within surface finish grooves or echoes from corners. Precise registration requires echo peak location rather than the less stable, more uncertain threshold TOF location.

Unfortunately, the amplitude clipping problem described above complicates peak localization. If the dynamic range of the acquired data were high enough, the reflected pulse would approximate a windowed sinusoid. An ADC with more than eight bits in its output would allow a lower gain and higher dynamic range. This would eliminate clipping while preserve the resolution at low amplitudes needed for detection. The Hilbert transform would then recover the window, including its peak, and it would be a simple matter to identify the location of the maximum envelope value. However, because existing hardware causes clipping, the Hilbert envelope contains multiple, spurious peaks in each surface echo, as evidenced by the green plot in Figure 3.3. These peaks are close to each other in magnitude, and their relative magnitudes fluctuate from one A-scan to the next. This makes the maximum hop erratically between the multiple peaks when A-scans are viewed in a sequence. This maximum hopping would cause artificial, irregular peaks and valleys in a surface feature based on the location of the maximum.

It is also infeasible to use a low-pass filter, either before or after the Hilbert transform as a low-pass filter would spread out the surface echo in time. This spreading would blur the peak location, which is precisely the information the peak detector intends to recover. A surface feature constructed with either the simple maximum or low-pass filter approaches would have a detrimental effect on registration.

An alternative procedure devised for the purpose of this thesis makes a more stable, robust peak localization. First, the peak detector records the cumulative maximum; i.e., the maximum value from the Hilbert envelope thus far at each time sample, in the A-scan. Then it finds the time at which the product of the cumulative maximum and the remaining time in the A-scan is maximized. The remaining time factor in this product makes a slight bias to favor the earlier of the multiple peaks. In other words, each point on the cumulative maximum dominates a rectangular area containing all points of lesser amplitudes and at greater times. The peak detector searches for the cumulative maximum with the greatest dominated area. This defines the peak location, and the peak amplitude is the amplitude at that point.

The goal for front surface feature extraction is to obtain the amplitude and location of the peak in the echo signal from each A-scan. The resulting grid of points then approximates a two-dimensional feature in three-dimensional space. This creates a surface that is either 1) a ring bounded by two circles within a plane or 2) a frustum of a cone, depending on the angle between the surface and the axis of symmetry.

Figure 3.4 shows the BA front surface (see Figure 3.1) constructed by this feature extraction method. Even though each point is extracted individually from a separate A-scan, the feature makes a flat and smooth surface. The smoothness is evidence that this feature extraction method does prevent the “maximum hopping” problem mentioned above. The color-mapped amplitude plot in Figure 3.4 shows that there is some variation in the enveloped amplitude over the surface echo. However, the two-dimensional front surface feature only takes the values at the peaks from each A-scan. This makes the distribution of amplitudes relatively close to constant, within three percent of the mean over the entire surface. The BA front surface is a good example of how the front surface echoes in the normal-to-surface scans give a strong signal to create reliable registration features.

The front and back surface echo feature extraction, together with the interpolation and reduction of these features in the objective function, are the slowest parts of the registration computation. The registration software written for this thesis uses two methods to accelerate this computation.

First, the fact that the front surface dimensionality is reduced from 3D to 2D greatly reduces computation time during registration metric calculations. Second, downsampling the points in the surface echo features is a very effective way to speed up registration. The echo peak extraction could use every A-scan on the surface to construct the feature. However, given the smoothness of the surfaces, there is significant redundancy among A-scans that are near each other. A simple downsample selects an evenly spaced subset of A-scans in both of the scanning directions. This results in a much smaller volume of data in the feature but with virtually the same registration information. All surface echo features used in this thesis use a downsampling of 20 to 1.

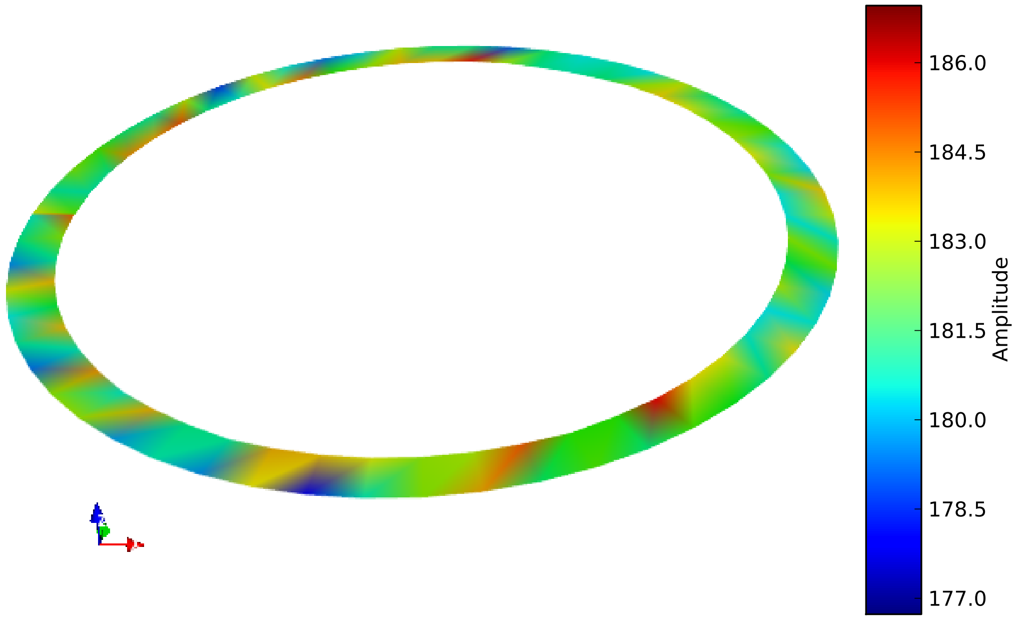


Figure 3.4: Front surface feature from the BA 0° scan, reconstructed from the ultrasound surface echoes. The front surface features are two-dimensional shapes in three-dimensional space. They take the shape of an annulus, as shown above, or the frustum of a cone, depending on the angle of the surface relative to the axis of symmetry. This feature is matched to the back surface feature from the TU 0° scan for registration.

3.1.3 Back surface echo

Extraction for the back surface echo is a similar procedure to that of the front surface in Section 3.1.2. For the back surface, however, the feature extraction takes a time window of data from each A-scan rather than just amplitude and location at the peak. It then constructs a three-dimensional feature, with a finite thickness that comes from the width of the time window. The finite thickness of the back surface is preserved in order to allow for interpolation, as Section 3.3.4 will clarify. Like the front surface echo feature, downsampling in the scan directions reduces the computational burden.

Figure 3.5 shows an isosurface plot of the extracted BA back surface. The constant-amplitude contours, gradated by color, show the variation in amplitude throughout the thin, three-dimensional region. The highest amplitudes are not visible in the

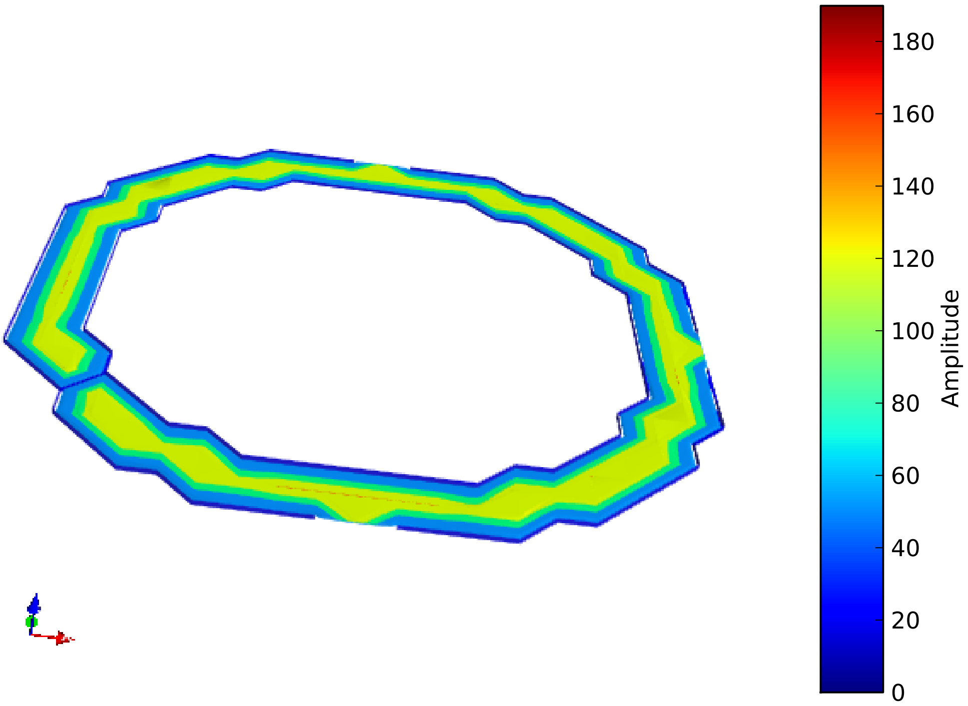


Figure 3.5: Back surface feature from the BA 0° scan, reconstructed from the ultrasound surface echoes. The back surface feature is three-dimensional, having a finite thickness due to the inclusion of multiple time samples from each A-scan. Like the front surface feature in Figure 3.4, it takes the shape of an annulus, as shown above, or the frustum of a cone, depending on the angle of the surface relative to the axis of symmetry. This feature is registered to the front surface feature from the TU 0° scan.

plot, as they occur in the interior of the feature. Amplitudes decrease to zero near the extremities of the feature. The 20 to 1 downsampling is more pronounced in Figure 3.5 than in Figure 3.4. The in-plane downsampling is evidenced by the jagged outline, caused by the interpolation inherent in contour plots.

3.1.4 Surface pair selection

Several considerations are important for selecting scans from which to extract surface echo features. First, detecting the echo signal is not as reliable for the ±5° angled scans due to specular reflection of the ultrasound beam away from the transducer. Only 0°, normal-to-surface scans are considered for surface echo registration.

A front surface feature is only useful when paired with a back surface feature. Therefore, each front surface needs to be on a part of the cross section where there is a good back surface echo. This requires the front and back surfaces at that location to be parallel or nearly parallel, because the echo is greatly reduced or absent when the back surface is at an angle relative to the front. For the same reason, it is also desirable if possible to have small distance between front and back surfaces. An increased front-to-back distance amplifies the effect of any angle between the two surfaces, and beam spread and attenuation both increase with depth.

It is important not only to consider which individual front/back surface pairs to use, but also the requirements for the full combination of surface pairs. There must be at least two pairs of front and back surfaces that are not parallel to each other. This is required for linearly independent information on the mapping parameters. Considering the typical gross shape of turbine discs, there will likely be suitable surfaces that are normal to the axis of symmetry. It is more challenging to find a pair of front/back surfaces that have a radial component in the surface normal vector, and that are both visible from the upper and lower sides of the disc. The surface pairs that meet these criteria for the Seeded Defect Disc are BA and TU, CB and ST, and GF and QR, which are marked in Figure 3.1.

Corners are the only other structural feature that generate a signal in the ultrasound data. Corners were evaluated for suitability in registration, but were rejected. Front surface corners are impossible to locate directly, since the inspection hardware scans each front surface and stops just short of the surface boundaries. It would be necessary to extrapolate the front surfaces to get front-side corner locations.

Echoes from back-side corners are often visible in the A-scans, but they are typically too difficult to locate reliably since surface waves and reflections from the adjacent surfaces often obfuscate the corner. Concave back surface corners can also create two-bounce echoes, even far from the corner itself. Linear interior features like back corners are also obscured by beam spread, more so than the planar surface features.

3.2 Geometry parameters and mapping

Mapping refers to the geometric transformations to one data set that allow it to better register with another. It is performed during registration by varying the mapping parameter values in order to match registration features. Once optimal parameter values are obtained, the same mapping applies after registration to all interior points, during the detection phase of automated inspection.

Mapping is intimately connected to the geometry of each NDT application. As with feature extraction, the design of mapping functions is necessarily application-specific, but there are principles that should apply in general. Many medical sensing applications use a very generic, arbitrary morphing or deformation to register soft tissue ultrasound images [26]. In NDT, the inspected parts are effectively rigid during the inspection. The part might undergo rigid body motion, and there may be part-to-part dimensional variation, but there will likely be no discernible deformation.

A related principle is that there is only a certain amount of independent registration information contained in the data. Too many parameters, allowing for very arbitrary deformations, could make an underdetermined problem. The mapping should therefore reflect the constraints imposed by the physics and geometry of the inspection and the structural rigidity of the part.

Understanding connections to other fields of study can assist the construction of appropriate mapping functions. Mapping, imaging, and deconvolution are really three different ways to conceptualize the same mathematical process. Imaging is deconvolution, in the sense that the signal received at the sensor is the transmitted signal, convolved with the impulse response of the structure. The image produced for visualization is an estimate of the inverse problem solution; i.e., the impulse response to the interrogating wave. Various application areas have explored this connection between imaging and impulse response. Laser vibrometry, structural health monitoring, seismology, remote sensing, and medical imaging are a few examples [43–49].

Imaging is mapping in the sense that it is defined by a transformation from one domain and codomain to another. Ultrasound scan data, for example, are recorded in a 3D domain consisting of the 2D probe location/scan surface axes and a time axis. The scalar codomain is the raw or suitably preprocessed amplitude. These data are then mapped to a spatial 3D domain in x, y, z coordinates with scattering potential as the codomain variable.

The quantities of interest in NDT are the structure's geometry and material properties. They are inferred by their effect on the impulse response. All the uncertain parameters describing the transducer beam are nuisance variables, since they affect the measured data but are of no interest in the end. The mapping function should capture the effect of nuisance parameters, to enable optimization to find the best way to compensate and remove their effect.

This view of mapping and deconvolution ties into recent work by Clarke on matching and comparing guided wave data collected at varying temperatures [50]. Previous work by Weaver and Lobkis [51] found that a linear phase shift was the dominant effect of temperature on guided waves in a plate. Other researchers, such as Lu and Michaels [52] and Croxford et al. [53], used the linear phase shift assumption to adjust baseline waveforms for damage detection. Clarke's Optimal Stretch algorithm extends that work to account indirectly for the effect of the unknown exogenous variable (temperature), though he did not actually care to determine its value. The algorithm compensates for the way temperature affects the impulse response of a structure and, by convolution, the received signal.

In Optimal Stretch, the Fourier transform of two pitch-catch signals, measured at different temperatures, are registered by scaling the frequencies of one of the signals [54]. The frequency-scaled signal multiplies the ratio of an unscaled to a frequency-scaled representation of the input signal, to account properly for the effect of the nuisance parameter (temperature). Clarke did not identify his procedure as deconvolution, but it is the best justification for the procedure. The same type of scaling is required in the frequency domain as in the time domain, so the Fourier transform gives no benefit to simplicity, linearity, or any other property that facilitates the temperature compensation. The main difference from a time-domain approach is that the impulse response is conveniently re-convolved with the unscaled excitation signal, by multiplication in the frequency domain. The registration mapping for the current application uses wave speeds as unknown parameters, performing effectively the same compensation as frequency scaling to the transform between time and space domains.

Long et al. provide another example of data domain transformation to remove nuisance variables. They used an ultrasound array with a conformable bladder of couplant to inspect irregular weld surfaces [55]. The shape of the front surface was not exactly known beforehand, and it was not of interest as a final result. It represented a nuisance variable because it had to be estimated to remove its effect on the data. Long et al. found the front surface in the ultrasound data, then used that estimated front surface for imaging [56]. In analogy to the mapping component of the present work, they estimated geometry using the same data set that subsequently benefited from improved imaging.

Blind Signal Separation

In general, the transducer beam amplitude can be a highly complex function in 3D space and time. Accounting for that beam profile could make deconvolution more accurate. It also probably requires a better beam model than is likely to be available, especially under industrial conditions. Statistical techniques might allow the wealth of data to make up for lack of knowledge about the beam profile.

Blind Signal Separation (BSS) is able to separate additively mixed sources, with no assumptions about the sources except that they are statistically independent. With this method, a physics-based beam model could serve as an initial guess, and the BSS solver could adaptively find one that better fit the data. Cardoso gives a good overview of BSS, alternatively referred to as Blind Source Separation or Independent Component Analysis (ICA) [57].

The method has improved remote sensing images as a preliminary step for feature detection [58]. BSS visual image processing has also improved visual images by removing reflections, using multiple images taken with slightly offset camera angles [32]. Ukai et al. extended the same principle of independence maximization to the problem of convolutive mixtures and adaptive beamforming [59]. BSS is based on statistical independence. This means that a fast and accurate estimator for Mutual Information (MI) would be needed for BSS. An MI estimator developed for registration could therefore also be used for BSS. Section 3.3 below defines MI and describes its possible application to registration.

Factorization imaging methods

Other imaging techniques fall into the category of factorization methods [60]. Many factorization methods use a multiple scattering model rather than the Born approximation for weak scatterers. Green's functions play a prominent role in these methods, again highlighting the connection between imaging and the impulse response. For example, Simonetti and Huang explain the connection between beamforming and diffraction tomography [61]. At least some of these methods offer the potential for super-resolution [62] and multi-target focusing [63]. In the context of this project, detecting the presence of a flaw is more important than resolving its shape. Multiple scattering could be a significant concern for some inspections of near net shape or completed parts to detect flaws near surfaces.

Inverse ray tracing

In the current implementation, the deconvolution uses a ray tracing model. The transducer beam is assumed to be an impulse function in time and space that propagates along a straight line between interfaces. This is the simplest possible model that can account for reflection and refraction, both of which are essential to the turbine disc application. The ray tracing model could extend for other applications to account for mode-conversion and diffraction by branching into multiple rays at the appropriate locations. There is no need to include these effects for the turbine disc application. The front surface incident angles are too small to cause significant mode conversion, and no data are analyzed beyond the back surface echo.

Some previous works on imaging and damage detection self-identify as a form of inverse ray tracing. Many conventional imaging techniques, such as plane swept B-scan, focused swept B-scan, sector B-scan, and the Total Focusing Method (TFM) with sensor arrays [64] or ellipse and hyperbola imaging with sparse sensor guided waves [65] use averaging to combine samples containing the response along rays from a hypothetical scatterer. Samples focused on a non-scattering region, though they might have reflections from other locations, tend to be uncorrelated and average out to zero. All of these imaging techniques require sensor arrays with pitch-catch capability on multiple receiving transducers. For turbine disc NDT, current inspection procedures call for a single transducer in pulse-echo mode. If inspection procedures changed in the future, or for other applications that already allow for these imaging methods, these conventional imaging techniques could benefit from the same registration procedure described in the rest of this chapter. With the present data, these imaging methods are not available.

Meksen et al. used a randomized Hough transform to detect automatically the hyperbolas that characterize crack responses in Time-Of-Flight-Diffraction [66]. Windsor and Capineri also attempt to detect hyperbolas in ground penetrating radar images [67]. The ray tracing model and the scan geometry predict the hyperbolic form in the measured data. In [68], Gough and Hawkins unify a number of synthetic aperture imaging algorithms developed in different fields such as NDT, medical, geological, and remote sensing. They accomplish this unifying framework by expressing the algorithms as mapping operators, which simplify reformatting the data from one sampling grid to another.

A ray tracing model is simple relative to other imaging techniques, but comes with a great advantage to computational efficiency. It allows the assumption that data points transport in a Lagrangian sense, preserving the sampled data amplitude. This is significant because the effect of nuisance parameters is then only a coordinate transformation on the data locations in physical space. A more complex deconvolution would require reprocessing the original amplitude data at each registration iteration. This would have a highly detrimental effect on computation times, especially considering the large size of automated NDT data files. The mapping functions in Section 3.2.1, which were derived for the turbine disc application in this work, are a form of ray tracing.

3.2.1 Mapping functions

Mapping transforms coordinates from the assumed, nominal values of the mapping parameters. The mapping parameters are a means to an end during both regis-

Table 3.1: Registration parameters. p_1 and p_2 are wave speeds. p_3 and p_4 express uncertain values in the inspection geometry. p_5 through p_{10} are the complete six degrees of freedom for rigid body motion.

variable	*description*
p_1	longitudinal wave speed in metal
p_2	longitudinal wave speed in water
p_3	water gap distance
p_4	shift in the axis of symmetry location
p_5	z coordinate of flip rotation axis
p_6	rotation angle about the axis of symmetry
p_7	eccentricity angular direction
p_8	eccentricity magnitude
p_9	tilt axis angular location
p_{10}	tilt angle

tration and detection. Because of this, mapping parameter values or their physical interpretation are not important unless the values are unreasonable. It is important that physically possible and likely distortions be in the function space of the mapping operator. Proper mapping function design considers how the data point locations will be distorted as the parameters deviate from nominal.

The amplitude data are in a structured format, with an amplitude sample at each vertex in a three-dimensional grid. It is convenient to define a coordinate system for each data set, based on its indices t, u, and v. The amplitude data samples, then, occur at each point in the regular grid defined by the the integral values of the t, u, and v coordinates. The coordinate transformation to physical x, y, and z locations is then defined as functions of t, u, and v and the parameters p_1 through p_{10}. Table 3.1 identifies the registration parameters. The mapping function descriptions below further define the physical meanings of each parameter.

The first data index, t, is the sample index within each A-scan. It is equal to the product of the acquisition sample frequency and the physical time since the transducer pulse.

The probe traces a circular path relative to the disc, due to the disc's rotation on the scanning platform. The u index counts the steps along the circumference of that circular path. u therefore advances by one unit for each A-scan. As a coordinate, u is periodic, starting from $u = 0$ at a common angular datum. It repeats after reaching the maximum value of $N_u - 1$, where N_u is the number of A-scans per circumference.

The last index, v, is the scan path number. v increases by one each time the probe moves a step in the axial/radial direction parallel to the front surface. Each v indicates a distinct circular (apparent) probe path.

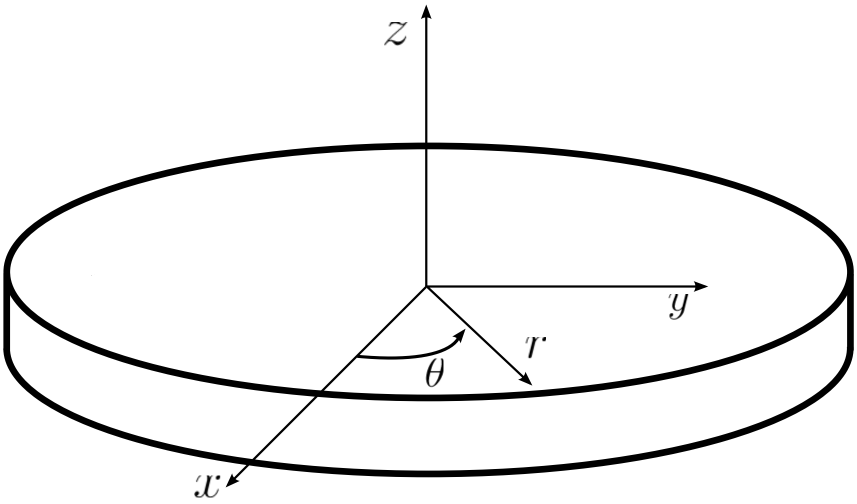


Figure 3.6: Coordinate systems used by the mapping functions. A circular cylinder represents the turbine disc schematically. The x, y, z and r, θ, z coordinate systems share the z axis. The z axis is also the axis of rotation for the immersion tank platform. The θ coordinate is measured relative to the x axis.

The probe moves in the $\theta = 0$ plane as the platform rotates around the z axis, according to the coordinate system shown in Figure 3.6. A three-point radial clamp holds the disc in place on the rotating platform. (See Figure 2.2.) It is reasonable to assume the disc does not move relative to the platform while clamped. However, there will generally be a small degree of misalignment that creates uncertainty in the disc's location and orientation. The misalignment will be different between the upper and lower scans because the disc must be unclamped and flipped over. It is also possible for the platform to be slightly tilted away from perfectly horizontal. This would cause an apparent wobble motion even when the disc and platform are perfectly aligned with each other. All of these effects are accounted for by a transform that applies a rigid body motion (RBM) to the lower scans, to allow them to match the upper scans.

In addition to the disc RBM parameters, there are two parameters representing other uncertainties in the scan geometry. The water gap distance, or distance between the probe and the disc front surface, is targeted to a certain value during the data acquisition, but it is not known with perfect certainty. There is also some uncertainty in the location of the axis of rotation relative to the encoder data. The scanning frame position encoders measure location relative to the edge of the immersion tank, not relative to the axis of rotation. One of the registration parameters is the difference between the assumed and realized axis locations.

Two other registration parameters are longitudinal wave speeds that affect beam propagation. Only approximate values for wave speeds in the water and the metal are known a priori. An error in either would lead to an apparent scaling along the A-scan trajectory, representable as an offset proportional to the A-scan time, t. If the probe is not oriented normal to the surface, the beam would also have an apparent location offset due to refraction.

The registration mapping functions are the system of equations that transform points in the index coordinate system to the physical r, z, θ coordinates. The ten registration parameters appear in these coordinate transform equations. Varying the registration parameters has the effect of translating all of the data point locations in physical space.

In computing the physical x, y, z locations it is most convenient to begin by computing the hypothetical location due to beam propagation, assuming the disc symmetry axis is perfectly aligned with the axis of rotation. Equation 3.3 below gives this intermediate, perfectly aligned location as θ_0, r_0, and z_0. The rigid body motions then translate the point from θ_0, r_0, z_0 to x, y, z. This sequence does mean that the effect that the tilt angle would have on refraction is neglected. Tilt angles encountered in practice are very small and do not have a significant effect on refraction.

The incident angle, φ, is fixed by the scan setup to be either $0°$, $5°$, or $-5°$. This incident angle is assumed to be exact. The unit vector in the direction of the

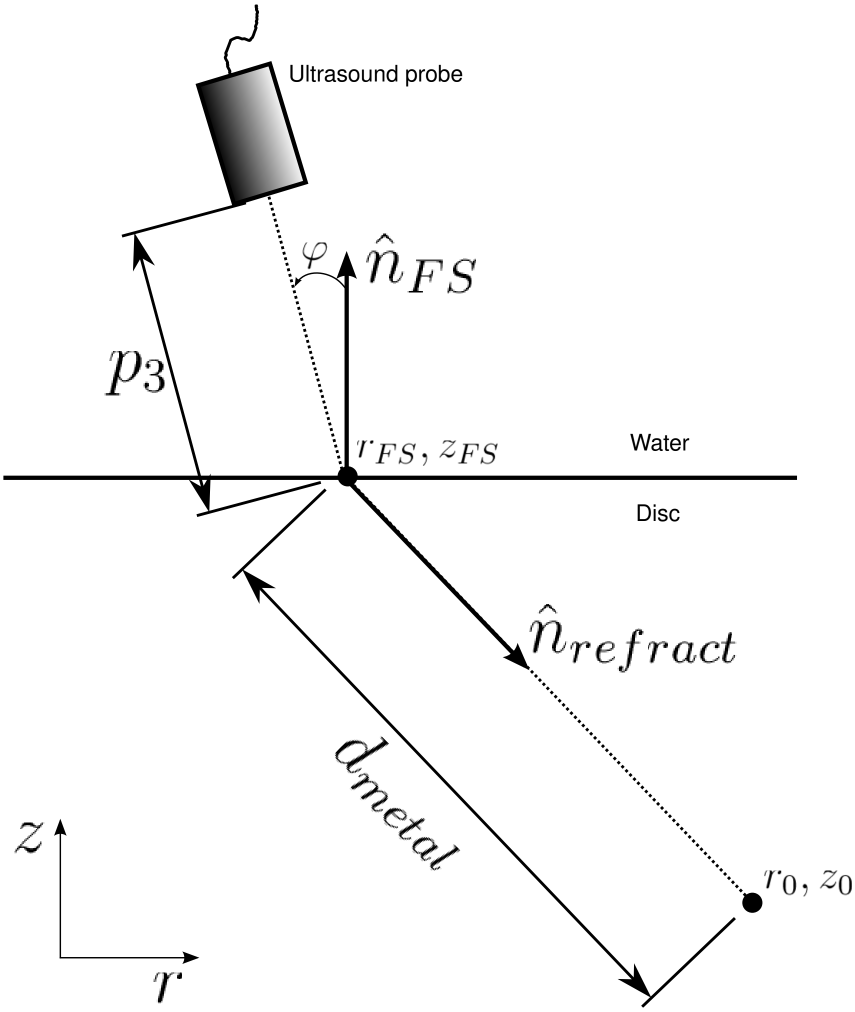


Figure 3.7: Beam propagation effects applied by the mapping functions. The ultrasound beam center line is incident at r_{FS}, z_{FS}. The beam propagates to a given data point, at r_0, z_0, and any reflections return to the probe along the same path. Refraction and propagation effects give a nominal location for the data point. This part of the mapping assumes that θ (not shown) does not change during the beam propagation, so that $\theta_0 = \theta_{FS}$.

refracted beam, $\hat{n}_{refract}$, is

$$\hat{n}_{refract} = -\begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{1-\left(\frac{p_1}{p_2}\right)^2 \sin^2\varphi} & \frac{p_1}{p_2}\sin\varphi \\ 0 & -\frac{p_1}{p_2}\sin\varphi & \sqrt{1-\left(\frac{p_1}{p_2}\right)^2 \sin^2\varphi} \end{bmatrix} \hat{n}_{FS} \tag{3.1}$$

The outward-pointing unit vector normal to the front surface, $\hat{n}_{FS}$, is taken from the nominal surface orientation, neglecting the effect of the tilt angle. Both $\hat{n}_{refract}$ and $\hat{n}_{FS}$ have components of θ, r, and z. The θ components are always zero, due to the symmetry of the disc and the scan motion. Figure 3.7 shows the geometric relationships between these quantities.

The one-way beam propagation distance in the metal, d_{metal}, is the distance from the point of incidence on the front surface to the data point location.

$$d_{metal} = \frac{t\,p_1}{2f_s} - \frac{p_1\,p_3}{p_2 \cos\varphi} \tag{3.2}$$

In Equation 3.2, f_s is the data acquisition sample frequency.

With $\hat{n}_{refract}$ and d_{metal}, it is possible to compute r_0, θ_0, and z_0.This requires the nominal, initially assumed locations for the front surface, $r_{FS}(v)$ and $z_{FS}(v)$. These nominal surface locations are functions only of v, due to the cylindrical symmetry. Their values are computed from information in the metadata file recorded by the scanner controller during data acquisition. Each disc surface has its own values for the three nominal front surface quantities $\hat{n}_{FS}$, $r_{FS}(v)$, and $z_{FS}(v)$. Together with φ, they contain all the information unique to each scan of each surface.

The pre-RBM location of the data point is

$$\begin{Bmatrix} \theta_0 \\ r_0 \\ z_0 \end{Bmatrix} = \begin{Bmatrix} -\frac{2\pi u}{N_u} \\ r_{FS} - p_4 \\ z_{FS} \end{Bmatrix} + d_{metal}\,\hat{n}_{refract}\,. \tag{3.3}$$

The first RBM is a flip rotation about an axis defined by the line $y = 0$, $z = p_5$. The flip rotation angle is fixed at π radians and accounts for the physical inversion of the disc by the inspector between the scanning of the upper and lower scans. The flip is followed by a rotation of p_6 about the axis of symmetry. This flip and rotation RBMs are illustrated in Figure 3.8

An eccentricity RBM is applied which translates the disc in the x, y plane in the angular direction p_7, by a magnitude of p_8. The eccentricity is the offset between the axis of disc symmetry and the axis of rotation. It occurs when the disc is not perfectly centered on the platform. Figure 3.9 shows the eccentricity RBM.

The tilt rotation RBM is illustrated in Figure 3.10. The tilt rotation axis is in the

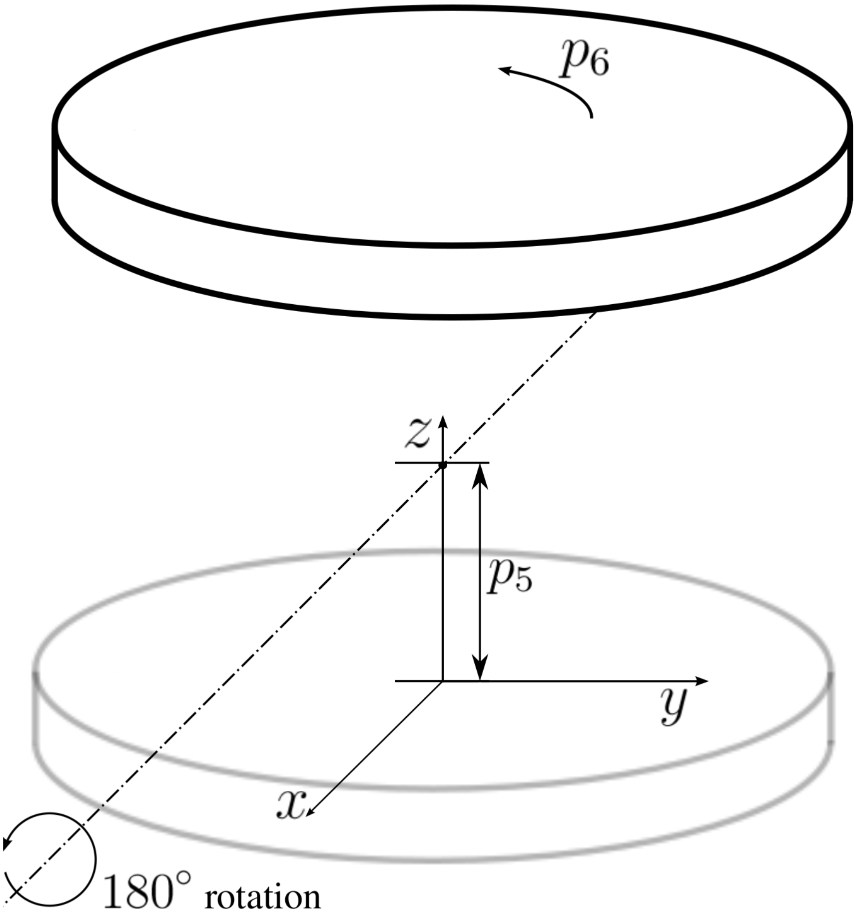


Figure 3.8: Flip and rotation Rigid Body Motions applied by the mapping functions. The data points from Figure 3.7 undergo a virtual flip, rotating 180° around an axis that is parallel to the x axis and at $z = p_5$. The grey disc shows the initial location, and the black-edged disc is the location afterward. After the flip, the disc rotates about the z axis by the angle p_6.

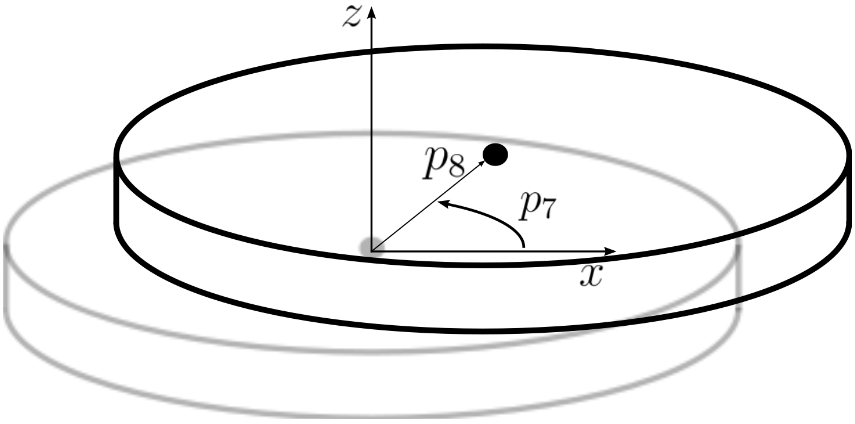


Figure 3.9: Eccentricity Rigid Body Motion. After the RBM in Figure 3.8, the disc translates in the x, y plane. The direction of this translation is p_7, relative to the x axis. The parameter p_8 is the magnitude of this translation.

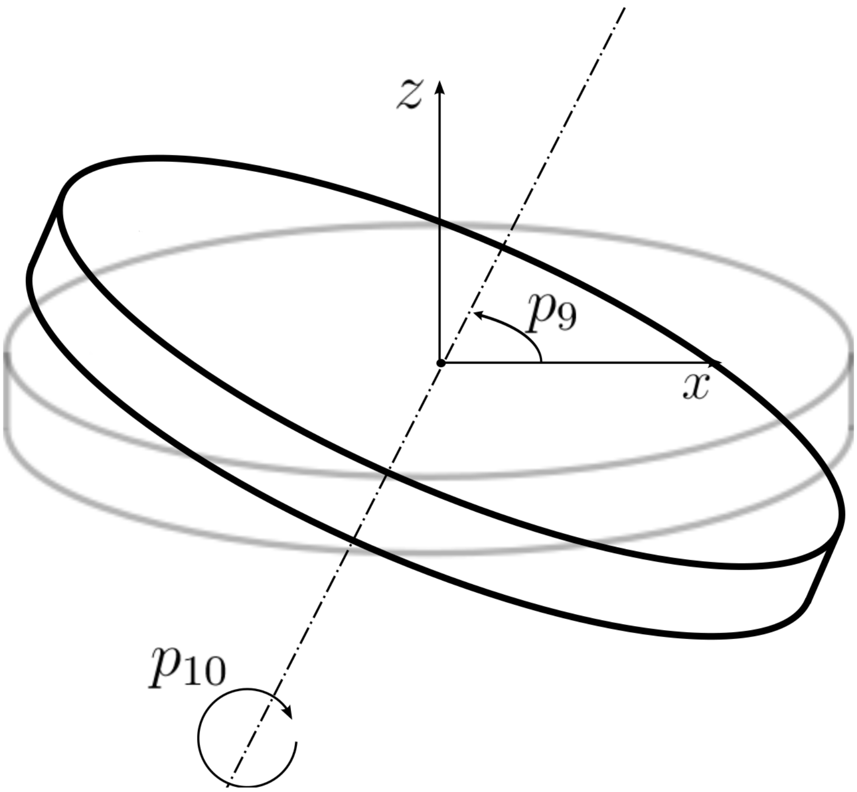


Figure 3.10: Tilt Rigid Body Motion. After the eccentricity RBM in Figure 3.9, the disc rotates in a tilting motion. The parameter p_9 is the position of the axis of rotation relative to the x axis, and p_{10} is the angular magnitude of the tilt.

$z = 0$ plane, but it could be at any angle relative to the x axis. The tilt rotation itself is most easily conceptualized as a sequence of three rotations. First, a rotation of p_9 about the z axis aligns the tilt axis with the x axis. A rotation of p_{10} about the x axis then applies the tilt displacement to the disc. Finally, a rotation of $-p_9$ about the z axis returns the tilt axis to its original location. The rotation matrix R_{tilt} applies the appropriate transformation when premultiplied to a vector with x, y, z components.

$$R_{tilt} = \begin{bmatrix} \cos p_9 & \sin p_9 & 0 \\ -\sin p_9 & \cos p_9 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos p_{10} & -\sin p_{10} \\ 0 & \sin p_{10} & \cos p_{10} \end{bmatrix} \begin{bmatrix} \cos p_9 & -\sin p_9 & 0 \\ \sin p_9 & \cos p_9 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.4}$$

The final location, after applying all beam effects, geometric uncertainty, and rigid body motions, is

$$\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = R_{tilt} \begin{Bmatrix} r_0 \cos\left(\theta_0 + p_6\right) + p_8 \cos p_7 \\ -r_0 \sin\left(\theta_0 + p_6\right) + p_8 \sin p_7 \\ 2p_5 - z \end{Bmatrix} . \tag{3.5}$$

Taken together, Equations 3.1 to 3.5 define a transform from the data indices t, u, and v to the spatial coordinates x, y, z. During registration, Equations 3.1 through 3.5 map the data points defining the front and back surface features as the optimization algorithm selects values for the parameters p_1 through p_{10}. After registration is complete, the same transform maps all points in the data set, using optimal parameter values determined by the registration. The detection algorithm then uses these registered data point locations.

3.3 Objective functions

The registration metric operates on extracted features from the data sets, one of which has been transformed in some way, and quantifies the quality of the match. The optimizers discussed in Section 3.4 search for the minimum of an objective function. A registration metric, then, must evaluate to a minimal value when the transformation is optimal. The metric value should gradually increase for mapping transformations that increasingly deviate from optimal.

Registration metrics usually operate on two data sets at a time, for simplicity in conceptualization and implementation. Such a binary operator is still useful in multi-data set registration, as long as it is invariant with respect to any mapping that is applied to both data sets. Consider an example with three data sets, in which Set 1 is mapped to the domain of Set 2, and 2 is mapped to the domain of 3. The metric computed between 1 and 2 in the domain of 2 should be equal to the same metric applied after both 1 and 2 are mapped to the domain of Set 3. A simple way to meet this requirement is to make no reference to domain except as an index for matching ordinate values. A sum over discrete data samples meets this condition. An integral over the domain would not since the integrand would need a Jacobian determinant when it is mapped.

The Seeded Defect Disc application separates the data sets into two groups, one of which will be registered against the other. This grouping creates a registration problem with only two data domains to register against each other. However, the registration metrics all use a sum of discrete data samples, such that the methods and software could be extended to handle three or more data domains, if necessary in the future.

The construction of registration metrics is application-specific, but it is possible to define registration metrics that are general enough to operate on data from many different NDT applications.

Mutual information

Many researchers have applied Mutual Information (MI) to image registration, particularly in the medical field; see, for example, [69–71] and the extensive survey by Pluim, Maintz, and Viergever [72]. This section describes and evaluates MI in the context of registration, where it has seen extensive use in other applications. For reasons outlined below, however, MI was ultimately found to be unsuitable for the turbine disc application.

Equation 3.6 defines MI assuming the data in f and g are random variables sampled from the joint Probability Density Function (PDF) p, with marginal PDFs p_f and p_g.

$$MI[f,g] = \int\int p(f,g)\ln\left(\frac{p(f,g)}{p_f(f)p_g(g)}\right)df dg \tag{3.6}$$

Roughly speaking, MI between two random variables quantifies how much the randomness of one is reduced by knowledge of the other. It is zero when f and g are independent. It is maximal when f is an invertible function of g, assuming f and g have a constant entropy. Therefore, a minimizing optimizer would need to use the negative of the estimated MI.

Much of MI's usefulness derives from its homeomorphic invariance [73, 74]. That is, if an invertible mapping exists between two pairs of random variables, the mutual information between each pair is equal, regardless of the complexity or nonlinearity of the mapping. In practical terms, this means that MI can match heterogeneous data types based on co-occurrence relationships, without needing to be forced into equivalence. It also means that it is an integral quantity that does not change under reparameterization.

Due to the relatively high computational cost of MI estimators, it is important to consider both accuracy and speed. Most MI estimators fit into two categories: kernel and histogram.

Kernel estimators spread each sample point into the surrounding random variable space with a kernel function. They frequently use a Gaussian kernel [75]. Kernel-based MI estimators require an assumed shape and variance for the kernels, and a number of techniques have been developed to set these parameters. The choice is still somewhat arbitrary, and the results can be sensitive to this choice.

Histogram estimators generate empirical probabilities of discrete random variables by sorting samples into bins [72]. Like the kernel estimators, histogram estimators suffer from arbitrary parameters in the bin sizes and locations. They also require a certain number of samples per bin to converge to each probability. Data-shuffling techniques can reduce the small-sample bias that is typical of histogram MI estimators [76]. Under certain conditions on the complexity of the underlying system, these techniques can also provide data-robust upper and lower bounds on MI.

Kraskov et al. use a fundamentally different approach to MI estimation [73]. Their k-nearest neighbor distance metric demonstrated very small bias by resolving the smallest scales in the data. This is particularly useful for testing independence, when the theoretical MI is zero but estimators typically give a positive result.

L_2 norm

The L_p norm of the residual vector between data sets f and g is defined as

$$L_p[f-g] = \left(\sum_i |f_i - g_i|^p \right)^{\frac{1}{p}} \tag{3.7}$$

The most commonly used are the L_1, L_2, and L_∞, equivalent to average absolute value, root-mean-square, and maximum magnitude, respectively.

Minimizing the residual L_2 norm is equivalent to maximizing the correlation between f and g. Bland and Altman made the point that the correlation coefficient by itself is not a valid objective measure for comparison, in a very widely cited article on medical measurement methods [77]. While they do not say so explicitly, the reason is clearly that the correlation coefficient is a test statistic. The significance level is the proper metric for null hypothesis evaluation, not the value of the statistic itself. The significance is a monotonic function of the statistic, however, and the L_2 norm is appropriate, if only as a relative indication of model fit. The one-to-one correspondence between the L_2 as a test statistic and its significance level means that either will give the same result when used as an optimization objective function.

The L_2 norm of the difference between two data sets is attractive for its simplicity and speed of computation. It is limited by the need to produce two data sets that, when properly registered, are linear functions of each other. This can be challenging or impossible with multimodal data from sensors of different types, producing data with different physical units, or if other characteristics, such as additive noise distributions, differ.

MI can be superior to the L_2 norm for some applications in a few respects. The L_2 is an estimate of statistical correlation, so it only tests for a linear relationship. Any nonlinearity in the correspondence between two data sets will misleadingly decrease a correlation estimate. In contrast, MI is equally sensitive to any invertible nonlinear relationship. MI requires no assumptions about noise probabilities, and is robust to non-Gaussian distributions as long as they are independent. MI is properly sensitive to errors that are uncorrelated but not independent, though this is probably uncommon in physical data.

An implementation of Kraskov's k-nearest neighbor MI estimator was tested for registration. The achievable registration accuracy was very good, but computational efficiency was poor. The MI estimator did not scale well with the volume of data, and it could only be tested on smaller subsets of the real data.

In addition to computational cost, mutual information has another obstacle. True

Mutual Information is defined in terms of an unknown probability distribution. With real data, however, only samples are available rather than a probability distribution, and only estimates of the true MI are possible. Different MI estimators can vary in sensitivity to this finite data effect.

Ultimately MI proved to be too slow, and its computation time increases at a rate faster than a linear function of data volume. This does not recommend MI for similar future applications, where the data volume may be even greater. The L_2 norm is much faster computationally. Since it is the RMS of the difference, it scales linearly with data volume. Both front and back surface features are extracted from the same ultrasound A-scans, so the requirement for monomodality with the L_2 norm is no restriction. Also, each of the features has a nearly constant amplitude over its surface. This justifies the assumption that maximum correlation between front and back surface amplitudes implies that the two are optimally aligned. Analogous statements are true for the datum hole features. For these reasons, both the datum hole objective function in Section 3.3.1 and the surface echo objective function in Section 3.3.2 use the L_2 norm.

3.3.1 Datum hole objective function

The datum hole objective function is simply the L_2 norm of the difference in amplitude between the two datum hole features. Figure 3.11 shows this metric as a function of the rotation shift (p_6) applied to the Upper datum feature in Figure 3.2. The L_2 operation makes a metric that is much smoother than the features from which it is computed.

It is possible to use a simple, gradient-based single-objective optimization algorithm to find the optimal p_6. This datum hole registration is a separate, preliminary step to the multi-objective optimization over the other nine registration parameters.

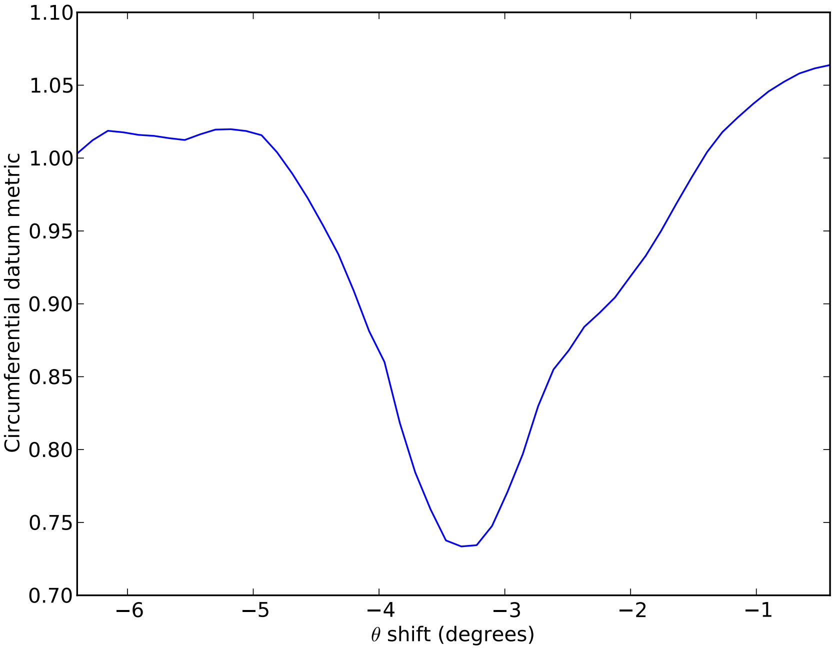


Figure 3.11: Datum hole metric, as a function of the rotation shift parameter. The metric quantifies the lack-of-fit between the datum hole signals from the BA 0° and TU 0° scans. The minimum occurs when the rotation shift parameter has the value that acheives the optimal overlap between the BA 0° and TU 0° datum hole features.

3.3.2 Surface echo RMS objective functions

The surface echo RMS metric is the L_2 norm of the difference in amplitudes between a matched pair of surface features. The registration mapping translates the points in the surface features from their t, u, v indices to a common, physical x, y, z coordinate system. The amplitudes in the three-dimensional back surface feature are interpolated to the locations of the two-dimensional front surface data points. This interpolation is a structured interpolation, as described in Section 3.3.4, so the inverse mapping that it uses is very fast.

After interpolation, the front and back amplitudes correspond to the same locations in physical space. The L_2 norm then takes the difference in amplitudes, squares the differences, and sums them to a single total, as expressed by Equation 3.7. The front and back surface echoes will not generally be equal in amplitude at their peaks. The difference in amplitude is not important since it is the location, not the value, of the metric minimum that matters. Each physical surface used in registration contributes a distinct surface echo metric as a separate optimization objective function. For the Seeded Defect Disc, there are six, as shown in Figure 3.1: BA, CB, GF, QR, ST,

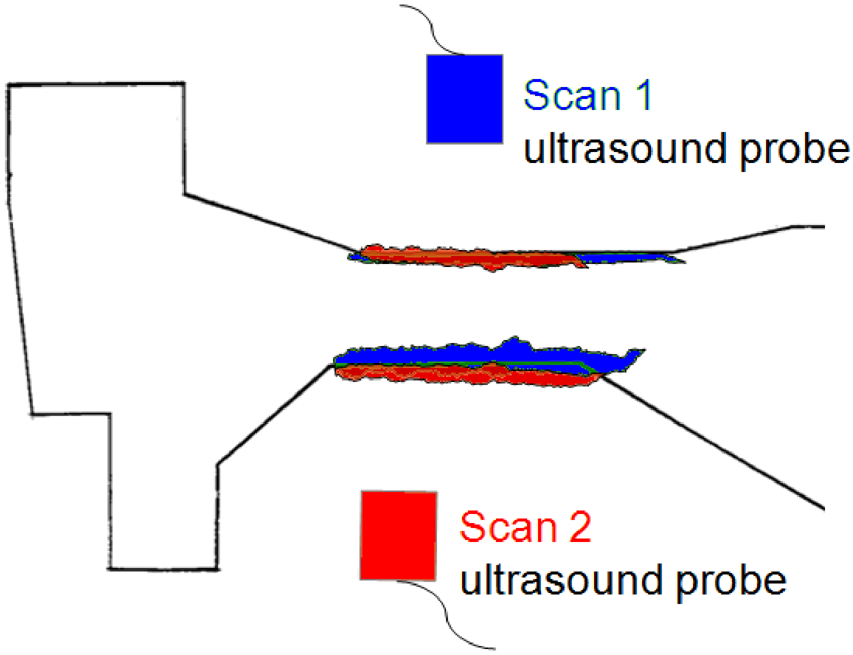


Figure 3.12: Schematic diagram showing the surface echoes used in the RMS objective functions. The front surface from the upper scan (in blue) is paired with the back surface from the lower scan (in red) and vice versa. Each pair of surfaces forms a separate RMS metric. The registration results for the full disc in Chapter 4 use six pairs of registration surfaces.

and TU.

In each A-scan used for registration, there is some uncertainty in surface echo time location due to the finite ultrasound pulse duration. The surface echo RMS metric accounts for the gradual drop-off in amplitude versus time, giving a gradual increase in the metric value. This effect is visible in Figure 3.13, which shows the RMS metric for the front and back echoes from the BA surface. The metric is plotted as a function of two registration parameters while the other seven are held constant. The two variable parameters in Figure 3.13 are the speed of sound in water (p_2) and the distance in water between the probe and part surface (p_3). Both of these parameters have the effect of scaling the relative surface echo arrival time. There is a roughly linear region in the figure with points that are all nearly minimal. This linear region shows the parameter values that give close to ideal registration for the BA surface. Moving orthogonal to the minima in these two parameters has the effect of shifting the two surface echoes away from each other, in the direction normal to the surface. The gradual drop-off in amplitude versus time in the surface echoes creates a finite distance over which the RMS metric increases, when moving away from the linear minima. At a sufficiently large displacement the two surface echoes no longer overlap, and the metric is locally constant with respect to the registration parameters.

The surface echo metric tends to have good sensitivity to parameter changes that move points in the direction normal to the registration surfaces. It is not very sensitive to changes that move or deform the surface in the tangent direction, within its own plane or conical surface. This is why it is important to include pairs of surfaces that are not all parallel to each other, as stated in Section 3.1.4.

This is also one of the reasons that it is important to use multi-objective optimization rather than a single objective function. Surfaces that are not parallel will have different sensitivities to the various parameters but also different values at their minima. There is no way to add these metrics into a single optimization objective while preserving sensitivity to all parameters.

Another reason to use multi-objective optimization rather than trying to combine metrics into one is that there are no comparable units between the surface echo metrics and the center of mass metric in Section 3.3.3.

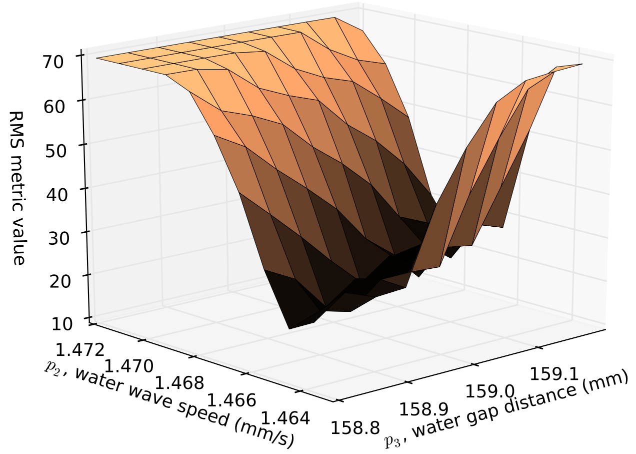


Figure 3.13: Root-mean-square (RMS) registration metric as a function of two of the registration parameters. The height of the plot surface shows the metric value for the lack-of-fit between the BA 0° front surface and the TU 0° back surface. (See Figure 3.1.) The RMS metric is shown as a function of: 1) p_2, the speed of sound in water, and 2) p_3, the distance in water between the probe and part surface. The other seven registration parameters are held constant. The minimum occurs over a roughly linear region in the parameter domain, rather than at a single point. The metric reaches a constant maximum away from the region surrounding the minima. Objective functions with constant-value regions like these can be very problematic for optimization solvers.

3.3.3 Center of mass objective function

Once registered, a pair of front and back surface features that are echoes of the same disc surface should obviously overlap. It is possible for such a pair to be widely separated, to the point that they do not touch. In this case, the surface echo metric in Section 3.3.2 is locally constant. This constant-metric region is visible in Figure 3.13, in the corner where p_2 is greatest and p_3 is least. The surface echo metric in that region has zero sensitivity to all registration parameters. The optimizer would have no information on how to improve the solution. It is necessary to use another metric with a longer reach, which can guide the optimization toward parameter values that cause paired surface features to overlap.

The center of mass (COM) metric allows the optimization effectively to pull surface echo features together, so the more sensitive surface echo metrics can find a more precise alignment. Figure 3.14 shows the COM metric as a function of p_2 and p_3,

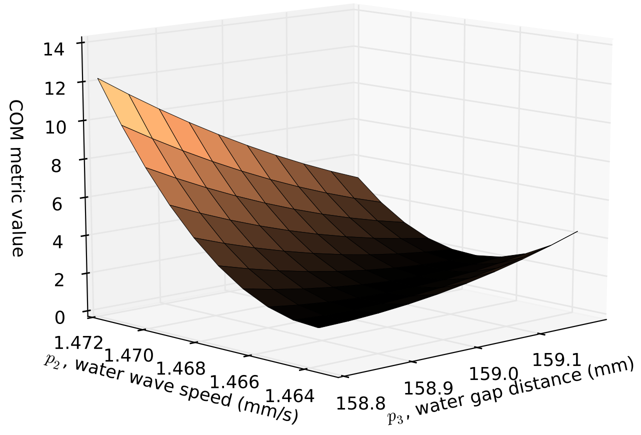


Figure 3.14: Center of mass (COM) registration metric as a function of two of the registration parameters. The height of the plot surface shows the metric value as a function of: 1) p_2, the speed of sound in water, and 2) p_3, the distance in water between the probe and part surface. The other seven registration parameters are held constant. There is a linear locus of minima, similar to the RMS metric in Figure 3.13. The COM increases monotonically away from a minimum, in contrast to the RMS. There is no plateau region that could trap the optimizer. However, it is also smoother and less sharply defined near the minima, which could lead to a less precise result if optimized without the additional information from the RMS metric. The COM and RMS metrics thus complement each other, and both contribute to efficient and accurate optimization.

over the same region in the parameter domain as for the RMS metric in Figure 3.13. The COM metric is close to minimal over the same general region as the RMS metric, though the minima are not as sharply defined. However, Figure 3.14 shows that the COM does not have the constant value region that is so problematic for optimizers.

The COM metric first takes the centroid in r, z plane of all the points in each surface echo feature. The r component of the centroid, r_{COM}, is computed as the mean of the data point locations, r_i, weighted by amplitude, A_i.

$$r_{COM} = \frac{\sum_i^{N_d} A_i r_i}{\sum_i^{N_d} A_i} \tag{3.8}$$

An analogous expression computes z_{COM}. Each pair of front and back surfaces contributes a distance, $\sqrt{r_{COM}^2 + z_{COM}^2}$, between their centroids. The COM metric is the sum of all these distances.

3.3.4 Interpolation

Registration and data fusion both require amplitude data samples at identical domain locations. Any nontrivial mapping of a data set will produce points that do not coincide with the points in another data set, requiring some form of interpolation.

Interpolation is potentially very computationally demanding with the large data sets generated by automated NDT scans. When data sets are 3D, the computational burden for interpolation can become a dominant factor, limiting the trade-off between registration quality and time to completion and driving some researchers to hardware-accelerated implementations [78].

A naïve algorithm that interpolates each of the N points in a data set using an interpolant defined by a second data set of comparable size would scale as $O(N^2)$, in both time and memory. The best scenario for interpolation is to limit interpolation to structured data. Algorithms that are as local and low truncation order as possible also generally enhance computational efficiency. This speed can come at the cost of accuracy and smoothness. If the interpolation errors are too large they can lead to Functions with Noise-Induced Multimodality (FNIM), an optimization pathology that can easily outweigh the time saved in interpolation [79–81].

Interpolation of unstructured data

If interpolation of unstructured data is needed or desired, it must be implemented with great care. A brute-force search for nearby points would be $O(N)$ for each point to interpolate, so interpolating a significant fraction of a data set would scale as $O(N^2)$. A hierarchical search, such as a k-D tree, would be optimal, scaling as $O(\log(N))$ in time and memory [82]. Unfortunately, this approach would create discontinuities in the interpolation at the boundaries between subregions on the tree.

One interesting alternative is a recently published technique using the Fast Multipole Method (FMM) [83]. This interpolator approximates a Radial Basis Function (RBF) which would otherwise require an $O(N^3)$ solution of a dense linear system. The original paper defined the RBF in terms of an Improved Fast Gauss Transform, though other types of RBFs can be used by substitution with the appropriate Taylor series in the FMM. It is still necessary to build a search tree, but this requires only $O(\log(N))$ operations instead of $O(N)$ and the resulting interpolation is smooth.

Interpolation of structured data

When data are structured, the samples are arrayed in a regular grid. It is then possible to define a coordinate system in which the amplitude data points are all located at integral values of the coordinates. This index coordinate system is, in a sense, the “native” geometry of the raw data structure. Interpolating points in these coordinates is fast and memory efficient. It is simple, in some cases trivial, to write code for the interpolation. For example, a nearest-neighbor interpolation requires only rounding the coordinates, then using the rounded-off coordinates as indices to lookup the amplitude.

There is never a need to search or build hierarchical tree structures. Any higher-order algorithm can obtain neighboring points by integer addition and subtraction after rounding, and it is trivial to compute distances between points. Interpolating all N points in a data set requires only $O(N)$ operations, even with higher-order methods, as long as they use a limited number of nearby points. The only difficulty in structured data interpolation is the transform from physical coordinate to the index coordinates. Even this step is facilitated by that fact that it is the inverse operation of the transformation from indices to physical coordinates. The forward transformation must already be implemented to define the mapping function. For example, Equations 3.1 through 3.5 define the forward mapping transformation for the turbine disc application.

Results from early experiments with unstructured interpolation were slow and, in some cases, numerically unstable. The computational and conceptual advantages of structured interpolation are clear, and it should be preferred over unstructured whenever it is possible to use. Because of these advantages, the software written for this work uses structured data for all interpolation. For each interpolated point, the operators use a first-order algorithm from the data at the corners of the bounding cube in the index coordinate system of t, u, and v that was defined in Section 3.2.1.

3.4 Multi-objective optimization

Once metrics are defined, data set registration can use optimization to find the minimal discrepancy. Realistic metrics will have multiple minima, requiring a global (as opposed to local) optimization. Simple local optimizers like Newton-Raphson are not suitable for global optimization unless they are wrapped by an algorithm that also seeks diversity and can escape a local minimum.

Computational efficiency requires registration metrics that operate on a subset of the full data. Often there are a number of metrics from which to choose, each with higher sensitivity to some registration parameters than others. Techniques exist to estimate these sensitivities [84], and it might be desirable to exploit known differences in sensitivity to decouple the problem into subspaces. In general, however, a multi-objective optimization solver is necessary to find simultaneously optimal solutions.

3.4.1 Optimization definitions and terminology

Optimization is one of the most computationally demanding parts of the analysis, making it worthy of careful implementation. Each sample point evaluated by the optimizer requires a call to all of the registration metrics. Optimizer inefficiency therefore multiplies the computation time for all of the mapping, interpolation, and registration metric operations. Therefore, a well-chosen multi-objective optimizer with efficient sampling is key to overall computational performance.

In single-objective minimization, there is typically one point in the parameter space, x_{min}, that is a global minimum. It strictly dominates all other points; that is, $f(x_{min}) < f(x)$ for the objective function f at all other x. In multi-objective optimization with multiple objectives f_i, there are points x_{dom} that are dominated by at least one other point, x_1. Mathematically, this means $f_i(x_1) \leq f_i(x_{dom})$ for all i and $f_i(x_1) < f_i(x_{dom})$ for at least one i. There can also be a point for which

$f_i(x_1) < f_i(x_2)$ on some i, but $f_i(x_1) > f_i(x_2)$ for other i. In this case neither x_1 nor x_2 dominates the other.

The set of all points that are not dominated by any other point is called the Pareto optimal set, or Pareto front [85, 86]. If there are N_f objectives that are continuous, well-behaved functions of the parameters, the Pareto optimal front will be a manifold with up to $(N_f - 1)$ dimensions within the objective function domain. Multi-objective optimizers search for this family of optimal solutions rather than a single point.

The registration results in Chapter 4 all use the Hypervolume Indicator (HI) optimization algorithm described in Section 3.4.2. The reasons for this selection will be made clear in Section 3.5. The field of multi-objective optimization is broad, and there are several considerations to use in choosing an algorithm for each application. Following is a discussion that gives some background for this rich field.

Perhaps the simplest approach to multi-objective optimization is to apply a single objective global optimization algorithm to a weighted sum of the objective functions [87]. This amounts to a parametric representation of the Pareto front in the objective function space, with the weights as the parameters. It is relatively easy to implement and can work well for simple, monotonic objective functions. It has difficulty when the Pareto front is nonconvex, when there are local minima near the global minimum, or when the density of points in the objective space varies too much over the parameter space. It can also be computationally inefficient, since each set of weights requires a separate call to the optimizer [87]. There is no reuse of information from previous calls, but each call is often unlikely to produce more than a small number of Pareto points in close proximity.

Given the requirement for a local solver, the work of Maes, Vandermeulen, and Suetens might prove valuable in the selection of an optimization algorithm. They evaluated both gradient- and non-gradient-based optimizers, such as the Powell, simplex, steepest-descent, conjugate gradient, quasi-Newton, and Levenberg-Marquardt methods for multiresolution medical image registration [88]. While their results are directed only toward single-objective optimization, their analytical derivative of MI ties together registration, interpolation, and optimization in a potentially useful way.

Genetic (GA) or multi-objective evolutionary algorithms (MOEA) form a family of optimizers inspired by the mutation and crossover of genetic material in biological reproduction [89]. Deb's work on genetic algorithms also has widely cited example objectives for testing and comparing multi-objective optimizers [90]. MOEA are very flexible in that they do not require an expression for the objective function gradients, nor do they require even that such gradients exist. They also tend to be robust to function evaluations that fail due to constraints or problems in the objective function code. Many implementations of MOEA can be difficult to initialize without problem specific prior knowledge [91]. They can also require many generational iterations to converge to a good representation of the Pareto front. More advanced MOEA often have many control parameters to fine-tune mutation rates and crossover, in order to balance exploration against local exploitation [92].

Particle swarm methods operate by analogy to herd or hive animals that spread out when searching for resources, then converge when some of them find a promising region. Yang catalogs many of the nearly innumerable variations on this basic idea, as well as a few other methods like tabu search and simulated annealing [93]. Poli, Kennedy, and Blackwell give a good overview of the developments with particle swarm methods since 1995 [94]. A particle swarm method can also serve as a secondary optimizer to spread outward from a Pareto point and fill in the optimal set [95]. Mostaghim and Teich apply particle swarm methods with a diversity criterion [96]. Their diversity metric is based on angular positions in objective space.

Most studies comparing multi-objective optimizers use thousands of function evaluations. Knowles and Hughes made a valuable contribution by comparing multi-objective optimizer performance at 100 and 250 function evaluations [97]. They tested two of their own algorithms: Bin_MSOPS uses a binary search tree to spread outward from points with high fitness [98]. ParEGO is initialized with a latin hypercube and then builds and updates a Kriging surrogate model to predict the largest expected improvement [99]. ParEGO was previously shown to be superior on a broad range of test problems than a widely-used evolutionary algorithm (NSGA-II), and it had significantly better performance than Bin_MSOPS on most of the limited-evaluation tests. These results are significant to the work of this thesis, given the need for rapid optimization results from relatively expensive objective functions. Their use of a Kriging model is also similar in spirit to the surrogate model developed for the Pareto density correction in Section 3.5.2, though in Section 3.5.2 it is for Jacobian estimation rather than as part of the optimization itself.

Łaniewski-Wołłk also used a Kriging surrogate model, in his case to predict HI improvement [100]. Comparisons with other criteria for expected improvement (expected improvement with fixed base values and probability of improvement) showed Expected Hypervolume Improvement (EHVI) to be the best for all attempted test cases for single-point sampling. In a related work on EHVI implementation, Emmerich, Deutz, and Klinkenberg developed a more efficient method for computing expected improvement in HI given a Gaussian predicted PDF for the to-be-evaluated point [101]. Extensions to the method presented in this thesis could incorporate either of these methods to the HI optimizer described in Section 3.4.2.

Another promising idea is to combine a variety of optimization methods under a higher-level supervisor. The supervising operator allows each optimizer to choose new points to sample and, with each iteration, favors the optimizers that have chosen more optimal points in the past. Vrugt, Robinson, and Hyman developed this idea in the context of single objective evolutionary optimization with A Multialgorithm Genetically Adaptive Method for Single Objective Optimization (AMALGAM-SO) [102]. They highlighted the fact that the No Free Lunch Theorem (NFL) [103], as well as many performance studies, show that it is impossible to develop a single optimizer that is efficient on a large range of problems.

An adaptive, higher level optimizer like AMALGAM-SO can avoid extensive manual trials to find the best optimizer for a given application while still achieving nearly optimal performance. It can operate on top of any other optimization algorithm, and it would ensure that any new optimizer that might be added as a future option could only improve performance. Vrugt et al. [102] specifically mention the ability of AMALGAM-SO to deal with non-smooth objectives, similar to the Functions with Noise-Induced Multimodality issues that can arise with registration metrics.

It is important to recognize that the optimization solver's purpose is not to find a single value for each of the registration parameters. For some parameters, this might be the least important result of the optimization. Offsets in geometry might occur that are significant in comparison to scan step size, due to part-to-part dimensional variability or human positioning of parts and sensors. However, relative distances and angles should be quite accurate once the scan is under machine control. It is possible that all of the Pareto registration parameter values would not map data points very far from their initial, nominal location. If the optima are too far from the nominal in these cases, the inspector might decide that a data set that is unusable and should be redone. The more important optimization result would then be the measurement of spread or uncertainty in the parameters.

This again points to the need to define how to translate multi-objective optimization results into results on a parameter distribution. Section 3.5 describes an original method to do just that, by computing consistent statistics on the Pareto-optimal set.

A related point is that, in contrast to most optimization applications, neither the objective functions nor the parameters are of any interest of themselves. The objective functions only serve to indicate likely values for the parameters. The parameters are important only insofar as they indicate the relative position between points in two data sets.

An ideal optimizer would be biased toward sampling points in the parameter space that gain the most information about these relative positions. Bader's recent work on hypervolume metric optimizers mentions a method for biasing the distribution of points along the Pareto front, according to a user-defined criterion [104]. Efficient sampling for information in the dependent quantities of interest would be a novel and very useful application of Bader's thesis. Appendix A shows an example of how this distribution biasing could work in a registration application. In the remainder of this thesis, however, only the unbiased method of Bader's HI optimizer is used.

Pareto set identification

Even the seemingly simple act of finding the dominant points among objective function samples can quickly become a computational burden. The trivial algorithm of comparing each pair for dominance scales very poorly, as N^2 where N is the number of objective function evaluations. In the author's experience with a previous version of the software, the time required produces a noticeable delay with a few hundred points. It becomes crippling at a few thousand points or beyond.

Fortunately, more efficient algorithms exist for this problem. Godfrey, Shipley, and Gryz provide an excellent reference to practical scaling issues, discussed in the context of the equivalent problem of database skyline look-ups [105]. The present work uses their Block Nested Loops 1 algorithm with unlimited window size, which performs very well in this application.

3.4.2 The Hypervolume Indicator optimization algorithm

There are many high-quality, well-tested single-objective optimizers that are readily available. The parametric multi-objective optimizers cited above are one way to leverage this body of work. However, a much more theoretically sound approach for using a single-objective optimizer is the Hypervolume Indicator (HI), also known as the Hypervolume Estimator. All consistent optimization algorithms will converge to points on the Pareto front as the number of iterations increases. However, as explained below, HI is the only optimization algorithm for which the statistical convergence behavior is known a priori [104]. This makes possible an extension to the HI optimization, contributed by this thesis in Section 3.5, to estimate integrals over the entire, theoretical Pareto set. The NDT registration results in Chapter 4 will use this new capability to calculate residual uncertainty in relative data locations.

HI is defined as the fraction of a bounded region of the objective space that has thus far been dominated by a set of evaluated points. It expresses the quality of a population of points with a single quantity, making HI a unary indicator. It is the only known unary indicator that converts multiple objectives to a single objective while preserving strict compliance with Pareto optimality [104, 106].

HI optimization simultaneously seeks diversity as well as local improvement, without any *ad hoc* control parameter needed to balance between them. It also provides an indication of completion. Early discoveries of Pareto points will increase the HI sharply, while later Pareto points that smooth out the optimal manifold will only dominate a small part of previously undominated space. As the number of points increases to infinity, the HI will converge asymptotically to a constant value. This stopping criterion is an independent application of HI, and is useful regardless of whether or not it is used to select points during optimization.

Asymptotic HI point distribution

The HI optimization algorithm has a unique property that is very important to the Pareto integral method in Section 3.5. Namely, the Pareto points that are obtained from HI optimization converge to a fixed distribution as the number of points on the Pareto front increases [107]. The form of this distribution depends only on the geometry of the Pareto front in the objective domain.

The Pareto front is a locus of points, f^*, in the domain of the objectives, f. When the Pareto front is a smooth function of the objectives, the distribution of optimal points discovered by the HI converges to an expected density, $\delta_F(f^*)$. Bader's thesis [104] gives the form for this density.

$$\delta_F(f^*) = \frac{1}{C} \sqrt[N_f]{w(f^*) \prod_{i=1}^{N_f} e_i^*} \tag{3.9}$$

C is a normalization constant, such that the integral of δ_F over the Pareto front is unity. The user can define an optional weight function, $w(f^*)$, to alter the density [108, 109]. The e_i^* are the components of the unit vector, $\hat{e}^*$, that is normal to the Pareto front at f^*.

HI is the only multi-objective optimization algorithm known to have this limit distribution property [110]. Most multi-objective optimization algorithms are designed to search the parameter domain in an effort to spread points out across the Pareto front. However, there is no way to predict the distribution they will obtain.

Bader showed that it is possible to weight the hypervolume integral in HI, to bias the asymptotic distribution in a user-defined way. A recently published paper by the present author shows one approach to this type of weight-biasing; this is reproduced in Appendix A. Unfortunately, the weighting greatly increases the computational burden for HI in two ways. First, the HI integral estimation requires Monte Carlo sampling. This is much slower than the exact computation method for the unweighted integral. Second, weighted HI requires an approximation of the Pareto front. It may be possible to perform this approximation in a similar manner to Section 3.5. In contrast to the application in Section 3.5, which only requires an approximate Pareto front at the conclusion of optimization, weighting the hypervolume requires an approximation for each generation of the HI optimizer.

Given the computational cost, such an effort to bias the points during optimization might not be worth the extra complexity unless the Pareto front is highly curved in the objective domain. For many applications it will be more effective to spend the same resources on sampling more points with the unweighted optimizer, rather than forcing the optimizer to distribute the points more evenly with a weight function. This is the approach taken in the development that follows in Section 3.5.

3.5 Statistics over the Pareto set

This section presents for the first time a method to estimate an integral over the Pareto front in a multi-objective optimization problem. The method is applicable for an unknown Pareto front, which is a necessary capability for any realistic application. This algorithm requires the points returned by an HI optimizer and the assumption or knowledge that the objectives are smooth functions of the optimization parameters. No other information about the problem is required.

The method computes any integral of a function over the Pareto front as a statistic of the Pareto points. In application to this thesis, this Pareto integral estimator is used to compute the mean and standard deviation of the Pareto set of registered point locations. The mean and standard deviation are the two statistics used in Chapter 4 to summarize the location and spread of the Pareto point distributions that result from registration.

The Pareto integral estimator requires that the optimal points come from the HI optimization algorithm described in Section 3.4.2. The fact that the points obtained by HI converge to a distribution with a known form is essential, and this property is unique to HI. Any other optimization method would be unable to predict any particular distribution. The points would not be guaranteed to cover the whole front, and the point density would be arbitrary and unknown.

HI does produce an asymptotic distribution, but it is not uniform. An integral approximated by samples from the Pareto front needs to account for the non-uniform density. The Pareto integral estimator compensates for the non-uniformity in a manner analogous to the Monte Carlo Importance Sampling method in Section 5.4. As with the Monte Carlo method, the integral over the continuous Pareto front is approximated by a sum over discrete points. Convergence depends on the Monte Carlo strategy of taking independent, identically distributed samples of the integration domain variable, as if it were a random variable. All of the Pareto points from the optimizer are in the desired integration domain, but the density is non-uniform. If the density can be estimated at each sample point, the effect that it has on the integral can be compensated for.

If the f_i^* points are sampled from the $\delta_F(f^*)$ density, the integral of a function, $g(f^*)$, in the objective domain is

$$E_{Pareto}\left[g(f^*)\right] = \int g(f^*)df^* = \int \frac{g(f^*)}{\delta_F(f^*)}\delta_F(f^*)df^* \approx \sum_{i=1} \frac{g(f_i^*)}{\delta_F(f_i^*)}\,. \tag{3.10}$$

Equation 3.10 shows how to estimate statistics of the response quantities. For example, if $\delta_F(f^*)$ were known it would be possible to estimate the mean of all Pareto-optimal mapped locations for any physical location, using $E_{Pareto}[r(p^*)]$ and $E_{Pareto}[z(p^*)]$ where $f^* = f(p^*)$. The mean provides a single best estimate that unifies all the information from the Pareto front.

Equation 3.10 can also estimate the second moments as the expected values of $r(p^*)^2$ and $z(p^*)^2$. The standard deviation of r^*, σ_{r^*}, from the means and second moments is then

$$\sigma_{r^*} = \sqrt{E_{Pareto}\left[r^{*2}\right] - E_{Pareto}\left[r^*\right]^2} \tag{3.11}$$

and equivalently for z. The standard deviation quantifies location uncertainty, for r and z individually. The results in Chapter 4 use this method to estimate mean and standard deviations for these quantities in the Seeded Defect Disc registration.

Using Equation 3.9, the Pareto density correction factor for the jth Pareto point in the sum in Equation 3.10 is

$$\frac{1}{\delta_{Fj}} = \frac{\sum_k \sqrt[N_f]{\prod_{i=1}^{N_f} e^*_{ik}}}{\sqrt[N_f]{\prod_{i=1}^{N_f} e^*_{ij}}} \tag{3.12}$$

where e^*_{ij} is the i component of the normal vector, evaluated at the jth Pareto point.

Bader's method also provides the user with the capability to weight the Pareto front with a weight function, $w(f^*)$ in Equation 3.9, in order to bias the achieved density during the optimization. Indeed, this Pareto biasing was the focus of Bader's work, and the application of known density to statistic estimation seems not to have been considered. By not applying a particular weight to the hypervolume integral, the weight function defaults to unity. See the paper in Appendix A for an example of how this weight function can be used for a capability similar to Pareto statistic estimation. All results in Chapter 4 use an unweighted HI optimization, with the density in Equation 3.9.

3.5.1 Pareto normal estimation

The primary obstacle to applying Equation 3.9 is that the shape of the Pareto front is unknown. It is necessary to approximate the unit normal, $\hat{e}^*$. The following is an original method devised for this work to estimate $\hat{e}^*$ and to use that estimate to compute statistical properties of the Pareto front according to Equation 3.9.

The first step in understanding the Pareto normal approximation method is to recognize that, from a Pareto point, all infinitesimal steps in the parameter domain are tangent to the Pareto front in the objective domain. Any movement away from the Pareto front is a second order effect in the parameter domain. This can be proven by contradiction.

Consider an infinitesimal parameter increment dp from an optimal point, p^*. The objectives $f(p)$ are approximately linear near p^*, such that $f(p^*+dp) = f(p^*)+df(p)$. If $f(p^*)$ strongly dominates $f(p^*)+df$, then $df_i > 0$ for all i. But the infinitesimal step in the opposite direction in the parameter domain, $-dp$, corresponds to the point $f(p^*) - df$ which strongly dominates $f(p^*)$. This negates the initial assumption that

$f(p^*)$ is Pareto optimal. Therefore, $f(p^*)$ cannot strongly dominate $f(p^*) + df$.

Now consider the possibility the case of two parameter increments in different directions, dp_1 and dp_2, from the same optimal point, p^*. In the objective domain, these points are $f(p^*) + df_1$ and $f(p^*) + df_2$. Posit a strongly dominant relationship: $df_{1\,i} < df_{2\,i}$ for all i. If this were true, then there would be a third parameter increment, $dp_3 = dp_1 - dp_2$, which by linearity dominates the original point, $f_i(p^*) + df_{3\,i} < f_i(p^*)$. This also negates the initial assumption that $f(p^*)$ is Pareto optimal.

Therefore, if p^* is a Pareto point, then so is $p^* + dp$ for an infinitesimal step dp in any direction in the parameter space. Because all $f^* + df$ are in the Pareto front all df are tangent, as vectors in the objective domain, to the Pareto front at f^*. That makes all df vectors orthogonal to $\hat{e}^*$. With $N_f - 1$ linearly independent df vectors, it is possible to find a vector that is orthogonal to all of them and therefore proportional to $\hat{e}^*$.

The differential objectives are expressible in terms of differential parameters.

$$df = J\,dp\,, \tag{3.13}$$

where the elements in the Jacobian matrix, J, are

$$J_{ij} = \frac{\partial f_i}{\partial p_j}\,. \tag{3.14}$$

Because every df is tangent to the Pareto front at f^*, its scalar product with the normal vector is zero.

$$df \cdot \hat{e}^*(f^*) = 0 \tag{3.15}$$

Now define a linear system

$$Ae^* = b \tag{3.16}$$

in which A is a $(N_p+1)\times N_f$ matrix. The matrix A in this system is a concatenation of the transpose of the Jacobian matrix and a vector of ones. That is, the elements in the first N_p rows are

$$A_{ji} = \frac{\partial f_i}{\partial p_j}\,, \tag{3.17}$$

and the elements in the last row are

$$A_{(N_p+1)\ i} = 1\,. \tag{3.18}$$

The vector b has length $N_p + 1$ and is all zeros except for the last component, which is one.

In multi-objective optimization there can be more parameters than objectives, so in general A will have more rows than columns. Equation 3.16 then represents an overdetermined system that requires an approximate solution. This is achieved with the Moore-Penrose pseudoinverse.

$$e^* = \left(A^T A\right)^{-1} \mathbf{1} \tag{3.19}$$

The solution to the linear system in Equation 3.19 provides, after normalization, the components of the Pareto normal vector for the density in Equation 3.9.

The solution to Equation 3.19 is well-defined at any point. This is important because it means the above procedure will work approximately, even when point is near but not exactly on the Pareto front. The solution to Equation 3.19 finds a vector that is close to orthogonal to all N_p differential objective vectors, df_j. "Close to orthogonal" here means that it is optimal in the least-squares sense. It is also computable for any number of parameters, as long as $N_p + 1 \geq N_f$. This allows an approximation to the HI density δ_F without requiring a model for the Pareto front shape as a function of f^*. The only remaining requirement is an approximation to the Jacobian.

3.5.2 Approximate Jacobian matrix

A finite difference calculation can approximate the Jacobian, if the objective functions are all computationally inexpensive. However, it is common in optimization problems for the objective functions to be very demanding of computing resources. Registration of large data sets is no exception. A finite difference Jacobian matrix would require $N_p N_f$ additional objective evaluations for every Pareto point. In many cases the number of additional function calls would be comparable to those during the optimization itself. Following is a method to approximate the Jacobian matrix in a much more computationally efficient manner.

The Jacobian approximation begins with a surrogate function for the objectives. The surrogate selected for this work is based on a radial basis function (RBF) for each objective as a function of the registration parameters. An RBF approximates a function from a collection of sample points, often referred to as nodes or nodal points, where that function has already been evaluated. Each nodal point is the center of a kernel. Each kernel has an amplitude, k_i, that depends on the distance from the sample point.

$$k_i(p) = a_i e^{\overline{\|p_i - p\|^2}^{-1}} \tag{3.20}$$

In the Gaussian kernel in Equation 3.20, the amplitude depends on the Euclidean distance in the parameter domain between the RBF node at p_i and the point under evaluation at p. A coefficient, a_i, scales the amplitude of each kernel. It is clear from Equation 3.20 that the influence of any node diminishes to zero at points far from it.

Ordinarily, the RBF approximation to a function is the sum of the kernels at all nodal points, p_i. The RBF approximation to an objective function, $f_{RBF}(p)$, is slightly different. The objective functions in this work are always greater than zero, and they are optimal at their minima. All of the Gaussian kernels go to zero as the distance $\|p_i - p\|$ increases to infinity. An ordinary RBF that used only the sum of the kernels would be erroneously minimal at points far from all the Pareto nodal points. Therefore, to improve accuracy near the convex hull of p_i, $f_{RBF}(p)$ includes a constant term, f_{max}, that is equal to the maximum value [2] of f.

$$f_{RBF}(p) = f_{max} + \sum_{i}^{N_{RBF}} k_i(p) \tag{3.21}$$

The amplitude at every point has a contribution from all N_{RBF} nodes. The kernel weights are determined by evaluating the RBF at the node locations and forcing $f_{RBF}(p_i)$ to equal $f(p_i)$. This produces a dense linear system that can be solved for a_i.

---

[2] Each surface echo L_2 norm metric has a constant, maximum value when its pair of surfaces do not overlap. The Center of Mass metric does not have a maximum value; its f_{max} is set at $2 \max_i f_{COM}(p_i)$.

Equation 3.21 generates an RBF approximation for each objective function, using the Pareto optimal points already discovered. This provides a smooth approximation to these functions that is differentiable everywhere. It is then relatively easy to evaluate the derivatives. The derivative of the objective function with respect to one of the parameters, p_j, is

$$\frac{df_{RBF}}{dp_j} = \sum_{i}^{N} 2k_i\left(p\right) \frac{p_j - p_{ij}}{\left\|p_i - p\right\|^4} \tag{3.22}$$

These are the terms that the Jacobian matrix in Equation 3.14 comprises.

3.5.3 Application of the Pareto integral estimator

The Pareto integral estimator in Section 3.5 can compute any integral of a function over the Pareto front. The method is very general, and is applicable not only to data registration but to any multi-objective optimization. It can compute quantiles, higher moments, or any other statistic that is the expected value of a function of the Pareto quantities. It requires no prior knowledge of the optimization problem or its outputs, other than smooth objective functions as described above. The estimator only requires the discrete Pareto points from the optimizer. As with practically all consistent numerical integration algorithms, it nevertheless converges to the true integral in the limit as the number of Pareto points increases to infinity.

It would be desirable at this point in the discussion to compare results from this Pareto integral estimator to another. However, to the author's knowledge, no other method has been devised that is capable of this type of estimation. The next chapter will demonstrate one example of its use in a practical application, with an unknown Pareto set. The registration results in Chapter 4 use this estimator to show the mean and standard deviation of the optimal mapped locations that result from registration.

3.6 Summary

This chapter outlined the multiple steps in development and computation required for registration of data from distinct data sets. Registration is an essential preliminary step for defect detection.

Section 3.1 defined feature extraction methods for datum hole reflections and the front and back surface echoes. Section 3.2 outlined the mapping functions used in the turbine disc application, with a total of ten registration parameters: six rigid body degrees of freedom and four other uncertain physical quantities. Section 3.3 defined the quality-of-fit metrics, for both the datum hole and surface echo features, and clarified their essential characteristics. The mapping parameters and fit metrics formulated registration as an optimization problem, becoming the optimization parameters and objective functions, respectively.

Section 3.4 explained why registration should use multi-objective optimization and briefly explored the background of this rich and varied field. The section also justified the selection of the Hypervolume Indicator optimization algorithm from among the many available, leading into the development of Section 3.5.

Section 3.5 presented an original contribution that may have much value and wide applicability beyond data registration. The capability to evaluate statistics on the Pareto set gives multi-objective optimization practitioners the ability to use their results in a new way. The Pareto integral algorithm in that section takes advantage of a unique property of the Hypervolume Indicator optimizer, with a recently published form for the asymptotic Pareto density. This new algorithm developed for this thesis can estimate arbitrary integrals over the full, theoretical Pareto front.

A radial basis function approximation to the objective functions was devised, in order to estimate the normal vector term in the Pareto density. This approximation enables a correction for the theoretical HI density, without the need to know the Pareto front beforehand.

When applying registration to a particular application, it is important to understand how accurate the registration needs to be. This will be depend on the details of each application. Clearly, an insufficient registration will have a negative impact on the subsequent detection. Beyond a certain tolerance, however, more registration accuracy is unnecessary and only wastes time and computation. Chapter 4 shows how the Pareto integral estimator developed for this thesis quantitatively estimates the registration accuracy. The registration results in Chapter 4 serve as an example of how a data-fusion NDT practitioner can ensure the registration is fit for purpose, while also reserving computational resources for detection.

Chapter 4

Registration results

This chapter shows results for registration of the data sets from the Seeded Defect Disc. The procedure follows the approach laid out in Chapter 3. Features are extracted from the data set, after which the data point locations are mapped to search for improved values of the registration metrics. Each physical surface used to register the data gives rise to a pair of front and back surface echoes in the ultrasound data, each from a separate scan from opposite sides of the disc. Each of these surface echo pairs contributes a distinct, unique registration object, representing an individual source of information about the registration fit. The back surface echo feature from each pair has a small yet finite thickness, corresponding to the temporal extent of the echo. The metric quantifying overlap between an echo pair therefore has a region of optimal parameter values, rather than a single best fit with zero error. Residual uncertainty will always remain for these physical reasons, and the registration procedure outlined above accounts for this.

Registration is a multi-objective optimization problem, with the metrics as objective functions. The registration parameters are the optimization decision variables. The optimization searches for parameter values that improve the registration metrics. Instead of a single set of parameters, the end result is a set of Pareto optimal points. Some Pareto points will have a lower value in one metric, and moving to another Pareto point will increase one registration metric but decrease another. In this way, the full registration results represent the trade-offs between these separate sources of information and the residual uncertainty that remains at the limit of what registration can achieve.

The results in this chapter begin with the most simple registration step. Section 4.1 gives results from the circumferential registration. Circumferential registration is a one-dimensional, single-objective optimization problem to find only the rotational offset parameter. The circumferential registration is a simple, straightforward process compared to the multi-objective optimization for the remaining parameters.

The set of Pareto points that result from any multi-objective optimization will dominate all other sampled points, but they do not dominate each other. In the context of data registration, this creates a multi-valued, one-to-many mapping for any given initial physical point. Each point in physical θ, r, z space maps to a “cloud” of multiple points, also in physical space, each of which is Pareto optimal. The Pareto integral estimation method, developed in Chapter 3, serves to quantify statistics on these Pareto point clouds. These statistical estimates give the mean and standard deviation in r and z directions for the Pareto set.

The work in this chapter puts all of the preceding registration steps together in a real application. Test cases with modified data sets will give a standard against which to evaluate performance. The test cases are designed to create the ideal condition for evaluation, when the distortions for which registration will compensate are known. Because the correct solution is known, errors can be computed exactly. These test case results will indicate how closely to expect registration to align the real disc data.

After circumferential registration, the surface echo registration results follow. Section 4.2 shows the results of test cases constructed by duplicating one side of real disc registration data. This produces tests with known registration solutions. The errors remaining after registration are shown for increasing magnitudes of applied distortion.

Section 4.3 has results for the real registration problem on the full data set from the Seeded Defect Disc. This is the registration problem that will be faced by every disc inspection under the automated NDT system proposed by this thesis. Under ordinary circumstances, the registration itself pulls in all available information on relative positioning between the data sets. There would be nothing left for an independent confirmation of registration success. The Seeded Defect Disc is different, in that there are known indications, which are nowhere used in registration. These indications are evident in the ultrasound data from multiple scans, on both sides of the disc. The indications therefore provide a means for evaluating the registration results for the full registration problem.

By the end of this chapter, the registration performed for Section 4.3 will have prepared the data, bringing all data sets into alignment and ready for detection. The Pareto statistical information on location accuracy will also provide the detector with an estimate of achievable registration distance. The detector needs that distance to set an appropriate resolution size.

4.1 Circumferential registration

Recall from Section 3.1.1 that the circumferential registration was simplified by separating the angular shift metric and parameter to a separate optimization problem. The problem then reduces to finding the minimum of the datum hole metric, which is plotted in Figure 3.11.

This is a single-objective optimization problem with only one parameter. There is only one minimum of the function in the vicinity of the initial parameter value. The objective function (see Section 3.3.1) tends to be quite smooth near the minimum, so the optimization converges very quickly. Unfortunately, no quantitative information about uncertainty is available like the multi-objective optimizations provide. Because this is a single-objective optimization, there is only one optimal parameter value in the solution.

It might be possible to revisit this lack of circumferential registration uncertainty quantification in future work. For example, a user could examine how quickly the metric increases away from the minimum. An expression for the parameter uncertainty would then need a quantified uncertainty for the value of the metric.

In the final analysis, the need for uncertainty quantification on the rotational offset parameter does not appear to be very great. The uncertainty in the r and z

directions are tied to characteristic lengths that are needed to define the detection approach. The same characteristic length will be applied to all three directions in space, including the circumferential direction, to divide the volume into small blocks that are nearly cuboidal. The circumferential length derived from the r and z uncertainty corresponds to a range of angular uncertainties, according to the range of r over the disc cross section. The angular uncertainty so derived already encompasses the region near the minimum of the datum hole objective function in Figure 3.11.

4.2 Surface echo registration with known distortions

To begin the registration performance analysis, it is important to use a test case with a known result. The registration process described in Chapter 3 is somewhat complex, with many different processing steps that operate on a large volume of data. An evaluation of a real registration application would be difficult or impossible, without first knowing where the correct answers lay.

This section shows a way to evaluate the full registration method and its implementation in software, using realistic data. A registration problem with a known solution is created by duplicating the data from the upper surface registration scans BA, CB, and GF (see Figure 3.1). The duplicate data sets are treated as if they were scans from the lower side of the disc, by nominally reversing the definition of front and back surfaces. For example, in the original BA scan, the echo from the BA surface is the front surface echo and the echo from TU is the back surface echo. In the duplicate BA scan, the surface echo for BA is treated as a back surface, and the TU echo is treated as a front surface. The original, non-duplicate scans are then distorted by applying known registration parameter values to the mapping functions. The parameter values are chosen randomly, but in such a way that the applied displacements equal target magnitudes, up to a maximum of 20 mm. This magnitude is greater than what is expected in practice, so the test results below should provide an upper bound on the effect of the distortion magnitude.

This method of generating test cases also avoids any possible error due to inadequacy or insufficient flexibility in the mapping functions. The mapping functions for the initial distortion are the same as the mapping functions used by the registration, to compensate for that distortion. In this way, the theoretically ideal registration is guaranteed to be in the function space of the registration mapping.

4.2.1 Construction of a known registration test

The registration test cases for this section use a simple yet powerful way to generate tests with known distortions. The same method should be possible to use for any registration application, with real data from that application. This is accomplished by taking two copies of the upper data sets, as mentioned previously. For the Seeded Defect Disc, only the BA 0°, CB 0°, and GF 0° from the upper set are used for registration, so only these three scans require duplication. One of the copies is labeled the “upper” and the other copy is labeled the “lower” data set. On the lower copy, the test case generation algorithm extracts a back surface registration feature from the front surface echo. It extracts a front surface registration feature from the back surface echo. These front and back surface registration features are then available for registration with themselves, from the upper data set copy.

The algorithm then chooses a set of random values for all of the eight parameters, excluding the rotational offset parameter. These random parameter values define a transformation that is applied to the data positions in the upper copy of the data sets, with a random distortion magnitude. The algorithm rescales all of the random parameter values together, iteratively, until the distortion reaches a targeted magnitude.

The upper data positions are then fixed by these initial random parameters. The lower data set copies vary their locations with the registration, with parameter values chosen by the optimizer as is normally the case in real registration. Because the original distortion is known, it is possible to evaluate exact errors and compare them to the uncertainty estimates based on the Pareto integral method derived in Section 3.5.

4.2.2 Known distortions in r

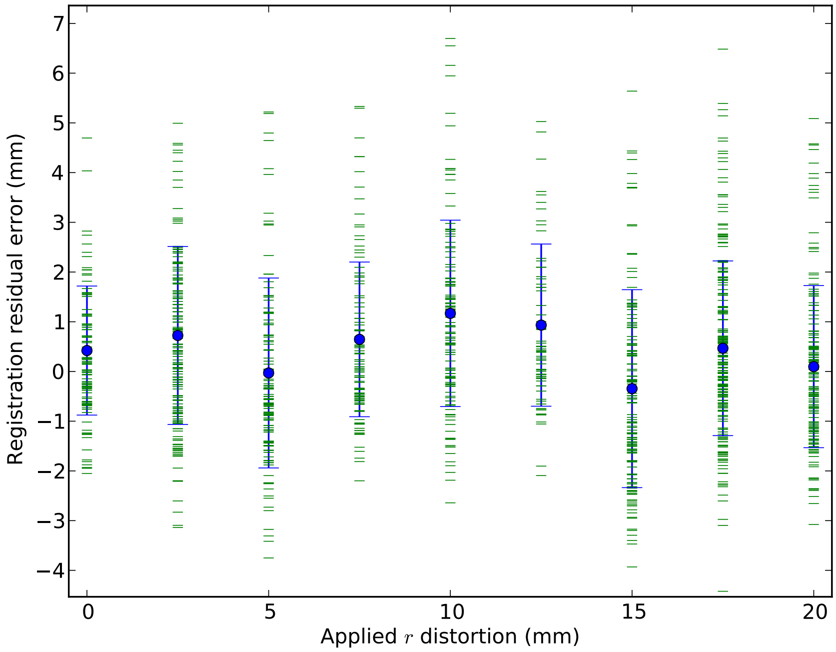


Figure 4.1: Registration error with an applied distortion in r. The distortion magnitudes are as measured at Corner J in the JI 0° scan, which is the extreme uppermost corner on the outer rim of the disc. The parameters in the upper registration scans are artificially perturbed in a random direction to produce displacements in r of the given magnitudes. The registration software searches for the perturbed parameter values, as if they were unknown, to bring the lower registration scans back into alignment with the upper scans. Green ticks mark all of the Pareto points from each registration optimization. Blue circles and error bars mark the means and standard deviations, respectively, for each set of the green Pareto points. The registration algorithm is able to match applied distortions up to at least 20 mm consistently, to within 1 mm for the Pareto mean. The registation quality does not discernably depend on the distortion magnitude.

In this section, the random parameter values are constrained to target the magnitude of the r component of the distortion. The z distortion components vary as well, but only the r direction is controlled for these tests. Figure 4.1 shows the results. There is a separate registration problem for each r distortion magnitude, from 0 to 20 mm in increments of 2.5 mm. The distortions are measured at the J corner, as located in the JI 0° scan. Corner J was chosen because it is at the maximum radius and distance from the mid plane of the disc, as shown in Figure 3.1. This maximizes the effect of any errors in rigid body rotations, and so gives a conservative estimate on the resulting accuracies.

Each of the horizontal green ticks in Figure 4.1 represents a Pareto point. The vertical position of each green tick in the plot shows the r component of the error for one of the Pareto points. The blue circle shows the error of the Pareto mean of r. Note that it is not simply based on the mean of the Pareto points. It is weighted by the Hypervolume Indicator density correction factor in Equation 3.12, computed by the method defined in Section 3.5.

All Pareto means are within 1.25 mm of the exact value, and all but one mean error is less than 1 mm. The standard deviations are approximately 1.5 to 2 mm. Most significantly, there does not appear to be any trend with increasing distortion magnitude, for either the mean error or the standard deviation.

4.2.3 Known distortions in z

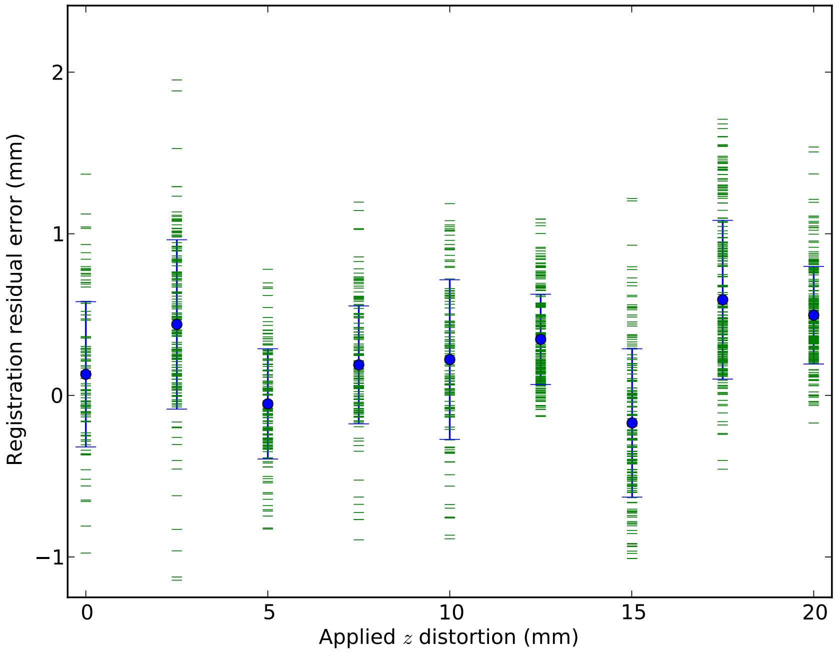


Figure 4.2: Registration error with an applied distortion in z. The distortion magnitudes are measured at Corner J in the JI 0° scan. The parameters in the upper registration scans are perturbed to produce displacements in z. As with the r distortion results in Figure 4.1, registration matches distortions of at least 20 mm to within 1 mm regardless of the distortion magnitude.

The same procedure to generate random distortions with a given magnitude as in Section 4.2.2 is applied again, in this case to the error in the z direction. Analogously to before, the distortion magnitude targets only the z component while the r component varies freely. Figure 4.2 shows the resulting Pareto points, the mean, and standard deviations.

As with the r error in Section 4.2.2, there is no discernible trend in the error with respect to the distortion magnitude. All errors are less than 0.75 mm. All but one are less than 0.5 mm. It is evident that the z component of the residual registration error is smaller than the r component. This trend continues throughout the rest of the test case results.

4.2.4 Known distortions in both r and z

Once again the random distortions, are applied to the upper set of scans, as in Sections 4.2.2 and 4.2.3. In this section, both r and z distortions are together targeted to the given distance. In other words, $\sqrt{r^2 + z^2}$ is required to equal the target distortion at the reference point.

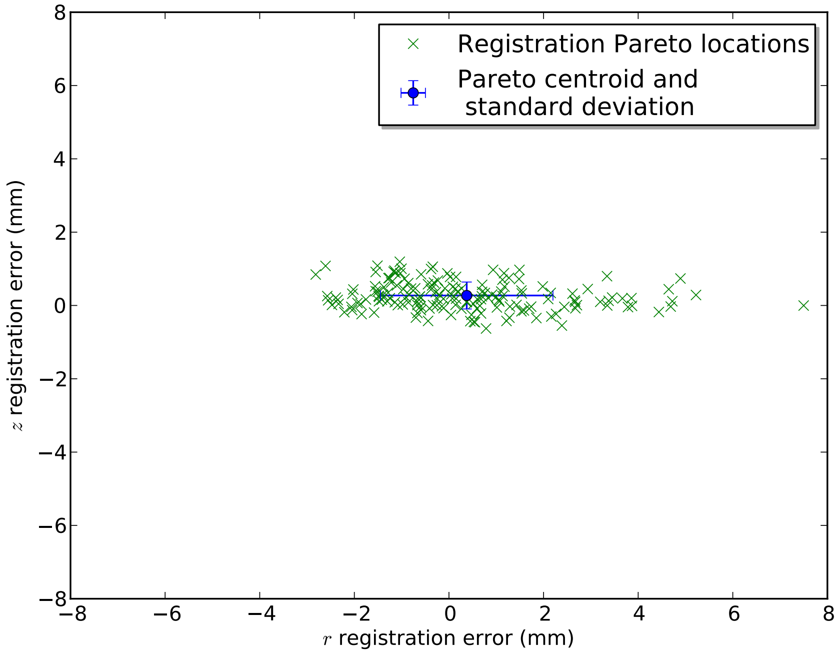


Figure 4.3: Registration error, with a 20 mm applied distortion in r and z. The distortion magnitudes are measured at J in the JI 0° scan. The parameters in the upper registration scans are artificially perturbed in a random direction to produce a displacement such that $\sqrt{r^2 + z^2} = 20$ mm. Each green × marks a Pareto point resulting from the registration optimization. The blue circle marks the mean location of the Pareto points. The blue error bars mark the standard deviations of the Pareto points in the r and z directions.

Figure 4.3 shows the results for such a registration. Each registration produces a set of bars, which are one standard deviation in the r and z directions. Note that the “68% rule” — that there is a probability of approximately 0.68 that a sample will fall outside one standard deviation from the mean — only applies to a Gaussian distribution. It does serve to illustrate the fact that there is a significant chance that a sample could be outside of one standard deviation from the mean.

In all of the following plots, the results come from running the same type of registration after an applied distortion. The registration is run multiple times with different applied distortion and different initial points in the optimizer. The results test the repeatability of the optimization results. They also give some statistical sense of the variability in the means and standard deviations. As always, these are estimated by the Pareto integral estimator from Section 3.5. On the plots to follow, there are no clouds of green Pareto points. For clarity, only the mean and standard deviation from each separate optimization are shown.

5 mm distortion at J

Figure 4.4 shows results from a 5 mm applied distortion, measured as before at the J corner in the JI 0° scan. The means all have errors in r that are within 1.5 mm. In z, the errors are within 1 mm.

The standard deviations are consistent among all registrations. They span a range from 0.5 to 0.75 mm in z. The r standard deviations have a consistently wider range. One standard deviation from the mean gives spans of approximately ± 1.5 to ± 1.75 mm in r.

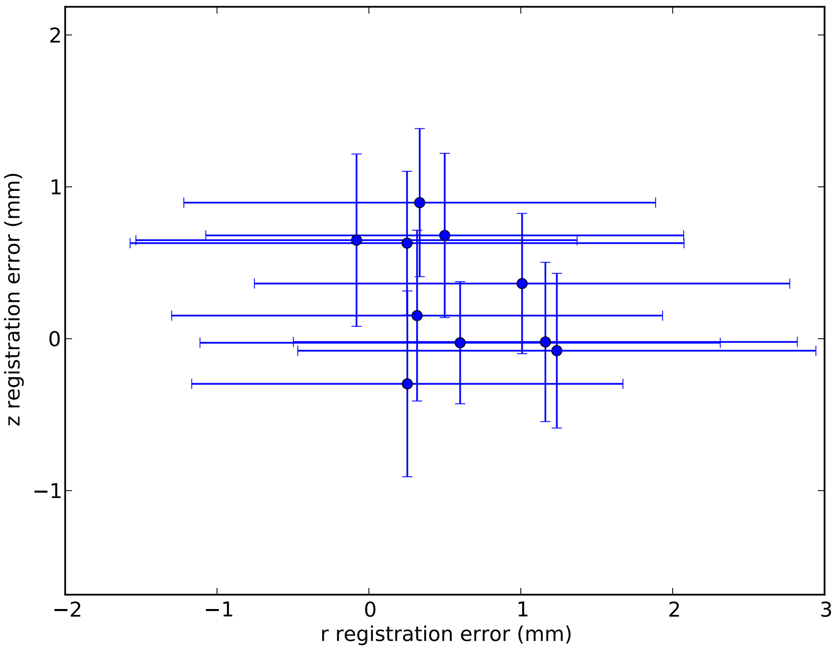


Figure 4.4: Registration errors from repeated runs with 5 mm applied distortions in r and z at Corner J. Parameter perturbations in random directions were applied, and the resulting registration errors were measured at J in the JI 0° scan, as in Figure 4.3. The procedure was repeated ten times with the same distortion magnitude but different sets of random parameter values, to test the consistency of the results. Each circle and and set of error bars show the mean and standard deviations, respectively, of one registration optimization. The residual errors of the means in r are less than 1 mm in all cases. Standard deviations of the r residual errors are typically between 1 and 2 mm. The true location in r is well within one standard deviation from the mean for all ten trials. The z error components are also less that 1 mm, with one exception. Standard deviations are smaller in z than in r, with values between 0.5 and 1 mm. This results in only seven registrations that bound the true error within one standard deviation. The other three are within two standard deviations.

10 mm distortion at J

Like Figure 4.4, Figure 4.5 shows the applied distortion at the J corner in the JI 0° scan. In this case the random applied distortion have a magnitude in the r, z plane of 10 mm.

Again, the Pareto means clustered within 1.5 mm in r and 1.25 in z. The standard deviations are about 1.5 to 2 mm in r, and 0.5 mm in z.

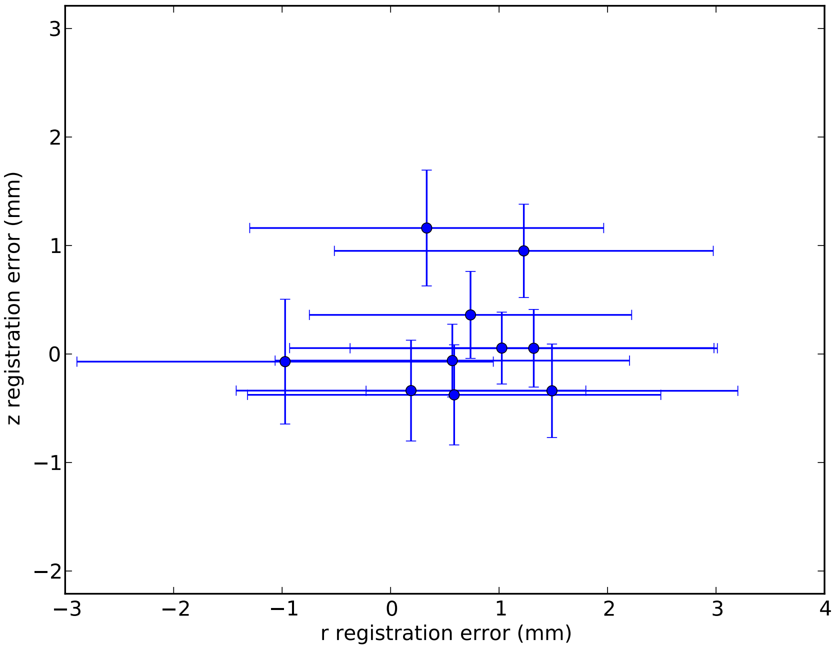


Figure 4.5: Registration errors from repeated runs with 10 mm applied distortions in r and z at Corner J. The same procedure applies as in Figure 4.4, but with a greater distortion magnitude. Registration error means and standard deviations at Corner J are similar in scale to those in Figure 4.4, despite the doubling of the distortion magnitude.

15 mm distortion at J

Figure 4.6 shows results from the same procedure as in Figures 4.4 and 4.5. Now the applied distortions have a magnitude of 15 mm.

Registration achieves an accuracy of 1.25 mm in r. In the z direction, the registration is within 0.75 mm. The standard deviation is again between 1.5 to 2 mm in r and 0.5 mm in z.

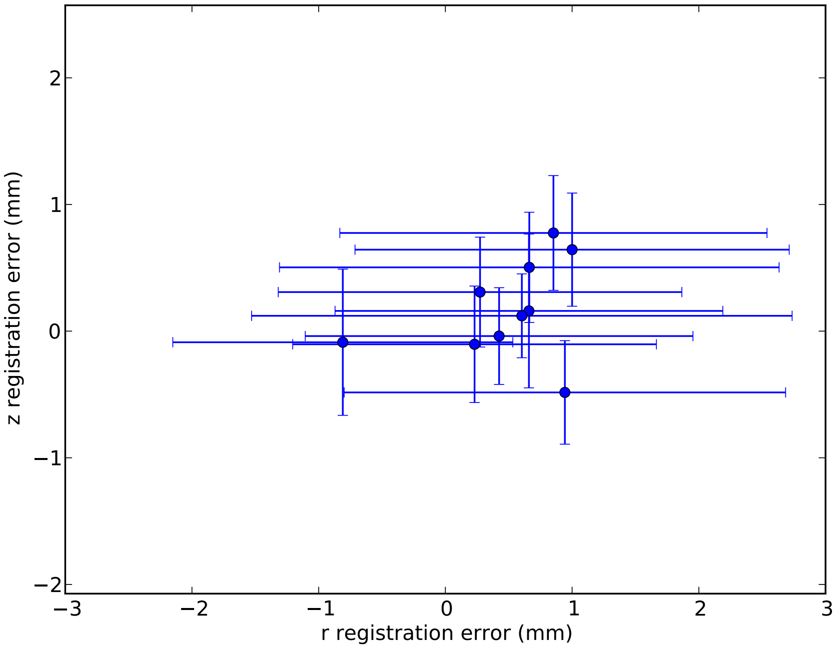


Figure 4.6: Registration errors from repeated runs with 15 mm applied distortions in r and z at Corner J. As in Figures 4.4 and 4.5, the plot shows means and standard deviations of Pareto points from multiple registrations, based on randomly distorted parameters. The registration error means and standard deviations in both r and z directions are very similar to those in Figures 4.4 and 4.5.

20 mm distortion at J

Figure 4.7 shows registration results for the maximum applied distortion of these test cases, of 20 mm. The original location of Corner J is recovered to within 1.25 mm in r and 0.75 mm in z. The standard deviations are between 1.5 to 2 mm in r, with one exception that is slightly greater. The z direction standard deviations are all approximately 0.5 mm.

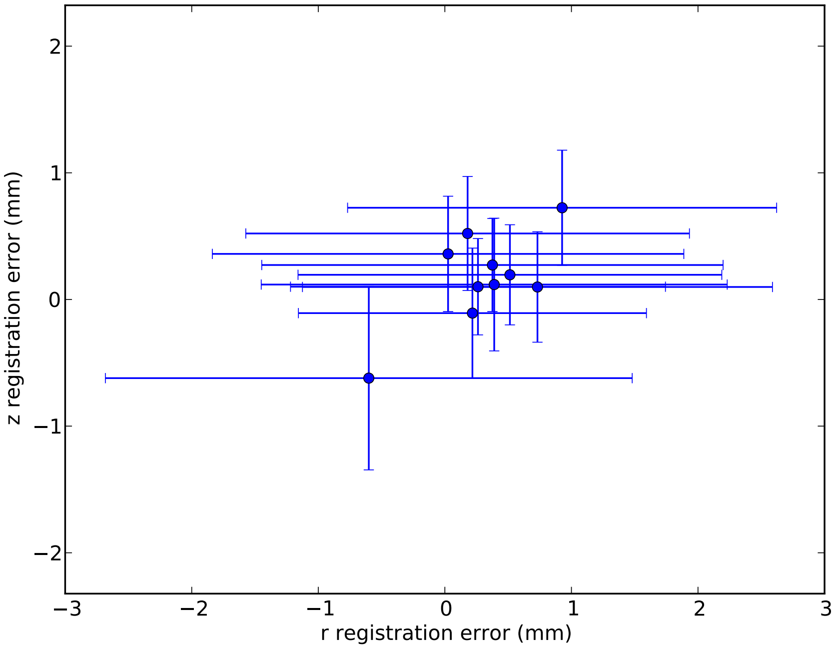


Figure 4.7: Registration errors from repeated runs with 20 mm applied distortions in r and z at Corner J. The registration errors come from the same procedure as in Figures 4.4 through 4.6, with greater distortion magnitude. The errors of the Pareto means are less than 1 mm in both directions. The Pareto standard deviations from each registration give a meaningful estimate of the residual error, with the r direction more conservative than z. All of these observations are consistent with the figures above showing results from 5, 10, and 15 mm distortions. There is no evidence of dependence of registration accuracy on the magnitude of the distortion. This is important to note because the true error magnitude is unknown when registering on real data.

20 mm distortion near U

All of the Figures 4.4 through 4.7 follow the distortion and corrections applied through registration at Corner J. Figure 4.8 shows registration results from another location. The same procedure is followed, but with distortions and registration error measured at the datum hole, near Corner U. Corner U is the opposite corner of the cross section from J. This makes it the point with the minimum radius, on the bore, as opposed to J on the outermost radius, on the rim. Figure 4.8 shows registration results after 20 mm applied distortions.

The errors are smaller near U, with 1.25 mm accuracy in r. All ten Pareto means are within 0.5 mm of the exact value in the z direction. In fact, with one exception, all are within 0.25 mm in z. All of the z standard deviations are below 0.5 mm, and standard deviations estimated in r are comparable to the results as J with the

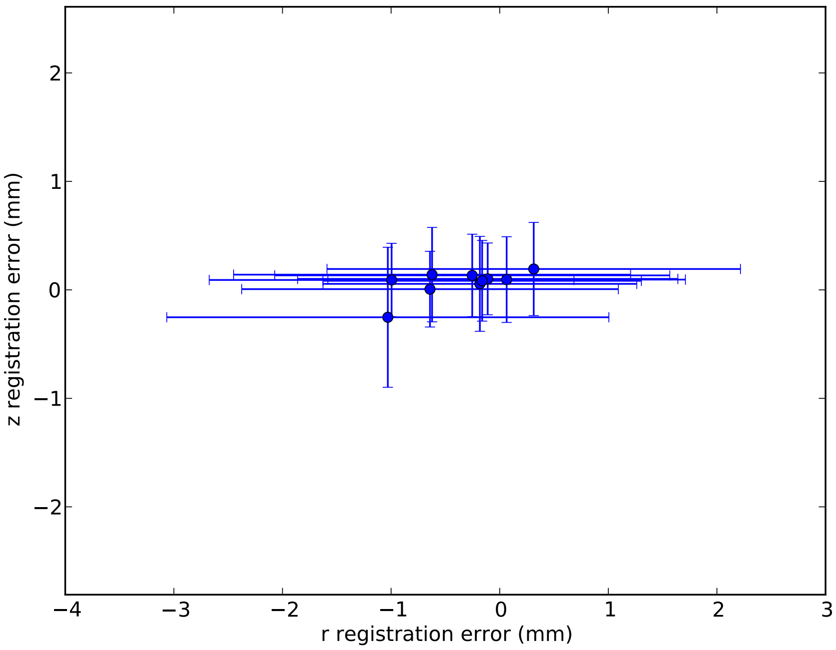


Figure 4.8: Registration errors from repeated runs with 20 mm applied distortions in r and z at a point near the disc bore. The registration errors come from the same procedure and distortion magnitude as in Figure 4.7, but with location error measured at different point. The point referenced in this figure is located at the midpoint between the BA and TU surfaces. This point has the same cross-sectional location as the circumferential datum hole. It is near Corner U, which is at the opposite corner of the cross section from J. This point also lies between two surfaces whose echo features are used in the registration. Registration accuracy closely follows the results at Corner J in Figure 4.7, except that the z errors are smaller. For these data, registration accuracy is not strongly dependent on the location within the disc.

15 mm distortion.

4.2.5 Known distortion summary

The test results from this section set expectations for accuracy with the registration results on real data, when registering upper and lower scans against each other. The first main conclusion from the results on known distortions is that there is no trend in accuracy versus applied displacement. The final results are just as accurate when initially distorted by 20 mm as when the distortion is zero. There is no need for concern about the magnitude of the initial distortion, which the registration needs to reverse. The residual uncertainty that remains after registration is due to the inherent, physical uncertainty in the data to be registered.

The standard deviations are almost all between 1.5 and 2 mm in r. The Pareto standard deviation seems to overestimate the remaining uncertainty in r, since errors of the mean are all less than 1.5 mm and 78 percent are less than 1 mm. Nearly all standard deviations in z are about 0.5 mm. All together, about 81 percent of the Pareto means are less than one standard deviation from the exact z value. The fact that both r and z have more than 68 percent of their samples within one standard deviation from the mean suggests that, for monomodal distributions, they are more sharply peaked than a normal distribution.

4.3 Surface echo registration with all scan data

The optimal registration parameters should map indications viewed from lower surfaces to align with that same indication as viewed from the upper surfaces. Nothing in the registration directs the indications to match each other directly. The alignment comes about as a byproduct of the surface feature registration because the same mapping that produced the successful registration maps all data, including any indications. The known indications in the Seeded Defect Disc therefore provide a way to evaluate the full machinery of the registration method and software implementation on real data.

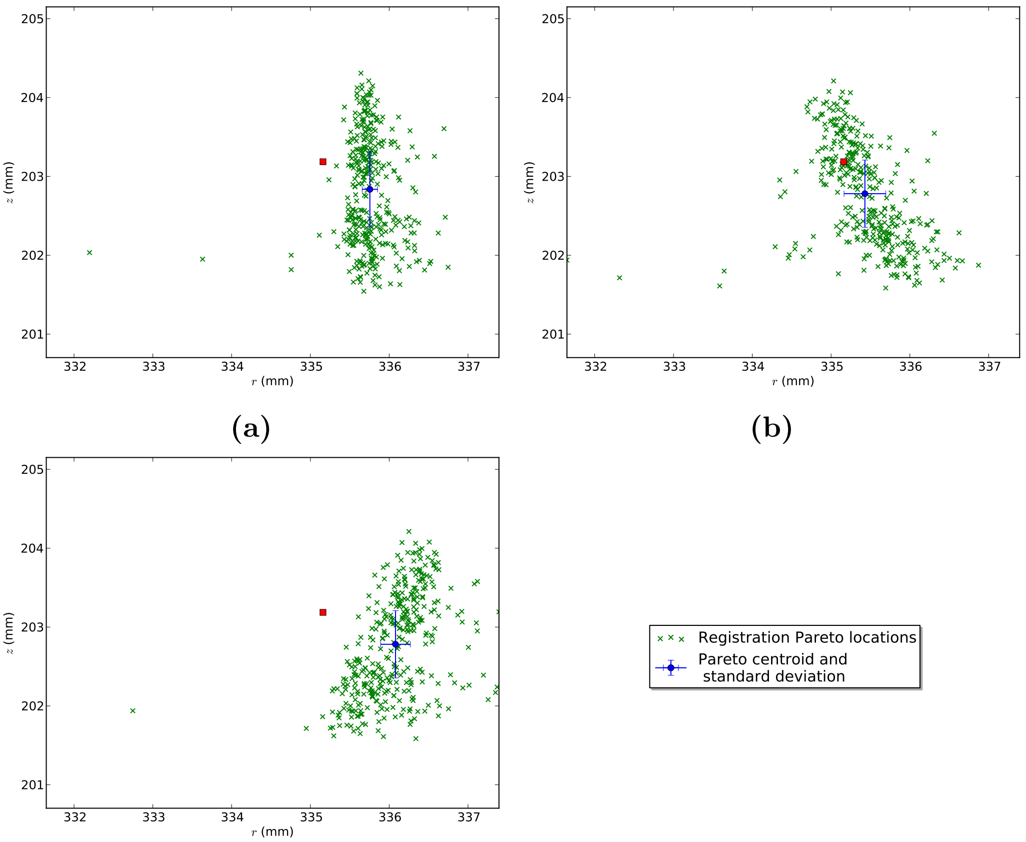
Registration results are shown below for the locations of Indication 1, Indication 3, and Indication 5. See Figures 2.4 and 2.5 for the locations of these indications. These three were chosen from among the five known indications because they are located near middle, outer, and inner radii, respectively. In terms of angular location, Indications 1 and 5 are separated by a wide angle, close to 180°. Indication 3 lies at an angle between Indications 1 and 5. Together these three indications span a wide range of locations. This spread in location makes these three points representative of the registration mapping throughout the disc.

The indications all have a finite size, and considering the beam spread from the focused probe, the signal created by these indications spans multiple A-scans within each scan. Indication 5, in particular, is significantly larger than the others. It might in reality be two defects located close together. The indications are also realistic in the sense that they surely have a surface texture and irregular shape, which can cause some directional dependance in the reflectivity. For the purpose of these comparisons, the point location for the indication in each scan is defined as the peak amplitude location. The finite size and non-point-like nature of the known indications means that the true relative locations cannot be determined with perfect precision. Nevertheless, the indications do provide a very valuable means to evaluate the registration accuracy for the full disc data.

Indication 1

Figure 4.9 shows the registration results on the full disc data, as evaluated at the location of Indication 1. The figure contains separate plots for each mapped scan, for easier readability. Three subplots in Figure 4.9 show the Pareto r, z locations for Known Indication 1 in each of the lower scans that view the indication from below. Each green ‘x’ marker in the plot corresponds to a Pareto point. The Pareto points discovered by the optimizer during registration map the original indication point to new locations. As in Section 4.2, the blue circles indicate the Pareto integral mean. The blue error bars show one standard deviation in each direction, estimated by the Pareto integral second moment. Red squares mark the location of Indication 1 as viewed from the upper scans.

Figure 4.10 shows all registration results with Indication 1 on one plot. The mean and standard deviations are identical to those shown in individual subplots in Figure 4.9. The clouds of Pareto points are omitted for clarity, and the means and error



(a) **(b)** **(c)**

Figure 4.9: Registration mapped point locations for Indication 1 in scans (a) QR 0°, (b) QR -5°, and (c) QR +5°. Registration Pareto points are marked in green, and the mean and standard deviations are shown in blue, as in Figure 4.3. The red squares are the locations of the indication in the upper scans. Locations of points in the upper scans remain fixed during registration. This example highlights the importance of the multi-objective optimization approach to registration. Registration found a distribution of points that represent the location uncertainty. A single-objective registration would have produced only one point, possibly on the opposite corner of the point distribution. There would also be no information regarding the spread or uncertainty of possible locations.

bars are shown together for easier comparison of all results.

The standard deviations are apparently underestimated in the r direction for QR 0° and QR +5°. The r standard deviations on Indications 3 and 5, below, are typically between 0.5 and 1 mm. This range of values would be a better indication of the uncertainty for Indication 1. The registration accuracy achieved, however, is very good for all three registered scans. After registration, all three lower scans are within 1 mm of each other and the upper scans.

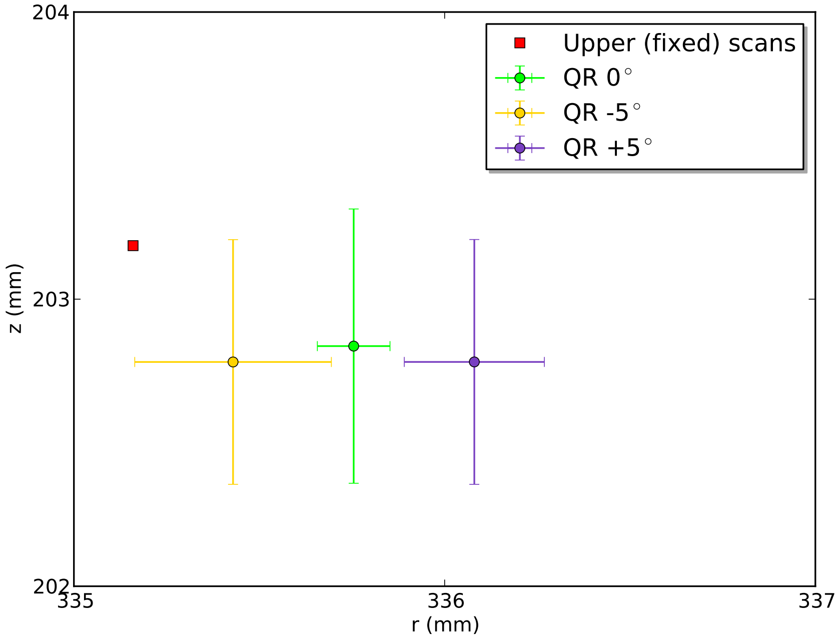
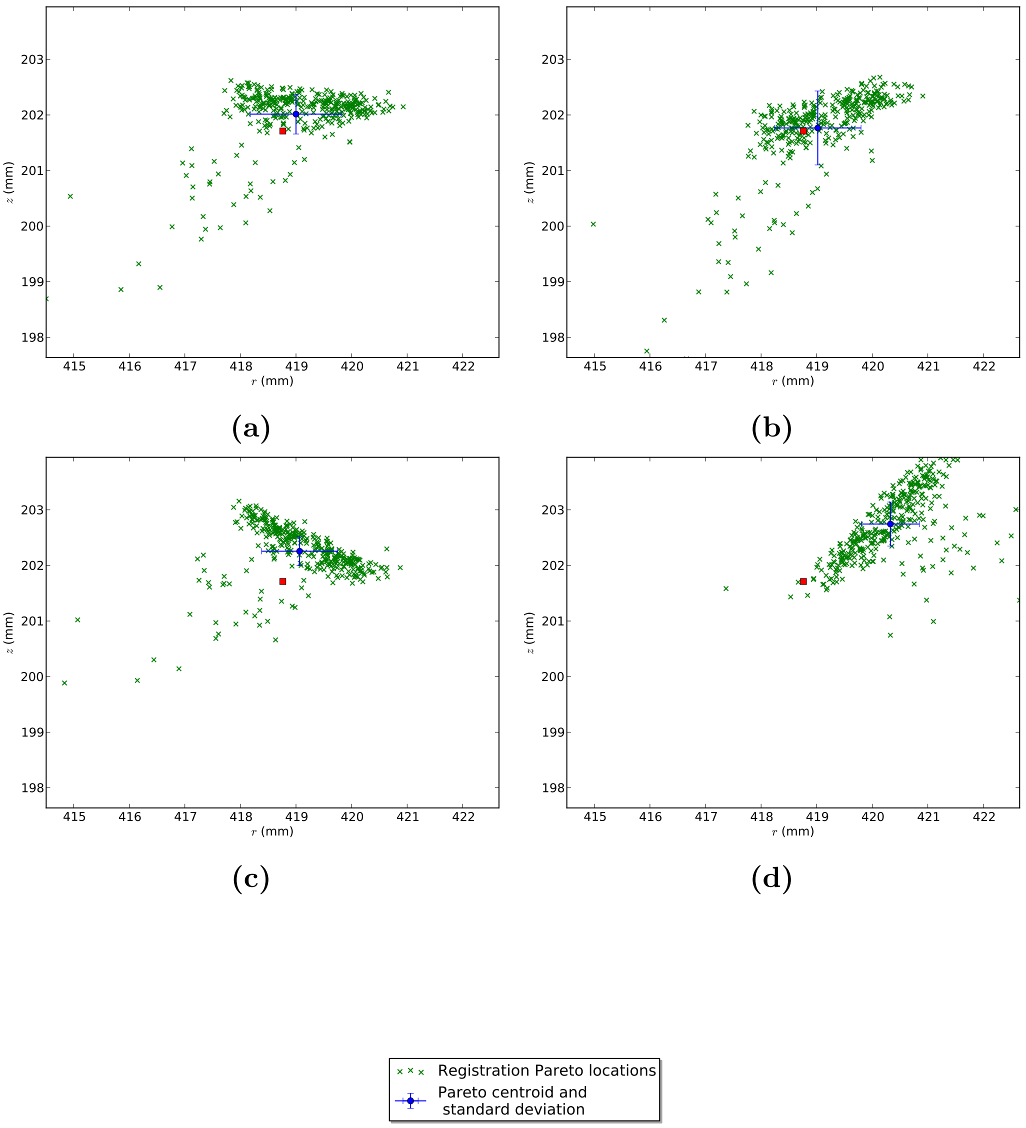


Figure 4.10: Registration mapped point locations for all moving scans viewing Indication 1. The circle and error bars show the Pareto means and standard deviations from each moving scan (from the lower group of data sets). The data are equivalent to those in the subplots of Figure 4.9, but plotted together here for comparison. All three QR scans are close to each other and have close to the same error with respect to the upper (fixed) scan. This is typical of scans covering the same surface but different angles.

Indication 3

All Pareto points from the four registered scans that view Indication 3 are plotted in Figure 4.11. Indication 3 has more points that are scattered farther from the main concentration of the point clouds. This is possibly due to the fact that Indication 3 is near near the maximum radius at the disc rim. Any error in a registration parameter that affects the angular orientation of the disc will have an amplified effect at greater radii. This also makes Indication 3 an example of how the Pareto integral approximation correctly weights all points, so that outliers do not have an undue effect on the estimated means and standard deviations.

All of the means and standard deviations are shown together in Figure 4.12. The KL scans all have excellent registration quality, aligning the indication to within 0.5 mm in both r and z. The ultrasound beam must propagate much farther from the PQ surface to reach Indication 3. Not surprisingly, the PQ 0° is off by more than the KL scans. It is still within 1.5 mm in both directions.



(a)

(b)

(c)

(d)

Figure 4.11: Registration mapped point locations for Indication 3 in scans (a) KL 0°, (b) KL +5°, (c) KL -5°, and (d) PQ 0°. Results are from the same registration as in Figure 4.9. The difference is that the point being mapped is Indication 3. These registration point distributions are good examples of the types of asymmetry that are possible. There is directional sensitivity in the location uncertainty, depending on the orientation of the surface and ultrasound beam in each scan. There are also differences in the spread of the distributions from different scans. In all scans, however, nearly all Pareto points fall within a 5 mm by 5 mm square. This fact will be useful later when setting an effective detection resolution.

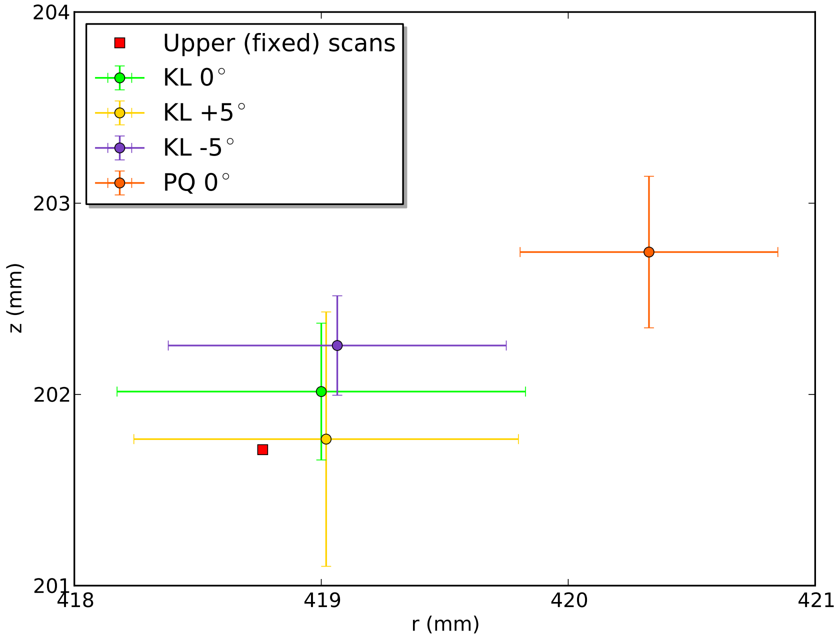


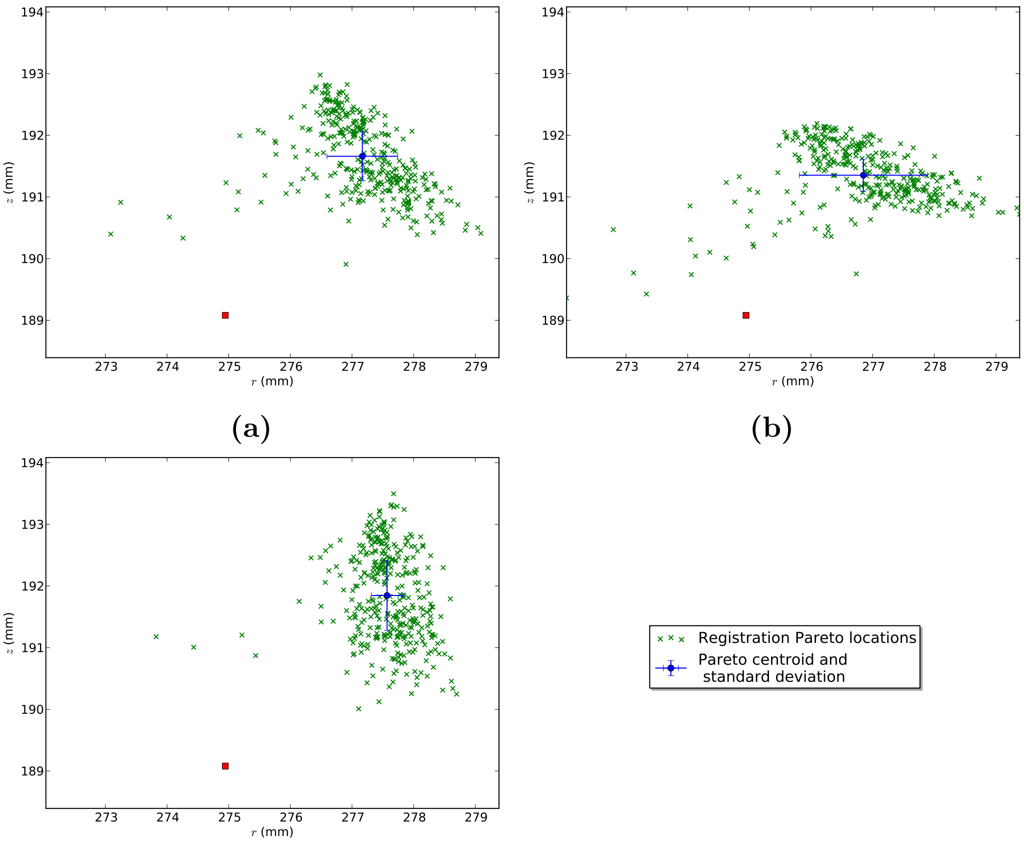
Figure 4.12: Registration mapped point locations for all moving scans viewing Indication 3. This is equivalent to the mean and standard deviations in Figure 4.11a-4.11d, but plotted together here. As in Figure 4.10, the three scans on the KL surface are very close to each other in mean location. The PQ 0° mean location is farther away from both the KL means and the upper (fixed) location. It also has more location uncertainty, as indicated by its standard deviations. The upper scan indication location is well within one standard deviation from all four lower moving scans.

Indication 5

Figure 4.13 has the mapped Pareto points from the three lower scans that view Indication 5. Indication 5 has the greatest post-registration discrepancy remaining between upper and lower scans. The distribution of the points in each scan has roughly the same spread as for Indications 1 and 3. However, the indication location in the upper scans is outside the convex hull of the Pareto points, in all three mapped scans.

Figure 4.14 shows all the means and standard deviations for the mapped lower scans, with the fixed upper scan location for reference. The distance between the two groups is approximately 3.5 mm. There are two reasons why this distance could be so large.

First, Indication 5 is located deep in the metal, far from any viewing surface. Any error in angle at the surface will be exacerbated by the propagation distance within



(a) **(b)**

(c)

Figure 4.13: Registration mapped point locations for Indication 5 in scans (a) RS 0°, (b) RS +5°, and (c) RS -5°. Indication 5 does not evidence as successful a match as Indications 1 and 3 in Figures 4.9 and 4.11, respectively. The upper (fixed) scan location does not lie within the point distributions of any of the three lower scans that move with the registration. This suggests that there is location uncertainty for Indication 5 that is not accounted for by this registration. A closer examination of the amplitudes from all of the scans covering this region reveals that Indication 5 might be large enough not to scatter as a point reflector, or that it might be two small indications close together.

the disc. This effect will have an impact whenever the region under investigation is distant from a surface. This effect has already been seen with the PQ 0° scan in Indication 3. However, this effect does not appear to be the greatest factor, since all three RS scans are very close to each other. This is true even though their relative angles are about ±20° inside the disc.

The second, and more likely dominant, effect is the relatively large size of Indication 5. From the A-scans of the RS surface, it is clear that Indication 5 stretches over a region spanning from the upper scan location to the RS mean locations in Figure 4.14. Despite these two effects, the registration aligns the two groups of scans to within 3 mm in both the r and z directions.

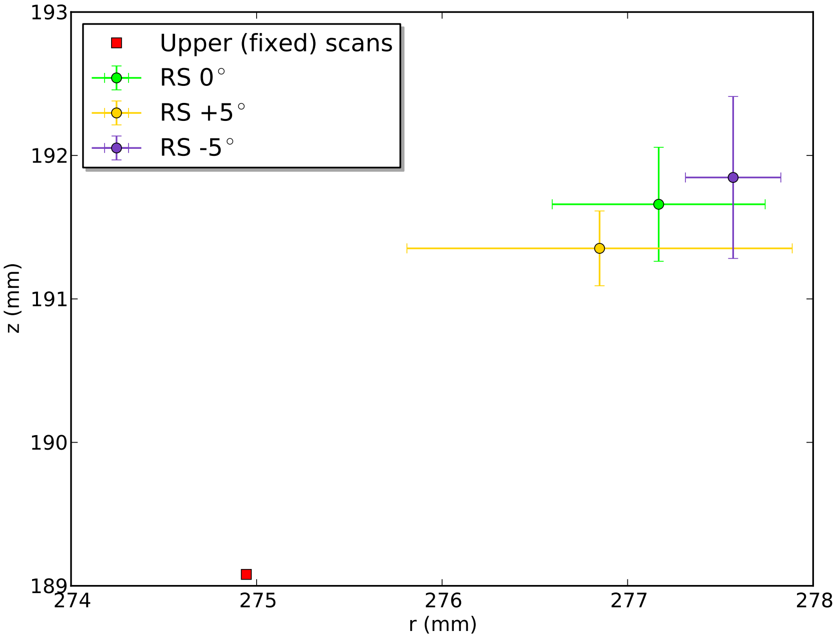


Figure 4.14: Registration mapped point locations for all moving scans viewing Indication 5. This is equivalent to the mean and standard deviations in Figure 4.13. The RS scans are within 2 mm in the r direction and 2.5 mm in the z direction. However, the standard deviations apparently underestimate the errors. This could be due to the finite size of Indication 5, such that ultrasound beams from different directions see a peak at different locations.

4.3.1 Temperature effect

Temperature is one nuisance variable that is not captured directly by registration. It is compensated for indirectly, by its effect on other registration parameters. The most sensitive parameter to changes in temperature is the speed of sound in water. This is an important registration parameter, because it affects not only propagation time but also the refraction angle. If the incidence angle is not exactly zero, uncertainty in water wave speed alone causes one point location to map to a 3D cloud of points.

The Seeded Defect Disc data for all of the lower surfaces were recorded in one day. The data for all of the upper surfaces were recorded on another day, several weeks later. In both cases, the scans were completed in less than two hours. The temperature should be very close to uniform for all scans within the upper and lower groups, but it is not likely to be the same between the two groups. During registration, the upper surfaces used the constant nominal longitudinal wave speed of 1475 meters per second. This corresponds to a water temperature of approximately 17.7° C [111]. Registration sought to compensate for the differences between registration features by varying the water wave speed (along with the other parameters) while mapping the lower scan points. In the Pareto points for the lower surfaces, the water wave speed ranged between 1452 and 1473 meters per second. This span of wave speeds corresponds to a temperature range of 11.2° C to 16.9° C. It is unlikely that the water temperature was as low as 11.2° C, but it might have changed by several degrees between scans. The fact that the registration optimizer did not narrow the range of apparent temperatures more tightly than 7.7° C shows that the registration metrics are not overly sensitive to small temperature changes. However, the initial registration uncertainty would still be reduced if the temperature were accurately measured or, preferably, controlled. One finding from these registration results is that, when moving to an automated NDT system, it is best to collect all data for a disc on the same day.

4.4 Summary

The results in this chapter demonstrated the ability of the automated NDT system to register turbine disc data, in preparation for detection. First, test cases were developed to evaluate the registration method. The test cases maintained a high degree of realism by using data from the Seeded Defect Disc inspection. The registration data were modified to use duplicated copies of one set of scans to register against each other. Random parameter values imposed distortions with targeted magnitudes. In this way, these registration test cases had a known solution to compare with the registration results.

The registration tests achieved successful registration to within 1.5 mm in both the r and z directions. These results were consistent, across all imposed distortions. There does not appear to be any dependence on the distance or magnitude of the distortions.

Results from the real registration for the full disc data were shown at three of the known indications in the Seeded Defect Disc. While finite size and directional reflectivity might cause some difference in apparent location, the indications provided a useful way to evaluate the registration results since they are visible in scans from both sides of the disc. Peak location comparisons for upper and lower sides showed that registration aligned the indications to within 3 mm in both r and z directions for the worst case (Indication 5). This is despite the possibility of a finite indication size interfering with the comparison. The other two indications register to within 2 mm.

These estimated limits on registration accuracy will tie directly into the work on detection, as will be explained further in Section 6.1. Accurate registration is an absolute requirement for detection; detection algorithms will not function if data are not aligned. However, there is always some residual uncertainty, even after the best possible registration is complete, due to physical limitations. The detection results in the following chapters overcome this problem by grouping the amplitude data into small regions. These regions, known as resels (resolution elements) must have a size at least as large as the registration accuracy for detection not to be impaired. Based on the results from this chapter, a resel size of 5 mm would be sufficiently large. The detection results in Chapter 6 will used this size for data resels, informed by the registration tests above.

The results, for both the known registration tests and the full disc registration, use the Pareto integral estimator from Chapter 3. The Pareto integral estimator in this case computes the mean and standard deviation of the full theoretical Pareto front. Like any application for the Pareto integral estimator, it uses only the discrete points from the optimization results. This application shows the estimator's usefulness for quantifying the location uncertainty remaining after registration. These results provide a demonstration of this newly developed capability.

Chapter 5

Detection approach

The ultimate goal of this work is to distinguish between parts with and without defects. Previous chapters have laid the groundwork by providing means to ensure the data are properly prepared and aligned, and to quantize the data into resels that account for the resolution limit of that alignment. This chapter provides definitions for the status quo detection method and the solution devised to improve it. A meaningful comparison between the two requires an analytical approach that encompasses both. This chapter describes a mathematical framework in which to define and evaluate detectors as hypothesis tests, following the work of the present author's collaborator, Brierley [19]. This chapter also describes a novel contribution to the computational implementation of Brierley's data fusion detector.

Section 5.1 begins the discussion by describing single-input detectors. This is an important class of detectors in its own right, with the status quo detection method as a significant example. Section 5.1.1 defines the status quo method that is currently in industrial use. The Local Empirical Noise detector is another single-input detector. It is proposed as an improvement upon the status quo detection method in Section 5.1.2. Single-input detectors also serve as a conceptual stepping stone to the multiple-input detectors that follow.

The concepts of data fusion and multiple-input detectors are introduced in Section 5.2. The discussion opens with a description of an approach that makes it possible to use single-input detectors, like the LEN, as building blocks in a multiple-input detector. The Discrete Fisher Detector (DFD) is defined in Section 5.2.3. The DFD is the multiple-input detector that will generate the final results in Chapter 6. Section 5.3 explores how the DFD compares to theoretical limits, using realistic data.

One of the contributions developed for this thesis to improve detection computation is laid out in Section 5.4. It is a numerical method that accelerates computation for the DFD by orders of magnitude. This improved technique makes data fusion practically feasible on the scale required for turbine disc inspection.

5.1 Single-input detectors

The intended approach for improving detection is to combine information from multiple, overlapping inputs. The analysis begins with single-input detectors, as they are a conceptually more simple prerequisite. Single-input detectors form the normalized inputs that are used by the multiple-input detectors described in Section 5.2. Additionally, the status quo inspection method may be properly described as a single-input detector. Quantification of the improvement that a multiple-input detector can make in the final results, therefore, requires an analysis approach that can bridge the gap between single- and multiple-input detectors. This section is therefore directly relevant to rest of this chapter as well as the detection results in the next chapter.

5.1.1 Amplitude Threshold Detector

The current inspection procedure uses an amplitude threshold as the detection criterion. Details of the procedural requirements would be laid out in each manufacturer's proprietary documentation. For the purposes of this analysis, the detection method may be understood as follows.

For each A-scan, the raw amplitude is first rectified with the absolute value function, in the same manner as described in Section 3.1.1. Any signal from the rectified A-scan amplitude that is greater than or equal to a predefined threshold triggers the inspection to halt. The inspector is required to investigate further when this happens, after which he or she makes a determination regarding the indication. The threshold crossing could be a false call, for example if the time gate is too close to a surface or corner echo, or if a rough finish spreads the echo signal over a wider time period. If the inspector cannot rule out the signal as a false call, the event is reported as a flaw and entered into the inspection record for that disc. The initial event that brings the region to the inspector's attention is when the amplitude crosses the predetermined threshold. This status quo inspection criterion is referred to in this work as the Amplitude Threshold Detector (ATD).

Performance evaluation for the ATD requires a histogram of amplitudes over the full disc. As will be further explained in Section 6.2 below, this histogram is necessary for estimation of two performance-related quantities: Probability Of Detection (POD) and Probability of False Alarm (PFA). For any hypothetical choice of amplitude threshold, the fraction of amplitudes over that threshold is the PFA, assuming no defects. POD estimation requires an assumption for the cumulative probability conditioned on the presence of a true flaw, $P(s)$, with signal amplitude, s, as a random variable [112–114]. POD is then equal to $1-P(s)$ with s set to the amplitude threshold.

5.1.2 Local Empirical Noise detector

The ATD in the previous section uses a constant threshold over the entire disc, even though noise magnitudes can vary greatly from one scan to another and even within a single A-scan. For example, Figure 5.1 shows enveloped amplitudes around the disc circumference, for two radial locations in the QR $+5°$ scan. It is clear from the plots in Figure 5.1 that the two radial locations have distinct noise distributions. Within each of the two plots, however, it is plausible that the amplitudes represent samples from a single distribution.

The Local Empirical Noise (LEN) detector is a single-input detector that addresses this issue by adapting to local noise levels. The LEN was developed by Brierley, contemporaneously with this thesis. For more information and a full description and analysis of the method, see Brierley's forthcoming thesis [19].

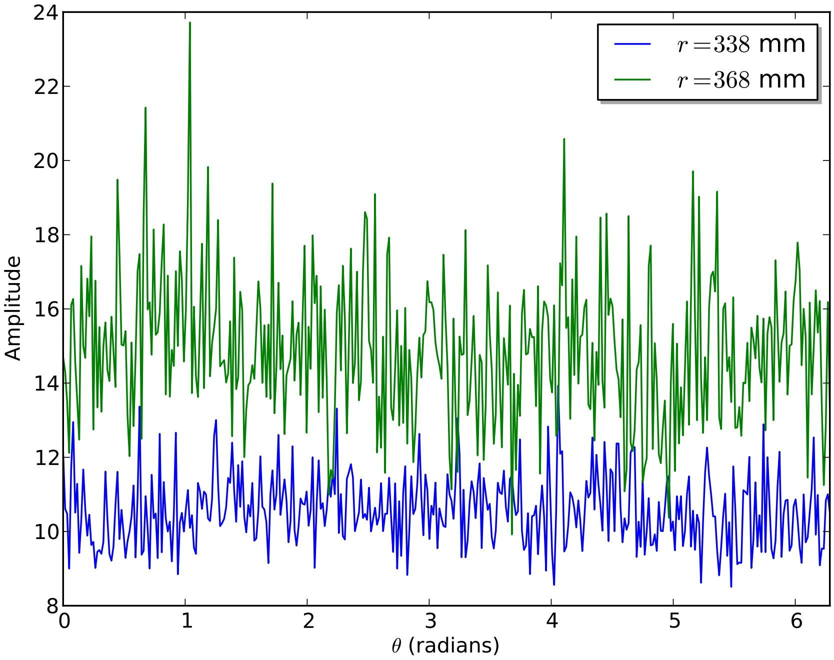


Figure 5.1: Enveloped amplitudes at two radial locations in the QR +5° scan. The two amplitude signals are taken from the same position in z and are plotted as functions of θ. The data are from the same scan and are separated radially by only 30 mm. However, the two signals are clearly sampled from two different noise distributions.

Some approaches to detector design and analysis assume an analytical distribution for the noise. It is important to have a very accurate noise model because of the way hypothesis tests are constructed. Every hypothesis test defines a Cumulative Density Function (CDF) [115]. This probability function characterizes the random behavior of a test statistic computed from test data, given that the null hypothesis is true. When used for testing, the sample data under test then compute a particular value for that test statistic. The value of the CDF at the test statistic just computed is called the p-value. In some hypothesis testing applications, the p-value is then compared to a predetermined threshold, known as the significance level.

In the context of NDT detection, the hypothesis testing framework operates by finding discrepancies between the noise model and the observed data. An erroneous noise distribution will therefore cause samples from the true noise distribution to appear as non-noise indication signals. This would drastically impair performance by increasing the PFA without improving the POD.

Previous work on ultrasound noise modeling for other applications used an analytical form, such as Rayleigh multiplicative noise [116] or the more general K and Rice distributions [117, 118]. Unfortunately, no histogram of noise samples from the turbine disc inspection data fit any analytical model very well. The noise distributions were also found to vary in different regions and with different viewing directions.

The LEN overcomes these challenges by constructing a noise distribution with minimal modeling assumptions. For most of the noise samples, it uses an empirical distribution. This distribution is a discrete probability mass function, taken directly from the lower ninety-five percent of the data. A hypothesis test based on such an empirical distribution satisfies its null hypothesis by construction. This is due to the fact that the p-value is determined by the same distribution that the samples come from.

If the distribution were entirely empirical, the biggest drawback would be that the p-value would have no sensitivity to Signal-to-Noise Ratio (SNR). The highest amplitude would have a probability of $1/N$, for N samples in the distribution, regardless of its magnitude relative to the rest of the distribution. LEN overcomes this problem by fitting the five percent highest-amplitude tail to an analytical curve, in this case a Pareto distribution [119]. The resulting discrete/continuous hybrid noise distribution balances null hypothesis fidelity against sensitivity in the tail.

In order to construct an empirical distribution, the LEN needs a set of amplitude samples. Clearly these samples must be reasonably assumed to be from the same distribution. Several physical and geometric effects cause very different amplitude distributions in the transducer signals. Within a single A-scan, different points in time have propagated through different thicknesses of material. The differences in propagation distance cause different magnitudes of attenuation, beam spread, and backscattering. Additionally, between scans or at different axial/radial positions in the same scan, the beam traverses regions of the disc with different nearby structural features. These differences in amplitude preclude agglomeration into an empirical distribution.

However, the geometry of the inspection procedure and the disc itself both exhibit axial symmetry. All of the physical effects described above that cause different noise amplitude levels are constant around a circle that is centered on, and in a plane orthogonal to, the axis of symmetry. It is therefore logical to conclude that the amplitude samples around each such circumferential ring are from a common distribution.

By formulating a noise distribution and comparing each amplitude to be tested to that distribution, the LEN detector operates as a hypothesis test. This has several advantages over an amplitude threshold detector. A distribution derived from the local samples makes the test satisfy the null hypothesis when there are no flaws, without depending on the accuracy of an assumed, analytical noise model. The hypothesis test formulation also has the convenient property that the p-value gives a quantitative expression of the flaw likelihood rather than a binary “yes” or “no” result. Finally, the biggest advantage of a hypothesis test detector in the context of this work is that the multiple-input detectors in Section 5.2 need p-values from independent tests as inputs. The p-values that the LEN generates meet this requirement.

5.2 Multiple-input consensus detectors

A consensus test is a hypothesis test that combines p-values from multiple other hypothesis tests into one resultant p-value. This makes consensus tests useful for data fusion. Every point inside the disc is viewed from different directions by multiple scans. A detector based on a consensus test could take single-input detectors from these scans and use their mutually supporting information to improve over-all detection performance.

Under the null hypothesis, H_0, each of the input p-values has a uniform probability distribution on the interval $[0, 1]$. The joint distribution for N input p-values is uniform on the hypercube $[0, 1]^N$ under H_0, since noise samples are assumed to be independent. Consensus tests compute a test statistic as a function of all the input p-values. The p-value of this test statistic value is the consensus p-value.

5.2.1 The OR consensus test

The OR test is a simple example of a consensus test. It is likely to be too inclusive for practical detection, but it is useful for comparison to others such as the DFD and Minimum P-Value described below.

Consider each input's p-value, p_i, to be the probability that the LEN test would accept the null hypothesis that there is no indication. The probability of an indication according to input i, then, is $1 - p_i$. A logical "or" operation on the inputs would detect an indication if at least one of the input tests rejected the null hypothesis of only noise. The p-value for the OR test, p_{OR}, is one minus the detection probability. Therefore,

$$p_{OR} = \prod_{i}^{N} p_i \,. \tag{5.1}$$

The OR test is an extreme case in permitting a single low input p-value to dominate all others in the output. One input test that rejects the null hypothesis will cause the OR test to reject with at least the same significance level. In physical terms, if a defect is called in any of the tests, it will be called in the combined test. This is desirable behavior if a signal in at least one test is clearly not explained by noise. It will create many false positives if all of the signal inputs are near the noise floor.

5.2.2 The AND consensus test

The AND test is another simple consensus test example that is useful for comparison purposes. AND detects an indication only if all input tests reject the null hypothesis. The p-value, p_{AND}, is

$$p_{AND} = 1 - \prod_{i}^{N} (1 - p_i) \,. \tag{5.2}$$

The AND test is an extreme case in requiring agreement among all input tests. Any single input that fails to reject the null hypothesis will cause the AND consensus test to do likewise, with at least the same p-value. For defect detection, the AND test requires all tests to call a flaw simultaneously. If any one of the tests cannot rule out noise, AND will fail to detect.

As mentioned above, neither AND nor OR is likely to be useful for practical data fusion. They are illustrative as bounding cases, since any other detector based on a consensus hypothesis test will have a p-value between the AND and OR p-values.

5.2.3 The Discrete Fisher Detector

The Fisher consensus test, or Fisher's Method, is a well-known consensus test [120]. The Fisher consensus test forms the basis of the main detection results presented in Chapter 6. Equation 5.3 defines the Fisher consensus test statistic, T_{Fisher}, in which p_i are the input p-values.

$$T_{Fisher} = -2 \sum_{i}^{N} \ln p_i \tag{5.3}$$

When the p_i p-values correspond to continuous random variables, T_{Fisher} has a chi-squared distribution [121]. The LEN is the single-input detector of choice for the application in Chapter 6, but it has a hybrid discrete/continuous noise distribution. The Fisher consensus distribution uses convolutions of the probability densities for its inputs. The discrete component of the LEN probability densities contain Dirac delta functions at irregular intervals. When these delta functions are convolved with each other, the result also contains irregularly spaced delta functions. The resulting consensus test distribution therefore is not continuous and does not converge to a chi-squared distribution.

Brierley has derived a discrete analog of the Fisher Consensus Test, called the Discrete Fisher Consensus Test (DFCT) [19], that properly handles the discrete nature of the LEN input distributions. The DFCT distribution is not chi-squared, and it is not conveniently expressed in analytic form. P-values from the DFCT must be evaluated numerically with the Monte Carlo method.

The Discrete Fisher Detector (DFD) is a data fusion detector based on the DFCT. It selects and prioritizes indications by increasing order of the DFCT p-value, the inputs to which are LEN p-values.

5.2.4 The Minimum P-Value consensus test

The Minimum P-Value (MPV) consensus test is useful for comparing the DFD to the status quo inspection method. Recall that, in any hypothesis test, a smaller p-value indicates a greater confidence in rejecting the null hypothesis. For a detector, the null hypothesis is that there is no flaw and the measured signal is only noise. A small p-value, therefore, indicates a higher likelihood of a flaw.

The MPV p-value, p_{MPV}, is simply the minimum of all input p-values.

$$p_{MPV} = \min_i p_i \tag{5.4}$$

A detector based on the MPV test takes the output of whichever input detector is most likely to flag an indication and ignores all other inputs. The status quo amplitude threshold (see Section 5.1.1) effectively takes the output of whichever input has the highest amplitude. The amplitude threshold is not a hypothesis test, so it is not expressible in terms of input p-values.

MPV is the hypothesis test most similar to setting an amplitude threshold. The similarity in operation makes MPV useful for comparing other consensus tests to something close to the status quo. In fact, MPV is an improvement over the status quo in the sense that the input LEN detectors use a local noise model that can adapt to varying signal-to-noise ratio. No detector based on a fixed amplitude threshold has the capability to adapt to varying SNR.

5.3 Consensus test comparison

This section explores trends for the consensus tests defined in Section 5.2 with respect to SNR, agreement or disagreement between inputs, and the degree of coverage. The examples use the same noise distribution for all inputs. This distribution is taken from a circumferential ring from real scan data. A histogram of this distribution is shown in Figure 5.2. P-values are computed for each of the four consensus tests as the input amplitudes vary, thereby plotting p-value versus SNR.

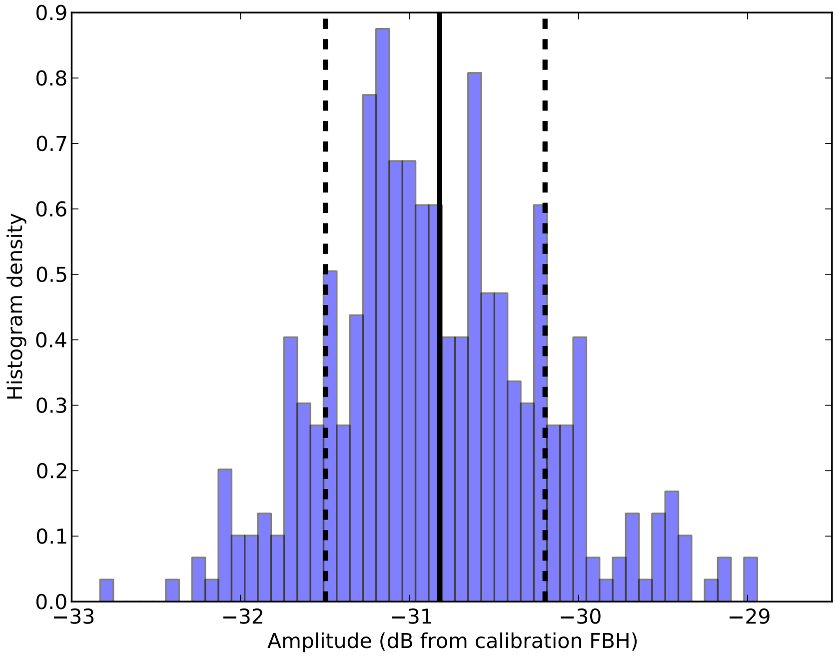


Figure 5.2: Noise distribution used by the consensus test comparisons. The amplitude samples are taken from the QR +5° scan, at the r, z location of Indication 1. These are the same samples shown as a function of θ in Figure 5.1, at $r = 338$ mm. The amplitudes are plotted in decibels, relative to the calibration block Flat-Bottomed Hole. The solid vertical line marks the mean amplitude. The dashed lines show one standard deviation above and below the mean.

Section 5.3.1 will show results for the case when all inputs to the consensus tests are equal and vary together. In Section 5.3.2, only one input has a variable amplitude; all other inputs are fixed at a low amplitude, below the mean of the noise distribution. Together these two sections define the extremes of mutual confirmation versus disagreement between inputs.

Each of the following subsections shows results for three inputs and for ten inputs. These choices for number of inputs to combine are significant to the target application. Three and ten are the minimum and maximum number of overlapping scans in the Seeded Defect Disc, as Section 6.1 will show.

5.3.1 Multiple inputs with equivalent SNR

This section analyzes the case in which multiple inputs share a common amplitude and p-value. The results show the effect of precise agreement between multiple inputs and how corroboration of the evidence can improve confidence in a signal, even when the SNRs of the individual inputs are relatively low.

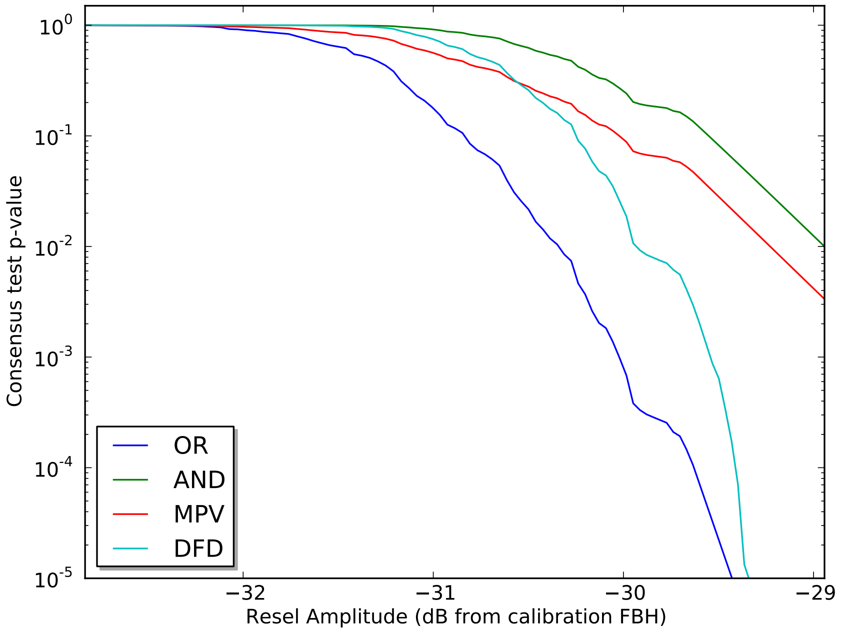


Figure 5.3: Comparison between multiple-input consensus tests, with all three inputs equal to each other. The AND and OR logical operator consensus tests form bounds for any other rational consensus test. MPV is the consensus/hypothesis test closest in operation to the status quo Amplitude Threshold Detector (ATD) method. Discrete Fisher Detector (DFD) is the detector proposed as a solution in this work. All four detectors have a lower p-value as the amplitudes increase. Low p-values indicate greater likelihood of a flaw. DFD is closer to AND at lower amplitudes and approaches OR at higher amplitudes. MPV follows the trend of AND, with roughly a factor of three reduction in p-value.

Figure 5.3 shows p-values from the OR, AND, MPV, and DFD consensus tests. Amplitudes are measured relative to the signal amplitude from a Flat-Bottomed Hole (FBH) in the calibration block as described in Section 2.1. Recall that the p-value returned from the consensus tests is the likelihood of such a result by random chance, given that there is only noise and no flaw to detect. A lower p-value therefore indicates greater certainty of flaw detection. The p-values from all tests converge asymptotically to unity when the three equivalent samples decrease to an amplitude of zero. All consensus tests decrease in p-value as the input amplitudes increase, but they do so at different rates.

The AND and OR consensus tests define upper and lower bounds, respectively, for any other consensus detector defined as a hypothesis test. MPV has the same trend as the AND test, with a shift down in p-value. This follows from Equations 5.2 and

5.4. For this case and linearizing Equations 5.2 for small input p-values,

$$p_{AND} \approx N p_{MPV} \,. \tag{5.5}$$

The factor of three difference in Figure 5.3 is visible as a vertical shift between the log-log plots.

The DFD test tracks the AND test at low amplitudes. This fact is significant because it gives DFD good rejection at lower SNR. That is, values that are noisy require all inputs that cover the region in question to agree before a flaw is indicated. The DFD then transitions to follow the behavior of the OR test at higher amplitudes. This would allow any single input with significantly higher amplitude to trigger detection.

All of the consensus tests trace irregular curves up to an amplitude between -30 dB and -29 dB. The irregular shapes are due to discretization in the LEN noise distribution. Above the ninety-five percent amplitude quantile, the noise distribution follows a continuous Pareto distribution of the form

$$p_{LEN} = \left(\frac{\mu}{A}\right)^{\sigma} \tag{5.6}$$

where p_{LEN} is the LEN p-value and A is the amplitude. The Pareto distribution parameters μ and σ are obtained by curve fit to the data above the ninety-five percent quantile.

Equation 5.6 facilitates trend analysis for consensus test behavior at higher amplitudes and lower p-values. The OR p-value from Equation 5.1 is applied to this case and expressed as a function of logarithmic quantities in Equation 5.7, for more convenient comparison to the log-log plots.

$$\log p_{OR} = -\sigma N \left(\log A - \log \mu\right) \tag{5.7}$$

The p_{OR} is linear at high amplitudes on the log-log plot in Figure 5.3, with a slope of $-\sigma N$.

The DFD p-value does not have an analytical form and must be computed numerically because of the discrete part of the LEN distributions. However, if the inputs can be approximated as continuous, the DFD distribution will be close to chi-squared. The p-value is then

$$p_{DFD} = \frac{\Gamma\left(\frac{N}{2}, \frac{T_{Fisher}}{2}\right)}{\Gamma\left(\frac{N}{2}\right)} \tag{5.8}$$

where $\Gamma(\cdot,\cdot)$ is the upper incomplete gamma function, and $\Gamma(\cdot)$ is the ordinary gamma function. The slope of the logarithms is

$$\frac{\partial \log p_{DFD}}{\partial \log A} = -\frac{(\sigma N)^{\frac{N}{2}} \left(\log \frac{A}{\mu}\right)^{\frac{N}{2}-1} \left(\frac{\mu}{A}\right)^{\sigma N}}{\Gamma\left(\frac{N}{2}, \sigma N \log \frac{A}{\mu}\right)} . \tag{5.9}$$

As A increases, the slope of $\log p_{DFD}$ converges to match the slope of $\log p_{OR}$ from Equation 5.7.

$$\lim_{A \to \infty} \frac{\partial \log p_{DFD}}{\partial \log A} = -\sigma N \tag{5.10}$$

Equation 5.10 confirms the observation from Figure 5.3 that DFD p-values parallel OR at higher SNR.

Figure 5.4 shows the four consensus tests results when there are ten identical inputs rather than three. The higher potential for mutual confirmation causes OR and AND to diverge more rapidly as the amplitude increases. In both cases the products in Equations 5.1 and 5.2 now have ten multiplicands instead of three.

MPV is unchanged, since it only depends on the minimum p-value which is unaffected by the number of inputs. MPV continues to follow the trend of AND, now with a p-value lower by a factor of ten. DFD also continues to exhibit the desirable property that it transitions from AND-like at amplitudes lower in the noise distribution to OR-like behavior as amplitudes increase. With ten inputs, this transition is much more abrupt than in the three input case.

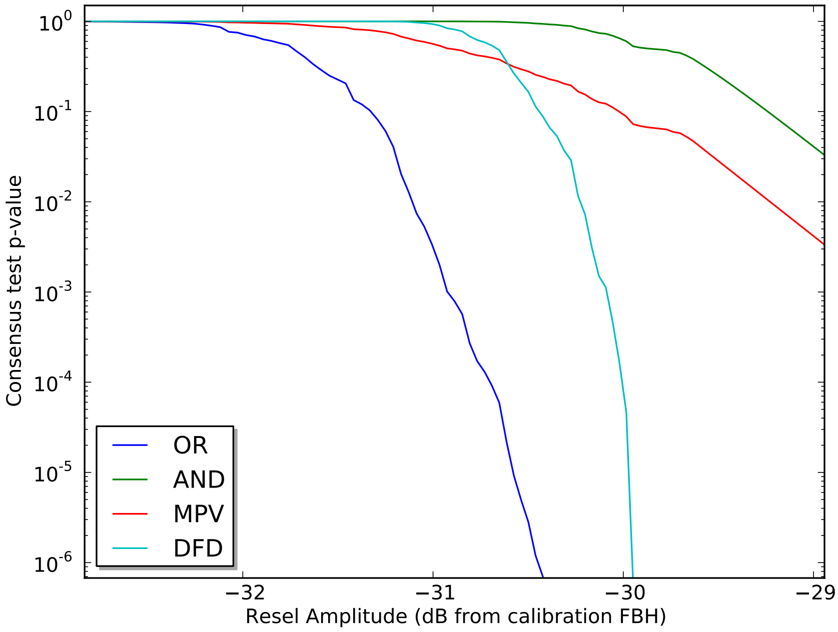


Figure 5.4: Comparison between multiple-input consensus tests, with all ten inputs equal to each other. All four consensus tests have p-value curves qualitatively similar to the three input case in Figure 5.3. AND p-values decrease more slowly with increasing amplitude, and OR p-values decrease more rapidly with ten inputs. MPV results are unchanged by the additional inputs and continue to follow AND with a reduction in p-value. DFD has the same tendency to transition from AND-like behavior at low amplitudes to OR-like behavior at higher amplitudes.

5.3.2 Multiple inputs with only one non-noise signal

Section 5.3.1 considered an extreme condition of agreement and mutually reinforcing information between the inputs. This section concerns the opposite extreme case of contradictory information. Its purpose is to test the conditions under which the consensus tests will indicate a defect when only one scan can detect it. If only the detecting scan were considered, detection would be assured by definition. The presence of additional scans in which the signal is below the mean noise level will make detection less likely with a consensus test. The higher the SNR in the one scan where the defect signal is apparent, the more closely the consensus test should approach the OR case.

As before in Section 5.3.1, all inputs are drawn from the same distribution. The amplitude from one input varies in the plots. The other input amplitudes are fixed at a value equal to one standard deviation down from the mean of the empirical noise distribution, illustrated by the first dashed vertical line in Figure 5.2

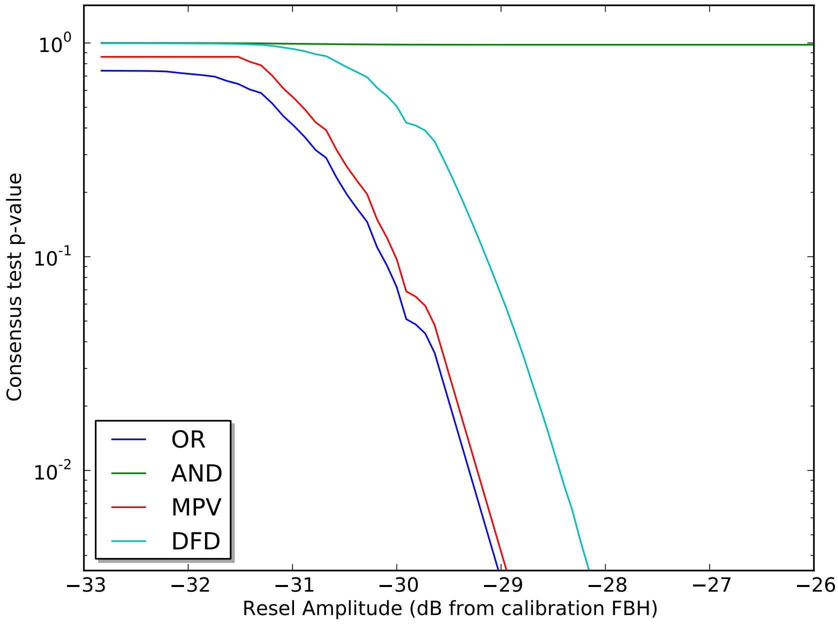


Figure 5.5: Comparison between multiple-input consensus tests, in which only one of three inputs detects a flaw. The data simulate a case in which a point inside the disc is covered by three scans, as with Figure 5.3. Here, however, only one input has a signal that varies while the other two inputs are fixed at an amplitude of the mean minus one standard deviation. AND has a low probability of detection, indicated by its p-value near unity for all amplitudes of the variable input. OR and MPV are very close, producing nearly identical results but with a shift in amplitude dB. DFD also follows the trend of OR, with a greater shift in amplitude. An amplitude about 1 dB higher in the detecting scan is required for DFD to reach the same likelihood as OR. This is due to the two noise-only scans exerting a contradictory effect in DFD contrary to the one scan that does detect a flaw.

The four consensus test p-values are plotted in Figure 5.5, as functions of the variable-input amplitude. AND maintains a p-value very close to unity, since the two inputs at the mean of the noise distribution will always impede detection. DFD follows the same trend as OR and MPV, but with a higher p-value. Alternatively, the relationship between DFD and OR can be interpreted as a logarithmic shift in amplitude. That is, DFD is qualitatively similar to OR under these conditions, except that it discounts the amplitude of the detecting scan by approximately 1 dB.

At higher amplitudes where the LEN noise distribution uses the Pareto tail, p_{OR} once again becomes a straight line on the log-log plot.

$$\log p_{OR} = -\sigma\left(\log A - \log\mu\right) + (N-1)\log p_{noise} \qquad (5.11)$$

There is now a dependence on the LEN input p-value of the noise inputs, p_{noise}, but the slope is determined solely by the Pareto shape parameter, σ.

Equation 5.8 is still valid for the DFD p-value under the continuous input p-values assumption. The difference is in the value of T_{Fisher}, which makes the logarithmic slope

$$\frac{\partial \log p_{DFD}}{\partial \log A} = -\frac{\sigma \left(\sigma \log \frac{A}{\mu} + (1-N) \log p_{noise} \right) \left(\frac{\mu}{A} \right)^{\sigma}}{\Gamma \left(\frac{N}{2}, \sigma \log \frac{A}{\mu} + (1-N) \log p_{noise} \right)} . \tag{5.12}$$

As before, the slopes of the logarithms are the same for p_{DFD} and p_{OR} in the asymptotic limit.

$$\lim_{A \to \infty} \frac{\partial \log p_{DFD}}{\partial \log A} = -\sigma \tag{5.13}$$

A shift of approximately 1 dB, or a factor of approximately 1.122 in increased amplitude, is required to achieve a given p-value with DFD versus the lower-bound OR test. In other words, the negative results from the two noise-only inputs causes DFD to require additional certainty from the one input that does detect a signal. The 12% required increase is hardly burdensome, given that DFD is overcoming the influence of two other inputs that are quite low in the noise distribution.

Figure 5.6 shows results from the same analysis as in Figure 5.5, but with coverage increased to ten. Nine inputs are fixed at one standard deviation below the mean of the noise, with only the tenth input varying in amplitude. The ten-input case has the same overall behavior as with three. AND is always very close to unity, since the nine noise inputs overwhelm the tenth signal input. The OR test gives somewhat lower consensus p-values due to the exponential effect of the seven additional noise inputs. The minimum of the input p-values is unaffected by the number of noise inputs, so MPV is identical to the three-input case. DFD again exhibits the desirable property of staying close to the AND upper bound when all inputs are at a low p-value, and following the OR behavior when at least one input has a higher amplitude. With nine noise-only inputs, DFD requires a 3.5 dB shift, or a factor of approximately 1.5 increase in amplitude, to achieve the same level of detection certainty as the lower bound OR test.

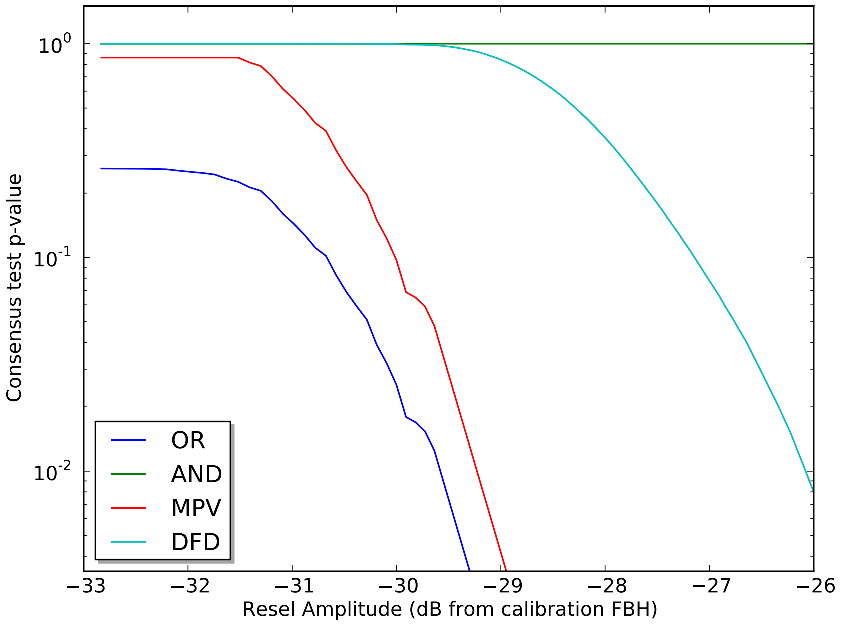


Figure 5.6: Comparison between multiple-input consensus tests, in which only one of ten inputs detects a flaw. The other nine inputs are held constant at one standard deviation below the mean noise amplitude. As with the three-input case in Figure 5.5, the AND test is dominated by the low-amplitude inputs and maintains a nearly constant p-value close to unity. The OR test p-values are lower by a constant factor, but MPV is the same with ten inputs as with three. DFD again transitions from AND at low amplitudes to an amplitude-scaled OR at higher amplitudes. In this case, the DFD p-value curve is shifted by approximately 3.5 dB, or a 50% increase in amplitude.

5.3.3 Consensus test comparison summary

The AND and OR consensus tests define useful bounds with which to analyze consensus tests. MPV is the consensus test based on hypothesis testing that is most similar to the status quo inspection method. MPV follows AND when inputs agree, and follows OR when only one input contains a detectable signal. This general MPV behavior does not depend on SNR. MPV therefore exhibits no ability to adapt its results when one or more inputs have a higher SNR than the others.

DFD, on the other hand, has results similar to AND when all inputs have a low SNR, whether or not they agree. DFD trends toward maximum detectability, as indicated by the OR curve, when at least one input has a higher SNR. Even when inputs disagree, a defect is still detected by a single input with high SNR. A higher amplitude by up to a few decibels is required to confirm the defect in light of the other inputs' negative result. This keeps the PFA for DFD down while still allowing detection.

5.4 Monte Carlo Importance Sampling for the DFD

Having shown that the DFD is an appropriate consensus test in this application, we now address its efficient calculation. Without some form of numerical acceleration, the p-value estimation for the DFD is so computationally expensive that it is impossible to use in a realistic inspection scenario. The following describes an algorithm that greatly accelerates these calculations, achieving speedups by orders of magnitude. This algorithm was devised as part of the work for this thesis, to enable the results computation on the full-disc data in Chapter 6. It could also be used with any other application of the DFD.

For the reasons described in Section 5.2.3, DFD requires Monte Carlo sampling to estimate its p-value. First the DFD software computes the consensus test statistic, T_{Fisher}, for the data under evaluation using Equation 5.3. It then takes a number of random samples, indexed by j, for the input p-values, $(p_i)_j$. Each sample is from a uniform distribution on the interval $[0, 1]$. These Monte Carlo samples approximate the null hypothesis p-value distribution, which is uniform by definition. Equation 5.3 then computes a test statistic value for each sample, $(T_{Fisher})_j$. The fraction of test statistic samples that are greater than T_{Fisher} is the estimated DFD p-value.

Computation time is very problematic for this Monte Carlo approach. The regions within the $[0, 1]^N$ input p-value space with a small consensus p-value are the ones that are of greatest interest for inspection, due to their higher likelihood of containing a flaw. Accuracy is therefore most important in these small p-value regions to distinguish between true defects and false calls. However, it is precisely when the p-value is small that Monte Carlo statistical error is greatest.

The probability of sampling a test statistic higher than T_{Fisher} is equal to the p-value for T_{Fisher}. An accurate estimate would require orders of magnitude more sample points than the inverse of the p-value to be computed. In several cases observed with real data, the p-values for known indications are smaller than 10^{-6}. Accurate estimation would require hundreds of millions of Monte Carlo samples to inspect a single point within a disc. This makes the approach prohibitively costly for practical applications. Furthermore, most of the computation time is wasteful because the vast majority of sample points convey very little information about the p-value.

5.4.1 Importance Sampling

Importance Sampling (IS) is a variance reduction technique for accelerating convergence of a Monte Carlo (MC) estimate [122–124]. IS works by sampling Monte Carlo points more densely in some regions of the integration domain, rather than with the default sampling distribution [125, 126]. For example, for a random variable x with a probability density p_x, the ordinary MC estimate for the expected value of $g(x)$ would be

$$E_{p_x}[g(x)] = \int g(x) p_x dx \approx \frac{1}{N_{MC}} \sum_{j}^{N_{MC}} g(x_j) \tag{5.14}$$

where the Monte Carlo points x_j are sampled from the p_x probability density. The sum in Equation 5.14 converges to the integral as the number of sample points, N_{MC}, increases.

If $g(x)$ is close to zero over part of the integration domain, MC sample points in that region will contribute comparatively little to the accuracy of the sum in Equation 5.14. These points represent wasted computational effort that would be better spent in regions where $g(x)$ has greater magnitude.

The idea behind Importance Sampling is to draw MC samples from a different probability measure. An appropriately weighted sum of these points recovers the original expected value.

$$E_{h_x}[p_x g(x)/h_x] = \int \left(\frac{p_x}{h_x}\right) g(x) h_x dx \approx \frac{1}{N_{MC}} \sum_{j}^{N_{MC}} w_j g(x_j) \tag{5.15}$$

where x_j are sampled from the IS density, h_x. The weights, w_j, are computed as the ratio of the two probability densities.

$$w_j = \frac{p_{x_j}}{h_{x_j}} \tag{5.16}$$

The key to accelerated convergence is to choose a distribution with a density appropriate to the integrand to be estimated. Ideally,

$$h_x \approx \frac{g(x)p_x}{\int g(x)p_x dx} \tag{5.17}$$

so that the result would theoretically converge with a single sample. In practice, equality in Equation 5.17 is not possible since the denominator is the quantity to be estimated. Equation 5.17 serves as an objective for designing an IS mapping to decrease wasted computation on unimportant sample points.

5.4.2 DFD Importance Sampling

In the context of Monte Carlo estimation for the DFD, the random variable x is a vector of input p-values. The default density, p_x, is uniform over the multidimensional x domain. The integrand, $g(x)$, is an indicator function that is zero for all values of x for which $T_{Fisher}(x)$ is greater than the $T_{Fisher}(p_i)$ already computed for the data being evaluated. It is unity when $T_{Fisher}(x)$ is less than or equal to $T_{Fisher}(p_i)$. The integral of $g(x)$ is the DFD p-value.

An optimal IS estimation for the DFD p-value would have an equal probability of producing $T_{Fisher}(x)$ above or below $T_{Fisher}(p_i)$. This would make each MC sample equally informative. The chosen approach for DFD importance sampling defines a new sample joint density, h_x, equal to the product of marginal densities. Each marginal is a function of only one input p-value. This treats each dimension separately with regard to mapping. The weight will then be the product of marginal weights.

There are two additional concerns for implementation that affect the choice of marginal IS densities. First, an important requirement for the IS density h_x is that its cumulative probability function must be invertible. This is due to the manner in which the computer generates pseudo-random numbers from arbitrary distributions. To draw a sample from a non-uniform distribution, the first step is to draw samples from a distribution that is uniform over $[0, 1]$. The uniform samples then pass as arguments to the inverse of the cumulative distribution function, $H^{-1}(x)$, where

$$H(x) = \int_{-\infty}^{x} h_{x'} dx'. \tag{5.18}$$

The resulting points are distributed according to the density h_x. Invertibility implies that $H(x)$ is a continuous function of x.

The second important consideration for the marginal densities concerns sensitivity in low probability regions. The IS sample density h_x should be small in less important regions, but it must not be too small or else the weight in Equation 5.16 will be very large. This magnifies the effect of statistical error in these low-density regions. The final result could then become sensitive to one or a few sample points in the uninteresting part of the domain. It is therefore important to keep h_x above a minimum value.

This work proposes the following DFD Importance Sampling mapping function for marginal densities:

$$\tilde{p}_i = \frac{m_i^2}{1 - 2m_i} \left(\left(\frac{1 - m_i}{m_i} \right)^{2p_i} - 1 \right) \tag{5.19}$$

In Equation 5.19, p_i is a MC point sampled from a uniform distribution, and $\tilde{p}_i$ is the mapped point, which would now be a sample from the IS distribution. The dimensions of the parameter domain are indexed by i. Each individual sample point passes through the mapping function in Equation 5.19 for each dimension separately.

The median of the distribution, $\tilde{p}_i(p_i = 1/2)$, is m_i. The user can set the median parameter for each i equal to the corresponding input p-value. That makes each MC sample equally likely to be above or below each input p-value, thus maximizing the importance of each sample.

The marginal weights, $\tilde{w}_i$, are

$$\tilde{w}_i = 2 \left(\tilde{p} + \frac{m_i^2}{1 - 2m_i} \right) \ln \left(\frac{1 - m_i}{m_i} \right) . \tag{5.20}$$

The weights are only guaranteed to be positive when $m \leq \frac{1}{2}$, but this is not really a limitation. Accuracy is most important when an input p-value, and therefore m, is very small. In the limit as m approaches $\frac{1}{2}$, the importance sampling distribution approaches uniform; i.e., effectively there is no importance sampling. The marginal mapping and weights should only be used when $m < \frac{1}{2}$. For $m \geq \frac{1}{2}$, the default uniform sampling with $\tilde{w}_i = 1$ should apply. This selective application of importance sampling can proceed for each dimension i individually, without affecting the computation of the final weight coefficient.

The final weight coefficient for the sample point, $\tilde{w}$, is the product of the marginal weights.

$$\tilde{w} = \prod_i \tilde{w}_i \tag{5.21}$$

The inverse cumulative distribution function, used to compute the mapping from uniform to importance sampling in Equation 5.19, is indeed continuous and invertible. This satisfies the first of the two IS criteria given above. The second, limiting the magnitude of weights, is satisfied for all practical purposes. The marginal weights in Equation 5.20 do increase without bound as m_i approaches zero, but only as $-\ln m_i$. For very small input p-values, $\tilde{w}_i$ therefore scales linearly with the p-value order of magnitude. Any inspection datum with a p-value which is many orders of magnitude less than unity will be flagged as an indication, making statistical error a very minor concern.

Sobol sequences

The final important detail for accelerating Monte Carlo convergence is the use of quasi-random sequences rather than random numbers [127]. Truly random samples from a uniform distribution will randomly form clumps and gaps in the domain. Sobol sequences are an example of a low-discrepancy sequence that optimally spreads points to avoid clumps or gaps with any finite number of samples [128].

Like random or pseudorandom numbers, Sobol sequences do not make undesirable lattice or regular grid patterns in the sample domain. For a given initial sample in the Sobol sequence, the samples that follow are deterministic rather than random.

Note, however, that the same is true of pseudorandom numbers that result from nearly every random number implementation for scientific computing.

The multidimensional Sobol sequences in this work come from an implementation based on the ACM TOMS Algorithm 659 [129]. These uniformly distributed quasi-random sequences are the p_i inputs to Equation 5.19.

5.5 Summary

The current state of practice in turbine disc inspection uses an amplitude threshold for detection. The Amplitude Threshold Detector (ATD) uses the same threshold for every A-scan and over the full propagation distance of each A-scan. The Local Empirical Noise (LEN) detector improves upon the ATD by rating amplitudes relative to a data-driven, adaptive noise model. It is therefore adaptive to local noise levels. Both ATD and LEN are single-input detectors and do not take advantage of the high degree of overlapping coverage from multiple scans.

A multiple-input consensus detector operates on the output from several single-input detectors to reach a single consensus result. A consensus detector can achieve better performance than its single-input operands in both POD and PFA simultaneously. The Discrete Fisher Detector (DFD) was recently developed for a similar ultrasound NDT application. The DFD can use LEN detectors as inputs, to take advantage of their noise adaptability. Results with simulated data show that the DFD adapts automatically to the quality of the information available from all of its inputs. The DFD operates much like a logical AND when all scans covering an inspected region are small amplitude relative to surrounding noise. The DFD behavior approaches a logical OR when at least one scan has a higher SNR. In contrast, amplitude threshold detection ignores any possibility of mutual confirmation and does not adapt to varying noise levels.

This chapter shows the compelling theoretical advantages that DFD has over the ATD currently in industrial use. However, the DFD algorithm requires a very computationally demanding Monte Carlo estimation. Section 5.4 describes for the first time an Importance Sampling transformation designed specifically for the DFD.

Computation time would be orders of magnitude slower and not practically feasible for realistic inspection scenarios without this transformation method. The Monte Carlo acceleration method in Section 5.4 is presented as an innovative contribution to the DFD technique.

Chapter 6

Detection results

The results in this chapter are the culmination of the automated NDT approach developed for this thesis. These are the ultimate flaw detection results that will be most important and most visible to the users and customers of NDT.

The results that follow give a comparison between the detection methods introduced in Chapter 5. These are the status quo Amplitude Threshold Detector (ATD), the Local Empirical Noise (LEN) detector that adapts to local noise levels within a given scan, and the Data Fusion Detector (DFD) that fuses the LEN results into a consensus verdict. The discussion opens with a simple scenario that is close to the status quo inspection, with data from a single scan examined in isolation. It then builds from that foundation to the full case, which uses all data available for every point in the disc. The evaluation uses real data at every stage, showing results for the five indications already known to exist in the Seeded Defect Disc that were first introduced in Section 2.1.

Each point in the disc is visible to multiple scans from different directions. The number of overlapping scans varies over the cross section. A data-fusion detection algorithm must be able to account for this coverage variability. It should take advantage of high coverage when it is available. It should also give meaningful and quantitatively compatible results when coverage is lower. An analysis of the degree of coverage is found in Section 6.1.

Section 6.2 defines the metrics for comparing the detectors. These metrics require an unknown distribution for the signal amplitudes. Section 6.2 also describes the assumption used to approximate this distribution. The detector's ability to cope with varying coverage allows for a natural progression in the discussion from that point. The analysis begins with one scan, examining detection results on the first known indication in Section 6.3.

The analysis then moves on to include one additional, overlapping scan, to show the improvement from the combined information in Section 6.4. Finally Section 6.5 will complete the development of data fusion detection with results that use all scan data. Section 6.5 concludes with the same full-coverage detection analysis for the other four known indications.

The results in Section 6.5 will show the ability of the data-fusion detector to identify the known defects with very high certainty. Section 6.6 expands on those results by testing the robustness of the detection algorithms for smaller or less reflective indications. Detection results are recomputed as the known indication amplitudes are scaled down to the noise floor, providing useful information about the limit for detectability.

One of the most compelling outcomes from this automated detection capability is the possibility of finding new indications that are not currently detectable. Section 6.7 shows evidence for possible indications in the Seeded Defect Disc that were previously unknown to the manufacturer.

6.1 Coverage

The internal region of the disc is divided into over 110,000 small regions for the purpose of detection. Each resel is a small cube-like region approximately 5 mm on each side. The term resel means “resolution element”, in an abbreviation analogous to the pixels (“picture elements”) that compose an image on a computer or television screen.

The resel is a basic unit of independently resolvable volume. Each resel therefore represents an independent inspection opportunity. Their size is tied to the physical resolution and registration accuracy, as described in Chapter 4.

The resel amplitude is the maximum amplitude contained within the resel. The reduction from data point amplitudes to resel amplitude takes the maximum, so that sharp peaks are not smoothed or smeared out. Other operations, such as the mean, would diminish a sharp peak. By using the maximum, the resel looses information about the shape of a peak that it might contain, but it does not loose information about the peak amplitude.

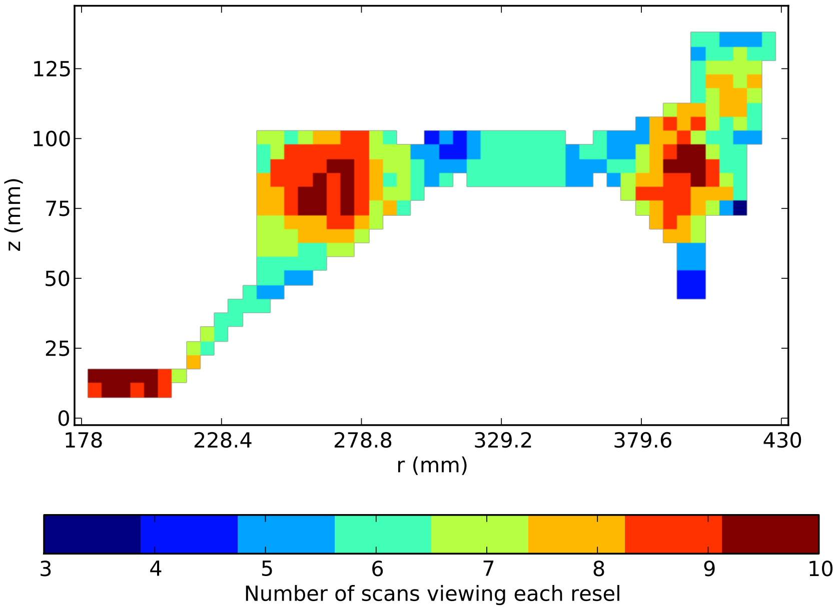


Figure 6.1: Scan coverage. The color indicates the number of separate scans that view each 5 mm by 5 mm resel in the cross section. For each resel, the consensus test will take data from all viewing scans as independent inputs. The boundary set of resels containing a front surface echo, the region exterior to that water/metal boundary, and the resels containing the datum hole are removed prior to the detection results to follow.

The coverage varies over the cross section due to the irregular disc shape. Figure 6.1 shows the number of scans that view each resel. The minimum coverage is three overlapping scans, which only occurs in a single resel in the cross section, near the L corner on the disc's outer radius. (See Figure 2.4 for corner labels.) The maximum coverage is ten, in three separate regions of the cross section that are accessible from multiple surfaces, on both sides of the disc. Most of the resels in the disc are covered by between five and nine scans. The maximum and minimum of ten and three are the same as the number of scans in the consensus test comparisons in Section 5.3.

6.2 Signal distribution and ROC estimation

Detector performance is measured in terms of Probability of Detection (POD) and Probability of False Alarm (PFA) [130, 131]. Figure 6.2 shows an illustration of these concepts for an Amplitude Threshold Detector (ATD). There is a probability distribution for the amplitude when there is only noise, shown in blue in the figure. The signal amplitude distribution is red. The vertical dashed line indicates the amplitude detection threshold.

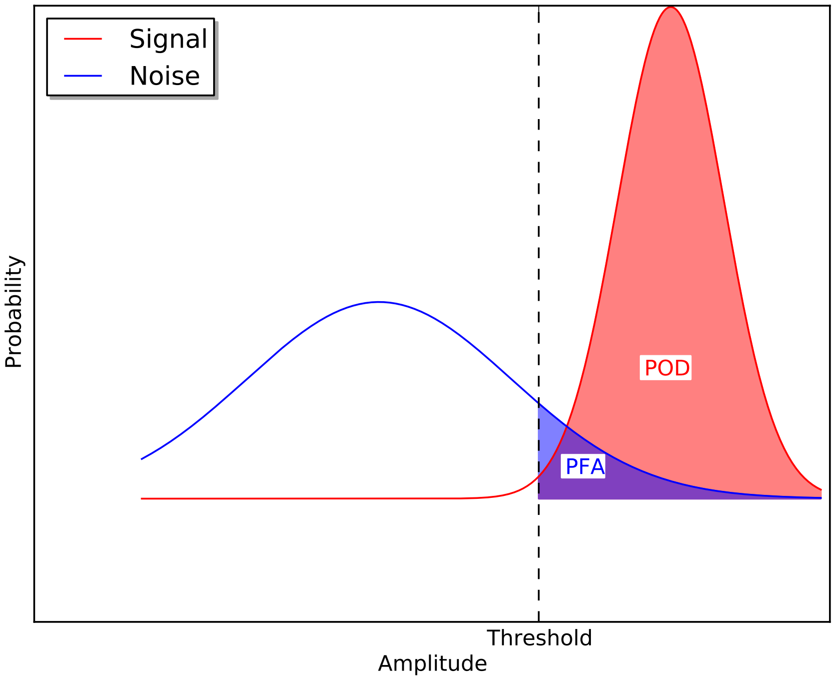


Figure 6.2: Illustration of Probability of Detection (POD) and Probability of False Alarm (PFA). Assume probability distributions for the amplitude measured when an indication is present (signal) and when no indication is present (noise) are known. Then, for any choice of threshold, the integral of the signal and noise distributions above the threshold are respectively POD and PFA.

POD is the probability that the signal will have an amplitude higher than the threshold. POD is therefore equal to the integral of the signal distribution, from the threshold to infinity. Figure 6.2 shows this graphically as the region shaded in red.

The PFA is the integral of the noise distribution from the threshold to infinity. This is the blue region in Figure 6.2. The event of a "false alarm" can have different definitions, depending on context. The term sometimes means the rejection of an entire part due to an incorrect NDE diagnosis. In all the results in this chapter, a false alarm event is defined to mean a single resel that is flagged by the detector as an indication where there is no defect. This definition of PFA makes a quantity that is much easier to estimate with data from a single disc, as opposed to such a rare event as a scrapped part.

This definition of PFA is also directly related to cost in at least two ways. First, the inspector's time is proportional to the number of possible indications output by the detection algorithm. Second, the most critical flaw may limit the service life of the part. There is therefore a financial cost to a prematurely terminated service life, if that critical flaw was incorrectly identified or overestimated in severity.

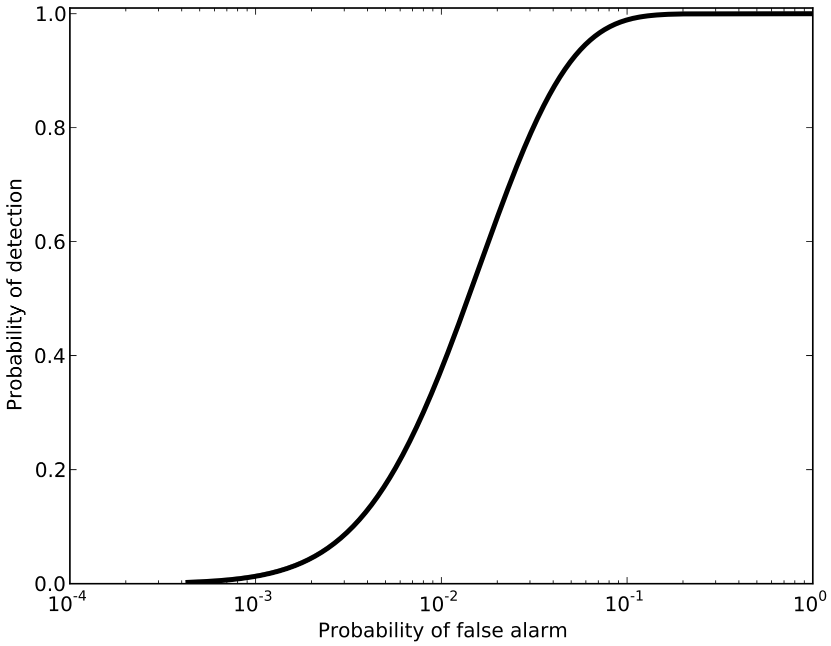


Figure 6.3: Relative Operating Characteristic (ROC) curve. The ROC curve shows the trade-off between the competing objectives of low PFA and high POD. PFA and POD are both functions of the threshold, as shown in Figure 6.2. The ROC is therefore a parametric curve in PFA vs. POD, computed by varying the threshold parameter.

Both PFA and POD decrease as the threshold sweeps from low to high amplitude. This makes PFA and POD into parametric functions of the threshold. The path they trace when plotted together is called a Relative Operating Characteristic (ROC) curve [85]. Figure 6.3 shows the ROC curve based on the POD and PFA from Figure 6.2.

The ROC curve shows the achievable trade-offs between the competing objectives of high POD and low PFA, as the threshold (or any other detector parameter) is varied. The ideal detector has POD of one and PFA equal to zero. Good detectors should come as close as possible to that point in the upper left corner of the ROC plot. The most useful part of the ROC curve is therefore the region where POD is close to one, especially if PFA is also small. Like Figure 6.3, all ROC curves in this chapter use a linear scale for POD and a logarithmic scale for PFA, to better illustrate behavior when POD is high.

6.3 Detection with a single scan

As a first example of detection, consider Indication 1 identified in Section 2.1. For convenience, Figures 6.4 and 6.5 reproduce the diagrams showing the Known Indication locations. Figure 6.6 shows the A-scan that contains the signal for this indication, as viewed by the QR +5° scan. In addition to the raw amplitudes and Hilbert envelope, horizontal red lines in Figure 6.6 mark the amplitude and extent of the resels containing this A-scan.

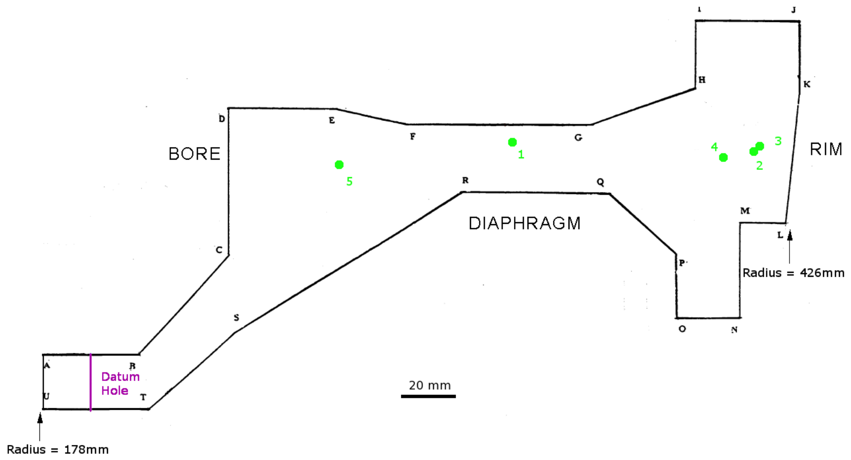


Figure 6.4: Seeded Defect Disc cross section, showing locations of the Known Indications in the r, z plane. Green circles mark the Known Indications, labeled 1 through 5. This figure is reproduced from Figure 2.4.

The amplitude is on a linear discretization scale, as it was recorded by the data acquisition hardware. The hardware records signals with an eight-bit Analog-to-Digital Converter (ADC). Therefore, the raw, unrectified signal in blue spans an

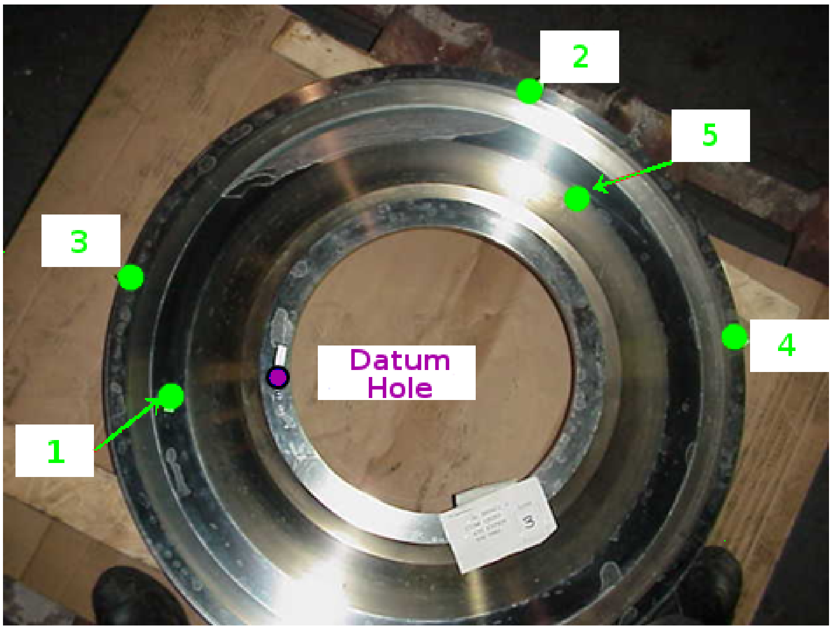


Figure 6.5: Seeded Defect Disc bottom view, showing locations of the Known Indications in the r, θ plane. Green circles mark the Known Indications, labeled 1 through 5. This figure is reproduced from Figure 2.5.

amplitude range of $[-127, 127]$. The amplifier gain is set just before the disc scan with a calibration block. The calibration block has a 0.05 inch diameter Flat-Bottomed Hole (FBH) drilled into block from the bottom. The inspector adjusts the amplifier gain such that the amplitude threshold is fifty percent of the maximum ADC range.

It is clear from Figure 6.6 that the amplitude at Indication 1 is much too low for the Amplitude Threshold Detector (ATD) with the default threshold of 64. The ATD would not detect Indication 1 on the QR $+5°$ scan unless the threshold were lowered at least to 14.5. Unfortunately, such a reduction in ATD threshold would also pick up many more resels which do not contain indications, raising the PFA.

An amplitude histogram can estimate the PFA for any amplitude threshold, by analogy to Figure 6.2. Figure 6.7 shows a histogram of all resels for the disc. This is the combined histogram of the full empirical distribution, agglomerating all resels, from all three angles, from all surfaces.

Figure 6.7 shows that the amplitudes have a very long-tailed distribution. There are even values well beyond the maximum unrectified amplitude of 127, due to the effect of the Hilbert envelope on some of the signals that are clipped at maximum

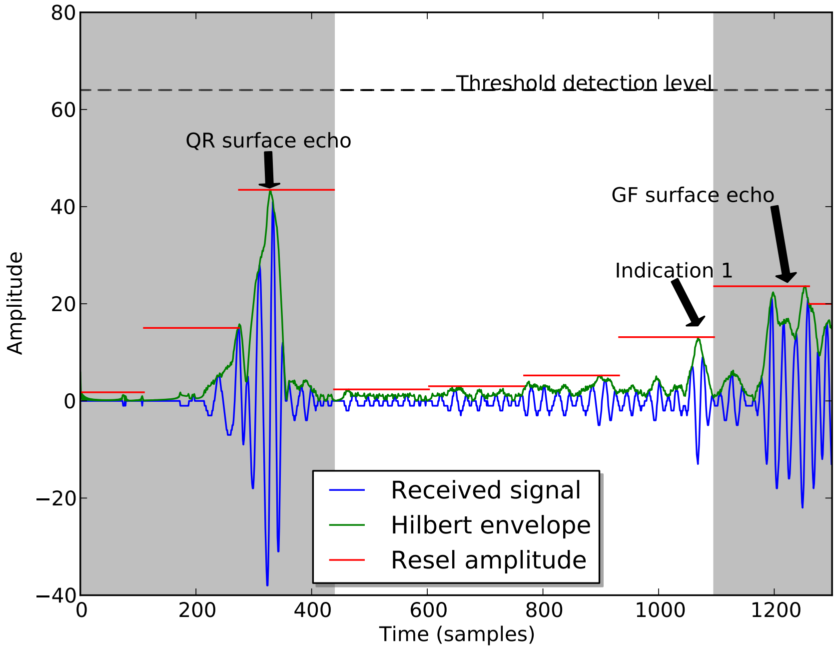


Figure 6.6: A-scan from the QR +5° scan. That is, the A-scan is from the QR surface (see Figure 6.4), with the probe rotated radially outward +5°. The Hilbert transform computes the evelope of the oscillatory signal received by the transducer. Red lines mark the time extent of the 5 mm resel cubes. Resel amplitudes are set to the maximum Hilbert envelope amplitude within each resel region. The reflection from Indication 1 is visible at time = 1080 samples, but it is well below the status quo threshold detection level of 64. Resels outside of, and including, the front and back surface echoes are excluded from all detection results.

amplitude.

Indication 1 has a relatively small amplitude in QR +5°, relative to the full amplitude distribution. The fraction of resels above that amplitude is 0.273. This means that, if the amplitude threshold in the ATD were lowered until it was just low enough to detect on the QR +5° scan, the PFA would have a high value of 27.3 percent. If the amplitude threshold were any higher, ATD would fail to detect Indication 1 in this scan. There is no way to adjust or adapt the ATD for that scan to detect the flaw without an unacceptably high PFA.

Figure 6.8 shows the histogram of resel amplitudes used by the LEN detector. All amplitude data are from a region that is constant in r and z, around the circular ring for $\theta \in [0, 2\pi]$. Regardless of the scan surface or direction, this region takes samples at the same radius and depth for each scan. Physical effects that affect the noise distribution, such as attenuation, beam spread, and proximity to structural

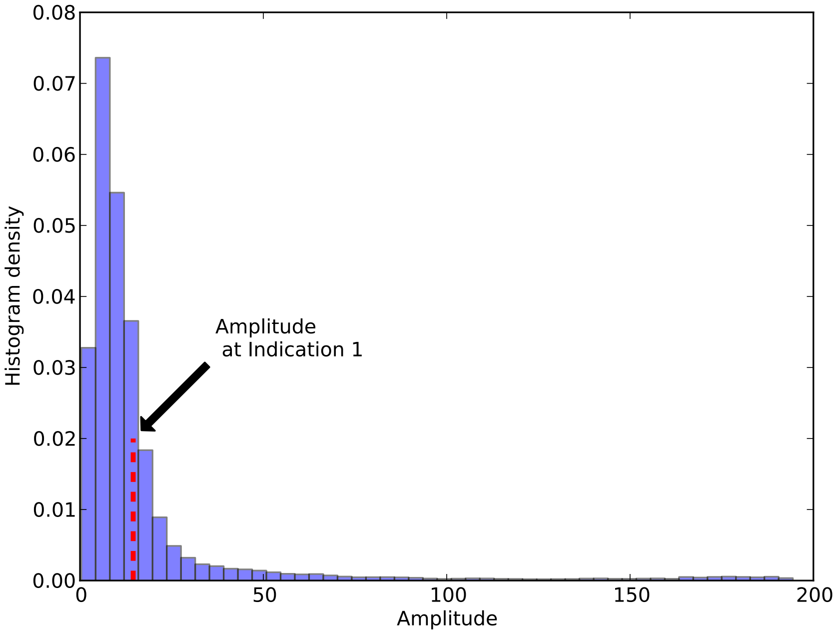


Figure 6.7: Resel amplitude distribution for all scans, at all three incidence angles, over the full interior of the disc. The amplitude at Indication 1 in the QR +5° scan is marked for reference. 27.3 percent of all resels have an amplitude greater than the indication amplitude of 14.5. Therefore, a threshold detector with a threshold low enough to detect Indication 1 from the QR +5° scan would have an undesirably high PFA of approximately 0.273.

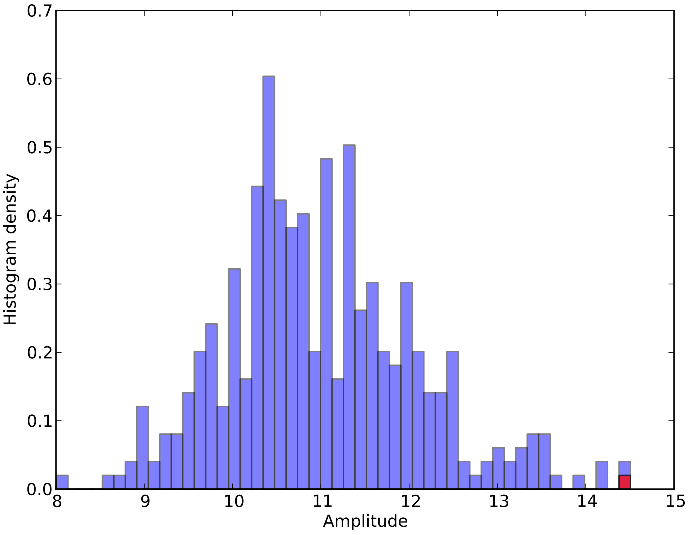


Figure 6.8: Resel amplitude distribution within a single circumferential ring from the QR +5° scan. The ring of resels contains Indication 1. The red bar marks the amplitude at the indication as one of two resels with the greatest amplitude in the distribution. The Local Empirical Noise (LEN) detector estimates a p-value for a resel based on the empirical distribution of resel amplitudes in that resel's circumferential ring.

features, are equal within the ring.

Figure 6.7 showed that the indication amplitude was not high by an absolute standard over the disc, but Figure 6.8 shows that it *is* high relative to the local noise. A locally derived threshold detector should outperform an amplitude threshold that is constant over the entire disc. The LEN effectively applies a local threshold, because it translates the amplitude to a p-value according to the local noise distribution. The inspector can then choose a threshold on these local p-values for the entire disc. The LEN will adapt that p-value threshold to local amplitudes for each ring.

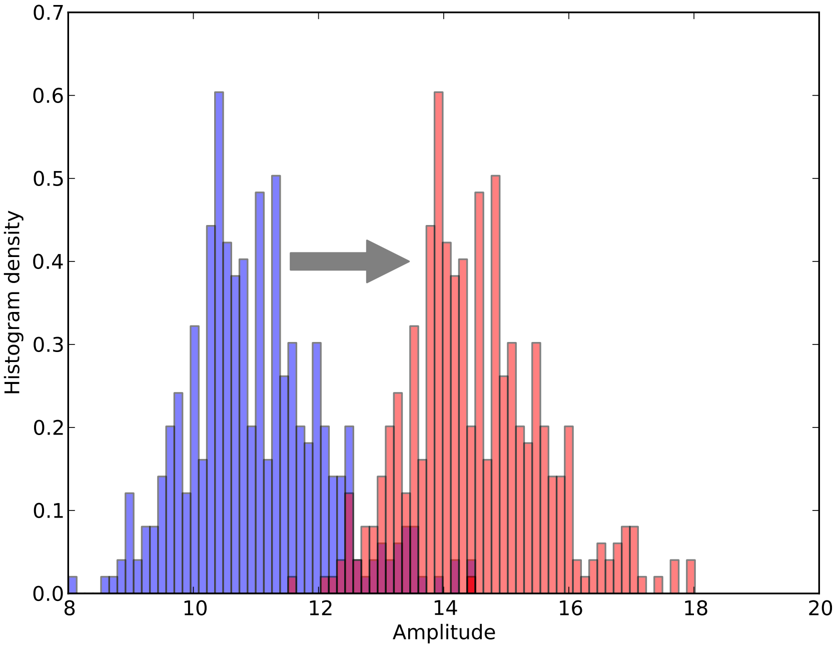


Figure 6.9: Noise-shifting procedure to approximate a distribution for the indication resel amplitude. The resel amplitudes around the circumferential ring containing the indication combine to form an empirical distribution for the noise, shown here in blue. In practice, no distribution is known for the signal as depicted in Figure 6.8; only the single sample at the hypothetical indication is available. The indication signal distribution (in red) is therefore appoximated with a copy of the empirical noise distribution, with the mean shifted to equal the indication resel amplitude. The signal distribution, together with the noise distribution of all resel amplitudes in the disc, determine the threshold detector ROC curve in the manner illustrated in Figures 6.2 and 6.3.

Nearly all amplitude data can be assumed to be noise, so a simple amplitude histogram gives a very realistic noise distribution for computing PFA for any choice of threshold, as in Figure 6.2. Estimating POD as a function of threshold requires a distribution for the signal, given that there is a flaw to detect. However, only the single resel under evaluation is available as a hypothetical signal. ROC calculation requires a method for estimating the probability distribution for a signal, given that a flaw is present.

Figure 6.9 illustrates the procedure adopted to estimate the signal distribution. A copy of the empirical noise distribution is shifted by a constant such that its mean equals the indication resel amplitude. This is equivalent to the assumption that the amplitude sampled is linear in the scattering potential, and noise is additive. The one signal amplitude sample is the theoretical signal without noise, plus a single sample of zero-mean noise. The most logical estimate of the distribution of noise that could have been added to the signal is a zero-mean version of the noise distribution.

For the ATD, the signal distribution generated by the above procedure and the noise distribution in Figure 6.8 together provide an ROC curve. The LEN and Data Fusion Detector (DFD) are based on p-values instead of amplitudes, as described in Sections 5.1.2 and 5.2.3. The Local Empirical Noise (LEN) cumulative distribution functions convert amplitude distributions to p-value distributions before computing the ROC curve. This conversion applies to both signal (at the indication) and noise (over the entire disc). The threshold is then interpreted as a significance level for the hypothesis test, rather than an amplitude threshold. Figure 6.10 plots both the ATD and LEN ROC curves.

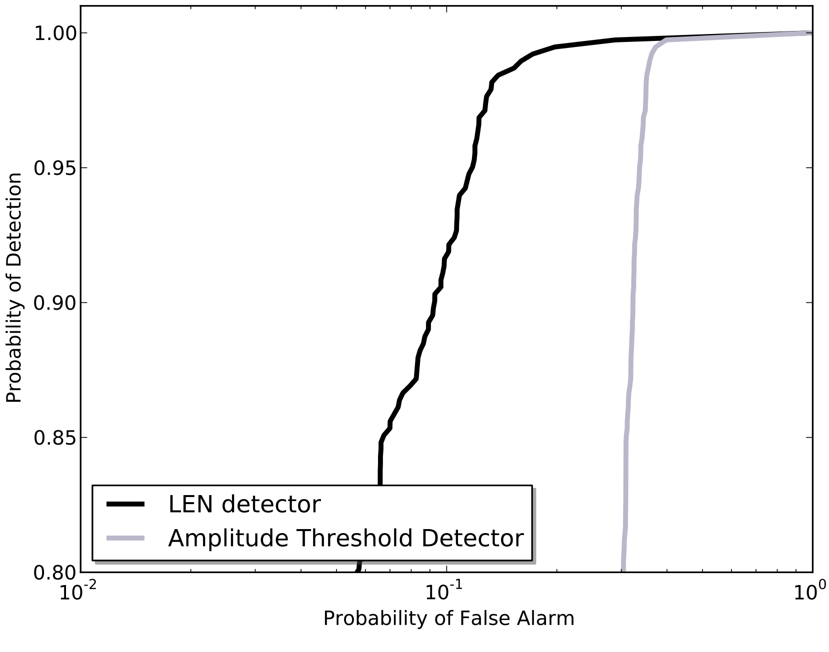


Figure 6.10: ROC curves for the Local Empirical Noise (LEN) and simple threshold detectors at Indication 1, QR +5°. See Figure 6.4 for the locations of Indication 1 and the QR surface. The LEN can detect Indication 1 from the QR +5° scan with a PFA up to four times smaller than the threshold detector. Both detectors have a undesirably high PFA, however, at more than 10 percent for any POD greater than 95 percent.

The locally derived threshold in LEN reduces PFA by approximately a factor of four. At this point, the detection still only uses information from a single scan, without any data fusion. It is possible to improve these results by combining data from multiple scans.

6.4 Detection with two scans

Now add a second scan, GF $-5°$, to the QR $+5°$ scan considered in Section 6.3. The GF surface is parallel and opposite to QR, as shown in Figure 6.4. Figure 6.11 shows the A-scan and resel amplitudes from GF $-5°$ with the signal from Indication 1.

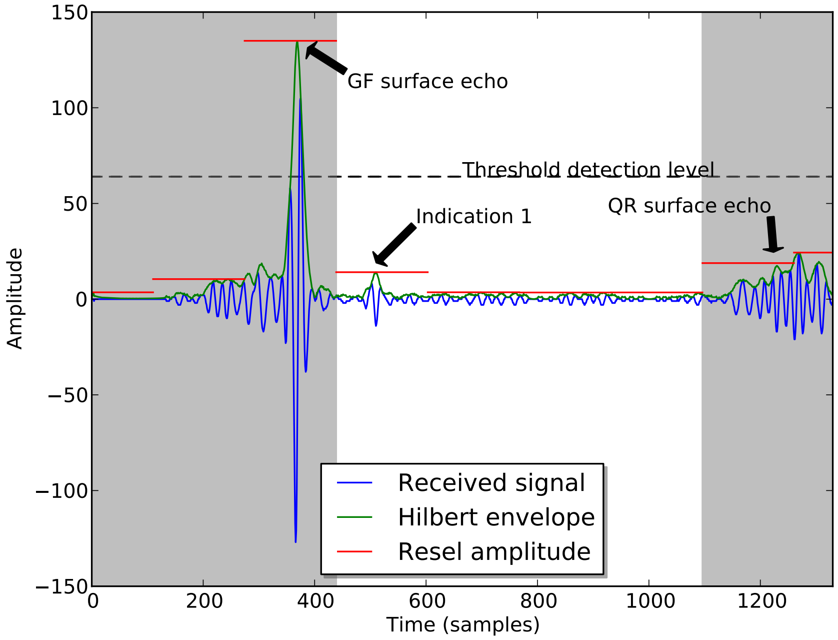


Figure 6.11: A-scan from the GF surface at -5° (probe rotated radially inward) for Indication 1. The reflection from Indication 1 is clearly visible at time = 510 samples. The resel amplitude at the indication is well below the status quo threshold.

The resel containing Indication 1 in GF $-5°$ also has a low amplitude. In fact, it is very close to the amplitude at the indication from QR $+5°$. However, the local noise levels are lower, possibly because Indication 1 is closer to the GF surface.

Figure 6.12 shows the relevant noise distribution for Indication 1 in GF $-5°$. The lower amplitudes in the noise distribution mean a lower p-value is expected at the indication, even though it has the same indication amplitude as QR $+5°$. Figure 6.13

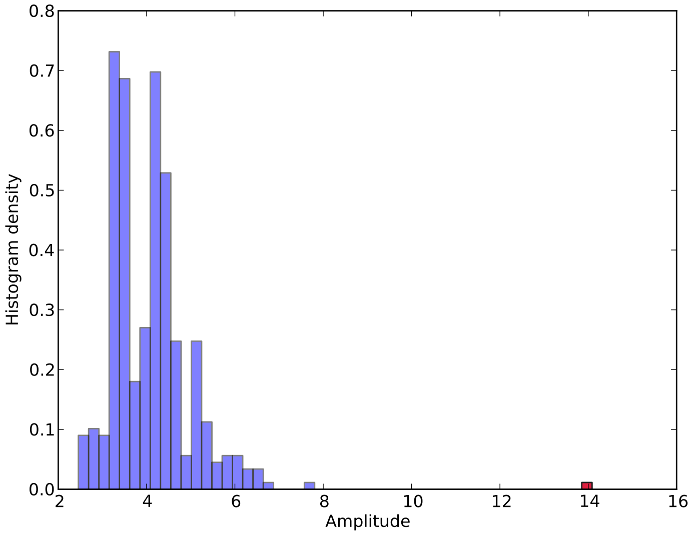


Figure 6.12: Resel amplitude distribution within a single circumferential ring from the GF -5° scan. The red bar marks the amplitude at Indication 1. The indication amplitude in this scan is close to the corresponding value from QR +5° (see Figure 6.8), but the noise in this scan is much lower. The LEN p-value rejecting the noise hypothesis is therefore significantly smaller.

shows that this is, indeed, the case.

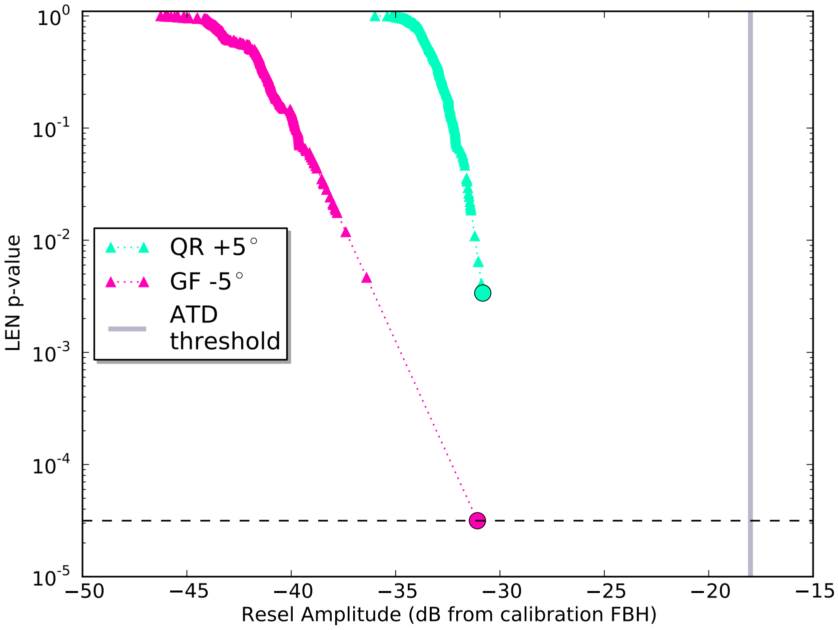


Figure 6.13: LEN p-values for the QR +5° and GF -5° scans at Indication 1. Both scans have low amplitudes that would require a low amplitude threshold to detect, which would lead to a high PFA elsewhere in the disc. However, both scans also have very low LEN p-values. A threshold on p-value could detect the indication in either QR +5° or GF -5° with a lower PFA than the amplitude threshold detector.

Figure 6.13 contains all information from the amplitude and p-value distributions for the two scans. The following sections contain several plots in the content-rich format of Figure 6.13. Following is an explanation to assist the reader in interpreting these plots and understanding the information they contain.

Each scan has a LEN distribution, represented by a series of triangular markers. Each triangle represents a resel. Circles mark the resels containing Indication 1, one from each scan. The distributions are plotted on a domain (x axis) of amplitude, with a logarithmic scale. The amplitudes are given in dB down from the calibration signal. Calibration blocks have a flat-bottomed hole (FBH) drilled up from the bottom side. During calibration, the gain on the amplifier is set such that the FBH reflected signal is 0 dB. Current inspection procedures call for an ATD default amplitude threshold of -18 dB, down from the FBH calibration signal. The -18 dB threshold on the logarithmic scale is equal to 64 on the linear eight-bit discretization scale from Section 6.3. This status quo threshold is indicated by the vertical gray line in Figure 6.13.

The codomain (y axis) indicates the p-value from the LEN detector, which is always a monotonically decreasing function of amplitude. The curve shape shows the form of the distribution. It also reveals the accuracy of the Pareto tail fit in the LEN distribution, as described in Section 5.1.2. The exponential term for the high-amplitude tail section of the LEN distribution will be linear in the log-log plot. If the Pareto tail assumption is accurate, it should be a smooth extrapolation from the lower-amplitude part of the LEN distribution. Recall that, as a statistical test for the noise-only hypothesis, a small p-value means the signal is unlikely to be noise. Figure 6.13 shows that there are small p-values at the indication in both GF $-5°$ and QR $+5°$. The purpose of data fusion is to combine the multiple p-values in a consensus test.

Figure 6.14 shows ROC plots for ATD, both LEN detectors, and the DFD that fuses the GF $-5°$ and QR $+5°$ LEN detectors. Both LEN detectors outperform the amplitude threshold. GF $-5°$ achieves a lower PFA than QR $+5°$ by approximately a factor of 25, because of its lower local noise level. DFD is better than either of its two LEN inputs. The improvement in DFD over both LEN detectors is due to their mutual confirmation and reciprocally supporting information.

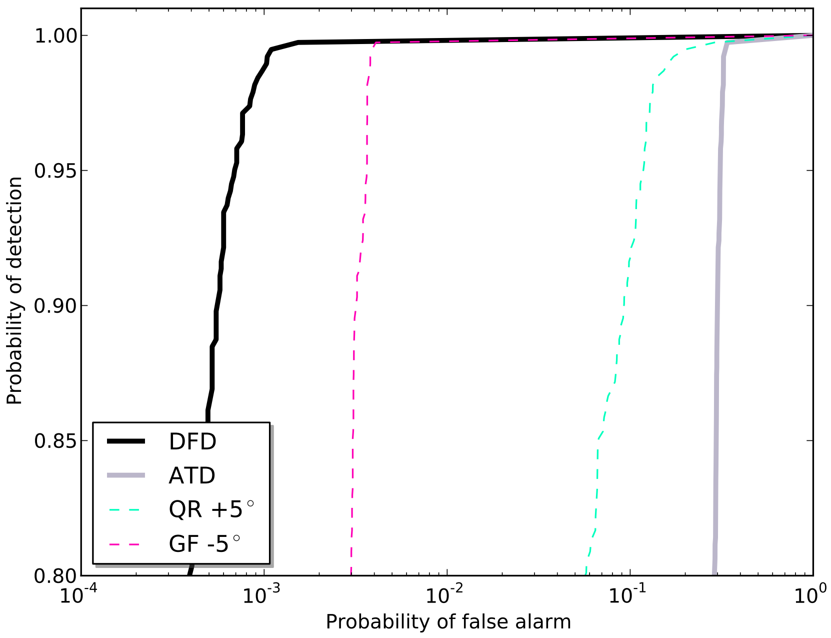


Figure 6.14: ROC curves for the Data Fusion Detector (DFD) and LEN detectors for scans QR +5° and GF -5° at Indication 1. The Amplitude Threshold Detector (ATD) ROC is shown for comparison. The GF -5° LEN detector achieves a lower PFA than the QR +5° LEN detector, due to its lower p-value. Both LENs are superior to the ATD. The DFD combines information from the LEN detectors on the individual scans using the Fisher consensus test and has the lowest PFA.

6.5 Detection with all scans

This section builds on the first steps into data fusion from Section 6.4. In the results to follow, all scans that cover the indication contribute to the data fusion detector. A p-value versus amplitude plot for the LEN distribution on each input, like Figure 6.13, shows the information available for detection at each of the five Known Indications first introduced in Section 2.1. ROC curves give detection comparison results, as in Figure 6.14: each input LEN, the consensus DFD, and, for reference, the ATD have ROC curves on the same plot to show relative improvement over the current inspection method.

Known Indication 1

Figure 6.15 shows the LEN p-values and amplitudes for all scans that cover Indication 1. Indication 1 is in the diaphragm of the disc, between the GF and QR surfaces. With three incident angles on each surface, a total of six scans cover this

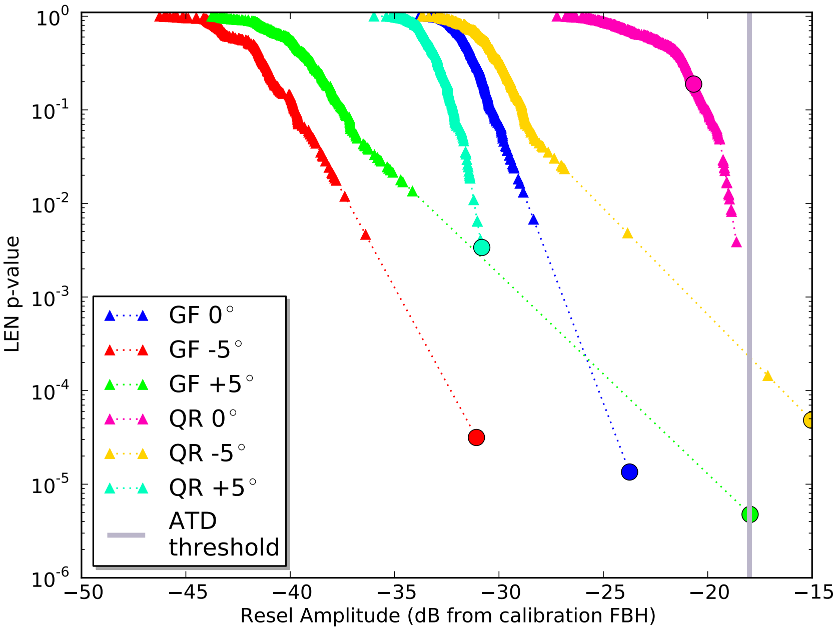


Figure 6.15: LEN p-values for Indication 1. The LEN detector is formulated as a hypothesis test, in which the null hypothesis is that the resel under consideration contains no indication and its amplitude represents only noise. LEN p-values equal the minimum significance level at which the null hypothesis is rejected; i.e., a small p-value expresses a high likelihood of an indication. The triangular markers show the distribution of amplitudes and LEN p-values for all resels in the ring containing Indication 1. The circles mark the values at Indication 1. Only the QR -5° and GF +5° scans have a detectable indication with the status quo detection threshold. Both GF 0° and GF -5°, however, have LEN p-values that are comparably low and well below their respective distributions.

region. Only two of them, QR $-5°$ and GF $+5°$, detect the indication by the amplitude threshold. GF $+5°$ only crosses the amplitude threshold by the narrowest margin.

Two of the scans have p-values that are likely too large to use for detection individually. The QR $+5°$ has a p-value that is only moderately low. QR 0° has a high amplitude at the indication, but it comes from a very noisy region. It is possible that the ring of resels feeding the LEN for QR 0° is picking up part of the back surface GF echo, and this is obscuring the indication echo.

The other four scans (GF $-5°$, GF 0°, GF $+5°$, and QR $-5°$) have a very low LEN p-value. Any one of these scans could detect the indication quite reliably with a p-value threshold. The DFD can combine these p-values with data fusion to enhance performance even further.

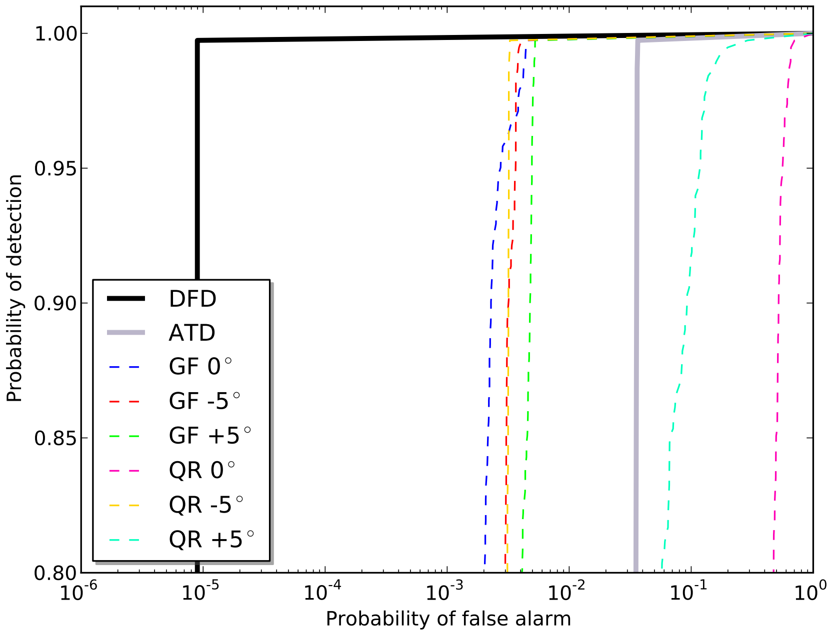


Figure 6.16: ROC curves, using all six scans that view Indication 1: GF 0°, GF +5°, GF -5°, QR 0°, QR +5°, and QR -5°. QR 0° and QR +5° always have a lower amplitude than at least one other scan, and the amplitude for QR 0° at the indication is low relative to the other resels in their LEN ring. The ROC curves for the individual QR 0° and QR +5° LEN detectors are therefore worse than the threshold detector, which uses the maximum amplitude across all six scans. However, the information from the other four scans is sufficient to give the data fusion detector very good discriminatory power.

Figure 6.16 shows the ROC curves for all of the detectors. For comparison to the current inspection method, the ATD ROC curve is plotted with a solid gray line. Each of the LEN detectors, one for each scan, is plotted with colored dashed lines. The DFD in solid black shows the result from combining the LEN detectors into a consensus output.

Not surprisingly, the ATD POD is greater than the QR +5° and QR 0° PODs at every PFA, showing that these locally derived threshold detectors are worse than a fixed amplitude threshold. However, each of the other four scans has a PFA about an order of magnitude lower than ATD. They combine to give the DFD a very high POD and very low PFA. Note that the DFD PFA drops suddenly at a value that is approximately the inverse of the total number of resels in the disc. In other words, the PFA is effectively zero at that point. It is the smallest number for PFA that can be estimated for these data.

It is worth considering whether or not a scan might be excluded when it has a high p-value relative to the others. For example, it would be possible to define a new detector which would disregard any inputs like QR 0° and QR +5° that detract from POD. Such a detector would indeed have higher PODs, but it would come at the cost of higher PFAs. Because this is a test case, it is known a priori that this resel contains an indication. Obviously that knowledge will not be available in general, and a realistic detector cannot fairly throw out the scans that contribute least to the DFD.

Known Indication 2

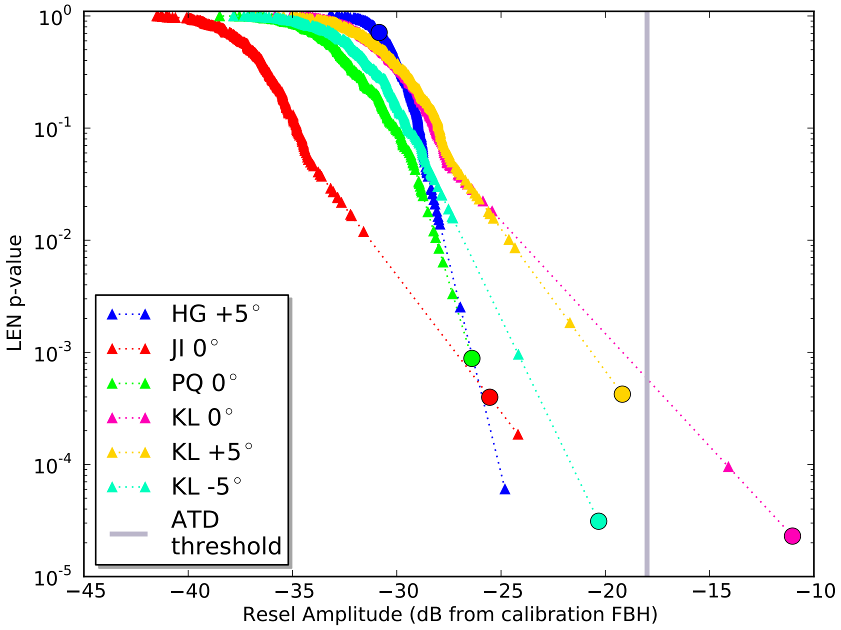


Figure 6.17: LEN p-values for Indication 2. Only one scan, KL 0°, is greater than the amplitude threshold. However, five of the six scans covering this location have quite low p-values. These other four low p-values would allow the Data Fusion Detector (DFD) to detect Indication 2 easily, even if the KL 0° scan were not available, if it had a lower amplitude, or if it were obscured by higher noise.

Indication 2 is in the rim of the disc, where it is covered by six scans. Refer to Figure 6.4 for the location of Indication 2. Figure 6.17 shows the LEN p-values and amplitude distributions for Indication 2. Only one scan, KL 0°, breaks the amplitude threshold. Five of the six scans, on the other hand, have quite low p-values. In contrast to the ATD, the other four LEN detectors would easily detect Indication 2. The indication would be reliably detectable even if the KL 0° scan were not available, had a lower amplitude, or were obscured by noise.

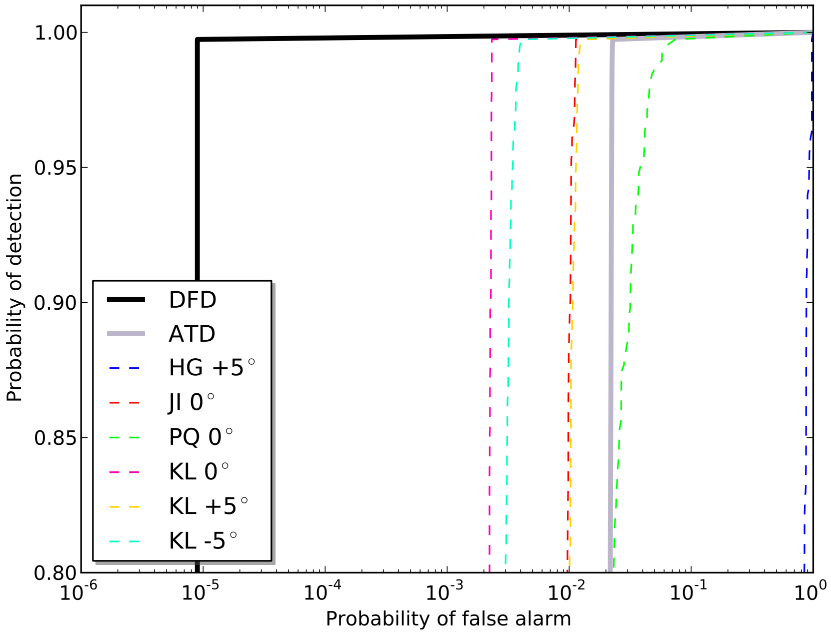


Figure 6.18: ROC curves for Indication 2 using all scans. Individually, the scans covering Indication 2 vary widely in the quality of information they provide. HG +5° has no detectable signal at the indication location, and PQ 0° by itself is worse than the ATD. The best individual LEN detector, KL 0°, could achieve an order of magnitude reduction in PFA from the ATD. The DFD reduces the PFA by more than an additional order of magnitude.

Figure 6.18 shows the ROC curves. HG +5° has no signal at the Indication 2 that is distinct from noise. PQ 0° by itself, i.e., its LEN detector, is worse than ATD. The best individual LEN detector is KL 0°, which is better than ATD by an order of magnitude in terms of PFA. DFD achieves an additional improvement in PFA over its best LEN input (KL 0°), by about a factor of forty.

Known Indication 3

Only five scans reach Indication 3 at its location in the rim, near the KL surface. Figure 6.19 shows that only one scan, KL 0°, is just high enough in amplitude for the ATD to detect it. Indication 3 was very close to being missed by the status quo inspection procedure. Both KL 0° and KL +5° have low p-values. None of the other three scans shows conclusive evidence of an indication.

Only the KL 0° and KL +5° LEN detectors outperform the ATD in terms of ROC, as shown in Figure 6.20. DFD reduces PFA by an additional order of magnitude

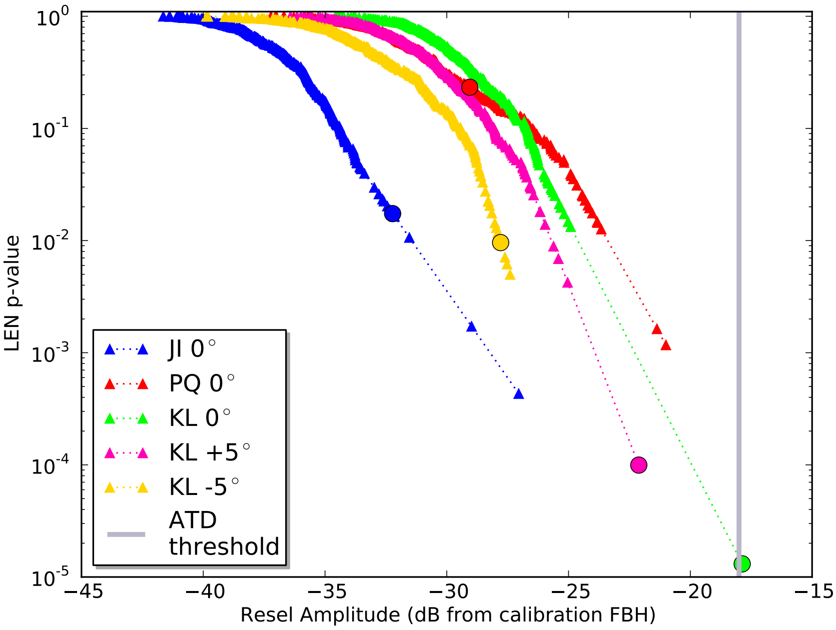


Figure 6.19: LEN p-values for Indication 3. KL 0° is just high enough to trigger the ATD threshold. Its high amplitude at the indication stands well above the noise, so the KL 0° LEN p-value is also very low. KL +5° also has a low enough p-value to make a significant contribution to detection by consensus. The other three scans have amplitudes that are much more difficult to distinguish from the noise.

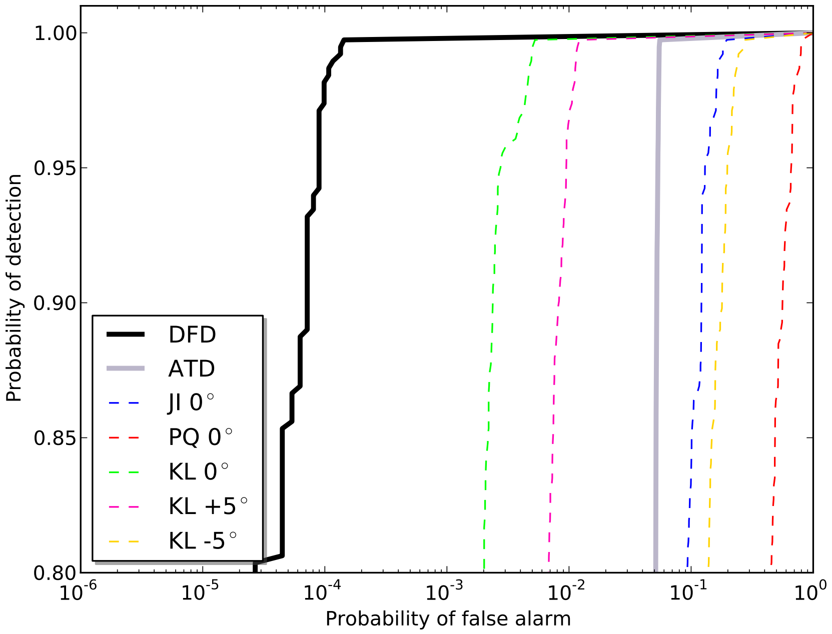


Figure 6.20: ROC curves for Indication 3 using all scans. Only the KL 0° and KL +5° LENs outperform the ATD. The mutual confirmation of these two LEN inputs is sufficient to push the DFD PFA down by two orders from the KL +5° LEN PFA.

from the best individual scan, KL 0°. It maintains this performance improvement over the best single input, even though it includes three of five inputs that are worse than the ATD.

Indication 3 is the only one of the five Known Indications to show some rounding of the ROC curve. For each of the other four Known Indications, the greatest consensus p-value is less than the smallest p-value from all other resels; there is no overlap. For the other four, their resel is always one of the first five resels detected, for all values in their signal distributions. Indication 3 has a few consensus p-values from its signal distribution that do overlap those of the noise-only p-values in the rest of the disc. This means that, for some values in the signal distribution, the consensus p-values from the Indication 3 DFD are higher than at least some other resels that do not contain a known indication. Even for this worst case, Figure 6.20 shows that the extent of the p-value overlap is so minor that it does not even begin until the PFA is as low as 10^{-4}. This remarkable performance shows how robustly the DFD detects the known indications.

Known Indication 4

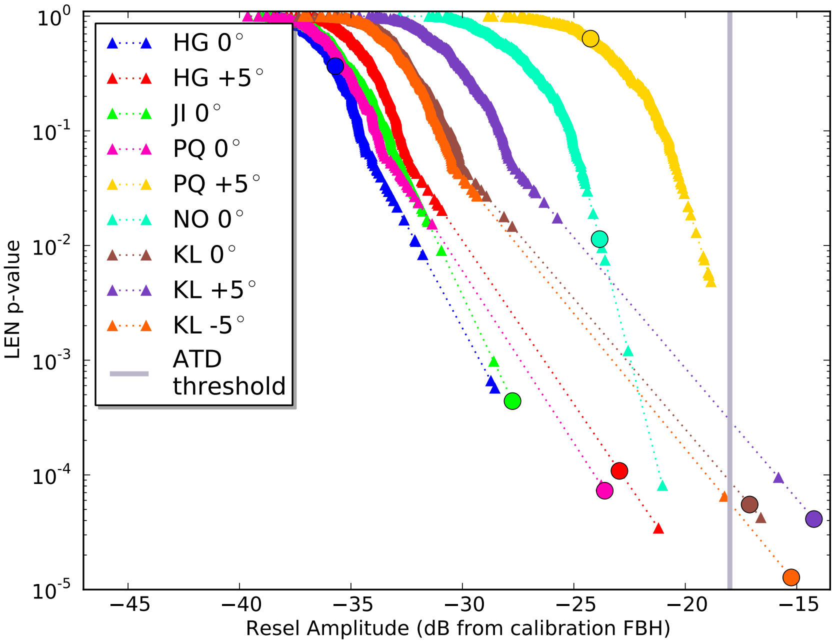


Figure 6.21: LEN p-values for Indication 4. The three KL scan amplitudes are above the ATD threshold. Three more (JI 0°, PQ 0°, and HG +5°) have lower amplitudes but also low p-values. The remaining three scans only see noise.

Known Indication 4 is also located in the rim, like Indications 2 and 3. (See Figure 6.4.) It is closer to the surfaces that border the diaphragm and rim, making it

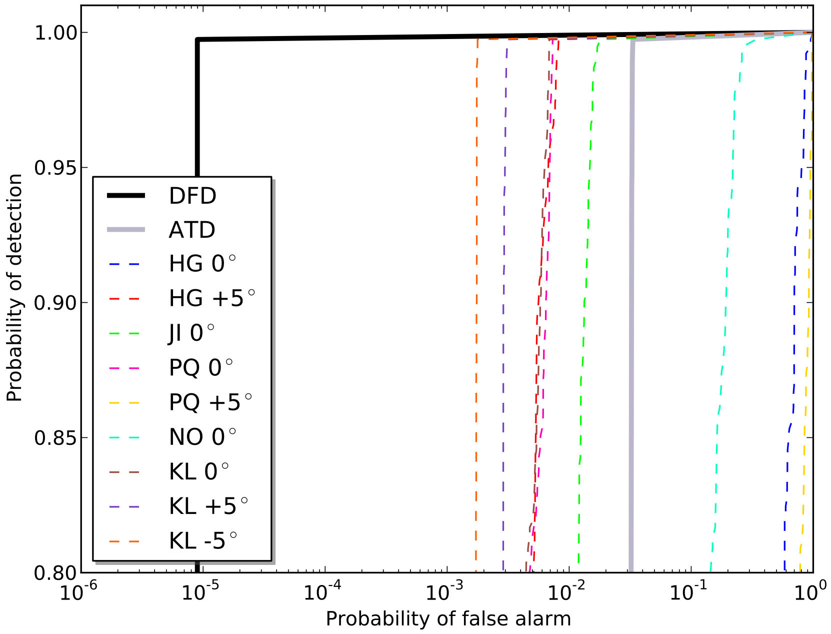


Figure 6.22: ROC curves for Indication 4 using all scans. This indication is the most easily detected by the ATD because all three KL scans have amplitudes clearly above the threshold. It is also easily detected by the DTD, since JI 0°, PQ 0°, and HG +5° also have LEN p-values low enough for the DFD to detect, even if the KL scans did not cover the indication.

accessible to more scans and increasing the coverage. LEN p-value and amplitude distributions are shown in Figure 6.21.

Indication 4 is detected reliably by ATD, with three of the nine scans (KL 0°, KL +5°, and KL −5°) above the −18 dB threshold. If those three scans were not available, or if they did not feature the detectable signal amplitude, LEN detectors for PQ 0°, HG +5°, and JI 0° would still have a low enough p-value to detect the indication. One scan, NO 0°, has moderately low p-value, and the remaining two only see noise.

The PQ +5° in particular is one that is problematic for any detection method based on an amplitude threshold. Many of the samples in its noise distribution are in a high amplitude region. A significant fraction are higher amplitude than the PQ 0° and HG +5° indication resels, though even the lowest p-values in PQ +5° are much greater than the p-values in those two scans. There is clearly no way for an amplitude-based detector to distinguish between the noise in PQ +5° and the indication signal in PQ 0° and HG +5°. With LEN detectors, on the other hand, the separation in p-value is by a comfortable factor of fifty.

Unsurprisingly, six of the nine individual LEN detectors outperform the ATD in the ROC plot in Figure 6.22. Indication 4 has both high coverage and high detectability in the majority of the input scans. When they are fused by the DFD, the result is as close to perfect detection as is possible with these data.

Known Indication 5

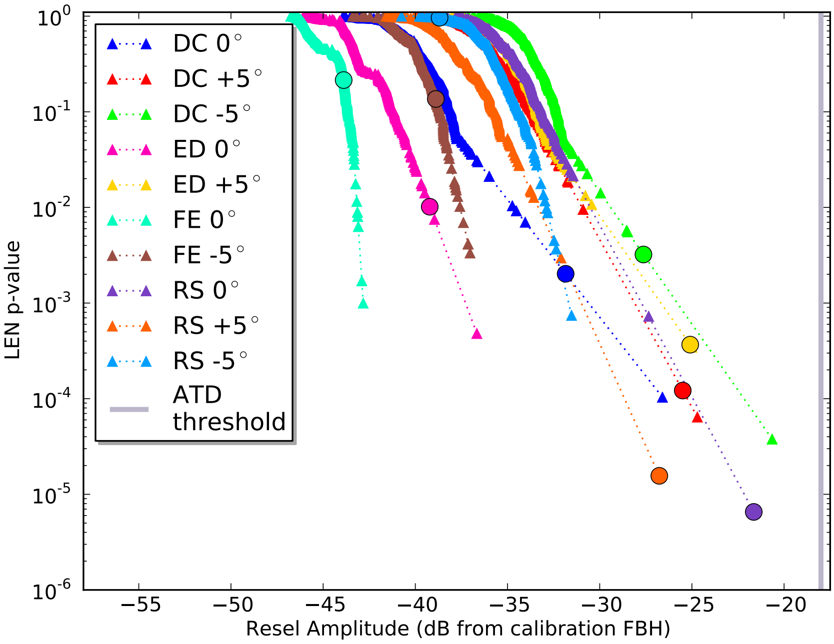


Figure 6.23: LEN p-values for Indication 5. This indication has the highest coverage of any of the known indications, with ten scans viewing its location. None of the scans has a very high amplitude, and there is a wide variation in the LEN p-values.

Indication 5 is located in the transition between the disc bore and diaphragm. This places it in the cross-sectional region with highest degree of coverage for this disc shape. Figure 6.23 shows[1] the LEN p-value distributions for the ten scans covering Indication 5.

Despite the high coverage, none of the amplitudes from any of the scans is high enough to cross the ATD threshold. Indication 5 would not trigger the detection built into the scanning software for the status quo inspection. For a disc in the

---

[1] For plots such as Figure 6.23 that have many curves or groups of points, it is a nontrivial problem to find a set of colors that are mutually distinguishable both on paper and on a projector screen. Glasbey et al. [132] give a very useful algorithm that was used to produce optimal color sets for these plots.

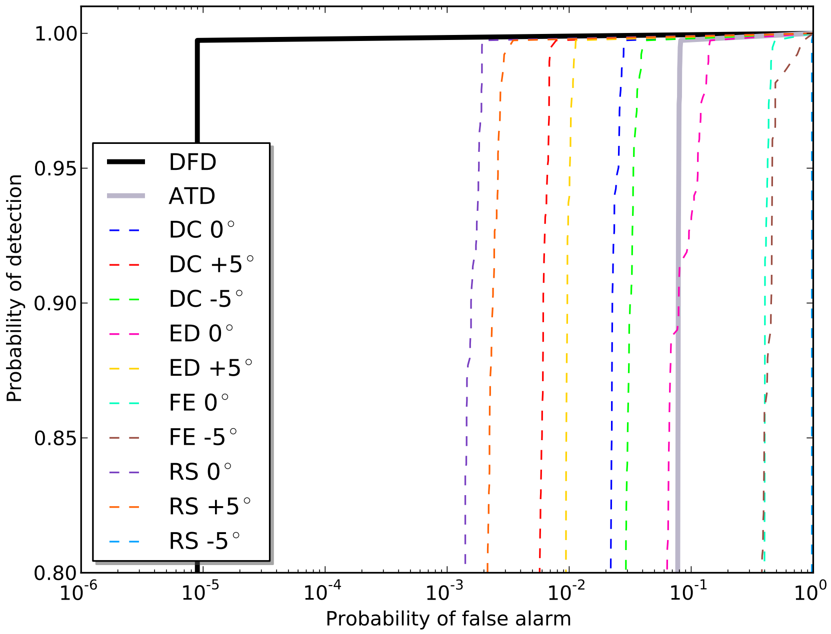


Figure 6.24: ROC curves for Indication 5 using all scans. The wide range of p-values from the LEN detectors is evident in ROC performance. RS −5°, FE 0°, FE −5°, and ED 0° (at high POD) are individually all worse than ATD. Two more scans, DC −5° and DC 0°, are moderately better than ATD with lower PFAs by about a factor of three. High coverage produces high detectability with DFD in this case, since there are four more scans covering Indication 5. All four of them have a PFA below 0.01 and the DFD combines them to give a PFA at the lower limit of what can be estimated from this data set.

production chain, if it were not known to be a disc in need of careful examination, an indication like this one could easily be missed.

In terms of LEN p-value, detecting scans range from very low (RS 0°) to moderately low (ED 0°), with RS +5°, DC +5°, ED +5°, DC 0°, and DC −5° in between. FE 0°, FE −5°, and RS −5° do not appear to detect any signal at Indication 5.

The ROC curves in Figure 6.24 show that six of the ten individual LEN detectors could outperform the ATD. Putting them together in the DFD improves PFA by more than two order of magnitude over RS 0°, which has the best individual LEN detector. Again, the DFD performance is overwhelmingly positive, to the point that the PFA is at the limit of what can be measured by the available data.

6.6 Detectability limit for low-amplitude flaws

Section 6.5 showed that the DFD detection performance for all of the Known Indications is extremely high. With the exception of Indication 3, the ROC curve for all DFDs takes the form of two straight line segments. The minimum PFA is equal to the inverse of the total number of resels in the disc, N_{resel}, in one line segment. The other segment remains between a POD of one and $1 - N_{resel}$. All samples from the signal distributions for Indications 1, 2, 4, and 5 produce a DFD consensus p-value that is higher than any other resel in the disc. Regardless of the sample selected from the signal distributions for these known indications, these indications are the four resels with the lowest p-value in the disc.

In the envisaged usage scenario for the automated inspection, the inspector would view resels in order of increasing consensus p-value. The previous section shows that, for all values in the signal distributions for the known resels, Indications 1, 2, 4, and 5 would be the first four resels selected. For Indication 3, the DFD ROC in Figure 6.20 shows detection performance that is not quite as extreme as the other four known indications. Its performance is still near the limit of what can be measured by this data set.

The results in Section 6.5 show the DFD's ability to detect, without any ambiguity, indications that are already detectable under current inspection procedures with the ATD. Of interest now is whether or not an indication can be detected, and how performance degrades, as the DFD encounters smaller or less reflective indications.

This section seeks the limits of detectability for the proposed DFD detector and contrasts them with those of the status quo ATD. The analysis maintains the realism of Section 6.5 by continuing to use the data from the known indications in the Seeded Defect Disc. The amplitudes at the indication from each of the scans, A_j, are scaled by a scale factor, α, maintaining the same proportions relative to the noise in their respective scans.

$$A_j(\alpha) = \alpha A_{j0} + (1 - \alpha)\,\mu_{noise} \tag{6.1}$$

A_{j0} is the resel signal amplitude at the indication, in the original data for scan j. The mean of the noise distribution is μ_{noise}. The same scaling factor, α, is applied for all scans covering the indication, while μ_{noise}, σ_{noise}, and relative SNR are different for each scan and determined empirically.

This proportional reduction in SNR across all scans intends to simulate smaller or less reflective indications, while preserving the same directional dependence, physical realism, and relevance. This procedure will reveal how DFD and ATD performance degrades as amplitudes and SNR decrease. At the same time, it attempts to make minimal assumptions about the data that would arise from these less obvious indications that would tax the ability of any detector.

For this purpose the SNR is defined for each scan at the indication resel.

$$SNR = \alpha \frac{A_0 - \mu_{noise}}{\sigma_{noise}} \tag{6.2}$$

The standard deviation of the noise distribution is σ_{noise}. When α is unity, Equation 6.2 recovers the SNR in the original data. When α is scaled down to zero, SNR is zero and the scaled amplitude A_j is equal to μ_{noise}.

In terms of amplitude decibels, the SNR is

$$SNR_{dB} = 20 \log_{10} \alpha + 20 \log_{10} \frac{A_0 - \mu_{noise}}{\sigma_{noise}} \tag{6.3}$$

All of the signals in the scans are shifted by the same distance in the logarithmic decibel space. The various scans have different initial SNR, and this scaling method preserves the ratios between their respective SNRs.

The plots in the remainder of this section show POD and PFA as functions of the average SNR_{dB}. Both ATD and DFD have a free threshold parameter that the user can use to choose an operating point of PFA, POD on the ROC curve: an amplitude threshold for ATD and a consensus p-value threshold for DFD. In the plots that follow, these thresholds are recomputed at every point in order to maintain a constant performance in the other performance metric. Specifically, in the POD plots, the thresholds vary for both detectors in order to maintain a constant PFA equal to 5 percent. In the PFA plots, the POD is maintained at a constant value of 99 percent.

Scaled-down Indication 1

Figure 6.25 shows the Probability of Detection for scaled SNR at Indication 1, for a fixed PFA equal to 0.05. The figure is produced by taking all scans that view Indication 1, scaling their amplitudes down while maintaining constant ratios of SNR between the various scans. Thresholds on amplitude and consensus p-value, respectively for ATD and DFD, are adjusted at each point on the plot in order to maintain the PFA at 0.05. All scans viewing Indication 1 have their amplitude varied in order to maintain the same reduction in SNR. The SNR uses the same noise distribution as the LEN. At each point in the Average SNR domain, the ATD and DFD are evaluated to determine a POD. The POD is plotted relative to the average SNR, averaged over the scans that view the indication location. The DFD outperforms the Amplitude Threshold Detector by detecting signals below 15 dB.

Figure 6.25 shows the POD for Indication 1, for both the DFD and ATD as functions of the average SNR of the inputs. The ATD experiences a sudden drop in POD near an average SNR of 15 dB. This is the point when the scan with the highest signal amplitude distribution slides past the amplitude threshold. Due to its nature as a threshold detector on the amplitude, the ATD response is very binary and depends almost exclusively on the single scan with the highest mean signal amplitude. These characteristics will follow the ATD for the rest of the results in this section, on the data from Indications 2 through 5.

In contrast, DFD maintains reliable detection down to 0 dB average SNR. This is largely due DFD's ability to behave similar to a logical OR test when at least one of its inputs has a higher SNR, almost regardless of the low SNR of the others. DFD also exhibits a more gradual roll-off in detection probability, due to its inclusion of information from multiple scans into the consensus test. Each input will fall below a detectable amplitude as the scale factor is reduced, but at different scale factor values.

This contrast in transition behavior is a significant difference. It indicates that the ATD can be very sensitive to a choice of PFA or amplitude threshold. The DFD is more forgiving, so there is less risk of an indication falling off a “cliff” if the detected amplitude is near a detectable threshold.

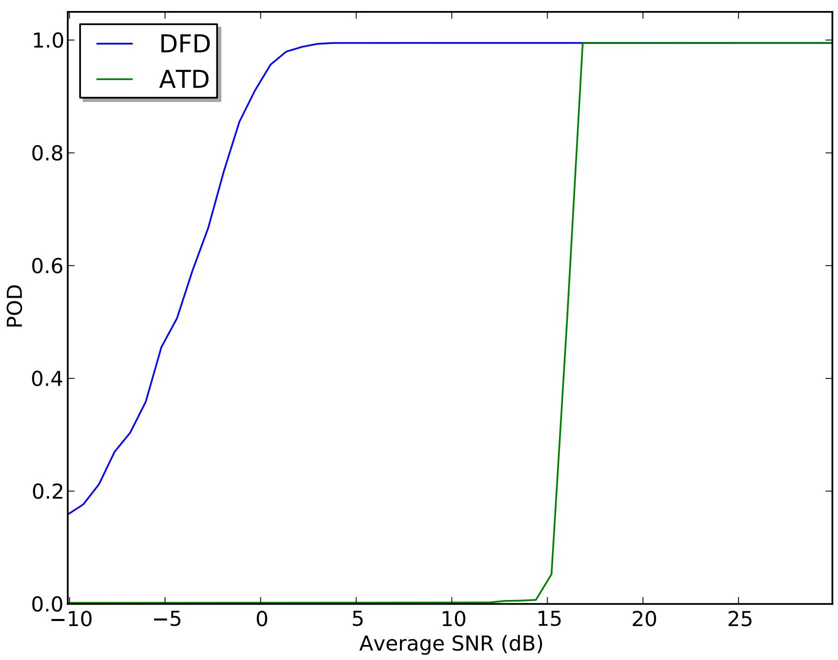


Figure 6.25: Probability of Detection for a scaled-down Indication 1. These curves show the hypothetical POD for indications that are identical to Indication 1 but with diminished amplitudes. Each scan has its amplitude at the resel containing the indication reduced while all other amplitudes in their respective noise distributions are held constant. The amplitude reductions are applied in such a way that the Signal-to-Noise Ratio (SNR) is reduced across all overlapping scans by the same factor. The average of the SNRs from each scan forms the plot domain. The Amplitude Threshold Detector (ATD) and Data Fusion Detector (DFD) produce a POD at each value for Average SNR. Detetor thresholds are adjusted to maintain a constant PFA of 0.05. The DFD outperforms the ATD over the entire domain, detecting signals with high reliability down to an Average SNR of 0 dB. The DFD also avoids the sharp drop in detection performance that is characteristic of the DTD.

Figure 6.26 shows Probability of False Alarm for scaled SNR at Indication 1, for a fixed POD equal to 0.99. The amplitudes of all scans viewing the indication are all scaled to maintain the same shift in SNR_{dB} relative to each other. PFA for both the ATD and DFD are plotted against the average SNR. The DFD achieves a lower PFA than the ATD when the SNR is above 0 dB.

DFD PFA transitions from noisy randomness to detection as average SNR increases. This transition occurs when the DFD signals emerge from the noise, and it happens suddenly. This is because there are few resels in the disc that have multiple overlapping scans with high amplitude and small consensus p-value. The ATD has a more gradual decline in PFA because the false calls are coming from the increase in amplitude relative to other amplitudes. It is only the highest amplitude among the

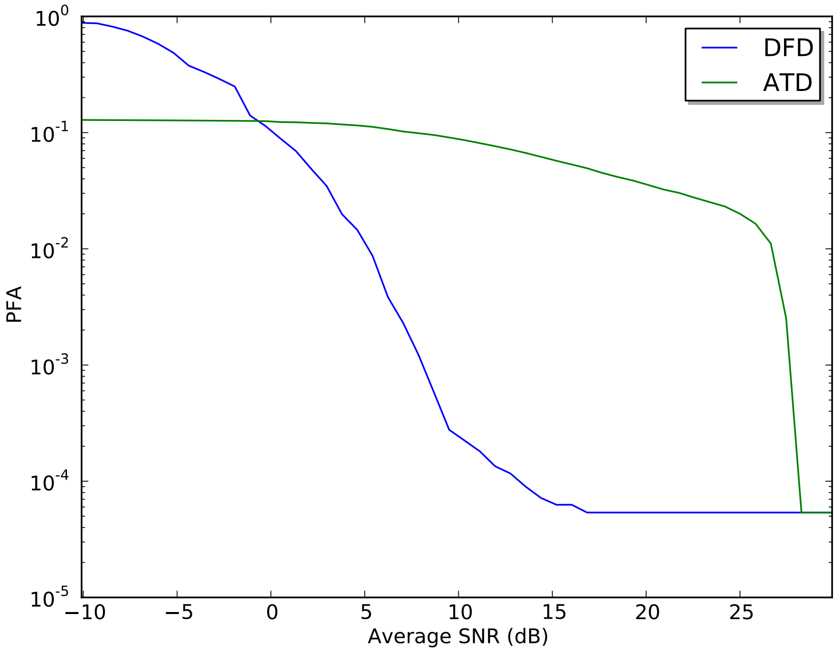


Figure 6.26: Probability of False Alarm for a scaled-down Indication 1. These curves show the hypothetical PFA for an indications that are identical to Indication 1 but with diminished amplitudes. The amplitudes from the scans viewing Indication 1 are scaled as in Figure 6.25. A PFA is estimated from the DFD and ATD detectors at each Average SNR, with POD held constant at 0.99. The DFD PFA is degraded below an Average SNR of 0 dB, but it greatly outperforms ATD at the more useful, higher SNR range where the indication is actually detectable (see Figure 6.25).

overlapping scans that controls the performance.

In terms of practical usage, the sharp transition in behavior in PFA is desirable because an inspector will start by investigating the resels with the highest DFD p-values. These resels correspond to the high SNR, low PFA region on the right side of the figure. There, PFA is very low, so any resel typified by this part of the curve will have a high likelihood of being a true indication. The inspector then works back through to lower SNR as he or she evaluates each highlighted resel.

Eventually the inspector will encounter the knee-like transition in the PFA curve, when the rate of false calls will increase markedly. The inspector will likely see this change, and will realize that more and more of the continued effort will be spent on what are clearly not defects.

Meanwhile, Figure 6.25 indicates that, not coincidentally, the POD decreases in this low SNR region where DFD PFA increases. This is the point where DFD is approaching the physical limit of detectability. POD and PFA simultaneously exhibit behavior that comes closer to randomness rather than detection when the SNR drops below that point. It rapidly becomes very inefficient to attempt inspection in that regime.

Scaled-down Indication 2

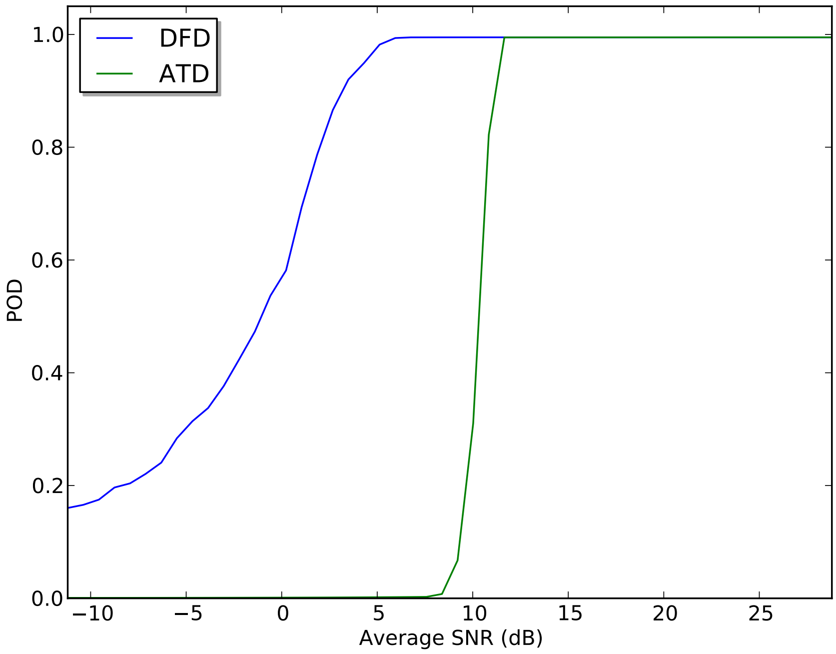


Figure 6.27: Probability of Detection for a scaled-down Indication 2. The ATD is able to detect Indication 2 down to a slightly lower Average SNR than Indication 1 (see Figure 6.25). This is due to the higher original amplitude of the KL 0° scan in Indication 2 compared to the original QR −5° amplitude in Indication 1. As with the scaled-down Indication 1 in Figure 6.25, the DFD outperforms the ATD. Compared to Indication 1, DFD POD starts to drop off at a higher Average SNR (5 dB vs. 1 dB). This is due to the fact that the initial SNRs for all of the scans viewing Indication 2 are closer to each other, so they all disappear into the noise simultaneously as their SNRs are reduced.

Figure 6.27 compares POD versus average SNR for Indication 2. As before, ATD has a sharp drop in POD, this time near 10 dB. In this case, the roll-off point for DFD is higher than on Indication 1, at about 4 dB, though it is still better than ATD. The higher roll-off is mainly due to the fact that Indication 1 has inputs with a wider variation in their initial SNR. As amplitudes scale down to a zero average, at least one or two of the input scans on Indication 1 still have a higher SNR. Indication 2 has inputs with initial SNRs closer together, so they tend to approach zero

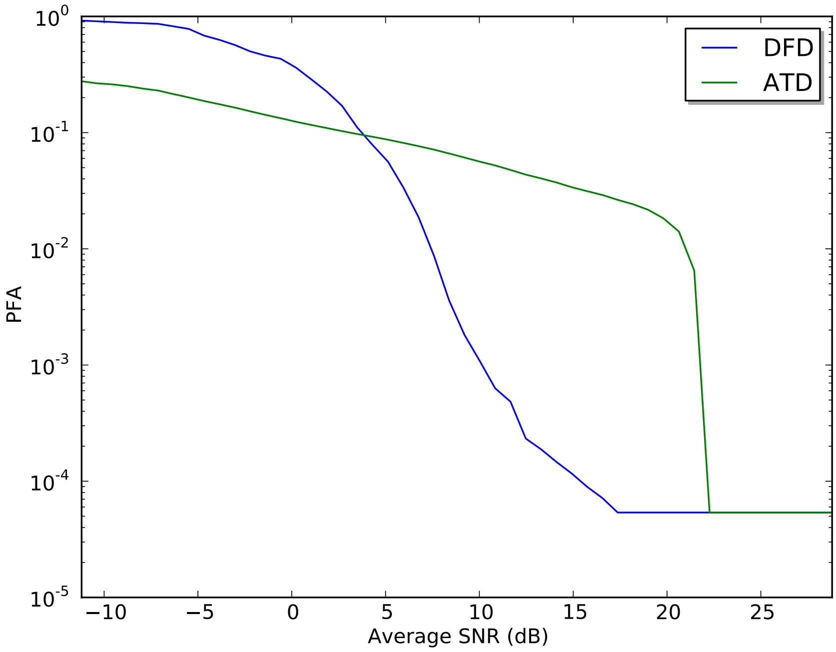


Figure 6.28: Probability of False Alarm for a scaled-down Indication 2. Indication 2 PFA shows features similar to the scaled-down Indication 1 in Figure 6.26. The DFD exhibits performance superior to the ATD over the detectable range of higher Average SNR. The crossover point between the two detectors is slightly higher than for Indication 1 (see Figure 6.26) and is close to the Average SNR at which the DFD POD starts to fall, for the same reason (see Figure 6.27).

simultaneously. There is no single input that can maintain a detectable consensus result for Indication 2, once average SNR drops much below 4 dB.

Figure 6.28 shows the PFA for scaled-down Indication 2. It is similar to the PFA plot for Indication 1 in Figure 6.26, with the difference that the crossover point is higher. In both cases the crossover point, where the PFA is equal between ATD and DFD, is close to the average SNR at which DFD rolls off.

Scaled-down Indication 3

The POD curves in Figure 6.29 are similar to those for the Scaled-down Indication 2. ATD suddenly fails to detect at about 10 dB average SNR. Detectability begins to drop for DFD at an average SNR of about 5 dB and declines gradually over a useful range of a few dB.

Figure 6.30 is very close in behavior to the PFA for Indication 2. The scans that view

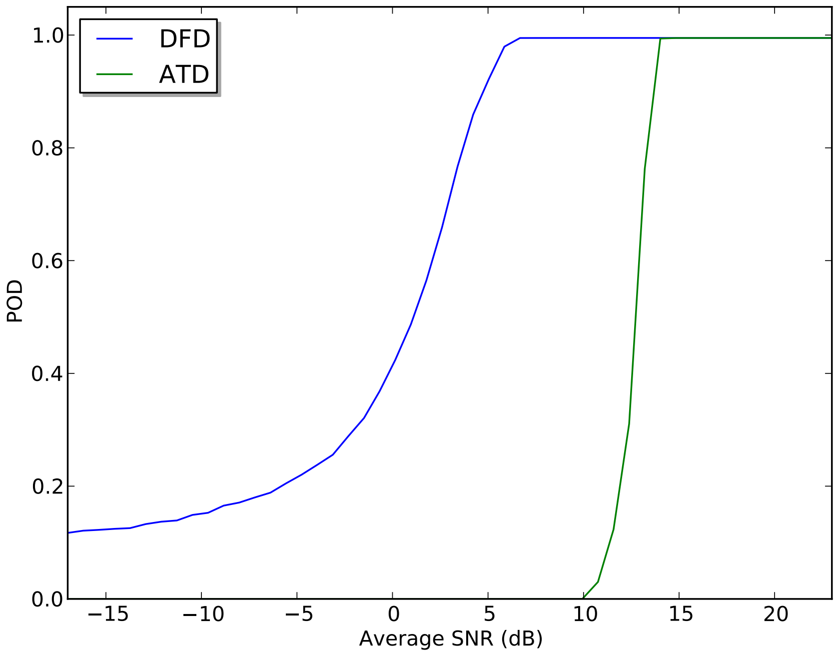


Figure 6.29: Probability of Detection for a scaled-down Indication 3. As with Indications 1 and 2, the DFD can reliably detect the indication down to a lower SNR than the ATD. Like Indication 2, the scans that view Indication 3 have initial SNRs grouped close together. The DFD consequently looses detectability near 5 dB Average SNR.

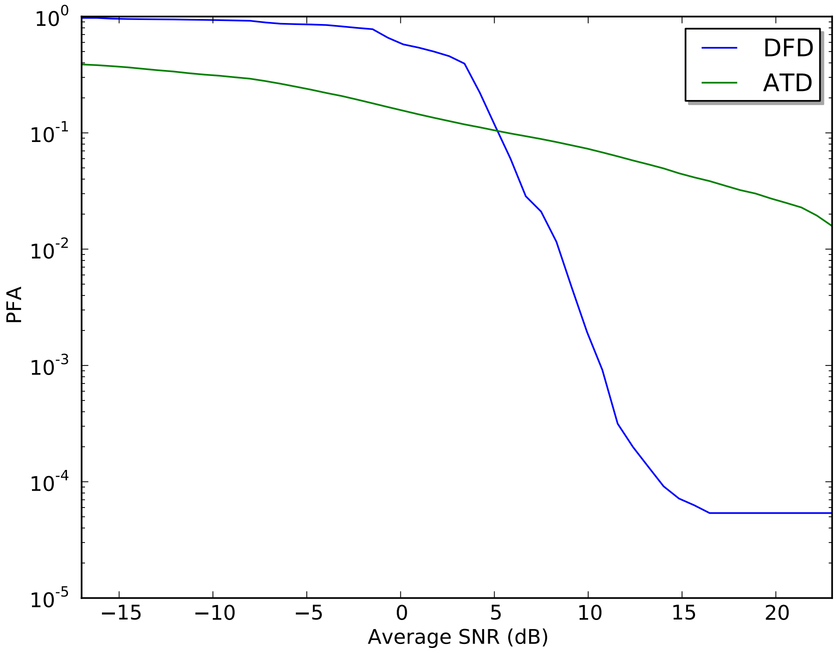


Figure 6.30: Probability of False Alarm for a scaled-down Indication 3. Results are similar to those for Indication 2 in Figure 6.28, with the two detectors crossing at an Average SNR close to where the POD begins to fall off. (See Figure 6.29.)

Indication 3 are close together in p-value, so they tend to lose detection together. This is reflected in the higher SNR level of the PFA crossover and POD roll-off, both of which are near 5 dB.

Scaled-down Indication 4

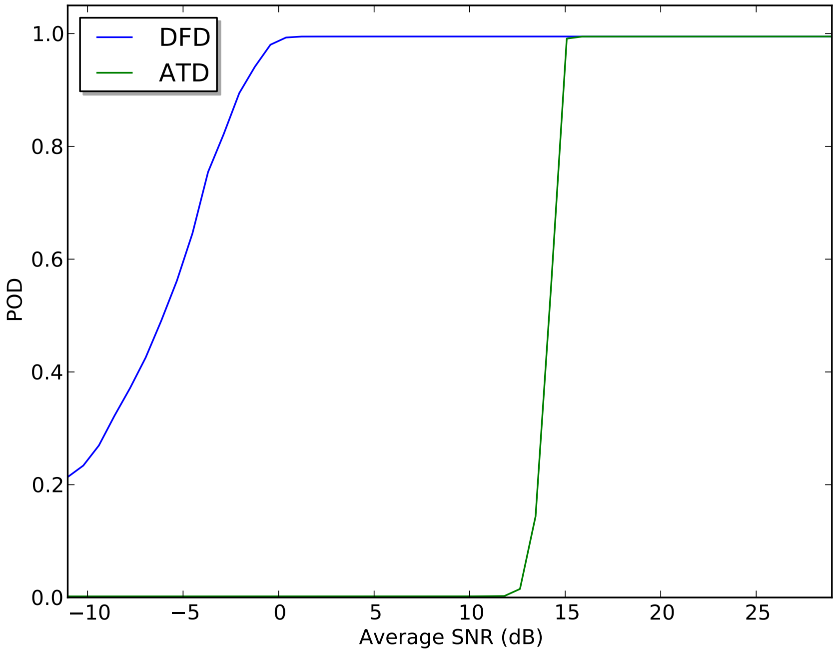


Figure 6.31: Probability of Detection for a scaled-down Indication 4. The DFD detects the indication down to an Average SNR of 0 dB, as with Indication 1 in Figure 6.25. The ATD suddenly fails to detect the indication once the scaled-down amplitudes drop below 15 dB.

Indications 4 and 5 show the greatest improvement for DFD over ATD. Figure 6.31 shows that there is a 15 dB difference in the amplitude detectability limit between the two detectors. In terms of amplitude, the DFD can detect an indication that is 5.6 times smaller than what the ATD can detect.

The two detectors have drastically different performance in PFA, as well. In Figure 6.32, the PFA for ATD is poor until an average SNR of 25 dB. DFD has an equivalent PFA at signals that are 20 dB lower. By about 13 dB, the DFD has already reached its very low minimum value.

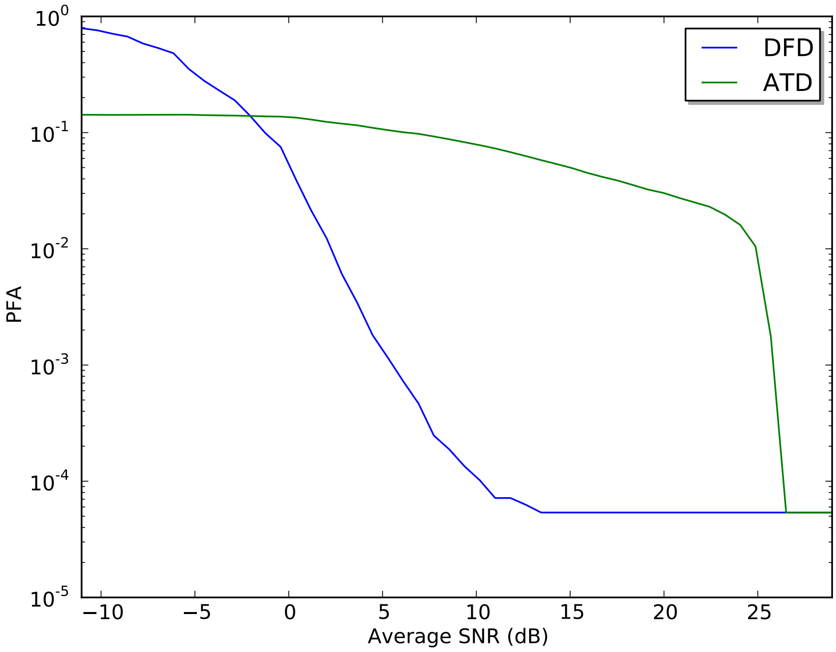


Figure 6.32: Probability of False Alarm for a scaled-down Indication 4. Results for both detectors are very similar to those for Indication 1 in Figure 6.26.

Downscaled Indication 5

Indication 5 has the highest coverage for any of the resels, with ten scans that view it. The high number of overlapping scans makes no difference to ATD, which still cuts out from detection at 18 dB average SNR in Figure 6.33. Indication 5 has inputs with a wide range of initial SNRs. This allows DFD to detect with high probability, down to an average SNR of 0 dB.

What is even more significant is the very low PFA, down to 5 dB average SNR, as shown in Figure 6.34. The PFA appears to flatten out at an asymptotic minimum at 11 dB, and the POD is already practically equal to one by the time the average SNR is that high. This means that, in terms of PFA and POD, the DFD detects an indication similar to Indication 5 with an average SNR at 10 dB as well as a similar indication with higher SNR.

Remarkably, DFD performs this well in both POD and PFA, even though the ATD detects nothing at all unless the average SNR is at least 8 dB higher. Even at that SNR, the PFA for such an indication is a full three orders of magnitude higher with ATD than with DFD.

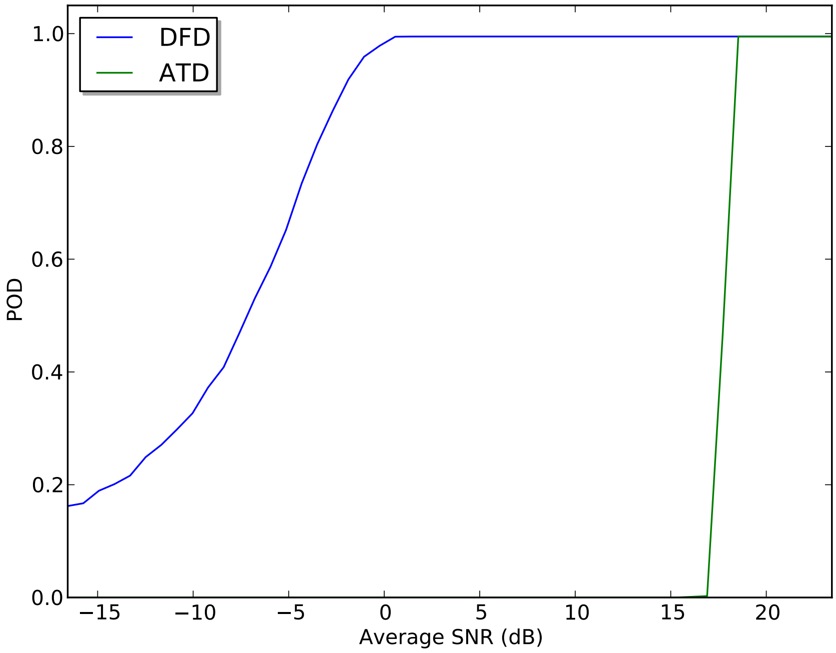


Figure 6.33: Probability of Detection for a scaled-down Indication 5. This indication has the greatest separation in dB between the detectability limits of the two detectors. The original amplitudes from the viewing scans are low, which means that the ATD detection limit is at a higher SNR than Indications 1 through 4. At least some of the input scans have a high SNR, however, so DFD is able to detect even when the Average SNR is close to 0 dB.

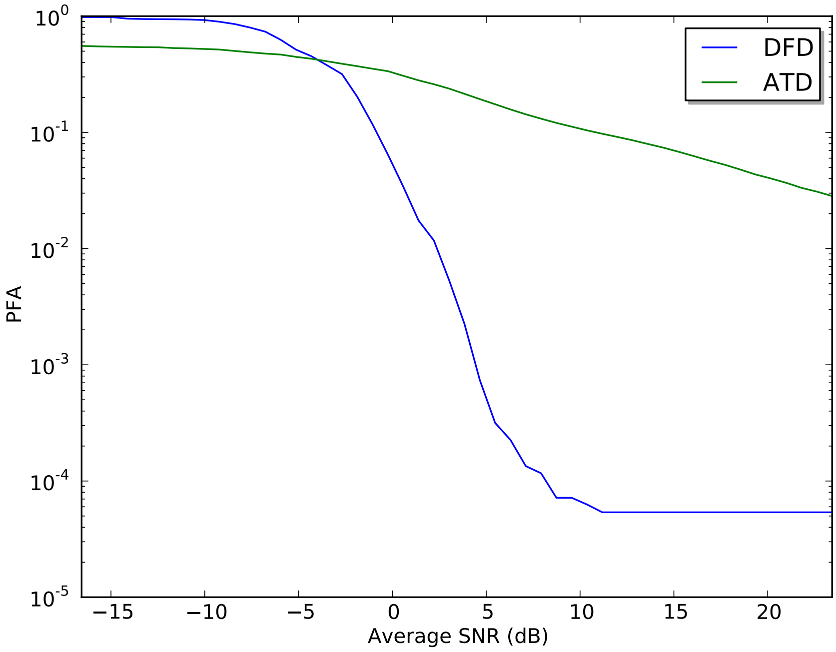


Figure 6.34: Probability of False Alarm for a scaled-down Indication 5. The ATD has a higher PFA with Indication 5 than with the previous indications. The scans viewing Indication 5 have lower amplitude, so the ATD requires a lower amplitude threshold to maintain a 0.99 POD. The PFA performance for DFD is not affected by the low amplitudes because the noise levels are also relatively low.

Downscaled Indication summary

One key characteristic of the DFD emerges from all of the scaled-down indications which sets it apart from the ATD. There is a clear range of a few dB in average SNR, over which PFA transitions from a useful, very low value to a high value indicating randomness. POD is simultaneously degraded over those very low SNRs, rolling off from a value practically equal to unity at the midpoint of the same range. This roll-off region likely indicates the physical limit of detectability.

Both POD and PFA have consistently good performance down to that detectability limit, but it would make little sense to continue once the automated inspection returned resels that are below that limit. The cost in terms of PFA quickly increases, in the sense that an increasing fraction of those resels would be false calls. The benefit in terms of POD quickly decreases. As POD rolls off from close to one it becomes very unlikely that the resels highlighted and prioritized will be true flaws. This rapid transition in both POD and PFA means that there is a relatively clear point at which to terminate an inspection.

In contrast, the ATD results for POD show a very sudden, harsh drop for a given PFA. The PFA, on the other hand, is nearly constant or slightly downward sloping with respect to average SNR. This means that, if resels to highlight for the inspector are prioritized by the ATD, there would be a false call rate nearly constant throughout the inspection process. Consider an inspection using the ATD, for a disc in which there were one flaw. At some point during such an inspection, the amplitude threshold would be lowered enough to detect the flaw, and the resel containing it would be presented to the inspector for evaluation. Before that point, however, the false call rate would have been roughly constant or only slowly changing. In that sense, the resels flagged for evaluation to that point would have been of a comparable likelihood to contain a reportable defect. There would have been no way for the inspector to anticipate that a true flaw would be coming from the detector.

The apparent limit to detectability means that the DFD performs at least as well as any other detector might. No other detector is likely to get any more out of the current data. This can tell the NDT staff designing the inspection procedures that, if they find a class of defects that DFD cannot detect well with the current data, the inspection hardware and/or procedures will need to change to improve the SNR. Also, if the inspection is improved in the future in any way that boosts signal amplitude or decreases noise, the DFD will automatically improve detection results via improved p-value separation, without needing to manually adjust any thresholds.

Again in contrast, the ATD PFA does not have a clear transition region, so it does not give the inspector any hint of when or at what point the ATD might detect a latent defect. This clearer transition with DFD is an advantage distinct from, and complementary to, the improved POD and PFA at a particular operating point.

6.7 Newly detected possible indications

Section 6.5 showed the outstanding performance that DFD achieves with the Known Indication data. That performance was shown to continue down to very low values for SNR in Section 6.6. This section continues to explore the capabilities of the DFD for the turbine disc, beyond improving reliability and lowering cost for detecting previously known flaws with the existing detection capability. Here the question of whether it is possible to detect indications that were not detected by the current, status quo detection procedures by searching for new defects within the Seeded Defect Disc beyond the five Known Indications.

If the indications detected in this section can be confirmed, it could point the way toward a potentially enabling gain of new capability for turbine disc manufacturers. Engine life estimates are based in part on the most severe known or assumed defect. If the inspection can provide the manufacturer with assurance that the worst defect is smaller, manufacturers could life the discs longer. This could reduce maintenance and replacement costs. It might also allow the use of new materials that are currently too difficult to inspect. Finally, there is even value in just giving NDT evaluation more, smaller defects against which to test future inspectors, techniques, and equipment.

The data to produce these new Possible Indications are exactly the same, since the data files come from the same inspection already executed for the detection analysis in Section 6.5. The physical inspection parameters are unchanged from the current procedure.

The goal is to make better use of the data, extracting information from the same data that have so far remained unused. The previous sections already demonstrated the ability of the DFD to adapt to local noise. Section 6.6 in particular showed that consensus can confirm a signal that, while small, is present in scans from multiple directions.

In this section are the five best candidates for possible new indications. They are labeled Possible Indications 6 through 10, ranked in order of increasing consensus p-value and following from the first five previously known indications. In the results to follow, the A-scans most relevant to evaluation for each of the Possible Indications are shown, just as they might be for the inspector using the automated inspection system. ROC curves for the LEN detectors on each scan, the ATD, and the DFD are generated in the same way as those in Section 6.5.

The resels with the lowest p-value over all of the model are from the five Known Indications. Some resels after those, with higher p-values, are resels that are adjacent to the five Known Indications. A few are resels that caught part of a corner or surface echo. These adjacent regions and structural features are excluded from the following as possible indications. Their data are included in the calculations for PFA, however, to accurately include the effect of this “bleed” from the structural features into the inspected region. Including these resels provides a very real example of the types of false calls that will arise as an inspector searches through resels with increasing p-value.

Figures 6.35 and 6.36 show the locations of both known and possible indications. Figure 6.36 shows that the indications are spread around the circumference of the disc. Interestingly, in Figure 6.35 it appears that all of both the known and possible indications, with the sole exception of Possible Indication 7, are in a narrow height range corresponding to the disc diaphragm. Also, all ten indications except Known Indication 5 are in the outer half of the cross section radially, from the diaphragm outward. This distribution of indications could be random, but might also be an artifact of the localization of impurities in the metal to this zone in r and z. Unfor-

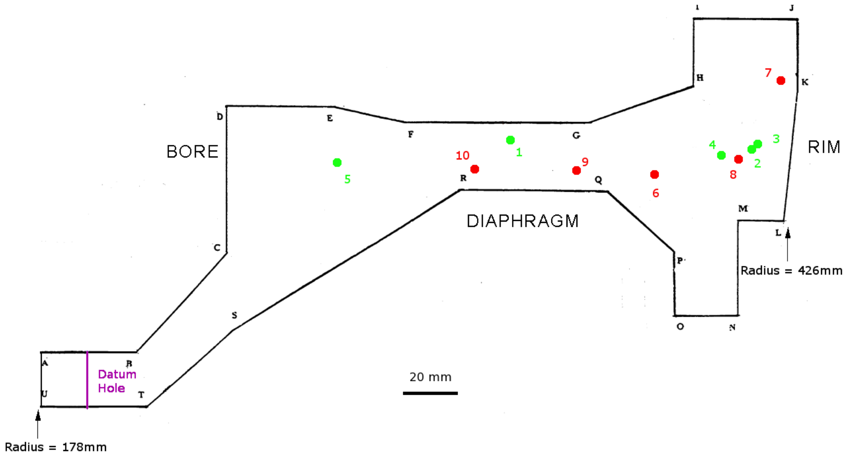


Figure 6.35: Seeded Defect Disc cross section, showing locations of the five Known Indications and five Possible Indications. Green circles mark the Known Indications, labeled 1 through 5. Red circles mark the new Possible Indications, labeled 6 through 10.

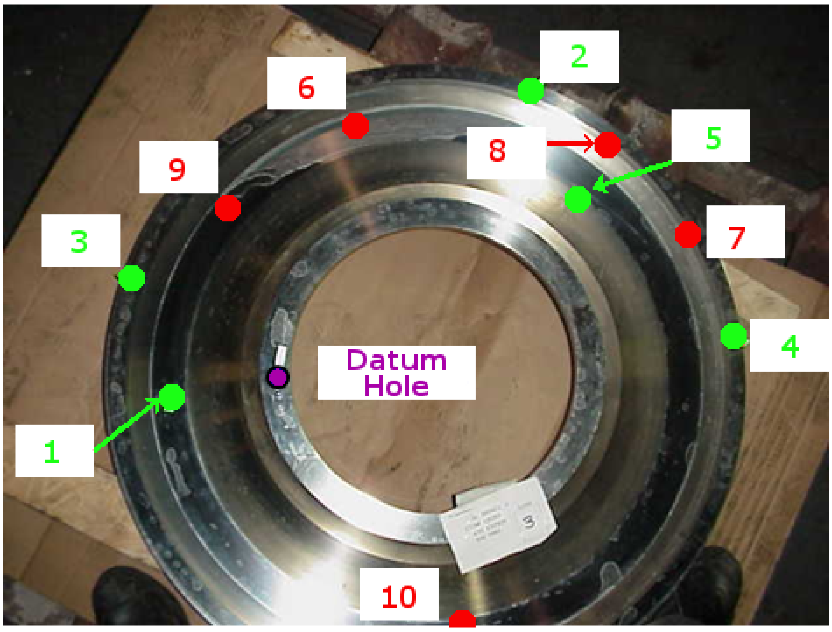


Figure 6.36: Seeded Defect Disc, showing locations of Known and Possible Indications in the r, θ plane. As in the cross sectional view in Figure 6.35, green circles mark the Known Indications, labeled 1 through 5. Red circles mark the new Possible Indications, labeled 6 through 10.

tunately, the author is unaware of any way at this time to determine the accuracy of this conjecture.

Possible Indication 6

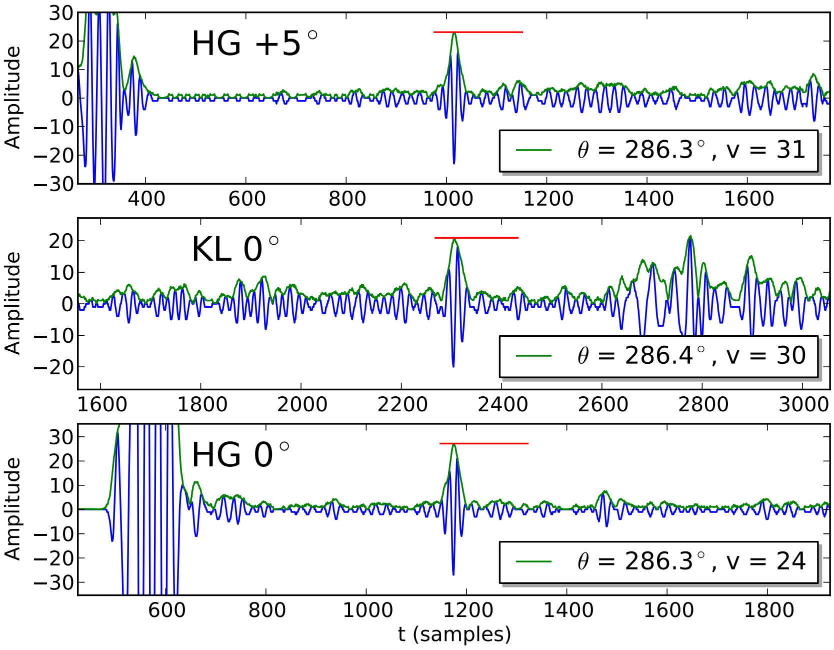


Figure 6.37: Possible Indication 6. The blue lines are the unprocessed A-scan amplitudes. Green lines show the Hilbert envelopes of the A-scans. The red bars show the amplitude and extent of the resel that contains the A-scan, for which the given A-scan determines the maximum amplitude. Angles are measured clockwise from the Datum Hole on the BA surface. The scan step is the number of 1.2 mm radial/axial steps from the starting corner to the circlar ring in which the A-scan is located. The linear amplitude scale is according to the discretization units of the data acquisition hardware. The eight-bit samples for the raw amplitudes have integer values between -127 and 127, and the gain is set such that the ATD threshold is 64. All six scans covering the resel marked with red have an amplitude well below the ATD threshold. Three scans show an amplitude that differs significantly from the local noise. This resel is the best candidate for a material flaw in the Seeded Defect Disc other than the five previously known indications.

Figure 6.37 shows the most promising of the possible new indications. Three of the scans covering this resel have a significantly low p-value. The A-scans from the HG +5°, KL 0°, and HG 0° scans are plotted in Figure 6.37. The HG −5°, PQ 0°, and PQ −5° scans also cover this resel. Their p-values are too large to contribute to the detection, and their A-scans are not shown.

In all three A-scans, there is a visible blip in the signal with a magnitude between 20 and 30 units, out of an ADC maximum of 127. Current inspection procedures call for an amplitude threshold of 64. All scans would pass over this possible indication without triggering the detection criterion, as they are all well below the threshold.

The amplitudes for the three scans plotted in Figure 6.37 are less than half the magnitude of the threshold, but they are significantly higher than nearby noise levels.

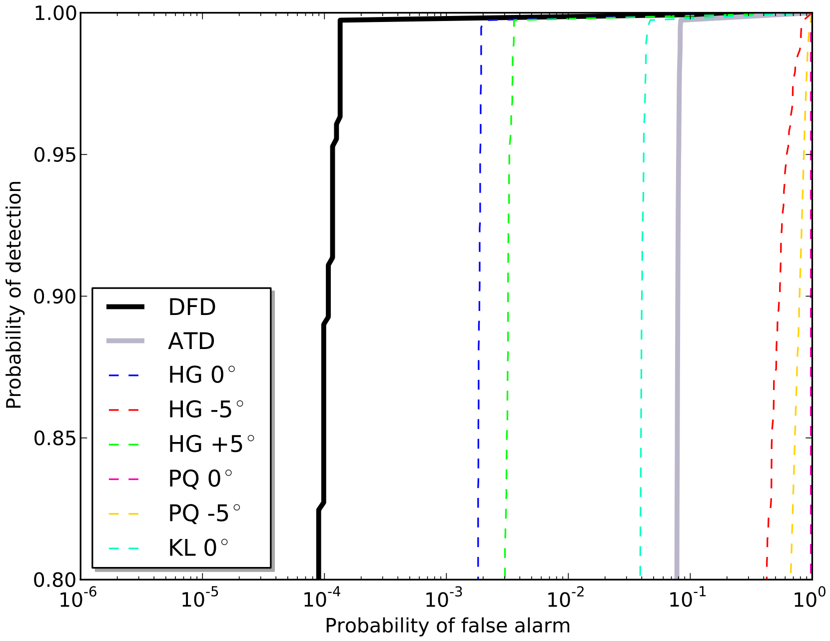


Figure 6.38: ROC curves for Possible Indication 6. Amplitude thresholding with the ATD has a relatively high minimum PFA of about 8 percent. Three of the individual scans viewing this location can outperform the ATD with their LEN dectors. HG 0° and HG +5°, in particular, indicate a low false-call risk for this possible defect, due to low local noise. The DFD significantly improves specificity over all other detectors. It is noteworthy that there is a steep drop in PFA for the DFD, just as there is with the known indications in Section 6.5.

Figure 6.38 shows the ROC plots for the LEN detectors on the six individual covering scans and the DFD consensus tests. The DFD achieves a low PFA, even up to high POD. The DFD PFA is not as low as the immeasurably small PFAs for the Known Indications. However, at 0.01 percent it would certainly be worth the inspector's time to examine.

Possible Indication 7

For the remaining four Possible Indications, the results are not as certain as for Possible Indication 6. Figure 6.39 shows the A-scans from the two scans with the lowest p-value for Possible Indication 7. The amplitude in the JI $-5°$ scan is actually relatively high for a Possible Indication. Unfortunately, it comes from a region with

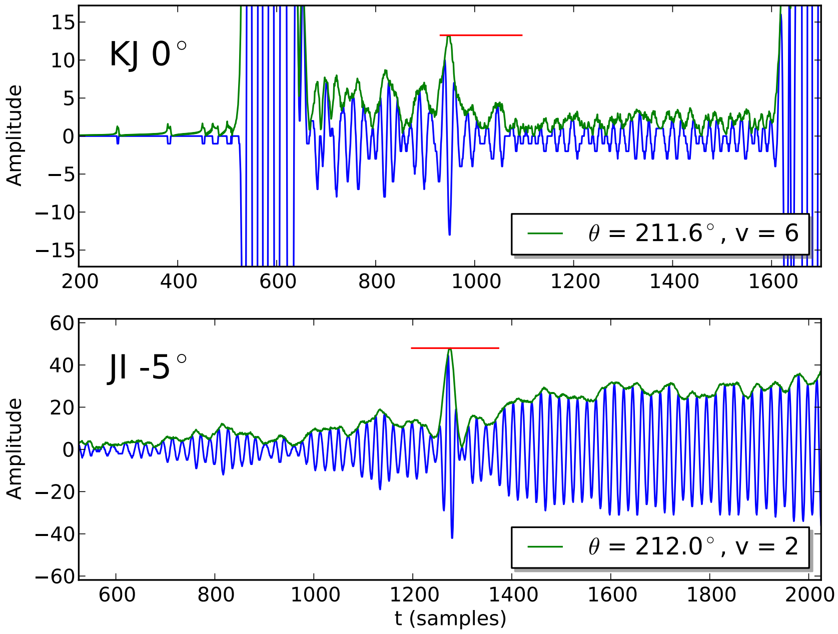


Figure 6.39: Possible Indication 7. Both scans are picking up some noise from proximity to Corner J, especially at later times in JI −5°. The amplitudes are still high enough to be distinguished from noise with high likelihood.

moderately high noise amplitude near Corner K, on the rim. It could be picking up corner and surface wave reflections causing higher background noise.

The second lowest p-value comes from the KJ 0° scan. The noise amplitudes are lower in the ring of resels from this scan, due to the different direction that avoids skimming the KJ surface.

Possible Indication 7 is located in a region viewed by scans from three different surfaces, for a total coverage of eight scans. It is clear from the ROC curves in Figure 6.40 that only the JI −5° really contributes to a defect conclusion. There are two orders of magnitude separation in PFA between JI −5° and KJ 0°. The other seven scans actually decrease the likelihood.

DFD shows a higher PFA and lower POD than the JI −5° scan alone. Possible Indication 7 is an example of the ability of the DFD to stay close to an OR test when only one input effectively flags it as a defect. Up to a POD of 94 percent, the difference in PFA between the JI −5° LEN detector and the DFD is less than 0.001.

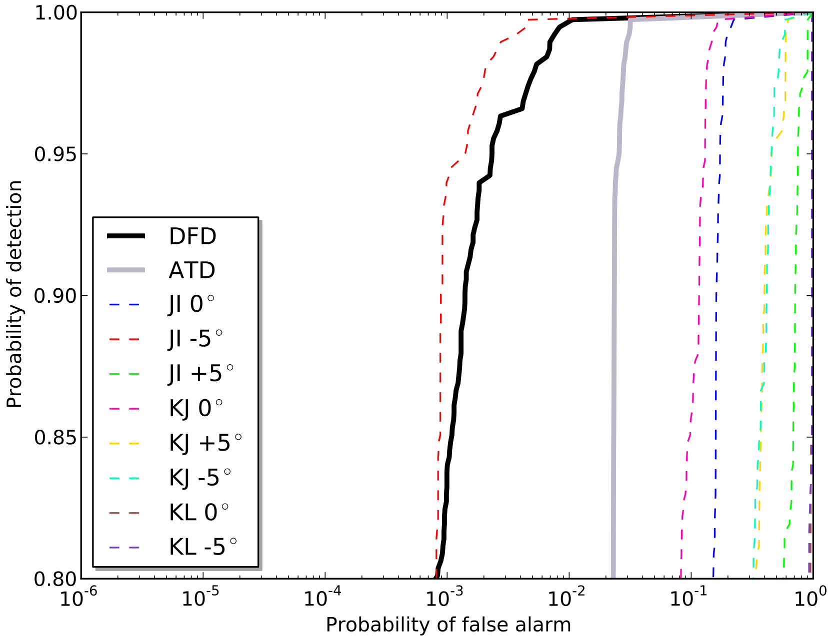


Figure 6.40: ROC curves for Possible Indication 7. Only JI $-5°$ gives significant evidence of a defect at this location. The other seven scans lower the likelihood of an indication, to the point that DFD actually has a lower POD or (equivalently) a higher PFA than the JI $-5°$ LEN detector.

Possible Indication 8

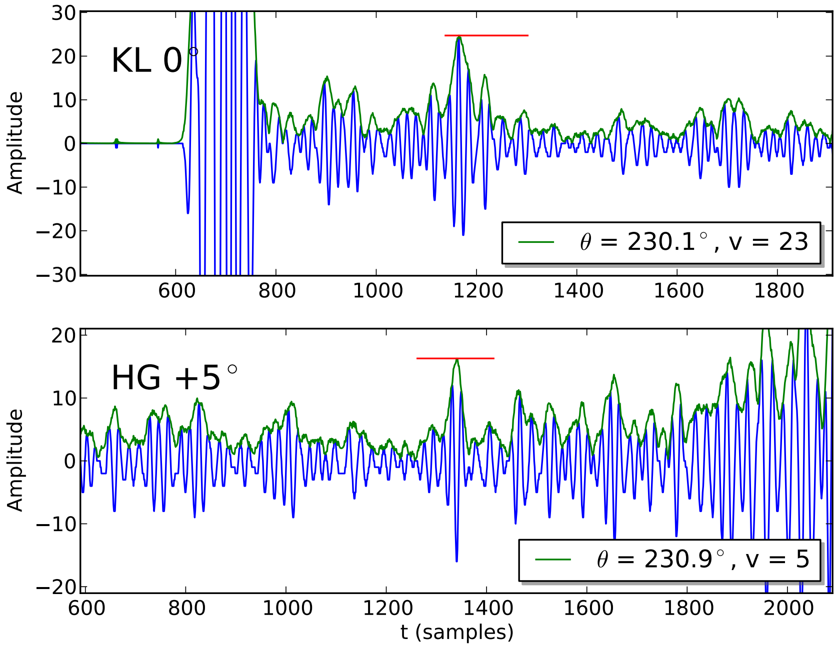


Figure 6.41: Possible Indication 8. The HG $+5°$ resel does not have a large amplitude, especially relative to the rest of the A-scan. The fact that it is coincident with a moderately high amplitude in the KL 0° scan nevertheless increases the consensus p-value from the DFD to identify this resel as an indication.

Possible Indication 8 is also in the rim. Figure 6.41 shows the A-scans for KL 0° and HG +5°. None of the scans that cover the location of Possible Indication 8 has a very high amplitude, even relative to their local noise levels. Many of them are in a moderate range, however, and it is the combined effect of these multiple inputs that drive detection with the DFD.

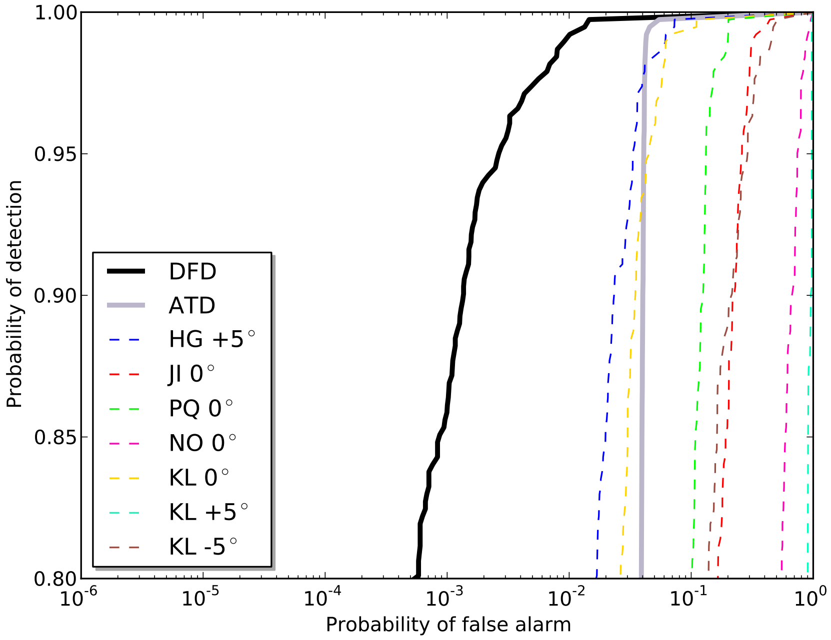


Figure 6.42: ROC curves for Possible Indication 8. None of the individual LEN detectors has a very low PFA. It is the combined effect of multiple, moderately low p-values from the LEN inputs that gives the DFD a low p-value and an ROC curve that extends to low PFAs.

Figure 6.42 shows that only KL 0° and HG +5° have individual LEN detectors with better performance than the ATD over any part of the ROC curve. Even for these two scans, they dominate the ATD ROC only at lower PODs.

The more rounded ROC curve for Possible Indication 8 illustrates the trade-offs between POD vs. PFA. At a 99 percent POD, the inspector must tolerate a false call rate of approximately 1 percent. At 95 percent POD, the false call rate drops to 0.3 percent. It is up to the NDT specialist at the manufacturing company to weigh the trade-offs in risk of non-detection versus cost of continued inspection, according to that curve.

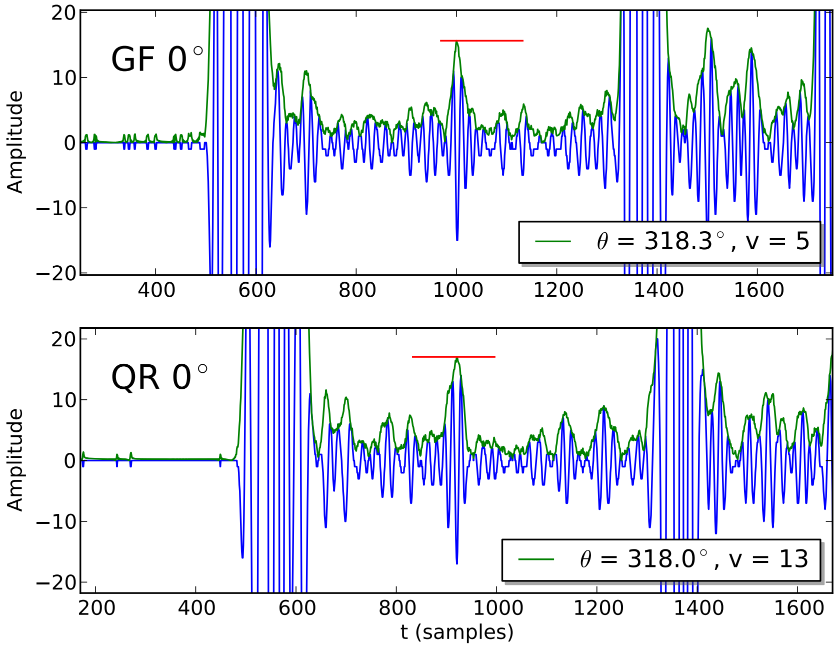


Figure 6.43: Possible Indication 9. Both of the scans contain an amplitude that in isolation would probably not be sufficient to flag an indication. The DFD, however, finds that it is unlikely for these signals to occur in the same place by chance. Possible Indication 9 is located within the thin diaphragm of the disc.

Possible Indication 9

Possible Indication 9 is in the diaphragm of the disc. Figure 6.43 shows the two normal-to-surface scans, GF 0° and QR 0°, that view this region.

Figure 6.44 shows that no LEN detector from an individual scan returns a highly conclusive result. Two (GF 0° and QR 0°) are simultaneously in moderately high likelihood of a defect. Like Possible Indication 8, it is the combination that increases overall likelihood in the DFD result.

This might be a case where an inspector would use the result from this automated inspection method to return to the disc for additional data acquisition. Possible Indication 9 is close to the QR surface, so beam spread and attenuation are relatively low. An inspector could try to gather more information on a marginal indication like this one, for example by using a higher frequency probe or changing the incident angle.

The automated inspection provides value by showing where to investigate further, even when its results are inconclusive on their own. There would be more work and

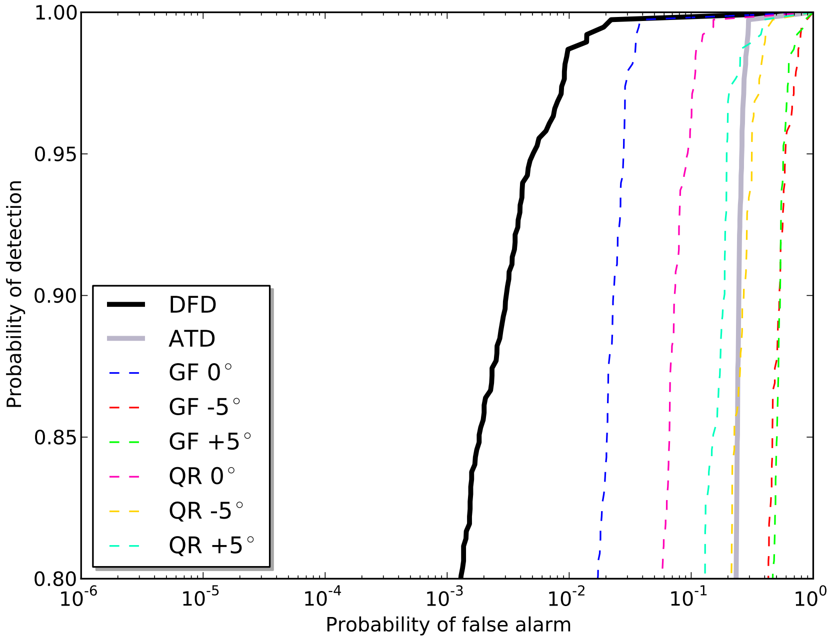


Figure 6.44: ROC curves for Possible Indication 9. As with Possible Indication 8 in Figure 6.42, a few moderately low LEN p-values together lead to detection by the DFD.

time required if this type of more in-depth analysis could not be planned for. It is the DFD that pointed the region out to the inspector, even if the data that were flagged are not perfectly sufficient.

Possible Indication 10

Possible Indication 10 is also in the diaphragm. As the last of the five Possible Indications, it has the lowest likelihood of representing a defect. Figure 6.45 shows A-scans from the two scans with the lowest p-value. Like Possible Indication 9, these are the two normal-to-surface scans, GF 0° and QR 0°.

The DFD ROC curve in Figure 6.46 shows that it is not possible to reach a PFA lower than 0.3 percent, even at a low POD. With such a high PFA for this possible indication, the human inspector viewing these possible indications might decide at this point to cease the evaluation. He or she could then make a determination for the disc, knowing that all resels yet to be viewed would have an even lower likelihood of containing a flaw.

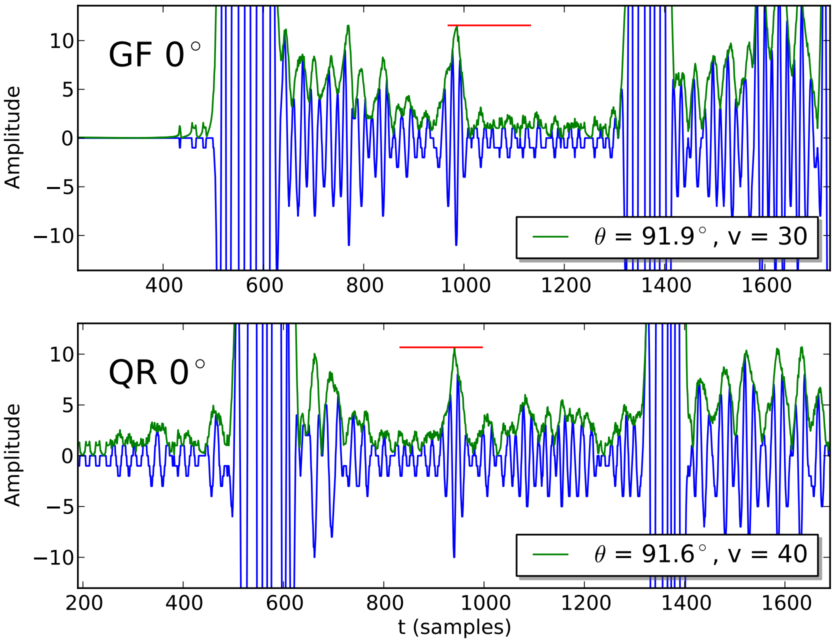


Figure 6.45: Possible Indication 10. This potential flaw has the lowest likelihood of the five Possible Indications. Like Possible Indication 9 in Figure 6.43, it is in the thin diaphragm of the disc.

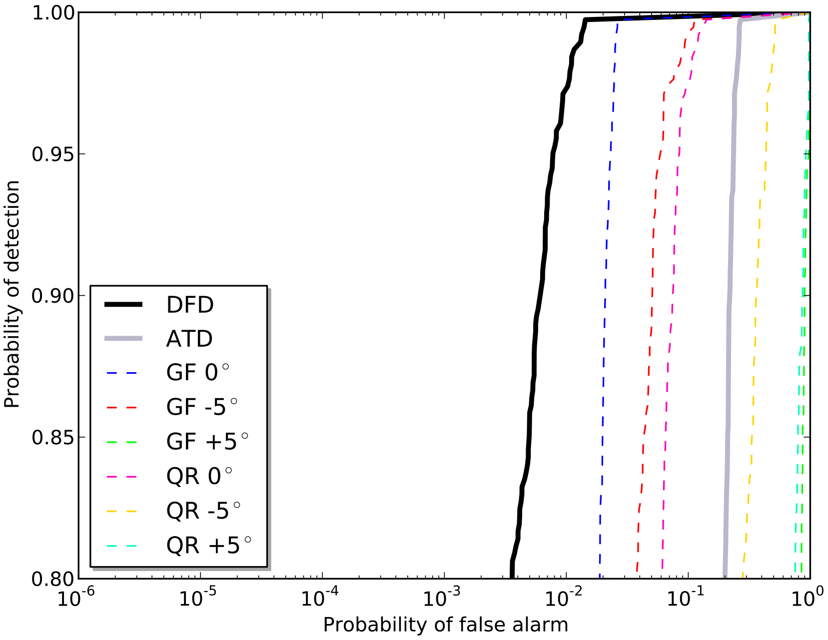


Figure 6.46: ROC curves for Possible Indication 10. The PFA for a DFD that could detect indications similar to Possible Indication 10 is as high as 0.5 to 1 percent. This might near the limit of what is useful for the human inspector to evaluate.

Possible Indications Summary

Possible Indication 6 has a sharp drop in ROC curve. For Possible Indications 7 through 10, the ROC curves vary in the steepness and curvature. In all of them, the POD starts to fall off to 99 percent at a PFA of about 1 percent.

Possible Indications 9 and 10 have a lower POD than the other three possible indications, at PFAs lower than 1 percent. However, they may be of interest for future work, since they are both in the web of the disc cross section. This is significant because this region is relatively thin and might allow for an independent means to confirm these two as true defects at a future time, using an alternative NDT technique. For example, current x-ray technology does not appear to be capable of detecting such a small flaw through a sizable titanium part. X-ray testing now would require sectioning the disc, which is not desirable for such a valuable test piece. As the available technology improves, it might some day be possible to use non-destructive x-ray inspection on the Seeded Defect Disc. If so, Possible Indications 9 and 10 will be first two to become accessible.

At this point it is important to recall the envisaged mode of operation with this automated inspection approach. Rather than forcing the inspector's attention to follow the scan in real time, as with the current procedure, the inspector will view resels after the automated detector has evaluated them, in increasing order of p-value. Each resel viewed in this sequence would represent a step along a theoretical, overall DFD ROC curve. Each of these steps increases POD and PFA for the resel to be inspected next. Eventually, the inspector decides there is no need to view any more possible indications. The ones he or she has recently viewed are clearly just noise, and the remaining, unseen resels have an even lower defect likelihood.

By the time the inspector reached an indication similar in nature to Possible Indications 7 through 10, their ROCs indicate how close the inspection is to a turning point. To have already seen indications like these with a 99 percent probability, the inspector would need to tolerate a false call rate of about 1 percent. In other words, he or she would typically have already viewed 1 percent of the resels in the disc, numbering about 1115. That may seem like a large number of resels to view. However, many of those resels will have been adjacent to previously detected indications, which the software might be improved to filter out. Also, if the inspector decided to quit at that point, there would be no need to see 99 percent of the disc, which saves the effort of inspecting almost 100 times as much data.

6.8 Summary

For the five Known Indications, the Data Fusion Detector can detect with very high POD and very low PFA. The results in Sections 6.3 and 6.2 showed the advantage of the LEN over the status quo inspection method, even without any data fusion. Sections 6.4 and 6.5 showed how the DFD effectively combines information from the individual scans that cover the same area. The DFD also adapts to local noise levels, giving credence to signals according to their amplitude relative to nearby noise.

Section 6.6 showed the limits of detectability for lower amplitude signals. The low-amplitude scenarios were highly relevant and realistic. The realism is due to the use of noise and signal amplitudes scaled down from real inspection data, taken from real indications. Results showed the DFD able to detect the scaled-down indications with high reliability until the amplitudes were comparable to the mean of their noise distributions.

Another very significant detection result is the identification of five new Possible Indications in Section 6.7. This finding has been reported to the disc manufacturer for possible confirmation. If any of the Possible Indications are confirmed, this will be a very important discovery. These new Possible Indications potentially indicate a capability to detect flaws with a smaller ultrasound signature than was previously possible, without any changes to the hardware or inspection procedure. This would be especially remarkable, considering this disc has been used for NDT evaluation and studied extensively for a number of years by competent inspectors. Any newly confirmed indications could be also be used in future NDT tests.

Chapter 7

Conclusions

7.1 Thesis review

This thesis demonstrated new capabilities for automation in Non-Destructive Testing. The results shown come from the application area of turbine disc inspection for aircraft engine manufacturers, but the development of the overall system and the specific techniques used were made to be widely applicable to a range of NDT areas. With an eye toward assisting and supporting the human inspector, this automated system analyzes the full data set from the entire turbine disc. It then prioritizes indications from the data by an automated detection algorithm, combining data from overlapping scans to increase certainty in the detection results and to better suppress false positives from various sources of noise.

Chapters 1 and 2 introduced the subject matter and the target application of aircraft engine turbine disc ultrasound inspection. Chapter 2 put forward an outline for the steps in the approach to follow.

Chapter 3 began the discussion of the approach, describing the concept of data registration, the challenges posed by practical registration of very large data sets, and the method developed to overcome those challenges. In particular, Section 3.5 described for the first time a method for computing the integral of an arbitrary function over the Pareto optimal set. The method can be used even when the Pareto set is unknown beforehand and is to be solved for by multi-objective optimization.

Chapter 4 contained the results from applying the registration methods to turbine disc data. The registration results showed that the turbine disc data could be registered to within 3 mm in all directions, with high reliability. In many instances, the registration accuracy was significantly higher. The results from the Pareto integral estimator demonstrated estimation of the mean and standard deviation of the Pareto set, providing a useful characterization of the residual location uncertainty in registration problems. These results represent a step forward for the academic field of registration between large scale, three-dimensional NDT data sets. The methods are sufficiently general and the results sufficiently accurate to recommend this work as a model for future NDT applications.

Chapter 5 returned to the development of the approach for this thesis. In this case the approach was for the stages of computation tied to defect detection, after registration had aligned the data sets with each other. The detection approach introduced two detectors developed concurrently by the author's colleague, Brierley: the Local Empirical Noise (LEN) detector and the Data Fusion Detector (DFD) [19]. It then described how they are to be applied to the turbine disc application. The LEN was shown to be an improvement over the status quo inspection method that uses an amplitude threshold. The discussion then explained how the data from scans coming into the same point from different directions could be combined, by merging multiple LEN detectors into the DFD.

Chapter 5 concluded with a description of a novel numerical technique, developed for this thesis, that allows practical implementation of the DFD on realistic inspection data. This Monte Carlo Importance Sampling transformation is an original contribution to the field, without which the computations involved in the DFD would be practically impossible.

Chapter 6 showed the final results for the turbine disc application. The chapter showed results for detection with data from the Seeded Defect Disc, a turbine disc with known internal, material flaws used for testing NDT techniques. The results showed that the DFD works extremely well on the five known indications. The results on scaled-down amplitude data showed that the DFD is able to maintain remarkably good detection, down to an average Signal-to-Noise Ratio of 5 dB. Finally, the detection results included the identification of five new indications, which are possible defects not previously detected. These indications still require confirmation before they can be declared a successful detection. Regardless of the outcome of that determination, the new possible indications are locations with the characteristics of small defects that had not been seen before. The possible indications and the ROC curves generated for them illustrated the capability of the automated NDT system to find questionable regions and bring them to the attention of the inspectors, with minimal human effort.

7.2 Summary of findings

The approach to automated NDT has many essentional components, but it is possbile to break them down into the two large stages of registration and detection. In Section 7.2.1 are some important considerations for registration that came to light during work on the turbine disc application. The recommendations will likely be helpful to anyone implementing a similar system for other applications, as well. Two novel contributions that were developed for this thesis are highlighted in Sections 7.2.2 and 7.2.3.

7.2.1 Registration requirements and capabilities

This thesis presents the argument that the appropriate way to perform registration is with the use of multi-objective optimization. One of the primary reasons for this is to find ranges on location uncertainty, rather than a single location value with no information on accuracy. Estimating location uncertainty requires the method for approximation of integrals on the Pareto front introduced in Section 3.5.

For a good back surface echo, registration needs surfaces that are close to parallel. If the back surface is angled more than a few degrees different from the front surface, the back surface echo drops significantly or disappears. This makes the back surface echo impossible to distinguish from the secondary echo noise in the A-scan, beyond where the back surface echo would be in time. The scans used for registration need to be scanned normal to the surface, a requirement that is not a problem in practice. It is also preferable that the front back surfaces not be too far apart, to limit attenuation and beam spread.

For good independence of information there must be at least one pair of surfaces with an axial component in their normal vectors, and at least one pair with a radial component. This is probably not an obstacle for most disc geometries. There only needs to be at least two pairs of front and back surfaces, with one pair not parallel to the other pair.

It is essential to have some type of circumferential datum. It is not reliable enough to depend on manually aligning the disc with the platform. For example, even 1° of difference causes about 7.5 mm difference in circumferential location near the disc rim. In other applications, a similar registration problem will arise when scan paths follow the structural symmetry. There must be at least one datum feature in the scan data that breaks that symmetry. It must be easily detected and located within both groups of data sets that are to be registered against each other.

7.2.2 Pareto integral estimation

Section 3.5 presented for the first time a method for estimating the integral of any function over the Pareto front. This method is potentially useful for many multi-objective optimization applications that require a characterization or summary of the theoretical, full set of Pareto optimal points. The technique makes use of a property that is unique to one type of multi-objective optimization algorithm, the Hypervolume Indicator (HI) algorithm. Previous work with the HI has used the uniquely derivable expression for the Pareto point density for other purposes. However, this is the first time that the Pareto density has been used to estimate a statistic over the Pareto set. All of the registration results in Chapter 4 use this technique to estimate the mean and standard deviation of the optimal mapped point locations for registered data. The registration accuracy information from the Pareto integral method informed the choice of resel size for detection.

7.2.3 DFD computation and importance sampling

Section 5.4 described an original contribution to the computation of Fisher's exact test for discrete inputs. The importance sampling transformation developed for this thesis enables the test to be applied when one or more of the input hypothesis tests has a p-value that is very small. In automated NDT, these are precisely the cases for which accuracy is most required, in order to prioritize properly the resels for human evaluation. Without the transformation, the consensus result would require a number of Monte Carlo samples that is orders of magnitude greater than the inverse of the smallest input p-value. The transformation creates a sampling distribution that is equally likely to produce a Monte Carlo point on either side of the marginal indicator function. This makes every sample maximally informative to the Monte Carlo summation. All of the detection results in Chapter 6 demonstrate the practical use of the method developed in Section 5.4.

7.3 Future work

Based on the results and findings from this thesis, there are certain modifications that are are either recommended from the lessons learned, or required in order to implement the automated inspection procedure. These changes to the inspection procedure, and the additions to the NDT automation software to make use of them, would be valuable directions for future work in this field.

7.3.1 Avoiding scan interruptions

The registration approach relies on the assumption that all scans from the upper side of the disc can be treated as a single registration group, and the scans on the lower side as a second group. This is likely to remain a valid assumption as long as all motion is under machine control and the disc is not disturbed, with the exception of the flip halfway through the inspection. All of the data used in this thesis were collected in this way.

It should not normally be a burden or impediment to put in place such a requirement, but it is not a requirement in the current procedure. The inspectors would need to become accustomed to the importance of this constraint. It would also be recommended to perform all scans for both sides of a disc at once, on the same day, to minimize temperature change effects. The data for this thesis were collected on separate days for upper and lower scans, some time apart. This made the change in ultrasound wave speeds a significant nuisance variable that registration needed to compensate.

7.3.2 Circumferential datum

As noted earlier in Chapter 3, it is necessary to register the discs in the circumferential direction, but both the geometry of the inspection and the disc itself are axisymmetric. There must be some type of marker to replace the datum hole, for discs to be inspected in the production chain. In other applications, parts lacking any inherent asymmetric features will have the same need for an artificial registration datum. Section 3.1.1 gives more information on the types of datum marker that might be used for this purpose.

7.3.3 Additional registration reference data

It might be possible to reduce the number of registration parameters or tighten the a priori uncertainty on them prior to registration with separate calibration routines. The data from these additional, per-disc registration calibrations would then augment the metadata that comes from the scanner position encoders. There probably is not much additional information that more data could provide on the rigid body motion uncertainty, beyond what is already contained in the inspection data. However, more targeted information could provide information on the physical and geometric parameters, other than disc orientation and position.

For example, a temperature sensor could measure and record the temperature of the water in the immersion tank, and associate those numbers with the ultrasound data files. This relatively simple step would require no other change to the inspection procedure. It could greatly reduce the uncertainty in the water ultrasound wave speed and, to a lesser extent, the metal wave speed uncertainty.

There are other measurements that are done periodically as part of the scanning equipment calibration. A simple version of some of these calibration measurements could be incorporated into an automated step during each disc inspection. These measurements would not be for the purpose of making physical adjustments to the hardware, as with maintenance calibration, but only for virtual adjustments via registration mapping. It would be too slow for a human inspector to do every time, but it would only require minimal additional time if the hardware were set up to run the routine automatically.

For example, to determine the location of the axis of rotation on the platform, relative to the tank coordinates used by the position encoders, the maintenance calibration calls for a metal sphere to be affixed to the center of the platform. The person performing the maintenance can then find the center of rotation by moving the probe to the location where the amplitude is maximal and the arrival time of the front surface echo does not oscillate while the platform rotates. An improved registration procedure could call for the calibration sphere to be left in place permanently. Then a brief, automated scan of the center of the platform would produce valuable data for constraining the rotation axis offset parameter. This could be incorporated into the automated inspection software during registration, requiring no additional time or effort on the part of the inspector.

A similar procedure could yield additional information on the platform tilt. If the probe were to continue scanning beyond the disc radially — either inward toward the center or outward beyond the rim — it could record the reflection from the platform under the disc. The analysis of the location of that echo could reuse the same software built into the registration for the disc front surface echo. This could find the degree and direction of any tilt miscalibration in the platform, which would not differ between the upper and lower scans. A small relative tilt could still exist between the disc and platform, due to foreign matter on the platform or a burr on the disc surface, which would still require registration to compensate. The advantage would be very good initial values and tighter bounds on the parameters.

As yet another example, the scanner could take a more direct approach to measuring water wave speed, not relying on a model for wave propagation speed as a function of temperature. The probe could record ultrasound data while moving toward or away from the disc or some other target. The speed would be measured by the time change in the front surface echo, while the water gap distance changes by a known amount. There would be no need for absolute distances, which are known with much less precision than relative distances.

A final recommendation for changes to inspection procedure would be to maintain an archive of all raw amplitude data and scanner metadata. A large volume of data is produced for every disc, and a full archive might have been infeasible in the past. The data are highly compressible, and data storage costs are continually decreasing. It is conceivable that further developments in automated inspection could improve reliability and detectability for smaller flaws, during the lifetime of a given disc. A full archive of raw data would open the opportunity to reinspect discs virtually, at any time in the future after the implementation of an improved analysis technique. These virtual reinspections would be possible even after the discs have proceeded on to finish manufacturing and entered end usage. This would allow revision to life estimates for discs after they are already in use and it is no longer possible to reinspect them physically. It could also provide extremely valuable retrospective information, to improve manufacturing and inspection in the event of a disc failure.

7.3.4 Additional scan data

Current inspection procedures call for three incident angles for each disc surface, whenever accessible: $0°$, $+5°$, and $-5°$. All three of these angles are defined such that the ultrasound beam remains within an r, z plane of constant θ. It would also be possible to scan a surface with the probe at an angle to that constant-θ plane. These additional scans from the side could be especially useful for areas with low coverage or to target resels with high flaw likelihood, to get data from more independent directions for the DFD. The data from this type of side-angle scan might have been more difficult to interpret in a manual inspection. An extension to the registration mapping functions would allow the data point locations to be determined without any effort by the human inspector, removing the possible disadvantage.

All of the modifications recommended above would require some amount of additional effort to implement. However, the approach to automated NDT outlined in this thesis was intended from the beginning to be an extensible framework, and the turbine disc application is only one of the many potential applications to which the methods could be extended with relative ease. The modifications listed above would require future work, but the effort would be mitigated by the fact that the framework is prepared to accommodate it.

Appendices

Appendix A

QNDE 2012 article

The following is a paper originally published in the *Review of Progress in Quantitative NDE* [133]. The weighted hypervolume optimization results discussed below represent a direction of work that is closely related to, but beyond, the registration methods used in the body of this thesis. The paper is reproduced here due to its relevance to the subject matter, as a claim of an innovative contribution and a possible direction for future work.

Data registration for automated Non-Destructive Inspection with multiple data sets

T. Tippetts, N. Brierley, and P. Cawley

Imperial College London, Mechanical Engineering, SW7 2AZ

ABSTRACT. In many NDE applications, multiple sources of data are available covering the same region of a part under inspection. These overlapping data can come from intersecting scan patterns, sensors in an array configuration, or repeated inspections at different times. In many cases these data sets are analysed independently, with separate assessments for each channel or data file. It should be possible to improve the overall reliability of the inspection by combining multiple sources of information, simultaneously increasing the Probability of Detection (POD) and decreasing the Probability of False Alarm (PFA). Data registration, i.e. mapping the data to matching coordinates in space, is both an essential prerequisite and a challenging obstacle to this type of data fusion. This paper describes optimization techniques for matching and aligning features in NDE data. Examples from automated ultrasound inspection of aircraft engine discs illustrate the approach.

Keywords: Reliability, Automation, Optimization

PACS: 02.50.Sk, 43.60.Pt, 81.70.Cv

A.1 INTRODUCTION

Automated data acquisition is increasingly common in industrial NDE. This is especially apparent in ultrasound testing, the technique central to much of this project, where the stored data now often consist of a full sequence of time-amplitude traces (A-scans), rather than a single amplitude-position (C-scan) image or the signal amplitude at a particular gate. The problem may be further exacerbated by multi-channel acquisition, using different techniques, viewing angles, or many-element array probes. All of these factors have contributed to a rapid increase in the volume of available data.

The collected data require analysis, and this is often done manually by a skilled operator sifting through countless images displaying subsets of the data. Given the expense involved in doing NDT, there is a high probability of an inspected part being fit for continued service; otherwise, it would in all likelihood be replaced without inspection. Consequently, the operator has to spend a large proportion of his/her time examining data where there are no defects to report. The resultant monotony can increase operator fatigue and hence the probability of a defect indication being missed. Additionally, typical UT software packages do not reduce the volume of data to be examined, prioritise regions of interest, fuse information in areas covered by multiple inspections or sensors, or identify changes since the last inspection. There is therefore scope to improve inspection reliability while reducing analysis time through automation, with significant potential cost savings.

A.1.1 Registration

Registration is the process of aligning datasets to a common coordinate system. Registration is essential for comparison or combination of data sets. There are many papers on registration, particularly for 2D images, and often in a medical context [1]. The publication [2] identifies the major steps in most registration procedures. The steps most relevant to the rest of this work are as follows:

Feature extraction: Many of the individual datasets are very large, and especially when processing several together, it is essential to select a subset for registration calculations. The extracted features should be associated with component geometry and therefore detectable across the different datasets.

Metric evaluation: A registration metric must be defined to quantify the quality-of-fit between matching features in multiple data sets. These metrics should be made as computationally efficient as possible. In practice their computation is often still the most time-consuming part of the registration, given the quantity of data under consideration.

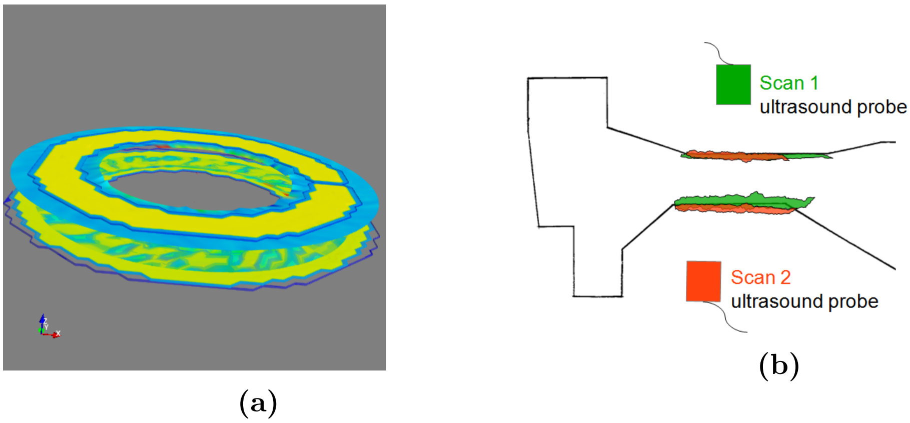
Optimization: The registration error metric must be minimized to find the best possible combination. The optimization must reach the global optimum and avoid any local minima. A wide range of generic algorithms may be used for tackling such problems, including multi-start [3], pattern search [4], and particle swarm optimisation [5]. The examples shown in this paper use a solver based on the hypervolume indicator [6].

Once registration is complete, the detection algorithm can combine data from multiple sensors or scans in overlapping coverage regions in several ways. Data subtraction, for example can detect changes over time or compare new data to a standard reference component as in [7]; an analogous problem is baseline subtraction in structural health monitoring (SHM) [8]. Other researchers have proposed techniques for enhancing flaw detection with data fusion, and have demonstrated the possible reliability gains [9,10]. A detection algorithm identifies indications in the fused or subtracted data and assigns to each indication an estimated probability that it represents a reportable defect. A second paper by the present authors in these proceedings discusses their approach to data fusion and detection. [11]

This paper focuses on one aspect of data registration, namely the optimization of registration metrics. In particular, there is a potential pitfall in the selection of points during registration. This poor point selection can cause data-fused detection to fail, and that failure is difficult or impossible to discover. This paper describes this pitfall, how it arises, and how to avoid it.

A.1.2 Application

Aircraft engine manufacturers inspect turbine discs at an intermediate stage of production by scanning each disc in an immersion tank. The disc is placed on a platform at the bottom of a tank full of water. The platform and disc spin while a computer-controlled arm moves an ultrasound probe above them, scanning each surface in a spiral pattern like a phonograph record player. A typical scan fires the ultrasound transducer in pulse/echo mode to cover each surface with 1 mm steps in both directions. The full A-scan time series is recorded at each step, and all data between the front and back surfaces are analyzed for full through-thickness coverage.



(a)

(b)

Figure A.1: Matching pairs of surface echoes for registration. (a) Upper and lower surface echo data, extracted from scans of each side of a turbine disc. (b) The front surface echo from Scan 1, shown in green, should match the back surface echo from Scan 2, shown in orange, and vice versa.

This procedure generates many gigabytes of data per disc. Under current procedures, an inspector takes hours to interpret and evaluate the A-scan data manually for each disc. There is clearly potential to make inspections faster with automated analysis.

Automation can also increase the inspection reliability by combining data from multiple scans. Considering a single given point inside the disc, it is clear that that same point may be visible from multiple directions, from scans of different surfaces. In the current practice of manual data analysis, the inspector views each of these scans individually. An automated detector could combine the information from overlapping scans to increase POD and decrease PFA simultaneously. However, registration is essential to align the data properly, using features present in the data to combine or fuse data successfully.

A.2 APPROACH

The registration features used in the turbine disc application are surface echoes. These surface echoes are extracted from of each of the hundreds or thousands of A-scans, and reconstructed into a 3D geometry. Fig. A.1 shows an example of two pairs of surfaces extracted from scans of a typical disc. To simplify the discussion, only two parallel, upper and lower surfaces are shown.

These surfaces are scanned from each side of the disc so the front surface of one scan should match the back surface of the other scan, and vice versa, as shown in Fig. A.1a. For each pair of surface echoes, a registration metric quantifies the lack-of-fit, which should be minimized. For example, the current version of the registration software uses a root-mean-square metric: the Hilbert-enveloped amplitudes from each front/back surface pair are subtracted from each other, and the metric is the root-mean-square of the residual differences.

One key point to understand is that there will be multiple pairs of surface echoes to align, and it is impossible to match all of them perfectly, simultaneously. Tradeoffs are possible between these metrics, in which one may be improved at the expense of another. Those tradeoffs represent the real uncertainty that is present in the data. This uncertainty cannot be eliminated by further processing and should not be ignored. Our approach is to account for these tradeoffs by leaving each metric as a separate objective and formulating the problem as a multi-objective optimization. This concept of tradeoffs in registration is a key aspect of this work, and the following section illustrates the point with an analogy.

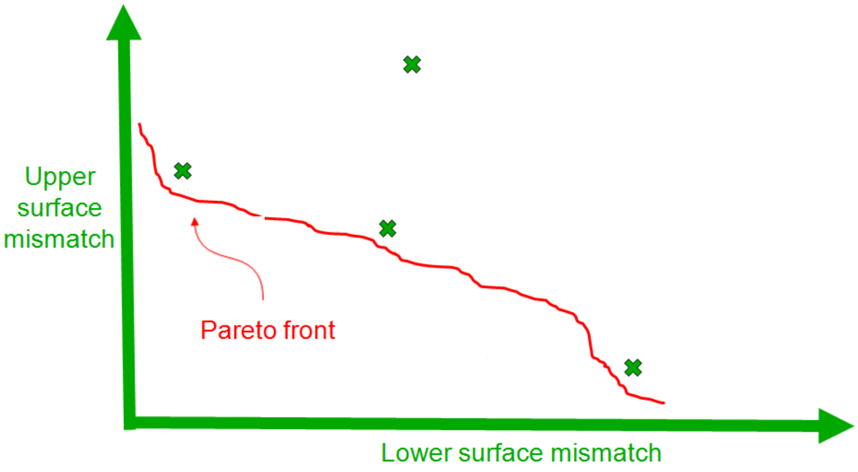


Figure A.2: Registration Pareto front example, showing the tradeoffs between separate metrics for the upper and lower surfaces. The metrics are optimally minimal.

A.2.1 Pareto Front

For registration, each pair of matching surface echoes has its own quality-of-fit metric, as shown in Fig. A.2. There are many points in the metric, or fitness, space that are dominated. In this context, dominated means that a point is worse, according to both metrics, than at least one other point. There is also a locus of points which are not dominated. These points are called the Pareto front [12].

It is clear from Fig. A.2 that there is no single best registration. Instead, there is a Pareto front representing the trade-offs between these metrics, and it is impossible to improve one without making the others worse. The process of registration is a search for points on or close to this Pareto front. In theory the Pareto front might be a continuous line, but it is only possible to sample at discrete points. Registration metric evaluation at each of these points is computationally expensive, so efficient sampling is important.

A.2.2 Registration Metric Evaluation

The forward problem takes a set of registration parameters and maps ultrasound data to locations in x, y, z space. Fig. A.3 indicates three parameters in blue: x shift, y shift, and incidence angle. By guessing values for x and y shifts and incidence angle, the registration software moves the surface echoes, indicated in green, until they match. The same registration mapping moves around any indications we might find, indicated in magenta. The goal of registration is that once the surfaces from various scans align, indications present in those scans will also be aligned. A detection algorithm can then use data fusion to improve both POD and PFA.

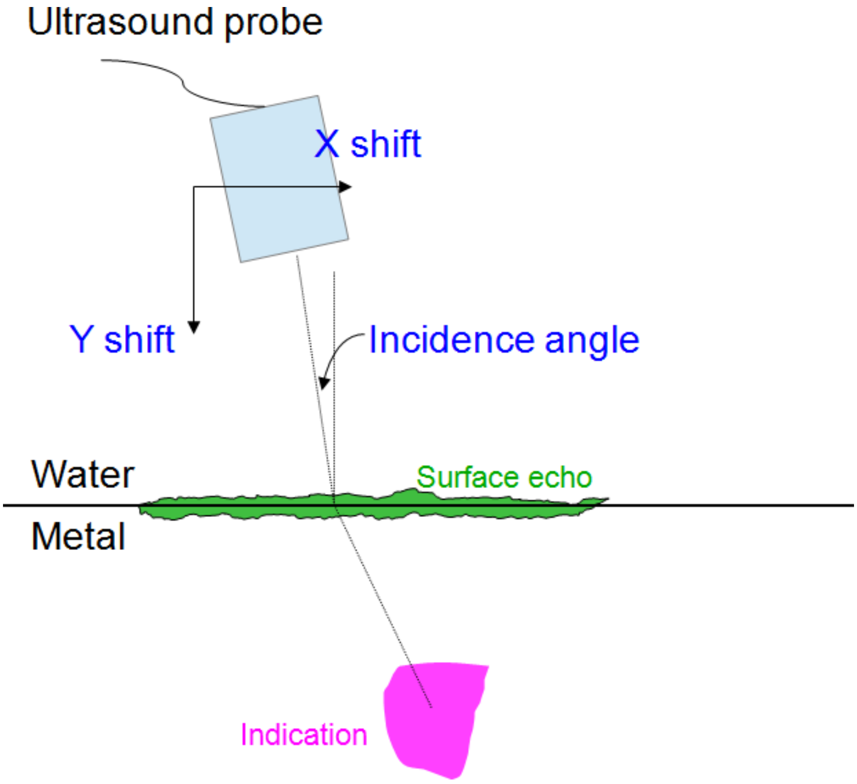


Figure A.3: Schematic diagram of registration variables. X shift, y shift, and incidence angle are registration parameters.

A.3 DENSITY BIAS PROBLEM

Fig. A.4 depicts this registration problem as a mapping between domains of variables. The (blue) registration parameters are the control variables. They are the independent variables whose values are freely chosen.

The (green) registration metrics, or fitness variables, are functions of the control variables. Each point in the parameter space maps to a point in the metric space. Fig. A.2 introduced the Pareto front for the fitness variables. The functional rela-

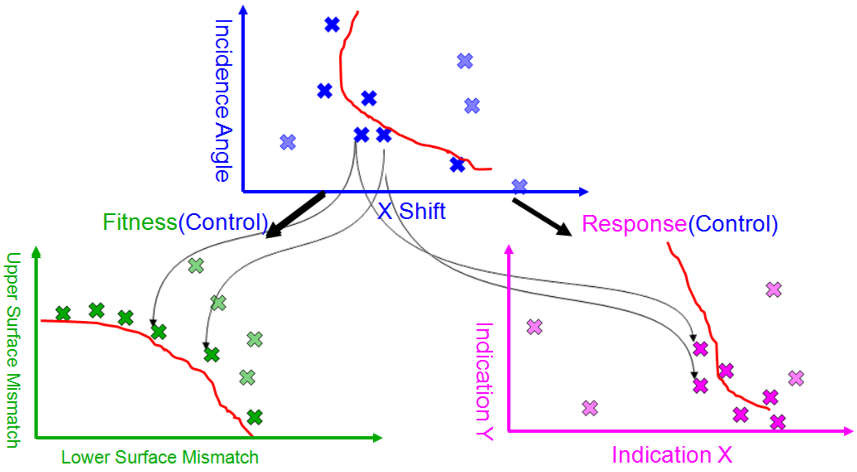


Figure A.4: Registration variable spaces. The registration parameters are the control variables shown in blue (top). Registration metrics are the fitness variables in green (bottom left). Indication locations are the response variables in magenta (bottom right).

tionships imply that there is a locus of points corresponding to the Pareto front in both the parameter and response variable spaces.

The (magenta) indication x and y values in Fig. A.4 are examples of a third set of variables. These location variables are the response, or final output quantities, and they are also functions of the blue control variables. They represent the location of a possible flaw, now mapped by the registration to match the location of data from a different data set.

Fig. A.4 illustrates another key point: the parameters and metrics are essential to registration, but once the registration is complete, neither of these variables really matters; they are only a means to an end. The only quantities that matter to the flaw detection algorithm are the x, y, and z location of an indication.

The registration solver must find points that trace out the Pareto front in the magenta response variables. The fitness and response variables are both nonlinear functions of the control variables. Therefore, the density of points along the Pareto front could be very different for the response variables than it is for the either the registration parameters or metrics. This means that even if there is a good distribution of points that characterizes the Pareto front very well in the fitness variable space, it could be a very biased distribution in the location response variables.

This distribution bias problem is depicted in Fig. A.4 by the cluster of magenta response points near the lower end of the red Pareto front, while the upper half of the Pareto front is devoid of point samples. Such an occurrence is very problematic in a real registration, since the true Pareto front is not known and must be approximated by the points that the registration solver discovers. The distribution bias could cause a failed registration, with results that confuse rather than assist the detection algorithm. With previous registration techniques, there is no way to prevent or even detect this situation because the response variables are not considered or even computed until after the registration has terminated.

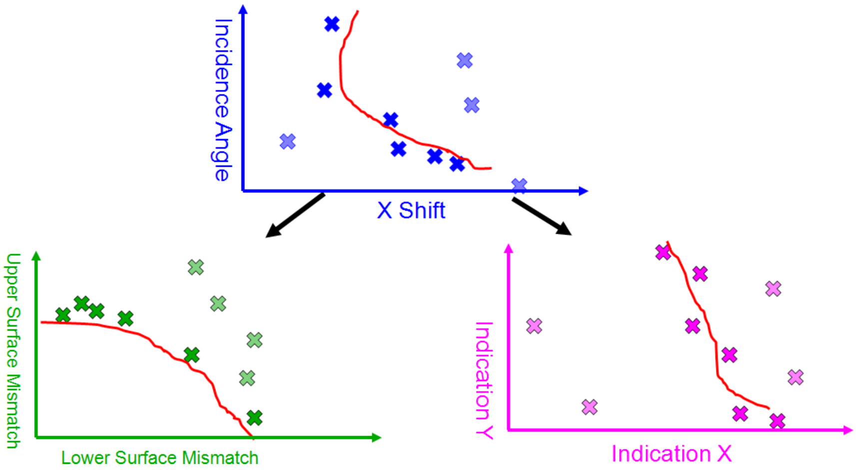


Figure A.5: Effects of the density correction. The samples points are distributed unevenly in the blue control (top) and green (bottom left) fitness variables, but that produces the desired uniform distribution along the red Pareto front in the magenta (bottom right) response variables.

A.4 DENSITY BIAS CORRECTION

The present work proposes a solution to the density bias problem. The basic idea is to works backwards from a uniform density target in the response variables. Conceptually, the solution is to map the uniform response distribution back into a non-uniform distribution in the control variables. That distribution is then mapped again to another non-uniform distribution in the fitness variables.

A complication arises from the fact that both fitness and response variables are functions of control variables, but they are generally not invertible. Therefore, the inverse mapping from response to control must be approximated. First a uniform distribution in the control variable space is approximated by taking a Latin hypercube sample (LHS). These points are then mapped to locations in the response variable space. Those response locations are used as nodal points in a Radial Basis Function (RBF) network. The RBF is defined to interpolate a function which is constant on the response variables. The RBF solution produces a nodal weight associated with each sample point. The same LHS points are then mapped from the control variables to fitness variables. A Gaussian Kernel Density Estimator (KDE) interpolates between LHS points in the fitness variable space, using the nodal weights from the response RBF. This produces a density in fitness variable space that corresponds to an approximately uniform distribution in the response variable space.

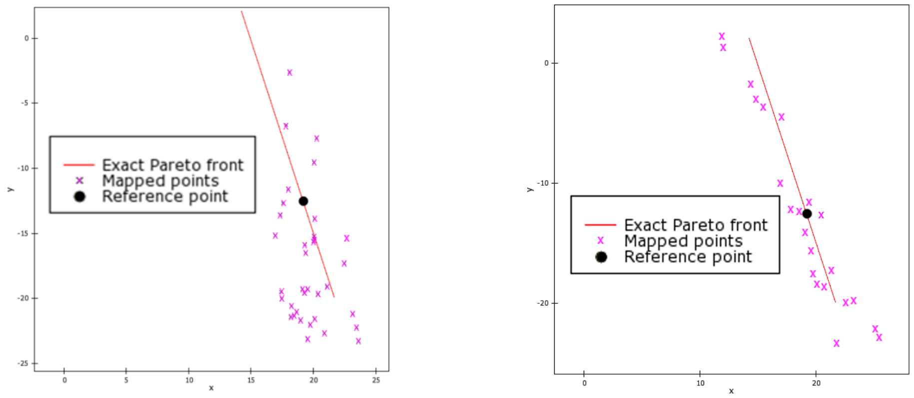
With a fitness point density target in hand, it is then necessary to use a solver that is able to make use of that information. All solvers are designed to search for the Pareto front. However, to the authors knowledge, the only type of solver that is capable of simultaneously targeting a density of points along the Pareto front is the weighted hypervolume indicator [6].

Fig. A.5 depicts the effects of the density correction. By applying the density target on the fitness variables, the solver still finds points near the Pareto front. They are spaced differently from before, perhaps very unevenly. That is, however, what produces the desired effect of a more even distribution in the response variables.

A.5 RESULTS

A simple example with five registration parameters will give a concrete demonstration of the density bias problem and its solution. The 2D model ultrasound probe from Fig. A.3 is assumed to have five registration parameters: x shift, y shift, incidence angle, water wave speed, and metal wave speed. Two registration metrics are simulated, based on the standard DTLZ1 test problem [13], so that the results can be compared to a known Pareto front. The same solver ran the registration both with and without the density correction. Results for the run without correction are shown in Fig. A.6a. Results with the density bias correction are shown in Fig. A.6b.

The plots show x and y physical locations in real space. In each figure, a solid black circle marks the initial guess for the location of a hypothetical indication. The red



(a) (b)

Figure A.6: Registration results with and without a biased point density. (a) Results from registration with a biased point density. The mapped points in magenta should represent the full Pareto front in red, but they are biased toward the lower right. (b) Results from registration with the bias density correction. The mapped points in magenta better represent the full Pareto front in red.

lines show the exact Pareto front solution of possible locations for that indication. In other words, the detection algorithm will fuse these data with a second data set, and that second data could have an indication anywhere along that red line. The software must recognize that those two indications are very likely two views of the same indication.

Each magenta point is the location of the indication after one of the registration mappings found after 100 iterations of the unweighted hypervolume indicator solver. These magenta points are therefore locations to which that indication could be moved, consistently with the registration metrics. A perfect solver would generate point samples covering the red line. No real solver does, of course, because only a finite number of iterations is possible.

Without the density correction, Fig. A.6 shows that the solver provides information for only some of the possible locations. The information is correct, but it is incomplete. The second image of this indication could be in the upper part of the Pareto front, and the solver would fail to match it. The biased density results can also underestimate uncertainty by making the spread of possible points look smaller than it really is.

With the density bias correction, targeting a uniform density in the X,Y space, the solver obtains a more uniform density of registration-mapped points. It is also less likely to miss points near the extremes and much less likely to underestimate the true uncertainty.

A.6 CONCLUSIONS

Advances in NDE technology have vastly increased the amount of available data through automated acquisition. Many applications need increased automation in the data analysis to avoid slow, human-intensive bottlenecks and to use potential correlation information from overlapping data sets. The authors are developing an automated processing methodology that screens and prioritizes data for evaluation by human inspectors.

Registration is an essential prerequisite for merging the overlapping data. There is uncertainty present in the registration that cannot be eliminated and should not be ignored. This uncertainty can be accounted for by using a solver that searches for the full Pareto front for all registration metrics. A correction for the point density bias in the response variables ensures an even distribution in the quantities that are relevant to detection.

A.7 ACKNOWLEDGEMENTS

The support of Rolls-Royce and the UK Research Centre in Non-Destructive Evaluation (RCNDE) is gratefully acknowledged.

A.8 REFERENCES

1. J. B. A. Maintz and M. A. Viergever, Medical Image Analysis, 2(1), 1-36(1998).

2. B. Zitova and J. Flusser, Image and Vision Computing, 21, 977-1000 (2003).

3. C. Voglis and I. E. Lagaris, Applied Mathematics and Computation, 213(1), 216-229 (2009).

4. C. Audet and J. J. E. Dennis, SIAM Journal on Optimization, 13(3), 889-903 (2002).

5. R. Poli, Swarm Intelligence, 1(1), 33-57 (2007).

6. J. Bader, Hypervolume-Based Search for Multiobjective Optimization: Theory and Methods, Ph.D. thesis, ETH Zurich, 2009.

7. R.J. Radke, S. Andra and O. Al-Kofahi, IEEE Transactions on Image Processing, 14(3), 294-307 (2005).

8. T. Clarke, Guided Wave Health Monitoring of Complex Structures, Ph.D. thesis, Imperial College London, 2009.

9. X. E. Gros, P. Strachan and D. W. Lowden, Research in Nondestructive Evaluation, 6, 227-236 (1995).

10. D. Horn and W. R. Mayo, NDT&E International, 33, 351-362 (2000).

11. N. Brierley, T. Tippetts, and P. Cawley, Data Fusion for Improving the Reliability of Automated Non-Destructive Inspection, in Review of Progress in QNDE, 32, edited by D. O. Thompson and D. E. Chimenti, AIP Conference Proceedings, American Institute of Physics, Melville, NY (2013).

12. X.-S. Yang, Nature-Inspired Metaheuristic Algorithms, Luniver Press (2008).

13. K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, Scalable Test Problems for Evolutionary Multi-Objective Optimization, TIK-Technical Report No. 112, ETH Zurich (2001).

References

[1] D.D.L. Hall and S.A.H. McMullen. *Mathematical Techniques in Multisensor Data Fusion.* Artech House Information Warfare Library. Artech House Incorporated, 2nd edition, 2004.

[2] Charles R. Farrar, Nick A.J. Lieven, and Matthew T. Bement. *An Introduction to Damage Prognosis.* John Wiley & Sons, Ltd, 2005.

[3] Baba C. Vemuri, Shuangying Huang, Sartaj Sahni, Christiana M. Leonard, Cecile Mohr, Robin Gilmore, and Jeffrey Fitzsimmons. An efficient motion estimator with application to medical image registration. *Medical Image Analysis*, 2(1):79 – 98, 1998.

[4] Charles R. Meyer, Jennifer L. Boes, Boklye Kim, Peyton H. Bland, Kenneth R. Zasadny, Paul V. Kison, Kenneth Koral, Kirk A. Frey, and Richard L. Wahl. Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. *Medical Image Analysis*, 1(3):195 – 206, 1997.

[5] D. Rueckert, L. I. Sonoda, C. Hayes, D. L G Hill, M. O. Leach, and D.J. Hawkes. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.

[6] N. Brierley, T. Tippetts, and P. Cawley. Improving the reliability of automated non-destructive inspection. In D. O. Thompson and D. E. Chimenti, editors, *Review of Progress in Quantitative Nondestructive Evaluation*, 2012.

[7] N. Brierley, T. Tippetts, and P. Cawley. Data fusion for improving the reliability of automated non-destructive inspection. In D. O. Thompson and D. E. Chimenti, editors, *Review of Progress in Quantitative Nondestructive Evaluation*, 2013.

[8] N. Brierley, T. Tippetts, and P. Cawley. Improving the reliability of automated non-destructive inspection. In D. O. Thompson and D. E. Chimenti, editors, *Review of Progress in Quantitative Nondestructive Evaluation*, 2014.

[9] Josef Krautkrämer and Herbert Krautkrämer. *Ultrasonic testing of materials.* Springer-Verlag, 4th edition, 1990.

[10] R. B. Thompson and E. F. Lopes. The effects of focusing and refraction on Gaussian ultrasonic beams. *Journal of Nondestructive Evaluation*, 4:107–123, June 1984.

[11] David Wright. Personal communication, March 2010.

[12] J. G. Biddulph. Ultrasonic inspection of disc forms. Rationalised Process Specification RPS705, Rolls-Royce, 2009.

[13] H.-J. Kim, L. W. Schmerr, and A. Sedov. Transferring distance-amplitude correction curves using ultrasonic modeling. In D. O. Thompson, D. E. Chimenti, L. Poore, C. Nessa, and S. Kallsen, editors, *Review of Progress in Quantitative Nondestructive Evaluation*, February 2004.

[14] Jianhua Jin, Xueying Liu, Maxim Kruschinski, and Diamond Han. Immersion multi-zone ultrasonic inspection system of large-diameter special billets. In *17th World Conference on Nondestructive Testing*, October 2008.

[15] Disc Inspection Staff at NDT Services, Derby, UK. Personal communication, July 2010.

[16] I. Yalda, F.J. Margetan, K.Y. Han, and R.B. Thompson. Survey of ultrasonic grain noise characteristics in jet engine titanium. In Donald O. Thompson and Dale E. Chimenti, editors, *Review of Progress in Quantitative Nondestructive Evaluation.* 1996.

[17] F.J. Margetan, I. Yalda, R.B. Thompson, J. Umbach, U. Suh, P.J. Howard, D.C. Copley, and R. Gilmore. Ultrasonic grain noise modeling: Recent applications to engine titanium inspections. In Donald O. Thompson and Dale E. Chimenti, editors, *Review of Progress in Quantitative Nondestructive Evaluation*. 1997.

[18] F. J. Margetan and R. B. Thompson. Microstructural noise in titanium alloys and its influence on the detectability of hard-alpha inclusions. In D. O. Thompson and D. E. Chimenti, editors, *Review of Progress in Quantitative Nondestructive Evaluation*, 1992.

[19] Nick Brierley. *The computational enhancement of automated non-destructive inspection*. EngD thesis, Imperial College, Department of Mechanical Engineering, 2014 (expected).

[20] Nick Brierley. EngD 12-month progress / transfer report: Improving the reliability of automated non-destructive inspection. Technical report, Imperial College, 2010.

[21] Nick Brierley. EngD 18-month progress report: Improving the reliability of automated non-destructive inspection. Technical report, Imperial College, 2011.

[22] Nick Brierley. EngD 24-month progress report: Improving the reliability of automated non-destructive inspection. Technical report, Imperial College, 2011.

[23] Nick Brierley. EngD 30-month progress report: Improving the reliability of automated non-destructive inspection. Technical report, Imperial College, 2012.

[24] Nick Brierley. EngD 36-month progress report: Improving the reliability of automated non-destructive inspection. Technical report, Imperial College, 2012.

[25] Richard J. Radke, Srinivas Andra, and Omar Al-Kofahi. Image change detection algorithms: A systematic survey. *IEEE Transactions on Image Processing*, 14(3):294–307, 2005.

[26] B.C. Vemuri, J. Ye, Y. Chen, and C.M. Leonard. Image registration via level-set motion: Applications to atlas-based segmentation. *Medical Image Analysis*, 7(1):1–20, 2003.

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

[28] I. D. Hall, A. McNab, and G. Hayward. Improved ultrasonic image generation through tomographic image fusion. *Ultrasonics*, 37, September 1999.

[29] A. Connor and A. K. Forrest. Removal of repeated objects in multi-channel images. In Yetai Fei, Kuang-Chao Fan, and Rongsheng Lu, editors, *Fourth International Symposium on Precision Mechanical Measurements.* SPIE, 2008.

[30] Barbara Zitová and Jan Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21, 2003.

[31] Guy L. Scott and H. Christopher Longuet-Higgins. An algorithm for associating the features of two images. *Proceedings: Biological Sciences*, 244(1309):21–26, 1991.

[32] R. Kaftory and Y. Y. Zeevi. Blind separation of images obtained by spatially-varying mixing system. In *15th IEEE International Conference on Image Processing*, pages 2604–2607, October 2008.

[33] Guido van Rossum. *Python Reference Manual*, 2011.

[34] Bjarne Stroustrup. *C++ Programming Language.* Pearson Education, 4th edition, 2013.

[35] David M Beazley et al. SWIG: An easy to use tool for integrating scripting languages with C and C++. In *Proceedings of the 4th USENIX Tcl/Tk workshop*, pages 129–139, 1996.

[36] Travis E Oliphant. *A Guide to NumPy*, volume 1. Trelgol Publishing USA, 2006.

[37] Haili Chui, Lawrence Win, Robert Schultz, James S Duncan, and Anand Rangarajan. A unified non-rigid feature registration method for brain mapping. *Medical Image Analysis*, 7(2):113–130, 2003.

[38] Leila MG Fonseca and BS Manjunath. Registration techniques for multisensor remotely sensed imagery. *PE & RS- Photogrammetric Engineering & Remote Sensing*, 62(9):1049–1056, 1996.

[39] Jean-Philippe Thirion. New feature points based on geometric invariants for 3D image registration. *International journal of computer vision*, 18(2):121–137, 1996.

[40] Jun-Wei Hsieh, Hong-Yuan Mark Liao, Kuo-Chin Fan, Ming-Tat Ko, and Yi-Ping Hung. Image registration using a new edge-based approach. *Computer Vision and Image Understanding*, 67(2):112–130, 1997.

[41] Gong-Jian Wen, Jin-jian Lv, and Wen-xian Yu. A high-performance feature-matching method for image registration by combining spatial and similarity information. *IEEE Transactions on Geoscience and Remote Sensing*, 46(4):1266–1277, 2008.

[42] Rui Xu and Yen-wei Chen. Wavelet-based multiresolution medical image registration strategy combining mutual information with spatial information. *International Journal of Innovative Computing, Information and Control*, 3(2):285–296, April 2006.

[43] Adelaide Duroux, Karim Sabra, James Ayers, and Massimo Ruzzene. Diffuse field interferometry for experimental Green's function estimation and damage detection. *The Journal of the Acoustical Society of America*, 124(4):2576, 2008.

[44] Eric Larose, Oleg I. Lobkis, and Richard L. Weaver. Passive correlation imaging of a buried scatterer. *The Journal of the Acoustical Society of America*, 119(6):3549–3552, 2006.

[45] Kees Wapenaar and Jacob Fokkema. Green's function representations for seismic interferometry. *Geophysics*, 71(4):SI33–46, 2006.

[46] E. Kerbrat, D. Clorennec, C. Prada, D. Royer, D. Cassereau, and M. Fink. Detection of cracks in a thin air-filled hollow cylinder by application of the DORT method to elastic components of the echo. *Ultrasonics*, 40(1-8):715–720, 2002.

[47] E. A. Marengo, F. K. Gruber, and F. Simonetti. Time-reversal MUSIC imaging of extended targets. *IEEE Transactions on Image Processing*, 16(8):1967–1984, August 2007.

[48] Anthony J. Devaney, Edwin A. Marengo, and Fred K. Gruber. Time-reversal-based imaging and inverse scattering of multiply scattering point targets. *The Journal of the Acoustical Society of America*, 118(5):3129–3138, 2005.

[49] Edwin A. Marengo and Fred K. Gruber. Subspace-based localization and inverse scattering of multiply scattering point targets. *EURASIP J. Appl. Signal Process.*, 2007, January 2007.

[50] Thomas Clarke. *Guided wave health monitoring of complex structures.* PhD thesis, Imperial College, Department of Mechanical Engineering, July 2009.

[51] Richard L. Weaver and Oleg I. Lobkis. Temperature dependence of diffuse field phase. *Ultrasonics*, 38(18):491 – 494, 2000.

[52] Yinghui Lu and Jennifer E. Michaels. A methodology for structural health monitoring with diffuse ultrasonic waves in the presence of temperature variations. *Ultrasonics*, 43(9):717 – 731, 2005.

[53] Anthony J Croxford, Paul D Wilcox, George Konstantinidis, and Bruce W Drinkwater. Strategies for overcoming the effect of temperature on guided wave structural health monitoring. In *Proceedings of SPIE*, volume 6532, pages 65321T–1 – 65321T–10, 2007.

[54] A. J. Croxford, P. D. Wilcox, G. Konstantinidis, and B. W. Drinkwater. Strategies for overcoming the effect of temperature on guided wave structural health monitoring. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 6532, May 2007.

[55] R. Long, J. Russell, P. Cawley, and N. Habgood. Ultrasonic phased array inspection of flaws on weld fusion faces using full matrix capture. *AIP Conference Proceedings*, 1096(1):848–855, 2009.

[56] J. Russell, R. Long, P. Cawley, and N. Habgood. Inspection of Components with Irregular Surfaces Using a Conformable Ultrasonic Phased Array. In *American Institute of Physics Conference Series*, volume 1096, pages 792–799, March 2009.

[57] J. F. Cardoso. Blind signal separation: Statistical principles. *Proceedings of the IEEE*, 86(10):2009–2025, October 1998.

[58] Fawzy Abujarad and Abbas Omar. Comparison of Independent Component Analysis (ICA) algorithms for GPR detection of non-metallic land mines. In Lorenzo Bruzzone, editor, *Image and Signal Processing for Remote Sensing XII*, volume 6365, pages 636516–1 – 636516–12. SPIE, 2006.

[59] S. Ukai, T. Takatani, T. Nishikawa, and H. Saruwatari. Blind source separation combining SIMO-model-based ICA and adaptive beamforming. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, March 2005.

[60] Matthew J. Fleming. *Far-field super resolution imaging.* PhD thesis, Imperial College, August 2008.

[61] F. Simonetti and L. Huang. From beamforming to diffraction tomography. *Journal of Applied Physics*, 103(10):103110, 2008.

[62] F. Simonetti. Multiple scattering: The key to unravel the subwavelength world from the far-field pattern of a scattered wave. *Phys. Rev. E*, 73(3):036619, March 2006.

[63] Claire Prada and Mathias Fink. Eigenmodes of the time reversal operator: A solution to selective focusing in multiple-target media. *Wave Motion*, 20(2):151–163, 1994.

[64] Caroline Holmes, Bruce W. Drinkwater, and Paul D. Wilcox. Post-processing of the full matrix of ultrasonic transmit-receive array data for non-destructive evaluation. *NDT & E International*, 38(8):701–711, 2005.

[65] J. E. Michaels, A. J. Croxford, and P. D. Wilcox. Imaging algorithms for locating damage via in situ ultrasonic sensors. In *IEEE Sensors Applications Symposium*, pages 63–67, February 2008.

[66] Thouraya M. Meksen, Bachir Boudraa, Redouane Drai, and Malika Boudraa. Automatic crack detection and characterization during ultrasonic inspection. *Journal of Nondestructive Evaluation*, 29:169–174, 2010.

[67] Windsor and Capineri. Automated object positioning from ground penetrating radar images. *Insight*, 40(7):482–488, 1998.

[68] Peter T. Gough and David W. Hawkins. Unified framework for modern synthetic aperture imaging algorithms. *International Journal of Imaging Systems and Technology*, 8(4):343–358, 1997.

[69] Frederik Maes, Andre Collignon, Dirk Vandermeulen, Guy Marchal, and Paul Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.

[70] William M. Wells, Paul Viola, Hideki Atsumi, Shin Nakajima, and Ron Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical image analysis*, 1(1):35–51, 1996.

[71] Philippe Thévenaz and Michael Unser. Optimization of mutual information for multiresolution image registration. *IEEE Transactions on Image Processing*, 9(12):2083–2099, 2000.

[72] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: A survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, August 2003.

[73] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Phys. Rev. E*, 69(6):066138, June 2004.

[74] A. Kraskov, H. Stögbauer, R. G. Andrzejak, and P. Grassberger. Hierarchical clustering using mutual information. *Europhysics Letters*, 70(2):278, 2005.

[75] Young-Il Moon, Balaji Rajagopalan, and Upmanu Lall. Estimation of mutual information using kernel density estimators. *Phys. Rev. E*, 52(3), September 1995.

[76] M. A. Montemurro, R. Senatore, and S. Panzeri. Tight Data-Robust bounds to mutual information combining shuffling and model selection techniques. *Neural Computation*, 19(11):2913–2957, 2007.

[77] J. Martin Bland and Douglas G. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 327(8476):307–310, 1986.

[78] P. Hastreiter and T. Ertl. Integrated registration and visualization of medical image data. In *Computer Graphics International*, pages 78–85, June 1998.

[79] H. G. Beyer and B. Sendhoff. Functions with noise-induced multimodality: A test for evolutionary robust optimization-properties and performance analysis. *IEEE Transactions on Evolutionary Computation*, 10(5):507–526, October 2006.

[80] J. E. Fieldsend and R. M. Everson. Multi-objective optimisation in the presence of uncertainty. In *IEEE Congress on Evolutionary Computation*, volume 1, pages 243–250, September 2005.

[81] Gerard Cornuejols and Reha Tütüncü. *Optimization Methods in Finance*. Number 5 in Mathematics, Finance and Risk. Cambridge University Press, December 2006.

[82] Jerome H. Friedman, Jon Louis Bentley, and Raphael Ari Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM Trans. Math. Softw.*, 3(3):209–226, September 1977.

[83] N. Memarsadeghi, V. C. Raykar, R. Duraiswami, and D. M. Mount. Efficient kriging via fast matrix-vector products. In *IEEE Aerospace Conference*, pages 1–7, March 2008.

[84] Raymond H. Myers, Douglas C. Montgomery, and Christine M. Anderson-Cook. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. Wiley Series in Probability and Statistics. John Wiley & Sons, 3rd edition, January 2009.

[85] R. M. Everson and J. E. Fieldsend. Multiobjective optimization of safety related systems: An application to short-term conflict alert. *IEEE Transactions on Evolutionary Computation*, 10(2):187–198, April 2006.

[86] Jürgen Teich. Pareto-front exploration with uncertain objectives. In *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization*, London, UK, 2001. Springer-Verlag.

[87] L. Dowhan, A. Wymyslowski, and R. Dudek. Multi-objective parametric approach to numerical optimization of stacked packages. In *International Conference on Thermal, Mechanical and Multi-Physics Simulation Experiments in Microelectronics and Micro-Systems*, pages 1–7, April 2007.

[88] Frederik Maes, Dirk Vandermeulen, and Paul Suetens. Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Medical Image Analysis*, 3(4):373–386, 1999.

[89] Sean Luke. *Essentials of Metaheuristics*. Lulu, 2nd edition, 2013.

[90] Kalyanmoy Deb. Multi-objective genetic algorithms: Problem difficulties and construction of test problems. *Evol. Comput.*, 7(3), 1999.

[91] A. Konstantinidis, Qingfu Zhang, and Kun Yang. A subproblem-dependent heuristic in MOEA/D for the deployment and power assignment problem in wireless sensor networks. In *IEEE Congress on Evolutionary Computation*, pages 2740–2747, 2009.

[92] Hui Li and Qingfu Zhang. Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II. *IEEE Transactions on Evolutionary Computation*, 13(2):284–302, 2009.

[93] Xin-She Yang. *Nature-Inspired Metaheuristic Algorithms*. Luniver Press, 2008.

[94] Riccardo Poli, James Kennedy, and Tim Blackwell. Particle swarm optimization. *Swarm Intelligence*, 1:33–57, 2007.

[95] S. Mostaghim and J. Teich. Covering Pareto-optimal fronts by subswarms in multi-objective particle swarm optimization. In *Congress on Evolutionary Computation*, volume 2, June 2004.

[96] Sanaz Mostaghim and Jürgen Teich. A new approach on many objective diversity measurement. In Jürgen Branke, Kalyanmoy Deb, Kaisa Miettinen, and Ralph E. Steuer, editors, *Practical Approaches to Multi-Objective Optimization*, number 04461 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2005. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

[97] Joshua Knowles and Evan J. Hughes. Multiobjective optimization on a budget of 250 evaluations. In Carlos A. Coello Coello, Arturo Hernández Aguirre, and Eckart Zitzler, editors, *Evolutionary Multi-Criterion Optimization*, volume 3410 of *Lecture Notes in Computer Science*, pages 176–190. Springer Berlin / Heidelberg, 2005.

[98] Evan Hughes. Multi-objective binary search optimisation. In Carlos Fonseca, Peter Fleming, Eckart Zitzler, Lothar Thiele, and Kalyanmoy Deb, editors, *Evolutionary Multi-Criterion Optimization*, volume 2632 of *Lecture Notes in Computer Science*, page 72. Springer Berlin / Heidelberg, 2003.

[99] J. Knowles. ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. Technical Report TR-COMPSYSBIO-2004-01, University of Manchester, 2004.

[100] Łukasz Łaniewski-Wołłk. Expected hypervolume improvement — Criterion for kriging based multiobjective optimization. In *19th Polish National Fluid Dynamics Conference*. Institute of Aeronautics and Applied Mechanics, Warsaw University of Technology, 2010.

[101] A. Emmerich. The computation of the expected improvement in dominated hypervolume of Pareto front approximations. Technical Report LIACS TR-4-2008, Leiden University, The Netherlands, 2008.

[102] Jasper A. Vrugt, Bruce A. Robinson, and James M. Hyman. Self-adaptive multimethod search for global optimization in real-parameter spaces. *Trans. Evol. Comp*, 13(2), 2009.

[103] D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, April 1997.

[104] Johannes M. Bader. *Hypervolume-Based Search for Multiobjective Optimization: Theory and Methods.* Doctor of Sciences, Dipl. El.-Ing., ETH Zürich, Computer Engineering and Networks Laboratory, December 2009.

[105] Parke Godfrey, Ryan Shipley, and Jarek Gryz. Algorithms and analyses for maximal vector computation. *The VLDB Journal*, 16:5–28, 2007.

[106] Tamara Ulrich, Johannes Bader, and Lothar Thiele. Defining and optimizing indicator-based diversity measures in multiobjective search. In *Proceedings of the 11th international conference on parallel problem solving from nature: Part I*, Berlin, Heidelberg, 2010. Springer-Verlag.

[107] Anne Auger, Johannes Bader, and Dimo Brockhoff. Theoretically investigating optimal μ-distributions for the hypervolume indicator: First results for three objectives. In *Proceedings of the 11th international conference on parallel problem solving from nature: Part I*, Berlin, Heidelberg, 2010. Springer-Verlag.

[108] Tamara Ulrich, Johannes Bader, and Eckart Zitzler. Integrating decision space diversity into hypervolume-based multiobjective search. In *Proceedings of the 12th annual conference on genetic and evolutionary computation*, New York, NY, USA, 2010.

[109] Eckart Zitzler, Dimo Brockhoff, and Lothar Thiele. The hypervolume indicator revisited: On the design of Pareto-compliant indicators via weighted integration. In *Proceedings of the 4th international conference on evolutionary multi-criterion optimization*, Berlin, Heidelberg, 2007. Springer-Verlag.

[110] Johannes Bader and Eckart Zitzler. Hype: An algorithm for fast hypervolume-based many-objective optimization. *Evol. Comput.*, 19(1):45–76, March 2011.

[111] Nykolai Bilaniuk and George S. K. Wong. Speed of sound in pure water as a function of temperature. *The Journal of the Acoustical Society of America*, 93:1609, 1993.

[112] Bryan D. Olin and William Q. Meeker. Applications of statistical methods to nondestructive evaluation. *Technometrics*, 38(2), 1996.

[113] Awf Volker. Modeling of NDE reliability: Development of a POD-generator. 16th WCNDT 2004 - World Conference on NDT, August 2004.

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

[115] Joseph P Romano. *Testing statistical hypotheses.* Springer, 2005.

[116] Boaz Cohen and Itshak Dinstein. New maximum likelihood motion estimation schemes for noisy ultrasound images. *Pattern Recognition*, 35(2):455–463, 2002.

[117] Vinayak Dutt and James F Greenleaf. Ultrasound echo envelope analysis using a homodyned k distribution signal model. *Ultrasonic Imaging*, 16(4):265–287, 1994.

[118] Li Weng, John M. Reid, P. Mohana Shankar, and Kawan Soetanto. Ultrasound speckle analysis based on the k distribution. *The Journal of the Acoustical Society of America*, 89:2992, 1991.

[119] D.A. Clifton, S. Hugueny, and 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*, pages 1–6, 2011.

[120] R . A. Fisher. *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd, 4th edition, 1932.

[121] Peter H. Westfall and Russell D. Wolfinger. Multiple tests with discrete distributions. *The American Statistician*, 51(1):pp. 3–8, 1997.

[122] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, New York, NY, USA, 3rd edition, 2007.

[123] Peter Jäckel. *Monte Carlo Methods in Finance.* John Wiley & Sons, 2002.

[124] Peter W Glynn and Donald L Iglehart. Importance sampling for stochastic simulations. *Management Science*, 35(11):1367–1392, 1989.

[125] David Siegmund. Importance sampling in the Monte Carlo study of sequential tests. *The Annals of Statistics*, pages 673–684, 1976.

[126] Michael Hohenbichler and Ruediger Rackwitz. Improvement of second-order reliability estimates by importance sampling. *Journal of Engineering Mechanics*, 114(12):2195–2199, 1988.

[127] Giray Ökten. Random sampling from low-discrepancy sequences: Applications to option pricing. *Mathematical and Computer Modelling*, 35(11):1221–1234, 2002.

[128] I. M. Sobol. Uniformly distributed sequences with an additional uniform property. *USSR Computational Mathematics and Mathematical Physics*, 16:236–242, 1976.

[129] Paul Bratley and Bennett L. Fox. Algorithm 659: Implementing Sobol's quasirandom sequence generator. *ACM Trans. Math. Softw.*, 14(1):88–100, March 1988.

[130] R Bruce Thompson. Using physical models of the testing process in the determination of probability of detection. *Materials evaluation*, 59(7):861–865, 2001.

[131] Dimitris Manolakis and Gary Shaw. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Processing Magazine*, 19(1):29–43, 2002.

[132] Chris Glasbey, Gerie van der Heijden, Vivian F. K. Toh, and Alision Gray. Colour displays for categorical images. *Color Research & Application*, 32(4):304–309, 2007.

[133] T. Tippetts, N. Brierley, and P. Cawley. Data registration for automated non-destructive inspection with multiple data sets. In D. O. Thompson and D. E. Chimenti, editors, *Review of Progress in Quantitative Nondestructive Evaluation*, 2013.

Publications

Some of the material presented in this thesis has been published. The details of these publications are listed as follows.

N. Brierley, T. Tippetts, P. Cawley. Improving the reliability of automated non-destructive inspection. In D. O. Thompson and D. E. Chimenti, editors, *Review of Progress in Quantitative NDE*, AIP Conf. Proc., 2012.

T. Tippetts, N. Brierley, P. Cawley. Data registration for automated non-destructive inspection with multiple data sets. In D. O. Thompson and D. E. Chimenti, editors, *Review of Progress in Quantitative NDE*, AIP Conf. Proc., 2013.

T. Tippetts, N. Brierley, P. Cawley. Data fusion for improving the reliability of automated non-destructive inspection. In D. O. Thompson and D. E. Chimenti, editors, *Review of Progress in Quantitative NDE*, AIP Conf. Proc., 2013.

N. Brierley, T. Tippetts, P. Cawley. Data fusion for automated non-destructive inspection. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science.* 470: 20140167, 2014.