PROBABILITY OF DETECTION FOR PULSED EDDY CURRENT INSPECTION OF SECOND LAYER CRACKS IN AIRCRAFT LAP-JOINT STRUCTURES

PROBABILITE DE DETECTION PAR UNE METHODE DE COURANTS DE FOUCAULT PULSES DE LA SECONDE COUCHE STRUCTURALE D'UN JOINT A RECOUVREMENT DANS UN ENGIN AERIEN

A Thesis Submitted to the Division of Graduate Studies
of the Royal Military College of Canada
by

Cole Omega Uemura

In Partial Fulfillment of the Requirements for the Degree of
Master of Science in Physics

January 2018

ROYAL MILITARY COLLEGE OF CANADA
COLLÈGE MILITAIRE ROYAL DU CANADA

DIVISION OF GRADUATE STUDIES AND RESEARCH
DIVISION DES ÉTUDES SUPÉRIEURES ET DE LA RECHERCHE

This is to certify that the thesis prepared by / Ceci certifie que la thèse rédigée par

Cole Omega Uemura

entitled / intitulée

PROBABILITY OF DETECTION FOR PULSED EDDY CURRENT INSPECTION OF
SECOND LAYER CRACKS IN AIRCRAFT LAP-JOINT STRUCTURES /
PROBABILITE DE DETECTION PAR UNE METHODE DE COURANTS DE FOUCAULT
PULSES DE LA SECONDE COUCHE STRUCTURALE D'UN JOINT A RECOUVREMENT
DANS UN ENGIN AERIEN

complies with the Royal Military College of Canada regulations and that it meets the
accepted standards of the Graduate School with respect to quality, and, in the case of a
doctoral thesis, originality / satisfait aux règlements du Collège militaire royal du Canada
et qu'elle respecte les normes acceptées par la Faculté des études supérieures quant à
la qualité et, dans le cas d'une thèse de doctorat, l'originalité

for the degree of / pour le diplôme de

Master of Science / Maîtrise en Science

Signed by the final examining committee: /
Signée par les membres du comité examinateur de la soutenance de thèse

_____, Chair / Président

_____, External Examiner / Examinateur externe

_____, Main Supervisor / Directeur de thèse principal

Approved by the Head of Department : /
Approuvée par le Directeur du Département : _____ Date: _____

To the Librarian: This thesis is not to be regarded as classified. /
Au Bibliothécaire: Cette thèse n'est pas considérée comme à publication restreinte.

_____
Main Supervisor / Directeur de thèse principal

Acknowledgements

I would like to thank Dr. Thomas Krause, my thesis supervisor, for his guidance and patience throughout the last two years. His dedication to the field of non-destructive testing (NDT) in combination with his instruction, encouragement and constructive feedback has greatly contributed to the success of this thesis.

I would also like to thank Dr. Ross Underhill for his unwavering support, assistance with designing LabVIEW software interfaces, programming computer algorithms, and invaluable guidance. The insights he provided from his vast knowledge of statistical methods were instrumental to the success of this thesis.

Additionally, I would like to thank Charles Annis, P.E., for providing the **mh1823 POD** software and providing guidance and R code for determining specimen POD coverage. His assistance and his passion for sharing his wealth of knowledge with respect to probability of detection is greatly appreciated.

Finally, I would like to thank my son, Apollo, for his patience and understanding throughout the past two years. Your love, support, and fabulous sense of humour has helped me to endure numerous challenges, and I could not have succeeded in this endeavour without you by my side.

Abstract

A thesis completed by Uemura, Cole Omega, in partial fulfilment of the requirements for a Master of Science in Physics from the Royal Military College of Canada on this 24th day of January, 2018, on *Probability of Detection for Pulsed Eddy Current Inspection of Second Layer Cracks in Aircraft Lap-Joint Structures*, under the direction of Dr. Thomas Krause.

Conventional eddy current testing is a non-destructive technique used extensively within the aerospace industry to detect surface and near surface defects in structures of aluminum aircraft components. Aging aircraft such as the CP-140 Aurora and CC-130 Hercules are susceptible to cyclic fatigue cracks at bolt hole locations in multi-layer aluminum lap-joint structures. Conventional bolt hole eddy current (BHEC) inspection techniques, however, require fastener removal to detect second layer cracks. A pulsed eddy current (PEC) technique has demonstrated the ability to detect second layer cracks without fastener removal. The probe design for this technique consists of a ferrite core surrounded by four pairs of differentially connected pick-up coils, which utilizes the ferrous fastener as a flux conduit, inducing eddy currents at greater depths within the material. The transient signal response is analyzed using modified principal components analysis (MPCA) to produce scores, which are then processed using cluster analysis. Since a large number of representative blank fastener signals are required to accurately assess the reliability of this inspection technique in the absence of in-service samples, a series of simulations was conducted using cluster analysis and the smallest half volume (SHV) analysis methodology. A probability of detection (POD) analysis of the PEC inspection technique is then performed, including the metric required, in the form of an $a_{90/95}$ value, to assess its reliability for the detection of second layer cracks at ferrous fasteners in aircraft lap-joint structures. Based on both the measured and simulation results obtained in this thesis work, it is determined that an $a_{90/95}$ value of 2.0 mm (0.08 inch) is associated with this PEC inspection technique. The average detection and false call rates obtained for the simulations are 93% and 11%, respectively, confirming a blind outlier detection capability. This research confirms an $a_{90/95}$ value for PEC that approaches that for BHEC (0.05 inch), but with only up to 15% of the fasteners being removed, as opposed to removal of 100% of the fasteners

for BHEC. The numerous advantages to the PEC inspection method are that 1) it does not require fastener removal, hence minimizing collateral damage on the surface structure, 2) reduces the cost associated with labour and human resources, and 3) shortens aircraft maintenance downtime. These results provide assurance that the PEC technique is capable of detecting second layer cracks in aircraft lap-joint structures with high confidence, reliability, and detectability.

Résumé

Thèse complétée par Uemura, Cole Omega, pour l'obtention d'une maitrise en science en physique au collège militaire royal du Canada ce 24 janvier 2018 sur la probabilité de détection par une méthode de courants de Foucault pulsés de la seconde couche structurale d'un joint à recouvrement dans un engin aérien, sous la direction de Dr Thomas Krause.

La méthode par évaluation non-destructive par courants de Foucault est largement répandue dans le domaine de l'industrie aérospatiale afin de déceler des défauts de la couche superficielle dans les composantes en aluminium d'un engin aérien. Les engins ayant plus d'années d'usure tels l'Aurora CP-140 et Hercules sont vulnérables aux fatigues cycliques près des trous de boulons dans le laminage en aluminium des joints à recouvrement. La méthode usuelle par courants de Foucault (MUCF) utilisées pour évaluer les trous des boulons nécessite le retrait des points d'attache. On a démontré que la méthode par courant de Foucault pulsée(CFP) pouvait déceler des fissures à la seconde couche sans avoir à enlever les points d'attache. La sonde de détection est un matériau de ferrite ayant quatre paires de bobines détectrices proche voisines qui sont branchées en mode différentiel. Ces bobines sont sensibles aux flux magnétiques induit dans les attaches fabriquées d'un matériau magnétique permanent au travers duquel un courant est induit. La réponse en régime transitoire est analysée par le truchement de l'analyse en composante principale modifiée(ACPM) qui nous donne un pointage et qui nous fournit ensuite une analyse de groupe. Puisque qu'un grand nombre de signaux sont nécessaires afin d'obtenir une évaluation fiable afin de mieux comparer avec des échantillons en service, une série de simulations fut dirigée pour estimer l'analyse de groupe ainsi qu'une analyse du plus petit demi-volume(PPDV). L'analyse de la

probabilité de détection(POD) de la méthode par CFP fut ensuite menée en se basant sur un critère spécifié sous la forme de $a_{90/95}$, qui permet d'évaluer la fiabilité d'une détection de fissure à la seconde couche d'une attache en ferrite dans un joint à recouvrement d'un engin. En se basant sur les résultats et simulations obtenues dans notre étude, il est recommandé qu'une valeur de $a_{90/95}$ égale à 2.0 mm (0.08 pouce) soit associée à cette méthode par CFP. Les taux moyens de détection et d'erreur lors de nos simulations étaient de 93% et 11% respectivement, ce qui confirme une capacité nette de bien détecter. Dans ce mémoire, nous montrons que la valeur de $a_{90/95}$ pour les CFP est proche de celle obtenue pour les MUCF (0.08 pouce). Ceci voudrait dire que seulement 15% des attaches devraient être enlevées au lieu de 100% pour l'évaluation par MUCF. L'avantage est que l'évaluation par CFP n'implique pas le retrait de tous les points d'attache et ainsi minimise les dommages à la surface d'une structure, réduit les coûts de main d'œuvre et diminue le temps de l'entretien. Les résultats obtenus sont concluants et nous permettent d'affirmer que la méthode par CFP est très performante lors de l'évaluation de la seconde couche de défauts dans les joints à recouvrement en ce qui a trait à la confiance élevée de détection ainsi qu'à la fiabilité.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ATESS   Aerospace and Telecommunications Engineering Support Squadron

BHEC   Bolt Hole Eddy Current

CAD   Computer-Aided Design

CFRP   Carbon Fibre Reinforced Polymer

DAQ   Data Acquisition

ET   Eddy Current Testing

EDM   Electrical Discharge Machined

ENIQ   European Network for Inspection and Qualification

FEA   Finite Element Analysis

GLM   Generalized Linear Model

ICA   Independent Components Analysis

KVL   Kirchhoff's Voltage Law

NAVAIR   Naval Air Systems Command

NI   National Instruments

MCD   Minimum Covariance Determinant

MD   Mahalanobis Distance

MLE   Maximum Likelihood Estimation

MPCA   Modified Principal Components Analysis

MVE   Minimum Volume Ellipsoid

NDT   Non-destructive Testing

NIH   National Institutes of Health

PCA   Principal Components Analysis

PEC   Pulsed Eddy Current

PFP   Probability of False Positive

POD   Probability of Detection

POD(a)   POD vs Size

RCAF   Royal Canadian Air Force

RHM   Resampling by Half-means

RIM   Range Interval Method

SCC   Stress Corrosion Cracking

SHV	Smallest Half Volume
SSR	Sum of Squared Residuals
SVM	Support Vector Machine
TJ	Technical Justification

List of Appendices

1. Introduction

1.1 Background

Many aircraft operated by the Royal Canadian Air Force (RCAF) experience structural deterioration due to fatigue cycling. During a typical flight, aircraft are subject to cyclic loading resulting in cyclic stresses that can lead to physical damage to the aircraft structure. Continued cyclic loading contributes to crack development and damage accumulation that can lead to failure, at stresses well below the material's ultimate strength. The damage and failure caused by cyclic loading is known as fatigue [1].

To address the effects of fatigue cycling, the aerospace industry has adopted a damage tolerance approach in aircraft maintenance and design. Damage tolerance analysis is used in the development of in-service inspection programs that contribute to maintenance and life extension of the structure or component. Fatigue crack growth prediction models have been developed to support these damage tolerance concepts. Fracture mechanics properties such as fatigue crack growth data and fracture toughness are essential to conducting damage tolerance analysis of primary aircraft structures [1]. Fracture mechanics techniques and fatigue crack growth prediction models are essential tools in preventing the occurrence of failures, but the most critical component of the damage tolerance approach is detecting the flaw or damage.

Detection of accidental damage, corrosion and fatigue cracking can be accomplished by inspection. The inspections can be destructive or non-destructive in nature. However, utilizing non-destructive techniques provides several advantages over destructive testing methods. Some of the advantages of non-destructive testing (NDT) include improved cost-effectiveness, reduced unscheduled maintenance requirements and increased scheduled maintenance intervals [2]. Due to the advantages associated with NDT, the aerospace industry utilizes NDT to identify defects, abnormalities or imperfections during manufacturing and as part of in-service inspections to detect damage, degradation and deterioration of critical aircraft components [3]. The conventional NDT techniques that are utilized for the inspection of aircraft components and structures are ultrasonics,

radiography, liquid penetrant, magnetic particle, and eddy current [4]. Each of these NDT techniques differ with respect to their capabilities and limitations. Correspondingly, the selection of a specific inspection technique is dependent on several factors including inspection requirements, access to the test site, ease of test method application and inspection costs [3]. Additionally, with each NDT method there are influential parameters whose variation can potentially influence the outcome of an inspection [5]. If these influential parameters affect the inspection outcome in such a way, whereby the inspection can no longer meet its defined objectives, then these parameters become essential parameters [5].

The CP-140 Aurora and CC-130 Hercules aircraft currently operated by the RCAF are susceptible to cyclic fatigue cracks around ferrous fasteners in the bolt holes of single and multi-layer aluminum alloy wing and tail structures. The current method utilized to detect and monitor fatigue cracks around ferrous fasteners in CP-140 Aurora and CC-130 Hercules aircraft uses bolt hole eddy current (BHEC), which requires fastener removal [6]. Fastener removal involves manually drilling out the steel fastener, which can also cause significant damage to the surrounding aluminum surface. If these cracks are not detected or if crack growth is not effectively monitored, the crack can reach critical crack length and catastrophic failure can result [1]. Consequently, it is vital that an effective NDT technique is selected to detect and monitor crack growth. Additionally, being able to quantify the effectiveness of a NDT technique is therefore, an essential component in the framework of the damage tolerance approach. Probability of Detection (POD) is a method used to determine the reliability and sensitivity of a NDT technique, as it provides the methodology for estimating the detection capability [7]. The POD vs. target size relationship provides the basis for determining the capability of the NDE technique, where the objective with respect to component reliability is to determine the largest notch size that the system does not detect [8].

Multi-layer aluminum alloy wing structures consist of lap-joints, as shown in Figure 1.
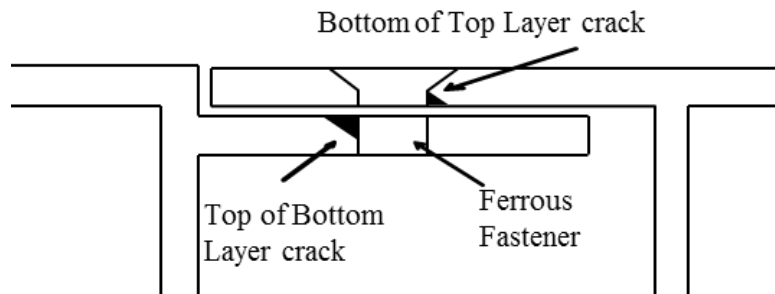
Figure 1: Cross-section of typical lap-joint configuration of multi-layer wing structure, displaying bottom-of-top and top-of-bottom layer cracks at the ferrous fastener location [9].

Second layer crack detection is currently assessed using bolt hole eddy current (BHEC) inspection methods, which requires fastener removal [6]. The fastener removal process consists of manually drilling out each fastener, which is labour and cost intensive and can result in significant damage to the surrounding aluminum surface [9], as shown in Figure 2 [10].

Figure 2: Example of typical damage that may be incurred by drill bit on aluminum surface of lap-joint structure during ferrous fastener removal process [10].

Aerospace engineering design focuses on a critical value identified as $a_{90/95}$, which is the target or flaw size that when inspected using NDT techniques, has a 90% probability of detection (POD), with 95% confidence [8]. Previous research identified the $a_{90/95}$ for conventional or BHEC technique on CP-140 Aurora aircraft to be 0.79mm (0.031 inch)

for electrical discharge machined (EDM) notches and 0.91 mm (0.036 inch) for fatigue cracks in as-drilled holes [6]. The United States Air Force identified $a_{90/95}$ for in-service flaw detection of edge/corner cracks in aluminum structures to be 1.27 mm (0.050 inch) [11]. Hence, this is the $a_{90/95}$ for conventional or BHEC technique currently being used on CP-140 Aurora aircraft. Pulsed eddy current (PEC) is one technique that was examined [12] as an alternative method to BHEC for detection of fatigue cracks in multi-layered wing structures. One main advantage of PEC as an inspection technique in this application is that fatigue cracks can be detected without fastener removal. The ferrous fastener is used as a conduit for the magnetic flux generated by the driving coil, which allows for deeper penetration of the eddy currents and corresponding response of the pick-up coils to greater depth [9]. As the $a_{90/95}$ for PEC detection of second layer notches has not been identified, this thesis work focuses on conducting a POD analysis to determine this value and the reliability of the PEC technique.

1.2 Eddy Current Testing

Eddy current testing (ET) is an electromagnetic NDT method that is used to detect flaws or discontinuities in structures that have electrical conductive properties. The ET technique utilizes a probe that consists of electromagnetic coils, which produce an alternating magnetic field. Eddy currents are induced in a conductor by this time-varying magnetic field (according to Faraday's Law), and flow in a circular pattern, which is oriented perpendicular to the direction of the magnetic field [4]. Specimen properties such as geometry, magnetic permeability, electrical conductivity and discontinuities impact the magnitude and distribution of these eddy currents [4]. Flaws or discontinuities disrupt the normal path and strength of the eddy currents. These changes to the eddy current signal can be observed using an impedance plane display, resulting in detection of the flaw or discontinuity present in the specimen.

One of the limitations of ET is that permeability variations inherent in ferromagnetic materials can be falsely identified as flaws [4]. Additionally, the high relative magnetic permeability in these materials also impedes the depth of penetration of eddy currents and consequently minimizes the effective depth of inspection [4]. Hence, when testing

ferromagnetic components, ET is only effective for detection of flaws at or near the surface of the specimen.

It has been shown that greater depth of eddy current penetration can be achieved, even in the presence of ferromagnetic fasteners, with the use of PEC [13]. Pulsed eddy current utilizes a square wave excitation, representing a spectrum of discrete frequencies, as opposed to conventional eddy current techniques that use a sinusoidal excitation. The frequency of the excitation current effects eddy current depth of penetration and based on skin depth relationships, it has been suggested that utilizing PEC should allow for deeper penetration as opposed to conventional eddy current methods [14].

1.3 Outlier Detection

Outliers, as related to this thesis work, are defined as an observation that shows characteristics that are inconsistent with the remainder of the data set [15]. Outlier detection indicates the presence of abnormal conditions, and as such is vital to safety critical applications such as in the identification of fatigue cracks in aircraft structures. There are three fundamental types of outlier detection – Type 1 (unsupervised clustering), Type 2 (supervised classification) and Type 3 (semi-supervised recognition) [16].

In the Type 1 outlier detection method, outliers are identified with no prior knowledge of the data set. A system model is then fitted to the remaining data in the set until no further outliers are detected. An alternative approach involves accommodation, where outliers are incorporated in the system model, which then utilizes a robust classification method [16]. Robust approaches such as the smallest half volume (SHV) method [17], establish a boundary around the majority of the data that is representative of normal behaviour, while accommodating outliers in the data [16].

Type 2 outlier detection methods, such as support vector machine (SVM) and artificial neural networks require prior knowledge of both normal and abnormal behaviour [16]. In this approach there is a requirement for an adequate spread of normal and abnormal data to cover the entire distribution so the algorithm can appropriately sort the data into

pre-defined groups [16]. This method is most suited to static data, because any shifts in the distribution may result in insufficient information to adequately categorize the data and the region may be classified incorrectly [16].

A Type 3 approach to outlier detection requires prior knowledge of data classified as normal. The algorithm is taught the normal classification and learns to recognize abnormality. Cluster analysis [18] is a Type 3 system that classifies a new data point as normal, if it lies within a predetermined boundary, and an outlier, if it lies outside this boundary. This method requires an adequate assortment of normal data points to determine classification effectively. For cluster analysis to be effective at identifying outliers, an appropriate algorithm must first be selected that can accurately model the data distribution and then a suitable threshold is identified to define the boundaries of normality [16].

This thesis work utilized the PEC probe with an 8 mm ferrite core and eight differentially paired pick-up coils highlighted by Butt [12] to be the most effective at detecting fatigue cracks in second layer aluminum structures. The resulting PEC signals are analyzed using the statistical analysis method known as modified principal components analysis (MPCA), described by Horan *et al* [19], which reduces the transient PEC signals to a series of eigenvectors and scores. The MPCA scores are analyzed using Type 3 outlier detection methods consisting of cluster analysis and the determination of a Mahalanobis distance (MD), followed by a smallest half volume (SHV) analysis, which is a Type 1 approach [17]. These approaches in concert with the boundary threshold, distinguish between the signals associated with electrical discharge machined (EDM) notches and those of blank fastener locations to identify locations with notches present [10] [20].

1.4 Research Survey

A literature review of the field of eddy current testing as related to aircraft structures and hit/miss POD analysis was performed. Specifically, analytical work related to PEC applications as related to the inspection of second layer aircraft lap-joint structures, upon which this research is built, is summarized. The current accepted methods used to analyze hit/miss data, involving likelihood ratio method and confidence bounds are

reviewed. Following this, articles related to statistical best practices with respect to POD studies to quantify inspection reliability of an NDT technique are reviewed. Finally, an article detailing the factors that influence hit/miss POD and recommended guidelines to assess the effectiveness of a binary response inspection system is examined.

Although extensive research has been conducted with respect to conventional eddy current testing for detecting flaws in aircraft structures, recent research [10][12][13][14][21][22] in PEC, specific to lap-joints and second layer detection, will be the focus of this review.

The basis for transient diffusion and crack detection in the presence of ferrous fasteners is explored. Additionally, PEC signal analysis using multivariate analysis techniques such as Principal Components Analysis (PCA), Modified Principal Components Analysis (MPCA), Mahalanobis Distance (MD), and Support Vector Machine (SVM) [23] are examined. These multivariate analysis techniques provide the basis from which crack detection in lap-joint and multilayer aircraft structures are viable. Following this, articles related to statistical best practices with respect to POD studies and analyzing hit/miss data to quantify inspection reliability of an NDT technique are reviewed. Finally, an article detailing the factors that influence hit/miss POD and recommended guidelines to assess the effectiveness of a binary response inspection system is examined.

1.4.1 Analytical Work

While defect identification using conventional eddy current techniques with sinusoidal excitation has been utilized for many years, development of applications associated with PEC methods have been the focus of recent research, as discussed in the following sections. One of the earliest examples of pulsed or transient eddy current, which utilizes a square wave excitation, dates back to 1921 with the work of Wwedensky [24]. His research described the time-dependent eddy current diffusion that results with the abrupt application of a uniform magnetic field into a long cylindrical conductor. Although Wwedensky's assumption that the uniform field exists out to infinity violates Maxwell's second law [25], his work formed the foundation for analysis of transient excitation. As a consequence of this assumption, the advancement of analytical solutions for transient

excitation has been limited, due to the eddy current field being confined to the cylindrical conductor.

The analytical and experimental work presented by Dodd and Deeds [26] concerned the sinusoidal excitation of a single coil surrounding a rod of conducting material under the assumption of constant amplitude alternating current excitation. With pulsed excitation there is a significant feedback effect from the sample on the drive coil [25], hence the approach by Dodd and Deeds [26] would not be applicable.

In 1972, Callarotti *et al* [27] developed solutions for models consisting of conductive and non-conductive cylinders surrounded by a conductive shell. A time-dependent magnetic field, oriented along the axis of the cylinder was applied. As a result of the surrounding conductive shell being finite, boundary conditions were applied such that there was continuity of tangential magnetic and electric fields between the inner and outer regions. A solution was obtained for the relative magnetic permeability of the sample, however only a thin-shell approximation was validated due to the complexity of the model. Hence, this work would not be applicable to an aircraft lap-joint configuration as the aluminum structure surrounding a bore hole has significant thickness and is not considered a thin shell.

In 2011, Desjardins [38] developed the first complete analytical model for the transient eddy current response in the presence of both ferrous and nonferrous conducting rods. Desjardins [38] developed stationary and transient solutions with one of the models consisting of a step function current applied to a bore hole in an infinite aluminum plate. In this application of magnetic diffusion, the system was considered with respect to the magnetic vector potential. The continuity of the magnetic vector potential across the boundary from inside the bore hole to the surrounding medium, gave rise to solutions being developed for both the stationary and transient response accounting for the three-way feedback effects between the driving coil, sample and pick-up coil [25]. Desjardins *et al.* [31] concluded that the presence of ferrous fasteners allowed for greater depth of penetration when considering multilayer aluminum structures, as compared to multilayer aluminum structures alone.

1.4.2 Detecting Defects in Multi-Layer Aircraft Structures

In 2013, He *et al.* [28] investigated the use of PCA and SVM methods to categorize defects in two-layer specimens. Support Vector Machine (SVM) is a machine learning method of classification that requires input of a training database that is necessary for the decision-making process [28]. Their paper notes that defects at different depths will affect the shape of the time-domain response of the PEC signal, and these amplitude changes may also be impacted by defect width. It was noted that the amplitude changes were weak in comparison to the maximum amplitude, and with normalisation of the signal it improves the ability of the PCA method to extract new features from the transient response. Without normalisation of the signal and using only the first two principal components, He *et al.* [28] achieved a defect detection rate, or classification accuracy, of 83.4%. With normalisation, the principal components were used as input parameters for the SVM classification method, and consequently, a classification accuracy of 100% could be achieved [28]. Results were also improved when considering the inclusion of variable lift-offs of between 0 mm and 1.4 mm, with classification accuracy being 61.4% and 91.7% before and after normalisation, respectively [28]. The results presented by He *et al.* [28] effectively showed that when normalisation is conducted prior to PCA, air gap and lift-off effects in multi-layer structures can be eliminated.

Also in 2013, Pan *et al.* [29] presented a method using PCA and selected frequency responses to eliminate air gap effects, when conducting experiments on two-layer structures using PEC. The focus of the research was on the classification of surface and subsurface defects as well as material thickness changes, while accounting for lift-off and interlayer gap effects that hindered this defect classification [29]. This paper demonstrated the feasibility of extracting additional information from the selected frequency responses of the PEC signal and using this information as an alternative PCA input, as opposed to previous works that only used the time-domain response [29]. It was recognized that since PEC utilizes a square-wave excitation signal, the Fourier series needed to be expanded to present the signal in terms of fundamental frequencies and harmonic components [29]. Pan *et al.* [29] concluded that frequency responses in the range of 3.7 to 5.4 kHz were best for classification of second-layer defects, whereas

frequency responses at the higher frequencies, ranging from 12 to 25 kHz were best for classification of first-layer defects. Their work presents a method for classifying both surface and subsurface first and second layer defects, in the presence of air gaps varying from 0 mm to 1.4 mm.

In the same year, He *et al.* [30] assessed PCA and Independent Components Analysis (ICA) methods of feature extraction using an SVM based algorithm for classification. The PEC signal associated with a multi-layer specimen was initially transformed from the time-domain to the frequency-domain and PCA or ICA was then used to extract features [30]. After defect classification was conducted using SVM, it was noted that PCA was not able to process signals that are independent of one another, whereas ICA separates the multivariable signal into additive subcomponents and thus can calculate independent components [30]. The results presented by He *et al.* [30] using only PCA, found that 82% of defects could be classified, and when the first two principal components were input to the SVM algorithm, the classification accuracy increased to 98.9%. Comparatively, when the ICA method was used alone, the classification accuracy was 85.4% and when they were input into the SVM algorithm, the classification results increase to 100% [30]. When the results were plotted graphically, the independent components also displayed a tighter cluster when compared to the principle components, which suggests that the ICA method is more suitable as a classification method [30]. He *et al.* [30] concluded that ICA produced superior results over PCA for feature extraction and when combined with an SVM classification algorithm, the results were further optimized. They also concluded that time-domain responses produced better results with respect to detecting bottom layer defects, whereas frequency-domain responses excelled were more effective at detecting top layer defects [30]

1.4.3 Detecting Defects in the Presence of Ferrous Fasteners

In 2010, Whalen [13] demonstrated that a PEC probe design consisting of a central driving coil and differentially paired pick-up coils could detect subsurface notches at the edge of a bore hole, in the presence of a ferrous fastener. Whalen [13] reasoned that when the ferrite core of the probe was aligned concentrically over the fastener head in a

multilayer sample, the ferrous fastener would act as a conduit that contributed to carrying magnetic field, and resulting eddy currents, to greater depths. Whalen [13] suggested that as the diameter of the ferrite core increased relative to the fastener head diameter, the flux and depth of penetration would increase, and consequently detection rates would improve [13].

In 2012, Desjardins *et al.* [31] tested an aluminum plate containing a bore hole using PEC and concluded that there is a larger flux transfer into the aluminum structure due to the magnetization of the fastener when transient fields are used. Their results also confirmed that when a steel rod was inserted in the bore hole, the induced currents were enhanced, which resulted in an increase in transient decay times [31]. Consequently, this resulted in an enhanced depth of penetration of the eddy current field into the plate perpendicular to the aluminum surface and improved the potential for increased defect detection when the eddy current field is disrupted by cracks emanating from the bore hole surface [31]. The work by Desjardins *et al.* [31] also included a probe design with a transmit coil and two differential receive coils positioned on either side. The aluminum sheet contained a ferrous fastener in the bore hole and the transmit coil of the probe was centered over the fastener. The result was enhanced flux concentration along the ferrous fastener and increased flux distribution that was evident when the steel rod was inserted in the bore hole. Desjardins *et al.* [31] determined that the probe design with the differential receive coils produced a large signal variation as a function of rotary angle when inspecting two cracks positioned 1.3 mm from the surface of the aluminum plate. These results suggested that the enhanced flux created by the ferrous fastener provided the necessary depth of penetration to potentially detect subsurface cracks in lap-joint structures [31].

In 2013, the research of Horan *et al.* [19] extended the PEC probe configuration presented in Desjardins' work [31] for the application of inspection of stress corrosion cracking (SCC) in CF188 Hornet inner wing spars. The experimental setup consisted of a PEC probe with driving coil wound around a ferrite coil with two differentially paired pick-up coils spaced 180º apart [19]. The specimen consisted of two layers, the top consisting of 13 mm layer of Nylon 6 polymer and underneath, a 3.2 mm thick layer of 7075-T6 aluminum, secured with ferrous fasteners. The sample consisted of both

blanks and notches, with the notches being of varying lengths and a constant width of 0.2 mm, extending from the edges of the bore holes oriented towards the adjacent fasteners. The probe was aligned over the ferrous fasteners in the simulated wing spar with an acrylic alignment tool with measurements taken from an equal number of blank and notched fastener sites. In this research, Horan *et al.* [19] presented the MPCA formalism that was used in the analysis of these measurements for defect classification of a simulated wing spar. Conventional PCA uses subtraction of the average response to produce scores, whereas MPCA does not [18]. Using the MPCA method described by Horan *et al.* [19], the first four eigenvectors were computed and these eigenvectors were used to produce scores that represent the PEC signal. The process of defect classification consisted of initially plotting the different scores, one as a function of another and observing clustering of the data. The results of Horan *et al.* [19] suggested that it is possible to distinguish between notched and blank fastener sites by plotting s_3 as a function of s_2. Although the acrylic alignment tool was used for probe placement, misalignment did occur, due to human error. It was observed that the transient response of the probe was sensitive to a slight misalignment of the probe with respect to the fastener, most notably in the direction of the differential pick-up coil pair, whereas misalignment in the direction perpendicular to the pick-up coils with displacements up to 0.5 mm had very little effect [19]. In this work, Horan *et al.* [19] also introduced the concept of Mahalanobis Distance (MD), used in conjunction with data clustering as a method to identify extreme outliers in the data set, and potentially flag data associated with probe misalignments.

In 2014, Babbar *et al.* [32] developed finite element analysis (FEA) models using a simulated test piece with geometry representing a CP-140 Aurora aircraft wing and transient signals from a PEC probe to simulate notch detection in lap-joint structures. The simulated PEC probe consisted of a central driving coil surrounded by eight differentially paired pick-up coils. The two-layer aluminum sample included varying crack sizes with varying orientations. The results presented by Babbar *et al*. [32] concluded that the presence of the lap-joint edge on the top layer at 270º and on the second layer at 90º resulted in significant differential signal responses even with no defect present. It was also observed that second layer cracks produced weaker signals,

which have peaks occurring at later times [32]. The FEA results confirmed the observations made by Horan *et al.* [19] with respect to probe off-centering. The signal responses were negligible when the shifts in the probe were directed perpendicular to the lap-joint edge and significantly different with respect to shape and amplitude when the probe was shifted in the direction of the lap-joint edge [32]. To differentiate between the information associated with cracks and blanks in the PEC signal, and account for variance in the probe placement, the signals generated by the FEA model were analyzed using PCA [32]. The modelling results presented by Babbar *et al.* [32] associated the first eigenvector with the probe displacement and the second eigenvector with second layer cracks. This research work demonstrated the potential for PEC to be used in the determination of second layer crack depth, orientation and detection of cracks in the presence of ferrous fasteners in multi-layered aluminum structures [32].

In 2015, Stott *et al.* [9] presented experimental results using the probe design previously simulated by Babbar *et al.* [32] to obtain PEC measurements of a multi-layer aluminum lap-joint structure representative of a CP-140 Aurora wing. The probe consisting of a central driving coil encircled by eight differentially paired pick-up coils was used to take measurements on a single sample consisting of both blanks and EDM notches of varying size (0.89 to 5.46 mm) and orientation. Due to the variation in crack sizes and orientations, discriminant analysis [18] was not successful at differentiating between notches and blanks, however it was identified that the MD could be used to calculate a relative distance between scores of notched and blank fastener sites [9]. This relative distance showed a direct correlation with crack size, suggesting the potential for obtaining the sizing of cracks via the application of PCA and MD [9]. The results obtained by Stott *et al.* [9] for the single sample demonstrated 97% detection rate of the notches with 99% confidence and 4% false call rate. When the false call rate was increased to 10%, a 100% detection rate with 95% confidence was achieved [9]. It was noted that the consequence of a false call would be the removal of the fastener and conventional BHEC inspection being performed [9]. The current method of BHEC requires removal of all fasteners, whereas even an elevated false call rate associated with a PEC inspection would result in fewer fasteners being removed, while maintaining a high level of detection. Hence, the work presented by Stott *et al.* [9] demonstrated the

effectiveness of the PEC system in crack detection of a multi-layer aluminum lap-joint structure and provided the basis for potential savings in both cost and inspection time.

In 2016, Butt *et al.* [12][36][37] presented experimental results that demonstrated the ability to detect second layer cracks in multi-layer aluminum wing lap-joint structures without fastener removal, using a PEC technique. Four different probe configurations, with varying driving coil core diameters were evaluated, to assess the correlation between core diameter and the magnetic flux transferred to the fastener, as postulated by Whalen [31]. Each probe configuration that was assessed consisted of a central driving coil wound around a single ferrite core, with eight pick-up coils positioned symmetrically around the driving coil, differentially paired at 180º from each other. Differentially pairing the pick-up coils in this configuration allowed for differences in their signals to be collected for post-processing [12]. An enhanced PEC probe design was then selected, which consists of a central driving coil wound around an 8 mm ferrite core, which is slightly larger than the ferrous fastener head (7.0 mm).   This larger core size allowed for increased magnetic flux to be transferred to the fastener, and consequently induces more eddy currents into the surrounding aluminum structure [31]. An op-amp in a noninverting configuration was connected to the probe and used to improve the input signal and resulting signal reproduction. A series of seven NAVAIR samples, based on the wing lap-joint structure of the Lockheed P-3 (CP-140 Aurora) aircraft were examined. The samples consisted of two sections of 2024-T3 aluminum plate joined together by a row of 22 to 24 ferrous fasteners in a lap-joint configuration. The plate thickness of the samples ranged from approximately 2.1 mm to 2.8 mm, the fastener to lap-joint edge distance varied by +/-1.8 mm, and the interlayer gap between top and bottom sheets ranged from 0 mm to 0.12 mm [12]. Each sample contained both blank and EDM notched fastener sites, with the notches positioned at the bottom of the top layer and the top of the bottom layer at a 45° angle to the edge of the bore hole, giving an aspect ratio of 1:1.   The EDM notches ranged in size from 0.76 mm (0.030 inch) to 6.10 mm (0.240 inch), at 45° intervals, oriented clockwise from the lap-joint edge. An acrylic alignment guide was used to position the PEC probe concentrically over the fastener head. The transient response signal was gated between 0.50 and 0.58 ms, as it was identified that this was the portion of the time domain signal that

contained second layer notch information [12]. Once gated, the signal responses were post-processed to produce eigenvectors and MPCA scores for each of the four differential coil pairs and cluster analysis was performed, which consisted of determining the covariance matrix from the scores of known blanks for each coil pair [12]. The MD was calculated for each fastener by coil pair and compared to a threshold value to determine whether a notch was present, such that if a hit was obtained on any coil pair, the result was recorded as a hit for that particular fastener [12]. The data was further processed to remove effects due to environmental conditions, repeat measurements and fastener to lap-joint edge distance [12]. Detection results obtained for the seven NAVAIR samples ranged from 81% to 99% with a false call rate of 5% [12]. The work presented by Butt *et al.* [12][36][37] demonstrated the effectiveness of the PEC system as a viable method for detecting cracks in multi-layered wing structures without ferrous fastener removal.

1.4.4 Multivariable Outlier Detection

In 1998, Egan *et al.* [17] presented two simplified methods for detection of multivariable outliers using robust statistics. They identified that traditional methods of outlier detection, that use mean and standard deviation could not deal appropriately with multiple outliers. Egan *et al.* [17] noted that the outliers distorted the mean but also had a significant effect on the covariance matrix, hence outlying observations could be missed when the MD was utilized. Since traditional methods of outlier detection were not adequate to deal with multivariate data, and hence should be replaced with more robust methods [17]. Egan *et al.* [17] noted that two robust methods, minimum volume ellipsoid (MVE) and minimum covariance determinant (MCD) often required significant computational time when dealing with higher dimensions. They presented two simplified methods, resampling by half-means (RHM) and smallest half-volume (SHV), which relied on a center for the data, rather than using the covariance matrix and the statistical relationship was developed with only 50% of the full data set [17]. The results presented by Egan *et al.* [17] suggested that not only were the RHM and SHV superior classification methods when accounting for multiple outliers, as compared to MVE or MCD, but also required less computation time.

The work presented by Butt *et al.* [12][36][37] as discussed in Section 1.4.3 also included results using robust statistics simulations. The SHV method was identified as a potential solution to the need for an in-situ calibration standard when the PEC inspection method was utilized in field applications [12]. Previous work by Underhill *et al.* [33] suggested a minimum sample size of 40 blank fasteners were required for desirable detection results when using the PEC inspection technique presented by Butt [12]. Consequently, randomly generated covariance matrices, with statistically similar properties to the original measurement data, were used to artificially inflate the number of blank fastener measurements to simulate real world detection scenarios [12]. The simulation technique used an SHV algorithm to classify the data, and as with the analysis for the experimental measurements of the single sample cases, a decision threshold was used to compare to the MD. It was determined that the best detection performance was achieved when a data fraction of 97.5% was used to construct the covariance matrix [12][36][37]. This confirmed the results of previous work obtained by Underhill *et al.* [20]. The blind detection results obtained by Butt [12] were 99% detection for Standards 22, 24 and 25 at 5% false call rate, whereas there were apparent detection issues with the remaining four samples. Additionally, when three samples were combined, the detection rate is substantially reduced, where Butt [12] attributes the loss of sensitivity to additional variables which were not being accounted for in the analysis process. The thesis work presented by Butt [12] suggested that self-calibrating blind detection of second layer cracks in aircraft wing structures was possible in the presence of ferrous fasteners, without prior knowledge of the sample being tested or the requirement for costly calibration standards.

1.4.5 Essential Factors for Hit/Miss POD Curves

In 2012, as a result of work in support of the European Network for Inspection and Qualification (ENIQ) Task Group on Risk (TGR), Annis and Gandossi [34] presented a technical document on the influence of sample size and other factors on hit/miss POD curves. The report highlighted the approach that ENIQ recommends for qualification of inspection procedures, utilizing both test piece trials and a technical justification (TJ) [34]. The TJ provided the evidence, mostly in the form of quantitative measure,

justifying that the inspection technique meets objectives with respect to an inspection. This measure provides the metric necessary to quantify probability of failure and risk reduction [34].

The report, built upon the statistical best practices for POD curves outlined in [35], identified a minimum of 60 targets for hit/miss POD analysis, and provided guidelines for assessing the effectiveness of an inspection system consisting of binary data, using POD vs size curves. Annis and Gandossi [34] highlighted the scope and limitations of using the **mh1823 POD** software [8], most notably that the specimens must have targets that provide a measurable characteristic, such as size, where there is an assurance of precision. They provided guidelines with respect to POD vs size analysis of binary response inspection systems that recommended a minimum number of 60 targets for hit/miss POD vs. size modelling. Additional recommendations were made with respect to a uniform target size distribution and that the target range should result in a specimen POD coverage of POD = 0.03 to 0.97 [34]. They observed that when sample size was reduced to less than 60, there was the possibility of producing numerically unstable results that affect the Generalized Linear Model (GLM) parameter estimation, and most importantly, that the confidence bounds could widen [34]. The results concluded that there are several influential factors that impact POD vs. size results in addition to sample size. These included target size range, distribution and location.

This research work will present the first POD vs. size analysis on the PEC system for the detection of second layer cracks in aircraft lap-joint structures.

1.5 Objective

Previous PEC work completed by Butt *et al.* [12][36][37] demonstrated effective notch detection capability for three NAVAIR samples using both cluster analysis and SHV analysis methodologies, with simulations producing blind outlier detection capability. Initially, the focus of this work is to further optimize the equipment and enhance the analysis techniques used for PEC inspection of aircraft lap-joint structures in the presence of ferrous fasteners to address deficiencies in the system encountered by Butt [12]. Essential parameters implicit in the cluster analysis and robust analysis methods,

such as eigenvalues, eigenvectors, covariance and correlation matrices will be analyzed to determine patterns or specific variables that may adversely influence detection results. The intent is to determine the factors that are implicit in optimizing the cluster analysis and robust statistics methods. Once all essential parameters are accounted for, the focus is then to prove the reliability of the PEC inspection technique by producing effective detection results when expanding the sample set to all seven NAVAIR specimens. The focus will also be to prove reproducibility by demonstrating effective detection capability across all NAVAIR samples, using the same experimental instruments utilized by Butt [12] to obtain PEC signal measurements. The robustness of the PEC inspection system will also be analyzed by increasing parameter variability with respect to off-centering and temperature of the probe, and distance from fastener-to-lap-joint edge. Effective detection capability without prior knowledge of unflawed behaviour is advantageous, especially in this application where it is difficult to obtain abnormal data, and it minimizes the requirement and high cost associated with calibration pieces. Hence, blind outlier detection methods will also be utilized in simulations involving cluster analysis and SHV methodologies. Confirmation of the reliability, reproducibility and robustness of the PEC system contributes to the viability of the technique to be utilized for in-service use.

The main goal of this thesis work is to then conduct a POD study to determine the reliability and detection capability of the PEC inspection technique. The POD analysis provides the metric needed, in the form of an $a_{90/95}$ value, to assess the reliability associated with PEC inspection technique in the detection of second layer cracks at ferrous fasteners in aircraft lap-joint structures. The analysis will consider both measured results and results obtained from simulations, where blind outlier detection methods are employed. Additionally, parameters that affect the $a_{90/95}$ value will be identified and assessed with the overall goal of optimizing the $a_{90/95}$ value and increasing the detection capability of the PEC system. The result of this work has significant implications associated with demonstrating the viability of this inspection technique as an approved NDT technique for this application, resulting in field deployment and in-service use.

The following objectives represent the areas of research that extend beyond the initial work presented by Butt *et al* [12][36][37], with the main objective being to conduct the first POD analysis of the PEC inspection system:

- Determine impact of increased precision on fastener to lap-joint edge distance on detection rate.
- Determine effects associated with variation in probe temperature on detection and false call rates.
- Assess impact of probe off-centering on detection and false call rates.
- Demonstrate reproducibility, reliability and robustness of PEC inspection technique across all seven NAVAIR sample series.
- Demonstrate blind outlier detection capability of PEC inspection technique across all seven NAVAIR sample series.
- Conduct a probability of detection (POD) analysis of the PEC inspection technique, and produce the metric required, in the form of an $a_{90/95}$ value.
- Determine optimum data fraction utilized in simulations involving cluster analysis, SHV, and bootstrap methodologies to produce best performance overall with respect to detection and false call rates, and $a_{90/95}$ value.
- Assess impact of variation in false call rates utilized in simulations involving cluster analysis, SHV, and bootstrap methodologies, with respect to detection rate and $a_{90/95}$ value.
- Determine effects of sample grouping size utilized in simulations involving cluster analysis, SHV, and bootstrap methodologies, with respect to detection rate and $a_{90/95}$ value.
- Assess the effects of notch size range and distribution, mis-located targets and specimen POD coverage on $a_{90/95}$ value.

1.6 Thesis Scope and Methodology

The following presents an outline of the remaining sections and methodology for this thesis.

Section 2 presents the electromagnetic theory that provides the foundation for eddy current inspection including Maxwell's Equations, magnetic permeability and electromagnetic diffusion. This is followed by a discussion of eddy current generation, and skin depth as it relates to flaw detection. Theory is then presented on pulsed eddy current generation in a conductive medium and equivalent circuit models. Following this, statistical theory related to analysis of PEC is presented, which includes MPCA. Cluster analysis and MD theory is presented as a Type 3 outlier detection method, which is utilized in second layer crack detection. Finally, the theory involving robust statistics methods such as SHV are presented as a Type 1 outlier detection method for an alternative approach to flaw detection.

Section 3 describes the guidelines for assessing the capability and reliability of an NDE system. Descriptions of the guidelines that are necessary to ensure a valid assessment of an NDE system in terms of POD as a function of target size are presented. Specifics with respect to test variables, physical characteristics of the test specimens, the inspection process and target size are summarized.

Section 4 provides an overview of the experimental setup utilized in the acquisition of data for this thesis work. The PEC probe used in this work is identified, and the specifications with respect to the core, driving coil resistance and differential pick-up coil pair spacing is presented. This is followed by a description of the data acquisition (DAQ) system and associated operational amplifier that were used to control and measure the critical PEC system parameters. The physical characteristics of the NAVAIR sample series, which detail the general dimensions, material properties, and most importantly notch length, orientation and location for each sample are then detailed. Finally, the probe alignment technique and associated equipment used in gathering experimental data is described.

Section 5 presents the signal response post-processing methods utilized to determine notch detection capability. Gating of the transient signal response is described, identifying the importance of extraction and analyzing only the information associated with the notch. Following this, the process of generating MPCA scores for each of the four coil pairs is presented. An overview of the cluster analysis method, which is applied

to the MPCA scores, and the requirement for information to be known about the blank fastener signal response, as it is a Type 3 outlier classification methodology is given. Environmental factors are discussed including probe temperature and its effect on the inspection results as well as measures that were taken to remove shifts in the data due to other unexplained environmental factors. Repeat measurement and probe off-centering effects are investigated, as well as the enhancement of the cluster based classification methodology that Butt *et al.* [12][36][37] presented to address these effects. A detailed discussion of the effects of fastener proximity to the lap-joint edge and the importance of measurement precision of this distance is given, as this parameter can significantly impact the flaw detection rate. Finally, the SHV robust statistics algorithm which is used as an alternative Type 1 outlier classification methodology for second layer crack detection, is described. This approach provides a method for blind outlier detection, which is advantageous for the eventual field deployment of this PEC system where calibration samples are not easily generated and advanced knowledge of blank fastener signals may not be practical. The effects of data fraction and sample size will also be discussed with respect to simulations using the SHV algorithm.

Section 6 outlines the methodology used to conduct the POD analysis on the PEC inspection system in the detection of second layer cracks at ferrous fasteners in aircraft lap-joint structures. The selection of an appropriate POD model that accommodates the binary data that is obtained through experimental measurements will be discussed. Following this, the limitations of the **mh1823 POD** software used to conduct the POD analysis will be discussed. It is determined that the POD model associated with the logit function provides the best fit of the measurement data, and the analysis to confirm this selection is presented. The effects associated with notch size range, mis-located targets and notch size distribution are then examined, detailing their impact on the POD results and the $a_{90/95}$ value. Following this, the effect of sample size on confidence bound widths as well as the probability of false positives is discussed. Finally, the rationale for refining the $a_{90/95}$ value obtained with the **mh1823 POD** software is discussed in detail. The process utilized to refine the $a_{90/95}$ value using ImageJ software is described.

Section 7 presents the experimental results obtained from measurements taken from all seven NAVAIR test samples, representative of the structure of the CP-140 Aurora (Lockheed P-3) aircraft. First, the results obtained for Sample 26 with revised fastener-to-lap-joint edge distances are compared with results using previously defined distances presented by Butt [12]. All of the results presented in this section utilize the revised fastener-to-lap-joint edge measurements, with increased precision. The single sample detection results for all seven samples are then presented. This is followed by the results obtained for four samples when varying probe temperature at 10ºC, 20ºC and 30ºC. Additional results are also presented on the experimental measurements for three NAVAIR samples, considering off-centering of the probe with respect to the fastener head. This is followed by simulation detection results, which utilized the SHV algorithm and are examined for the two cases consisting of a sample size of either 40 or 80. The results compare variation in sample size, a data fraction of 95%, 97.5% and 100%, and decision thresholds that were linked to 5%, 10%, and 15% false call rates. Finally, POD vs. size analysis results are presented for the combined NAVAIR sample series consisting of all seven samples using the **mh1823 POD** software. The analysis results include POD vs. size results for both 40 and 80 samples, the three data fraction percentages and three false call rates. Finally, the $a_{90/95}$ results obtained when processing the POD vs. size repeat measurement plots with ImageJ software are presented. These results include simulations conducted with 80 samples, the three data fractions and their associated false call rates.

Section 8 discusses the significance of the results obtained from evaluation of the seven NAVAIR sample series specimens. The impact of precision with respect to fastener to lap-joint edge distance measurements on detection rate is discussed, which includes a comparison of the detection results obtained prior to, and after edge distance revision. The impact of sample size, notch range and distribution on the POD vs. size and $a_{90/95}$ results obtained for simulations utilizing cluster analysis and SHV methodologies are explored next. The impact of environmental factors such as probe temperature and off-centering on detection rates and $a_{90/95}$ results are examined. Additionally, the variation observed between measured and simulation results as well as specimen characteristics are investigated to emphasize that there are material parameters and simulation

algorithm behaviours that require further study. Irregular fastener spacing and damage as well as sources of error are presented. A discussion of the reliability of the PEC inspection technique for this application is presented. The rationale for considering a 10% false call rate to be associated with this technique is also presented. This is followed by a comparison of the $a_{90/95}$ results obtained using both the **mh1823 POD** and ImageJ software packages, and a discussion with respect to determining an $a_{90/95}$ value that is to be associated with this inspection technique.

Section 9 summarizes the results of this work, providing not only the metric necessary ($a_{90/95}$), but also an assessment of this PEC system for the inspection of aircraft lap-joint structures. Recommendations for future work to expand the assessment and optimize this technique are also outlined, to facilitate this technique towards being accepted and ultimately deployed for in-service second layer crack detection in aircraft structures.

2. Theory

2.1 General

The theoretical concepts associated with eddy current inspection and POD analysis are presented in this section. These include electromagnetic field theory, material-related magnetic properties, magnetic field diffusion, eddy currents, skin depth, hit/miss analysis, test variables, false positives and the $a_{90/95}$ standard.

2.2 Electromagnetic Theory

Foundational to electromagnetic theory are Maxwell's field equations, which describe how electric and magnetic fields propagate, interact and are influenced by material parameters. Maxwell's field equations, presented in vector form in the International System of Units are shown in equations (2.1) to (2.4) [38] below:

$$\boldsymbol{\nabla} \cdot \boldsymbol{E} = \frac{\rho}{\varepsilon} \tag{2.1}$$
$$(Gauss'\,Law)$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0 \tag{2.2}$$
$$(No\,Monopoles)$$

$$\boldsymbol{\nabla} \times \boldsymbol{E} = -\frac{d\boldsymbol{B}}{dt} \tag{2.3}$$
$$Faraday's\,Law$$

$$\boldsymbol{\nabla} \times \boldsymbol{B} = -\mu\boldsymbol{J} + \mu\varepsilon\frac{d\boldsymbol{E}}{dt}. \tag{2.4}$$
$$(Amp\grave{e}re's\,Law\,with\,Maxwell's\,Correction)$$

The electromagnetic fields, **B** (magnetic flux density) and **E** (electric field) are established by the current density **J** and the charge density ρ. The material parameters presented in Maxwell's equations include the medium's magnetic permeability μ and the electrical permittivity ε.

Additional constitutive relations describing the electrical response of various materials, considering characteristics such as permeability and conductivity, are also needed to formulate general solutions to electromagnetic systems. Ohm's Law, expressed in equation (2.5),

$$\boldsymbol{J} = \sigma\boldsymbol{E}, \tag{2.5}$$

relates the current density to the electric field [38], where σ is the medium's electrical conductivity.

Alternatively, the current density can be related to the magnetic vector potential $\boldsymbol{A}$, and this is developed below. A relationship exists between the magnetic flux density $\boldsymbol{B}$ and the curl of the magnetic vector potential $\boldsymbol{A}$, such that

$$\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A}. \tag{2.6}$$

Using this relationship, equation (2.3) can be expressed as follows:

$$\boldsymbol{\nabla} \times \boldsymbol{E} = -\boldsymbol{\nabla} \times \frac{d\boldsymbol{A}}{dt}. \tag{2.7}$$

From Ohanian [39], since a field is derivable from a potential, where

$$\boldsymbol{E} = -\boldsymbol{\nabla}V, \tag{2.8}$$

then $\boldsymbol{\nabla} \times \boldsymbol{E}$ vanishes. Correspondingly, since

$$\boldsymbol{\nabla} \times (-\boldsymbol{\nabla}V) = \boldsymbol{0}, \tag{2.9}$$

equation (2.7) can be rewritten as

$$\boldsymbol{E} = -\frac{d\boldsymbol{A}}{dt} - \boldsymbol{\nabla}V. \tag{2.10}$$

The relationship presented in equation (2.10) can be substituted into equation (2.5) to provide the expression,

$$J = -\sigma \frac{dA}{dt} - \sigma \, \nabla V, \tag{2.11}$$

that relates the current density to the magnetic vector potential.

For steady state or quasistatic fields, the optimum gauge for the magnetic vector potential is the Coulomb gauge, as defined below:

$$\nabla \cdot A = 0. \tag{2.12}$$

Using the above relationships in equation (2.4), the full wave equation is obtained as

$$\nabla^2 A = \mu\sigma \frac{dA}{dt} + \mu\varepsilon \frac{d^2 A}{dt^2}. \tag{2.13}$$

Using the identity

$$\nabla \times (\nabla \times B) = \nabla (\nabla \cdot B) - \nabla^2 B \tag{2.14}$$

and Maxwell's Equations I to IV, the modified wave equation for the magnetic field **B** is as shown below:

$$\nabla^2 B = \mu\sigma \frac{dB}{dt} + \mu\varepsilon \frac{d^2 B}{dt^2}. \tag{2.15}$$

Similarly, the modified wave equation for the electric field **E** is as follows:

$$\nabla^2 E = \mu\sigma \frac{dE}{dt} + \mu\varepsilon \frac{d^2 E}{dt^2}. \tag{2.16}$$

Eddy current diffusion in conducting metallic materials operates at low frequencies (typically below 12MHz), and with good conductors the conductivity σ, is very large. Correspondingly, the second derivative in equations (2.15) and (2.16) may be neglected relative to the first due to the large conductivity value and the magnetoquasistatic approximation [40], which is discussed in more detail in Section 2.3. This term arises because of displacement currents but the approximation considers the system to be in equilibrium, as the currents are changing at a sufficiently slow rate. Hence, the parabolic diffusion equation for **B** may be expressed as follows:

$$\nabla^2 \boldsymbol{B} = \mu\sigma \frac{d\boldsymbol{B}}{dt}.$$ (2.17)

Similarly, for the electric field **E** this can be written as

$$\nabla^2 \boldsymbol{E} = \mu\sigma \frac{d\boldsymbol{E}}{dt}.$$ (2.18)

2.2.1 Magnetic Permeability

Another constitutive relation considers the magnetic permeability μ in the association between the auxiliary magnetic field **H** and the total magnetic flux **B**. In ferromagnetic materials, the time varying field generates an additional magnetization as seen below:

$$\boldsymbol{B} = \mu_0(\boldsymbol{H} + \boldsymbol{M}).$$ (2.19)

The total magnetic flux **B** will consist not only of the external magnetic field **H**, but also accounts for the large internal magnetization field **M** [40]. Hence the expression reveals that the total magnetic flux density is proportional to the sum of the applied field and the magnetization field, where μ_0 is the magnetic permeability of free space.

The initial permeability at low flux densities is much less than the permeability at higher flux densities [41]. The behaviour of ferromagnetic materials at low fields can be described by the Rayleigh law, such that the initial permeability is shown to be linear when the applied field is sufficiently small [42].

2.2.2 Quasistationary Magnetic Fields

Maxwell's equation and the full wave equation (2.15) have been simplified, as a consequence of the time rates of change of the displacement currents being sufficiently slow, hence the system is considered to be in equilibrium. The quasistationary approximation requires that the magnetic diffusion time and the charge relaxation time have the following relation:

$$\tau_m \gg \tau_e.$$ (2.20)

The magnetic diffusion time τ_m, is defined as the characteristic time required by an electromagnetic field to diffuse into a conductor. The charge relaxation time τ_e, is

defined as the characteristic time after which a variation of the electric charge density settles to steady-state conditions [38]. In applications with electric conductors, the effects of magnetic fields predominate over those of electrostatic fields and the relationship presented in equation (2.20) holds. Correspondingly, the whole system is subject to the same quasistationary field, and fields generated by current densities can be assumed to propagate instantaneously [43].

Ohm's law, as presented in equation (2.5) is only considered valid for time scales much larger than the average time between collisions of the free electrons in the conductor [44]. For the quasistationary approximation to be valid in a conductor, the magnetic diffusion time must be much longer than τ_r, the relaxation time or mean free time between electronic collisions. Copper, typically has a relaxation time of 2.4 x 10^{-14} s, whereas typical quasistationary magnetic dominated phenomena have a magnetic diffusion time ≥ 10^{-6} s, so the required relationship indicated below holds [45];

$$\tau_m \gg \tau_r.$$
(2.21)

2.3 Eddy Current Theory

Eddy current inspection is a non-destructive testing technique that is based on the concept of electromagnetic induction, hence it is used in inspection applications that involve electrically conductive ferromagnetic and non-ferromagnetic metals. Examination of time harmonic excitation by magnetic fields can provide insights for the interpretation of pulsed eddy current signal response, with which this thesis is concerned.

Electromagnetic induction, the basis for eddy current inspection, is the process of generating a voltage across a conductor whereby a changing magnetic field induces an electric field [38]. In conducting paramagnetic materials such as aluminum, the time-varying magnetic field causes circulating currents as per Faraday's Law. These circulating currents (Figure 3) generate a magnetic field that opposes the primary field, which changes the voltage or current in the coil.
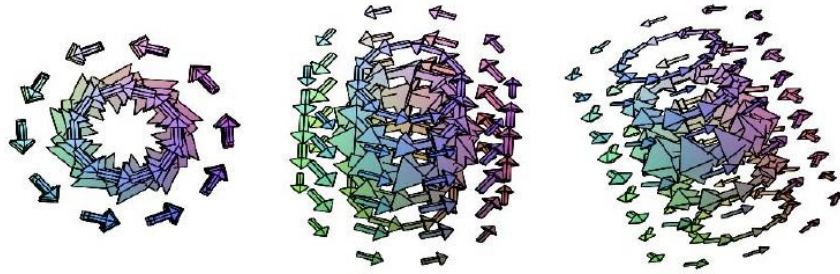
Figure 3: Circulating currents generate a magnetic field, as represented in multidimensional space [40].

Maxwell's Equation III as shown in equation (2.3) and shown below applies to a coil of wire;

$$\nabla \times \boldsymbol{E} = -\frac{d\boldsymbol{B}}{dt}.$$

Taking the integral over the area of the coil of equation (2.3), the left-hand side (LHS) of the equation corresponds to a line integral around a coil loop, by Stokes' theorem, and the right-hand side (RHS) of the equation corresponds to the integral of $\frac{d\boldsymbol{B}}{dt}$ over an area enclosed by the coil loop. Hence, equation (2.3) can be written as

$$\oint \boldsymbol{E} \cdot d\boldsymbol{l} = -\frac{\partial}{\partial t} \int_{\substack{coil \\ area}} \boldsymbol{B} \cdot d\boldsymbol{S}. \qquad (2.22)$$

The LHS of equation (2.22) represents the voltage across the coil and the RHS is the total magnetic flux through one turn of the coil. Hence Faraday's Law can be written for a coil with N turns as

$$V = -N \frac{\partial \boldsymbol{\Phi}}{\partial t}, \qquad (2.23)$$

where V, is the voltage induced in the circuit by the changing magnetic field, N is the number of turns of the coil, and $\boldsymbol{\Phi}$ is the magnetic flux. Lenz's Law states that when a voltage is induced by the change in magnetic flux, the polarity of this voltage produces a

current whose magnetic field opposes the change that produced it. Hence, this sign change is reflected in equation (2.23).

In conventional non-destructive testing applications, eddy currents are induced in a conductor by generating a time-harmonic alternating current (AC) through a coil that is in close proximity to the conductor, as shown in Figure 4. When the coil is placed near the conducting material, the eddy currents that are induced in the material via Faraday's law are similar in form but with opposing direction to the AC current, in accordance with Lenz's law. These eddy currents that are generated by the electromagnetic coils in the test probe are monitored by measuring the probe's electrical impedance.
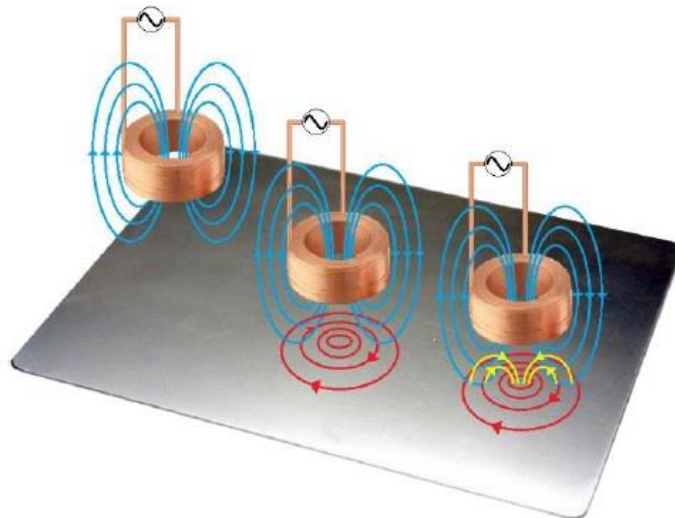
Figure 4 Eddy currents induced in a conductor [38].

Figure 4 shows the effect that is created when a time-varying magnetic field (blue) induces eddy currents (red) in a nearby conducting structure, and the opposing magnetic field (yellow) that is generated [40]. The eddy currents produce a secondary magnetic field, which opposes and interacts with the primary magnetic field and correspondingly, reduces the net flux change. The induced eddy current paths are modified in the presence of discontinuities, which permits flaw detection within the material. The interaction of the two magnetic fields can also provide information on the size and location of the discontinuities within the material.

The applications of eddy current testing are far-reaching and go well beyond flaw detection. Eddy current testing can be used in determining measurements associated with metal thickness, conductivity, and nonconductive coatings [46]. Measurements such as these can be used in numerous applications, including the identification of corrosion within aircraft skin and thinning of pipeline walls. Eddy current testing does have limitations that restrict its application to metallic conductors. Additionally, the sensitivity of conventional eddy current testing systems depends on associated parameters such as depth and volume of the discontinuity. Most notably, the reliability of eddy current testing is reduced by the perturbing effects created by magnetic materials [47]. One such perturbing effect is feedback, which is associated with the electromagnetic coupling of the sample with the exciting coil. The magnetic field that induces the eddy currents contributes to feedback, or 'back-emf' within the coil. Consequently, the applied field is then modified. The feedback effect is enhanced in the presence of ferromagnetic materials and in fact the magnetic component dominates [48].

2.3.1 Skin Depth

The eddy currents induced due to the changing magnetic field tend to concentrate near the surface of the test material, adjacent to the excitation coil. This phenomenon is known as the skin effect. The magnetic field produced by the eddy currents opposes the primary magnetic field according to Lenz's law, and consequently, there is a net reduction in the magnetic flux and a corresponding decrease in current flow and field as the depth increases [49]. This weakens the effects of the coil's field at greater depths into the test material.

The characteristic depth, δ, is known as the skin depth. An expression for skin depth can be derived from the wave equation for the electric field **E** shown in equation (2.16), assuming an infinite plane sheet as follows [50]:

$$\boldsymbol{\nabla}^2 \boldsymbol{E} - \mu\sigma \frac{\partial \boldsymbol{E}}{\partial t} - \varepsilon\mu \frac{\partial^2 \boldsymbol{E}}{\partial t^2} = 0, \qquad (2.24)$$

where μ is the permeability, σ is the conductivity, and ε is the permittivity. Equation (2.24) consists of both a conduction current term and a displacement current term, as

the first and second derivative terms, respectively. When the test material being considered is a good conductor, at frequencies below 10^9 Hz [26], the conduction current dominates over the displacement current. Correspondingly, the displacement current can be neglected, and the diffusion equation expressed in equation (2.18) is represented as shown below [49]:

$$\nabla^2 \boldsymbol{E} - \mu\sigma \frac{\partial \boldsymbol{E}}{\partial t} = 0. \tag{2.25}$$

Considering the current density $\boldsymbol{J}$ and Ohm's Law given in equation (2.5) that relates current density to the electric field, equation (2.25) can be written as:

$$\nabla^2 \boldsymbol{J} - \mu\sigma \frac{\partial \boldsymbol{J}}{\partial t} = 0. \tag{2.26}$$

The magnitude of the eddy current density decreases exponentially with distance below the metal surface. The current density $\boldsymbol{J}$ can be expressed in the frequency domain as follows:

$$\boldsymbol{J}(t) = \boldsymbol{J_0} e^{i(kz - \omega t)}, \tag{2.27}$$

where ω is the angular frequency, t is the time, z is the depth in the conductor and k is the wave number. Differentiating equation (2.27) yields:

$$\frac{\partial \boldsymbol{J}(t)}{\partial t} = -i\omega \boldsymbol{J_0} e^{i(kz - \omega t)}. \tag{2.28}$$

Since

$$\nabla^2 \boldsymbol{J}(t) = \frac{\partial^2 \boldsymbol{J}(t)}{\partial z^2} = -\boldsymbol{J_0} k^2 e^{i(kz - \omega t)}, \tag{2.29}$$

equations (2.28) and (2.29) can be substituted into equation (2.26) as follows:

$$k^2 \boldsymbol{J_0} e^{i(kz - \omega t)} = \mu\sigma i\omega \boldsymbol{J_0} e^{i(kz - \omega t)}. \tag{2.30}$$

Subsequently, this can be simplified as shown below:

$$k^2 = \mu\sigma i\omega. \tag{2.31}$$

An equivalent representation of *i* is determined as follows:

$$i^2 = -1 = \left(\frac{1}{2} + \frac{1}{2}i\right)^2 \tag{2.32}$$

$$i = \frac{1}{2} + \frac{1}{2}i \tag{2.33}$$

$$\sqrt{i} = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i \tag{2.34}$$

$$\sqrt{i} = \frac{\sqrt{2}}{2}(1 + i). \tag{2.35}$$

Substituting equation (2.35) into equation (2.31), the result is:

$$k = \frac{\sqrt{2}}{2}(1 + i)\sqrt{\mu\sigma\omega} \tag{2.36}$$

$$k = (1 + i)\sqrt{\frac{\omega\mu\sigma}{2}}. \tag{2.37}$$

The square root has the dimensions of an inverse length that is characteristic of the medium and the frequency, such that

$$k = \frac{1 + i}{\delta}. \tag{2.38}$$

The real part of the reciprocal of the wave number *k* is the value of the skin depth, *δ*. Hence,

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}}. \tag{2.39}$$

Hence, the skin depth is the quantity used to select the frequency of the applied magnetic field for detection of defects at a certain depth in a test material with known permeability and conductivity.

The conductivity σ, angular frequency ω, and permeability μ, can be represented as follows:

$$\sigma = \frac{1}{\rho} \tag{2.40}$$

$$\omega = 2\pi f \tag{2.41}$$

$$\mu = \mu_r \mu_0, \tag{2.42}$$

where ρ is the resistivity in $\Omega \cdot$ m (Ohm-meters) and f is the frequency in Hz, and the permeability of free space is

$$\mu_0 = 4\pi \; x \; 10^{-7} \; \frac{N}{A^2} \; . \tag{2.43}$$

Substituting equations (2.40) through (2.43) into equation (2.39) provides

$$\delta = \sqrt{\frac{2\rho \; (1 \; x \; 10^{-8})}{(2\pi f)\mu_r(4\pi \; x \; 10^{-7})}}, \tag{2.44}$$

and the skin depth can be rewritten as follows [49]:

$$\delta = \; 50 \sqrt{\frac{\rho}{\mu_r f}}, \tag{2.45}$$

where the resistivity is in in $\mu\Omega \cdot$cm (micro-Ohm-centimeters), the frequency is in Hz, and the resulting skin depth is represented in mm.

The current distribution beneath the surface of the test material also varies with conductivity of the material and angular frequency of excitation, and hence, skin depth. To represent the current density with respect to skin depth, equation (2.38) is substituted into equation (2.27) as shown below:

$$\boldsymbol{J}(t) = \boldsymbol{J_0} e^{\frac{-z}{\delta}} e^{i\left(\frac{z}{\delta} - \omega t\right)}. \tag{2.46}$$

The real and imaginary parts of equation (2.46) can be represented as follows:

$$\frac{\boldsymbol{J_z}}{\boldsymbol{J_0}} \propto \; e^{\frac{-z}{\delta}}, \tag{2.47}$$

and

$$\frac{J_z}{J_0} \propto e^{i\left(\frac{z}{\delta}-\omega t\right)}. \tag{2.48}$$

The relation provided in equation (2.47) demonstrates that as the depth increases, the current density decreases exponentially. The relation provided in equation (2.48) reveals that as the depth increases, there is an increasing time or phase lag of the sinusoidal signal. Therefore, the skin depth for a specific material and frequency, corresponds to the dept h in the material at which the current density has decreased to 1/e, or 37% of the surface density [50]. This is referred to as the standard depth of penetration, *1δ*, and as the eddy currents penetrate deeper than one standard depth of penetration, eddy current strength decreases rapidly with depth (Figure 5). At 3δ, eddy current strength has decreased by 95% from that induced at the surface of the sample.
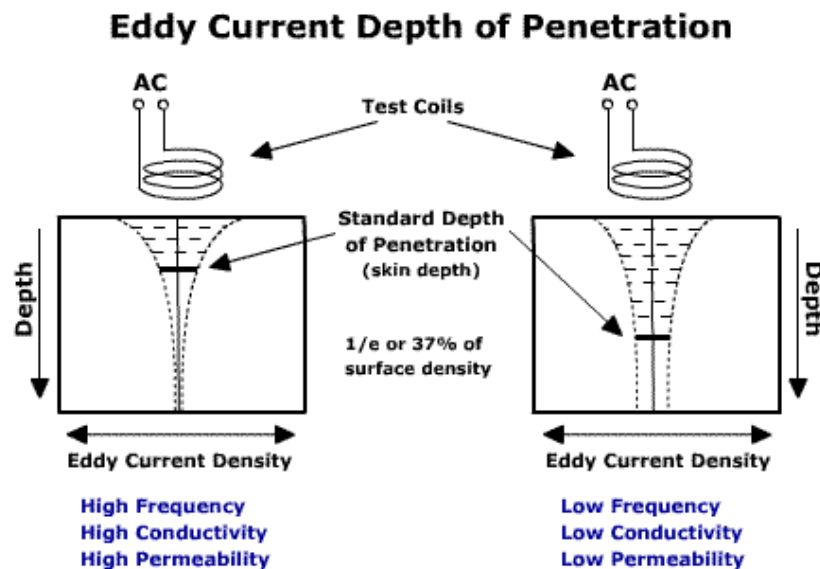
Figure 5: Eddy current depth of penetration [50].

In summary, the depth at which eddy currents penetrate the test material is dependent on the frequency of the excitation current and the electrical conductivity and magnetic permeability of the specimen. As shown in Figure 5 and equation (2.39), the depth of

penetration decreases with increasing frequency and increasing conductivity and magnetic permeability.

2.3.2 Flaw Detection

The preferred method for most aircraft inspection applications, including fatigue crack detection in both engine and airframe components, identification of hidden corrosion in lap-joint structures, and thickness measurements are single frequency eddy current techniques [3]. It has been determined that in crack detection, generally surface-open cracks that are 1 mm or larger in size can be detected [3].

In conventional eddy current applications, the presence of a flaw, crack or material discontinuity causes a change in the net secondary field experienced by the test coil. Hence, there will be a change in the system impedance, which can be detected and measured, and is interpreted as flaw detection. Some important parameters that affect the response signal consist of instrument frequency, gain and threshold, as well as the probe's proximity to the test piece and material discontinuities [3].

The size of the eddy current signal that is visible on the impedance plane display will depend on many factors. Some of these factors include the ratio of the material resistivity to the flaw resistivity, frequency and strength of the applied field and the volume of the flaw experiencing the applied field. Another factor involves the position of the flaw relative to the metal surface, which is in contact with the coil. It is important to note that both the magnitude and phase of the eddy current reading are affected by depth. As the driving frequency increases, the ability to penetrate into the material decreases, due to eddy current concentration close to the surface. This causes larger response signals at the pick-up coil and higher amplitudes and larger phase angles for surface-breaking flaws. As conductivity of the test specimen increases, the real component of the signal that is related to reactance decreases, while the imaginary component remains almost unchanged [3]. This results in a lower amplitude and clockwise phase change [3]. When probe lift-off increases slightly, the signal phase remains almost unchanged while the signal amplitude decreases [3].

2.4 PEC Theory

The effects of skin depth and the presence of ferromagnetic fasteners in lap joint structures severely limits the effectiveness of second layer crack detection by conventional eddy current, without the additional step of fastener removal. In the case of pulsed eddy current inspection, a larger applied field, lower frequency excitation components, long-time transient response and multi-variate signal analysis capability provides PEC greater potential for inspection of thick lap-joint wing structures without fastener removal. This section outlines some of the theory of pulsed eddy current.

2.4.1 Equivalent Circuit Analysis

The mathematical theory highlighted below enables solutions to, not only time-dependent systems but also systems of increasing complexity, in the frequency domain to be obtained. Mathematical theory, in combination with foundational electrical theory, provides the tools necessary to create accurate analytical models for complicated electromagnetic systems.

Kirchhoff's voltage law (KVL) is a fundamental law of electrical theory, which utilizes the powerful concept of conservation of electrical energy. This law states that the sum of the electrical potential differences, or voltages, including those associated with emfs and resistive components, around any closed loop in a circuit is zero [51]. This concept can be expressed in mathematical form as shown in equation (2.49), where N represents the number of voltage sources in the circuit:

$$\sum_{k=1}^{N} V_k = 0. \tag{2.49}$$

Another result of the conservation of electrical energy from Kirchhoff's voltage law is that the voltage between any two nodes in the circuit are the same, no matter which path is taken.

A circuit that includes a resistor and inductor is referred to as a R-L circuit. A representation of a typical R-L circuit is shown in Figure 6 below.
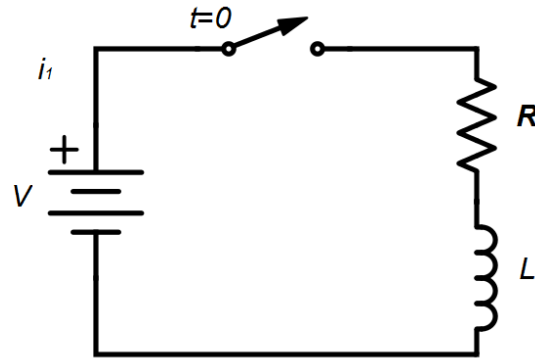
Figure 6: R-L circuit representative of a typical PEC setup for square wave excitation [52].

The equation for the source voltage in a R-L series circuit can be written using Kirchhoff's voltage law, as

$$v(t) = v_R(t) + v_L(t). \tag{2.50}$$

This equation reveals that the sum of the voltages across each of the circuit components is equivalent to the source voltage that is driving the electrical current through the circuit. In pulsed eddy current systems, a square pulse (step function) excitation is used in the driver circuit. The individual voltages across each of the circuit components can be represented with the constituent resistance, inductance, and current values as shown in equation (2.51). The circuit in Figure 6 can be solved using KVL and equation (2.49), resulting in the following equation:

$$V(t) - Ri(t) + L\frac{d}{dt}i(t) = 0. \tag{2.51}$$

As indicated in Figure 6, a boundary condition can be applied to equation (2.51), such that at $t = 0$, $V = 0$, which results in the following solution:

$$i(t) = \frac{V}{R}\left(1 - e^{-\frac{R}{L}t}\right). \tag{2.52}$$

At the instantaneous moment that the switch is closed in the circuit, the current rises to a steady-state condition, as shown in Figure 7, which is a graphical representation of equation (2.52) with τ as the relaxation time.
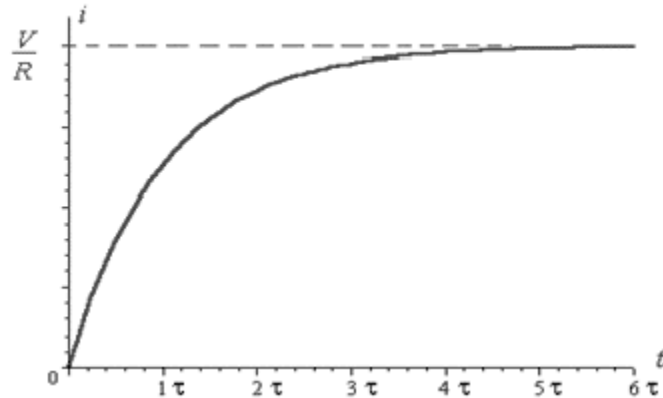
Figure 7: Transient current in driver circuit following application of square wave excitation.

The steady-state response of the system can then be determined by taking the limit as t approaches infinity,

$$\lim_{t \to \infty} i(t) = \frac{V}{R}\left(1 - e^{-\frac{t}{\tau_r}}\right) = \frac{V}{R}, \tag{2.53}$$

with the relaxation time defined as:

$$\tau_r = \frac{L}{R}. \tag{2.54}$$

The PEC probe used in this work consists of a central driving coil with an array of eight surrounding pick-up coils that are differentially paired and symmetrically oriented 180° from each other, positioned around the central driving coil. Correspondingly, the equivalent circuit for the PEC testing evaluated in this work includes a second loop, which represents the pick-up coil response to the emf generated by the driving coil (Figure 8). In this figure, variables with subscript '1' represent parameters associated with the driving coil and variables with subscript '2' are associated with the pick-up coil, with **M** representing the mutual inductance between the driving and pick-up coils.
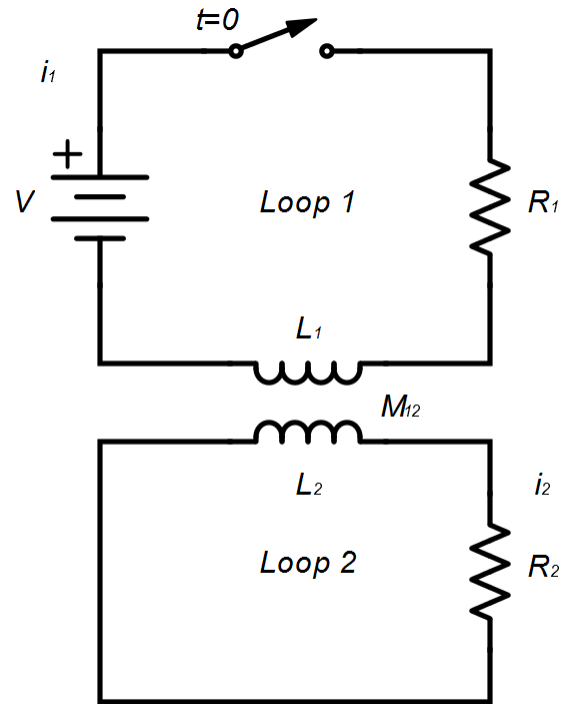
Figure 8: Equivalent circuit diagram for PEC driving coil and single pick-up coil [53].

The mutual inductance between the driving coil and pick-up coil is such that,

$$M_{12} = M_{21}. \tag{2.55}$$

The mutually induced emfs, $\mathcal{E}_1$ and $\mathcal{E}_2$, are then represented by the following equations [51]:

$$\mathcal{E}_1 = -M_{12}\frac{di_2}{dt} \quad and \quad \mathcal{E}_2 = -M_{21}\frac{di_1}{dt}. \tag{2.56}$$

Applying KVL to each closed loop of the equivalent PEC circuit in Figure 8 produces the following governing equations, where $U(t)$ is the step function [53]:

$$L_1\frac{di_1}{dt} + R_1 i_1 = M_{12}\frac{di_2}{dt} + V_o U(t), \tag{2.57}$$

and

$$L_2 \frac{di_2}{dt} + R_2 i_2 = M_{12} \frac{di_1}{dt}.$$ (2.58)

The Laplace transform of equations (2.57) and (2.58) are then taken, and the resulting equations are rearranged, to solve for the current in the pick-up coil, i_2. This results in the following expression [53]:

$$i_2(t) = \frac{M_{12} V_o (e^{-\alpha_2 t} - e^{-\alpha_1 t})}{(\alpha_1 - \alpha_2)(L_1 L_2 - M_{12}{}^2)},$$ (2.59)

where α_1 and α_2 represent inverse relaxation times as follows:

$$\alpha_1 = {}^1\!/_{\tau_1} \quad and \quad \alpha_2 = {}^1\!/_{\tau_2}.$$ (2.60)

Solving for α_1 and α_2 yields

$$\alpha_1, \alpha_2 = \frac{(L_1 R_2 + L_2 R_1) \pm \sqrt{(L_1 R_2 + L_2 R_1)^2 - 4 R_1 R_2 (L_1 L_2 - M_{12}{}^2)}}{2(L_1 L_2 - M_{12}{}^2)}.$$ (2.61)

The transient current flowing in the pickup circuit, defined in equation (2.59) is displayed graphically in Figure 9.
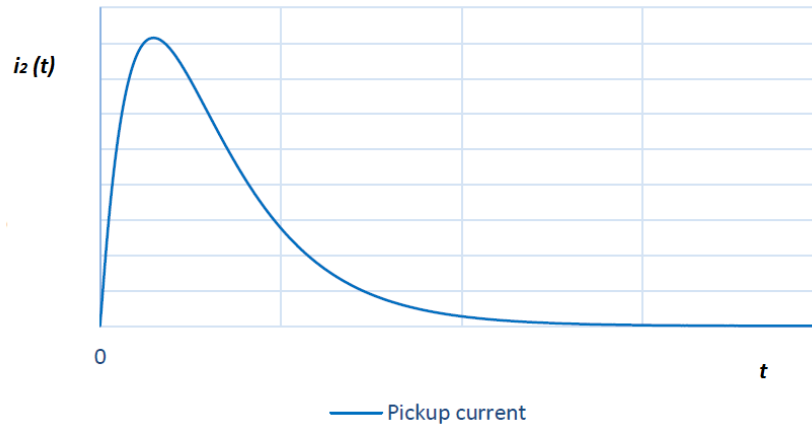
Figure 9: Transient current response to square wave excitation.

The experimental pulsed eddy current technique used in this work includes not only a PEC probe with a driving coil and pickup coils, but also includes the sample being inspected. Consequently, the addition of the conductive sample would result in an additional mutual inductance relationship, as shown in Figure 10. *Desjardins et al* [54] have determined the analytic solutions that account for simple rod geometries and the corresponding electromagnetic interactions and mutual inductance relationships shown below.
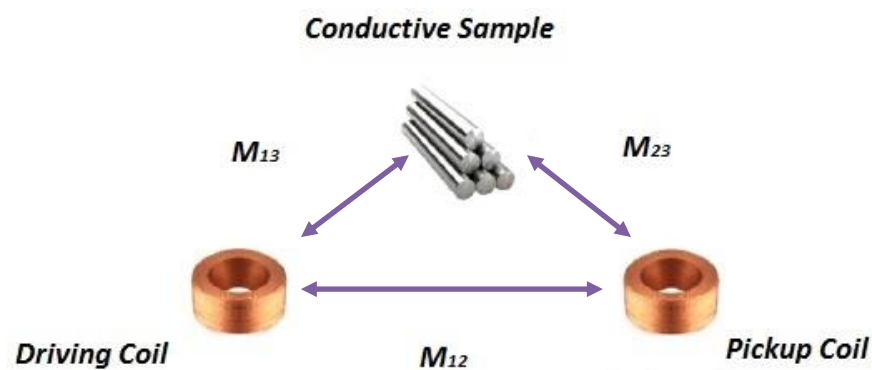
Figure 10: Graphical representation displaying mutual inductance relationship between a PEC driving coil, a single pick-up coil and the conductive sample.

2.5 Analysis of PEC Signals

Introduction of additional parameters, beyond that of the basic bore hole, including the presence of ferrous fastener and second layer conductor, preclude a simple analytical formulation of PEC phenomena that would facilitate a direct mathematical analysis of the transient signal response. Therefore, multivariate statistical analysis techniques have been developed for PEC signal analysis, as described in this section. In validating the PEC inspection technique for detecting second layer cracks at ferrous fasteners in aircraft lap-joint structures, it is necessary to analyze the PEC transient response, to identify and quantify the variation caused by the presence of flaws.

Section 2.5 outlines the theory associated with modified Principal Components Analysis (MPCA), cluster analysis and robust statistics methods, as well as covariance matrices. The statistical methods presented in this section provide the tools necessary to analyze PEC signals and ultimately provide a means for effective flaw detection.

2.5.1 Principal Components Analysis (PCA)

Principal Components Analysis (PCA) is a statistical signal analysis technique used in applications where processing results in vast arrays of data that do not intuitively reveal significant variation between the information presented in the data sets [55]. The strength of PCA is that it provides a mean of re-orienting high-dimensional multivariate data sets into a reduced number of dimensions that are not only manageable to analyze but contain almost all of the original information [18]. The PEC signal consists of several hundred data points with only very subtle differences arising between signals, resulting in high-dimensional data that is difficult to analyze. PCA reduces the data to a few numbers (scores) and provides a means of measuring those subtle differences.

Assuming the data is highly correlated, the utilization of PCA methods can isolate signal variation associated with the one desirable variable that is to be assessed. In this application, the one independent variable that causes the signal variation is the presence of a notch or flaw in the sample. Signal variation can also be caused by additional factors, due to the inherent sensitivity of PEC. Some of these factors that are observed in the samples in this work consist of skin thickness variation, non-uniformity in fastener spacing and distance to lap-joint edge. The largest uncontrolled variable is the placement of the probe over the fastener head. Transforming the transient response signals to a space containing single points of the relevant principal components, representing signal variation caused by the presence of defects, allows for the effects of probe misalignment and other irregularities to be removed. Minimizing and removing these artefacts results in an effective method of flaw detection

2.5.2 Modified Principal Components Analysis (*MPCA*)

The Modified Principal Components Analysis method finds a linear combination of basis vectors that represents the original data set optimally in a least squares sense. This method reduces PEC signals to a series of eigenvectors and scores, which represent the maximum variance between measured signals [18] [19]. Unlike conventional PCA, the MPCA method does not subtract the average response [19]. The assumption is made that there is a set of *p* signal measurements, denoted by *S*, with each measurement having *n* data points. Hence *S* forms a matrix with *n* rows and *p* columns, with column vectors *S_i*, each of which represents a single measurement. A column vector *v*, that does the best job in the least squares sense of representing *S*, is consequently the desired solution. This can be written as follows [19]:

$$T_j = s_j \mathbf{v}, \tag{2.62}$$

where *T_j* is the column vector resulting from scaling *v*, and s_j is the principal component score. For the matrix *T*, to best describe *S*, *v* is chosen to minimize the sum of squared residuals (SSR), such that,

$$SSR = \sum_{i=1}^{n}\sum_{j=1}^{p}\left(S_{ij} - s_j v_i\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{p}\left(S_{ij}\right)^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{p}S_{ij}s_j v_i + \sum_{i=1}^{n}\sum_{j=1}^{p}s_j^2 v_i^2, \tag{2.63}$$

where v_i is the i_{th} element of **v**.

The coefficient s_j in equation (2.62) is the principal component score of *S_j*, and as represented below is the scalar product of the transpose of *S_j and v:*

$$s_j = \sum_{k=1}^{n} S_{kj} v_k, \tag{2.64}$$

where S_{kj} is the kth element of *S_j* and v_k is the kth element of *v*. Equation (2.64) can be expressed in matrix notation as shown below:

$$s = S^T \mathbf{v}. \tag{2.65}$$

To ensure a unique solution for **v**, it is normalized as shown in the following expression [18]:

$$\sum_{i=1}^{n} v_i v_i = \mathbf{v}^{\mathrm{T}}\mathbf{v} = 1. \tag{2.66}$$

Substituting equation (2.64) into equation (2.63) yields:

$$SSR = \sum_{i=1}^{n}\sum_{j=1}^{p} S_{ij}{}^2 - 2\sum_{j=1}^{p} s_j{}^2 + \sum_{i=1}^{n}\sum_{j=1}^{p} s_j^2 v_i^2. \tag{2.67}$$

Using the normalized condition and Equation (2.66), the expression for the residual sum of squares is as expressed below:

$$SR = \sum_{i=1}^{n}\sum_{j=1}^{p} (S_{ij})^2 - \sum_{j=1}^{p} s_j{}^2. \tag{2.68}$$

The focus of PCA is to create a set of column vectors v, that do the best job at representing the data set in a least squares sense. Hence, there is a requirement for the residual sum of squares to be minimized. This is achieved by maximizing the second term in equation (2.68). In order to minimize the SSR, we must maximize the second term in Equation 2.36. Considering equation (2.65) and assuming the normalization represented in equation (2.66), the second term of equation (2.68) can be expressed as follows:

$$\sum_{j=1}^{p} s_j{}^2 = \mathbf{s}^{\mathrm{T}}\mathbf{s} = \mathbf{v}^{\mathrm{T}}\mathbf{S}\mathbf{S}^{\mathrm{T}}\mathbf{v}. \tag{2.69}$$

Solving this optimization problem can be achieved by using the method of Lagrange multipliers [18] whereby,

$$\Lambda = \mathbf{v}^{\mathrm{T}}\mathbf{S}\mathbf{S}^{\mathrm{T}}\mathbf{v} - \lambda(\mathbf{v}^{\mathrm{T}}\mathbf{v} - 1), \tag{2.70}$$

where λ is the Lagrange multiplier that enables the normalization condition of **v** as shown in equation (2.66). Using the theorem [56],

$$\frac{\partial}{\partial v} = \left(x^T A x \right) = 2Ax \tag{2.71}$$

in matrix notation, and taking the derivative of Λ with respect to **v** yields the following [19]:

$$\frac{\partial \Lambda}{\partial x} = 2SS^T v - 2\lambda v. \tag{2.72}$$

To maximize Λ, equation (2.72) is set to zero, which subsequently results in the standard equation for eigenvectors and eigenvalues:

$$Av - \lambda v = 0 \tag{2.73}$$

where,

$$A = SS^T. \tag{2.74}$$

This reveals that an eigenvector solution of SS^T will provide the optimum solution for representing the column vectors in **S**, compared to any other vector. Hence, the eigenvector with the largest eigenvalue will describe more of **S** than any other vector. The first eigenvector provides the smallest residual sum of squares. Once the first eigenvector has been removed, the second-largest eigenvalue does the best job of reducing the residual left over in S, denoted as S', which is represented as follows [19]:

$$S' = S - vs^T. \tag{2.75}$$

Similarly, each successive eigenvector describes the largest amount of residual error remaining that has not been accounted for by the previous eigenvectors. Hence, the first m eigenvectors will do the best job of representing **S** in a least squares sense when compared to any other possible combination of m basis vectors [19]. The sum of the

product of the scores and the basis vectors can be used to represent the original data, as shown in the expression as follows:

$$S = s_1 V_1 + s_2 V_2 + s_3 V_3 + \cdots \tag{2.76}$$

It has been observed that by using a relatively small number of eigenvectors (3-5), greater than 99% of the original signal can be reproduced with a high level of accuracy [9]. The scores obtained through signal processing can then be used for further processing.

2.5.3 Cluster Analysis Method

Cluster analysis can be defined as organizing objects based on their similarities where similar objects are positioned such that they are grouped together, whereas dissimilar objects are found outside this grouping. With respect to this PEC application, it is expected that the signals from blank fasteners be all approximately the same and grouped together, whereas the signals from fasteners with notches present would reveal differences that would place them outside this grouping. The Mahalanobis Distance, MD, is a cluster analysis distance which quantitatively describes the proximity of a point, y, from the centroid, $\bar{\mathbf{x}}$, of a group of points, while adjusting for covariance in the data. The MD can be represented by the following expression [18]:

$$MD = \sqrt{(\mathbf{y} - \bar{\mathbf{x}})' \Sigma^{-1} (\mathbf{y} - \bar{\mathbf{x}})}, \tag{2.77}$$

where Σ is the $m \, x \, m$ covariance matrix of the data matrix X, that has dimensions $m \, x \, p$. The column vectors of length, m, in the data matrix, represent the individual measurements and $\bar{\mathbf{x}}$ represents the row average of the data matrix that corresponds to the centroid of the measurements.

In PEC analysis, the data matrix X consists of the MPCA scores that are obtained from the signal processing of just the blanks. The scores for the blank fasteners are then used to compute the covariance matrix. This covariance matrix is then used to calculate the MD for all the experimental data, including both notches and blanks. In MPCA space, the MD represents a normalized measure of the notch distance from the center of the blank cluster in standard deviations. The MD has a chi-square distribution with m

degrees of freedom (DOF) equal to the number of variables or scores utilized. A decision threshold is determined based on a specific false call rate that is to be achieved. The MD is then compared to this decision threshold to determine whether the fastener hole has a notch present [9].

The plot in Figure 11 [12] depicts MPCA blank fastener scores s_2 and s_3 for data collected using three P-3 Orion lap-joint samples. In this figure, the blue ellipse represents the threshold limit. Inside the ellipse are the blank fastener measurements, representing data matrix X, from three lap-joint samples (standards 22, 24 and 25). The average can be used to compute the centroid $\bar{x}$. Points that fall outside the ellipse exceed the threshold limit and thus signify that a defect is present. The red line shown in Figure 11 is the MD from the centroid to the point outside of the blue ellipse, representing a fastener site containing a 5.08 mm EDM notch [12].
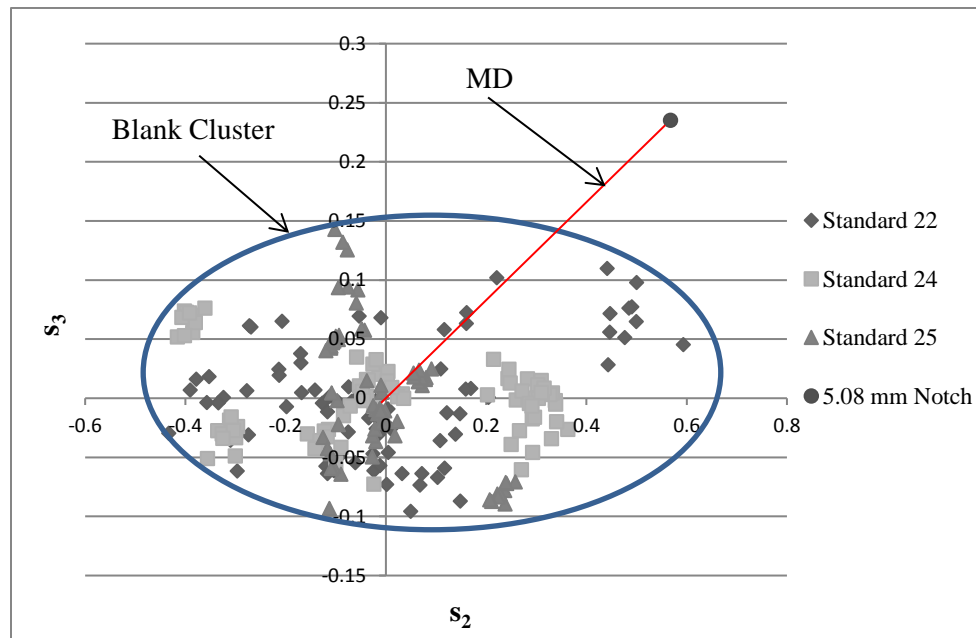
Figure 11: Plot showing blank fastener scores s_3 vs. s_2 for measurements taken from three P-3 Orion samples with blank cluster, MD and a single defect identified [12].

2.5.4 Robust Statistics Method

As indicated in Section 2.5.3, the value of the MD is compared to a decision threshold to determine if there is a flaw detected at the fastener site. Equation (2.77) reveals that

knowledge of the covariance matrix Σ, obtained from the blank fastener scores, and the data centroid, $\bar{\mathbf{x}}$, must be known to compute the MD. Outliers can affect both the centroid and the covariance matrix. Since the covariance matrix is constructed such that it depends on the square distance between each point and the centroid, the impact can be significant.

The smallest half volume (SHV) algorithm presented in [17] is a robust statistical method that provides an alternate method to determine the data centroid and covariance matrix. The data is scaled by columns, representing the scores, such that the median value is subtracted out and then divided by the mean average deviation [57]. This results in the data being scaled, in a manner akin to z scores in conventional statistics, which has the robustness to accommodate outliers. The length, L_{ij}, between each pair of scaled data points, i and j, is then computed as follows [17]:

$$L_{ij} = \sqrt{\sum_{k=1}^{m} \left(x_{ik} - x_{jk} \right)^2}, \tag{2.78}$$

where the summation is over all m dimensions, representing the scores. The vector lengths between observations are calculated, stored in an $m \times m$ distance matrix, and each column is sorted in ascending order. For each column, representing one of the observations, the first $p/2$ smallest distances are summed. This distance matrix provides a method of determining which group of $p/2$ observations are most similar [16]. Instead of $p/2$ observations, Butt *et al.* [12][36][37] used an arbitrary data fraction, h, $(1 \geq h > p/2)$ to select the cluster of data points that were used in determining the tightest group of scaled data. The mean and covariance matrix are then computed and used in the MD calculation to identify outliers.

2.5.5 Randomly Generating Covariance Matrices

To compensate for having a limited number of samples available for testing, and the inflated number of notched fastener sites in these samples, there are two methods to increase the number of blank fastener measurements. One approach utilizes the blank

generation method outlined by Butt [31], which generates random data with the same statistical distribution as the real measured data. The other approach, used in this thesis work, utilizes the bootstrap method [20] which randomly selects blanks from real blank measured data to produce an increased test data set to simulate an increase in the number of blanks in the sample.

2.5.5.1 Blank Generation Method

Given a $p \times m$ data matrix, $\mathbf{S}$, composed of p data measurements each column of which is composed of p numbers with zero mean, the $m \times m$ covariance matrix, $\mathbf{V}$, is as follows [33]:

$$\mathbf{V} = \mathbf{S}^T\mathbf{S}, \tag{2.79}$$

The covariance matrix is a real symmetric positive semi-definite matrix. Any such matrix can be decomposed using a process known as Cholesky decomposition [58], as defined below:

$$\mathbf{V} = \mathbf{C}^T\mathbf{C}, \tag{2.80}$$

such that

$$\mathbf{C} = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,m-1} & c_{1,m} \\ 0 & c_{2,2} & \cdots & c_{2,m-1} & c_{2,m} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & c_{m-1,m-1} & c_{m-1,m} \\ 0 & 0 & \cdots & 0 & c_{m,m} \end{bmatrix}. \tag{2.81}$$

Now consider the $m \times m$ matrix W, where $\mathbf{W} = \mathbf{N}^T\mathbf{N}$ and $\mathbf{N} = \mathbf{PC}$, $\mathbf{P}$ is a $k \times m$ matrix whose columns are normally distributed random numbers with zero mean and unit variance and where k can be any number greater than or equal to m. The expected value of the matrix $\mathbf{W}$ is expressed as follows:

$$E(\mathbf{W}) = E(\mathbf{C}^T\mathbf{P}^T\mathbf{PC}). \tag{2.82}$$

The C matrices are constant, thus the expected value as expressed in equation (2.82) can be rewritten as shown below [33]:

$$E(\mathbf{W}) = \mathbf{C}^T E(\mathbf{P}^T\mathbf{P})\mathbf{C} = \mathbf{C}^T\mathbf{C} = \mathbf{V}, \tag{2.83}$$

since the expectation value of $\mathbf{P}^T\mathbf{P}$ is the unit matrix by construction. Hence $\mathbf{N}$ has the same covariance matrix as $\mathbf{S}$. The only difference between the two matrices is that the column mean of $\mathbf{N}$ is zero, whereas the column mean of $\mathbf{S}$ is non-zero. Hence, one can derive a data matrix with the same distribution and means as $\mathbf{S}$ by adding the column means of $\mathbf{S}$ to the values of $\mathbf{N}$ [33].

2.5.5.2 Bootstrap Method

The bootstrap method involves the random selection of real blank fastener data with repeats, to increase the size of the test data set and consequently produce covariance matrices, which contain statistical properties that are identical to real experimental measurements [20]. The increased test data set produces covariance matrices that simulate real world detection scenarios where there are only a small percentage of flawed sites in the sample being inspected. As with the random generation method, once the covariance matrices are produced, the simulation technique utilizes the cluster analysis and robust statistics methods as outlined above to determine notch detection results.

2.6 POD Theory

Probability of detection studies are conducted to determine the detection capability of a specific NDT system with an associated flaw site location. The detection capability of the NDT system is determined by analyzing the system response to flaws or discontinuities of varying sizes and orientations. Representative specimens are produced and are examined under conditions that are representative of the field environment. The specimen flaws are produced in a manner such that the sizes and locations are known to the individual conducting the study. It is important to ensure that there are an appropriate number of unflawed sites in the specimen set, as this aids in the estimation of the rate of false indications. The specimens are then examined by qualified inspectors, representative of those who would be carrying out the inspection, although the size and location of the discontinuities remain unknown to the inspectors. The measurements and data recorded by the inspectors can be analyzed with statistical

techniques and specialized software to determine the POD and correspondingly, the capability and reliability of the specific NDT system being evaluated. It is important to note that the detection rate assumes that all notches are equally detectable, however POD accounts for differences in detectability, based on flaw size. Therefore, POD is taken as the main metric used to measure the reliability of the system.

The purpose of conducting a POD study of a specific NDT system is to evaluate the reliability and capability of the NDT system with flaws and inspection conditions that can be expected in field applications. Data that is obtained in a laboratory environment is not completely representative of data obtained under in-service conditions. As such it normally represents a best-case scenario, and it is a measure of the capabilities of the equipment. In the field, there are additional factors such as the environment, human factors and access that could influence the measurements and data collection. If actual inspectors are unable to replicate the results, then there is a need to determine if the issue is the inspector or the conditions under which the inspection is taking place. In the former case, more training may be required and in the latter, it may be possible to improve the inspection conditions. It is also important that the discontinuities produced in the specimens resemble flaws that would be encountered in-service. Electrical Discharge Machined (EDM) notches are used to simulate fatigue cracks and are therefore representative of discontinuities that are seen in real world applications. These EDM notches can be machined to specific sizes, shapes and locations with the use of an electrode. In evaluating the capability and reliability of a NDT system, it is important to ensure that the system response for the fabricated EDM notches produces similar results to that of a fatigue crack response.

Variation in system response for the same discontinuity can occur for several reasons. Different responses can be obtained under repeated inspections of the same flaw due to the variation associated with the calibration and setup of a specific NDT system and the judgement of the inspector. Additionally, material properties, flaw geometry and orientation are factors that can affect responses obtained from discontinuities of the same size. The probe coil size, orientation and type can be selected to optimize responses for specific flaw geometries and orientations. Human factors are also an important consideration with respect to uncertainty associated with the inspector's

analysis of the signal response, and this can correspondingly affect the estimate of the probability of detection. The 2nd European-American Workshop on NDE Reliability, 1999 [59] provided a definition of human factors that was more specific to NDT applications. It stated that, "Human factors are the mental and physical make of the individual, the individual's training and experience, and the conditions under which the individual must operate to influence the ability of the NDE system to achieve its intended purpose" [59].

2.6.1 Categories of Data

The data obtained through the inspector examination of the test specimens for the POD study can be grouped into two categories: those with only a binary output (hit/miss) or those that provide information as to the apparent flaw size ($\hat{a}\ vs\ a$).

Data obtained from $\hat{a}\ vs\ a$ systems can then be categorized as either uncensored or censored. Uncensored data is the most common type of data, and represents data where the value of each sample unit is observed or known. When inspecting for discontinuities in test samples the signal response can be outside the instruments' range of sensitivity so that it is not adequately registered, hence this type of censored data is usually encountered in $\hat{a}\ vs\ a$ systems. All $\hat{a}\ vs\ a$ systems have two censoring values, both left censored and right censored, where a right-censored value represents the maximum possible signal, greater than or equal to 100% screen height [8], such that it is too large for the system to handle. Left-censored values are values that are below the noise floor. The presence of either left or right censored data invalidates the ordinary least squares approach. The focus of $\hat{a}\ vs\ a$ plots is on sizing the discontinuity. In this project only hit/miss data will be analyzed, hence a POD versus size, or POD(a) curve will be produced. Hit/miss analysis may be followed by BHEC inspection, as it provides the means to size the defect [60].

2.6.2 POD Function

Analyzing binary results such as hit/miss data using histograms requires a significant quantity of data to improve the resolution in POD and the bounds are much weaker than those obtained with a link function [8]. The link functions are *logit, probit, cloglog, and*

loglog, which map the range into [0,1], the possible range of values for p, the probability of detection [8]. It was assumed that underlying this binary data was a mathematical relationship between POD and size. An estimate is then made of the model's parameters using a maximum likelihood method [8]. This provided the framework for hit/miss POD modeling.

The probability of detection at a discontinuity of size **a** is defined as the average POD of all discontinuities of that size [3]. Hence, the detectability of flaws is dependent on many factors, but most notably the flaw size. An assumption can be made that each flaw has a range of POD values, hence, the mean POD curve can be defined as

$$POD(a) = \int_0^1 p \cdot f_a(p)\, dp\,. \tag{2.49}$$

where *p* is the proportion of flaws detected and f_a is its density function [3]. When sufficient hit/miss and associated flaw size data is available, then a POD/size relationship can be obtained. Several statistical functions have been posited for calculating POD and confidence bounds, with two being the logistic or log-odds model, and the log normal (probit) model.

2.6.2.1 Logistic (Log-Odds) Model

After examining various methods of modeling NDT data, Berens and Hovey [61] concluded that the log-logistic model provided the most consistent distribution for determining a POD curve as a function of defect size. The expression for the logistic distribution is as shown below:

$$P_i = \frac{\exp[\alpha + \beta ln(a_i)]}{1 + \exp[\alpha + \beta ln(a_i)]}, \tag{2.50}$$

where P_i is the probability of detection for the *ith* flaw, a_i is the size of flaw *I*, and *α* and *β* are location and scale parameters respectively that define the curve. Berens and Hovey [62] assessed two approaches for determining the constant location and scale parameters, the Range Interval Method (RIM) and the method of Maximum Likelihood Estimators (MLE).

2.6.2.2 Range Interval Method (RIM)

The range interval method uses regression analysis to estimate the location and scale parameters necessary to define the logistic POD curve shown in equation (2.50). It is assumed that the detection within a small defect size range follows a binomial distribution and the variability of POD within this range is small [3]. For the RIM, the flaw data is divided into t intervals with equal length. The probability of detection is then calculated for each interval, resulting in t data points. This is defined as the ratio of flaws detected to the total number of flaws in the specific interval. Linear regression is then performed on the t data pairs of POD and flaw size to obtain the intercept and scale parameters, α and β of equation (2.50). The linear regression analysis is performed using the following transformations on the logistic distribution function shown in equation (2.50):

$$Y_i = ln\left(\frac{P_i}{1 - P_i}\right)$$

and

$$X_i = \ln(a_i), \tag{2.51}$$

where P_i is the proportion of flaws detected and a_i is the size of the i_{th} defect. The result of the transformations is a set of points that are fit with the line:

$$Y = \alpha + \beta X. \tag{2.52}$$

The parameters α and β are then substituted into equation (2.50) and a POD curve is constructed for a range of flaw sizes. In the analysis, flaws are grouped into size-intervals such that a_i is the average defect size in interval i, and n is the total number of intervals. It is important to note that if the estimated POD value for an interval is 0 or 1, then the transformation is undefined. Hence, for a POD of 0, a value of $1/(n + 1)$ is used, and for a POD value of 1, $n/(n + 1)$ is used. Using the approximation of POD of 0 overestimates the POD at small flaw sizes and underestimates the POD at large flaw sizes where the POD of an interval is 1. The 95% confidence bound, $Y_{95\%}$, can be found in the transformed domain as follows:

$$Y_{95\%} = Y_{mean} - 1.96\, S_{Y/X}, \tag{2.53}$$

where Y_{mean} is the sample mean, 1.96 represents the standard normal variate for 95% of the area for the normal distribution curve, and $S_{Y/X}$ refers to the variance and covariance of the transformed variables.

2.6.2.3 Maximum Likelihood Estimation (MLE)

Another approach used to estimate the α and β parameters in equation (2.50), is the Maximum Likelihood Estimation method. The likelihood L for a single observation is expressed as follows [3]:

$$L(P_i; a_i, x_i) = P_i^{x_i} \cdot (1 - P_i)^{1-x_i}, \tag{2.54}$$

where P_i is the probability of detecting the ith flaw of size a_i and x_i is the inspection outcome with 0 being a miss and 1 being a hit. The overall likelihood of observing a series of independent observations or inspections is the product of the individual inspections. This can be expressed as shown below:

$$L(P; a, x) = \left[\prod_{i=1}^{h} P_i \right] \left[\prod_{j=1}^{n-h} (1 - P_j) \right]. \tag{2.55}$$

where L is the overall likelihood, n is the total number of flaws, h is the number of flaws detected and $n - h$ is the number of flaws missed. To express equation (2.55) as a series of sums instead of a series of products, the logarithm of equation (2.55) is taken as shown below [3]:

$$lnL(P; a, x) = \sum_{i=1}^{h} lnP_i + \sum_{j=1}^{n-h} lnP_j. \tag{2.56}$$

The values of the α and β parameters in equation (2.50) can be chosen to maximize the likelihood in equation (2.56). Because a logarithm is a monotonic function, the maximum of the log likelihood will occur for the same parameter values as the maximum of the likelihood [3]. To determine the maximum likelihood estimate of α and β, denoted as α' and β', the log likelihood with respect to α and β is maximized [3]. Equation (2.56)

is differentiated with respect to α and β from equation (2.50) and the derivatives of the log likelihood are set to zero as shown below:

$$\sum_i^n \frac{\exp[\alpha' + \beta' ln(a_i)]}{1 + \exp[\alpha' + \beta' ln(a_i)]} - \sum_i^n x_i = 0 \qquad (2.57)$$

and

$$\sum_i^n \ln(a_i) \cdot \frac{\exp[\alpha' + \beta' ln(a_i)]}{1 + \exp[\alpha' + \beta' ln(a_i)]} - \sum_i^n x_i \cdot \ln(a_i) = 0, \qquad (2.58)$$

where x_i again is the inspection outcome with 0 being a miss and 1 being a hit, with the sum calculated over a set of flaws from 1 to n.

2.6.2.4 Probit Model

The log normal distribution has been utilized for modeling POD data. The cumulative log normal distribution is expressed as follows [63]:

$$P_i = 1 - Q(z_i), \qquad (2.59)$$

where

$$z_i = \frac{ln(a_i - \mu)}{\sigma}, \qquad (2.60)$$

such that $Q(z)$ is the standard normal survivor function, z_i is the standard normal variant, and μ and σ are the location and scale parameters of the POD curve [63]. Using the maximum likelihood method as discussed previously, equation (2.56) is differentiated with respect to the location and scale parameters μ and σ. The derivatives are set to zero and the resulting equations are solved simultaneously for μ and σ. Cheng and Iles [64] derive a method of determining the confidence bound on the probit POD curve, but this will not be discussed here as it does not use the MLE approach.

Hit/miss data can be analyzed using both the logistic and probit methods in concert with the Range Interval Method (RIM) or Maximum Likelihood Estimation (MLE) approaches. It has been found that the logistic distribution can be readily implemented on large sample sizes and tends to provide conservative POD results [3]. In determining the

estimation of the μ and σ parameters, the RIM approach requires significant dependence on interval size. This approach also requires an approximation of POD of 0, which overestimates the POD at small flaw sizes and underestimates the POD at large flaw sizes where the POD of an interval is 1. In contrast, the MLE approach only requires actual inspection data and does not consider interval size. Hence, the MLE approach is the preferred method to estimate the location and scale parameters. In the analysis of the hit/miss data for this project it was found that the best results were obtained with the **mh1823 POD** [8] software with the log-odds model and this will be discussed in greater detail in Section 2.9.4.

2.6.3 False Positives

A false positive, also known as a false call, occurs when a NDT method produces a system response that is interpreted as detecting a discontinuity when none is present [8]. There is a cost associated with the false positive rate. In some applications, false positives should be reported based on a group of items inspected, whereas in other cases if one false call would retire that component, it is reported by component [8]. The cost of false positives can be lowered if the false positives can be eliminated through allowable rework and repair. However, if false positives are exceedingly costly, the inspection should demonstrate high specificity and indications below the threshold should be penalized in a quantitative manner [8]. In this case, specificity is defined as the probability of a true negative, whereby the presence of no target is correctly identified [8]. Additionally, a high level of false calls can bias a POD curve because there is a probability that a portion of the detected flaws were identified by chance. To ensure an accurate modeling of the true POD, a false call rate of more than 5% should be accounted for [3].

The functions expressed in Section 2.6.2 do not account for the probability of false calls. The Spencer model was developed to account for false calls and miss rates that were independent of flaw size [3]. The Spencer model is expressed as

$$POD(a) = p_h + \left(1 - (p_m + p_h)\right) \cdot F(a; \mu, \sigma), \qquad (2.61)$$

where *POD(a)* represents the probability of detection at the flaw size a, p_h is the probability of a false call, p_m is the probability of missing a defect independent of flaw size, and $F(a; \mu, \sigma)$ is the distribution used to fit the data [3]. The distribution F can be modeled as before with the logistic or probit curve and all the parameters in equation (2.61) can be determined by the maximum likelihood estimation method. The Spencer model reduces sensitivity to outliers and provides reasonable values for small and large flaw sizes that previously provided issues with other methods [3]. When large values are obtained for p_h and p_m, this reveals that there are problems associated with the inspection procedure or inspector competency, which can result in large variability of the signal at the decision threshold, an insufficient number or distribution of flaw sizes, or there is poor correlation with flaw size and the POD [3].

The inspection threshold, $\hat{a}_{th}$ is the smallest value of $\hat{a}$ that the NDT system records, below which the signal is indistinguishable from noise [8]. The decision threshold, $\hat{a}_{dec}$, is the value of $\hat{a}$ above which a signal is interpreted as a hit and below which the signal is interpreted as a miss. The decision threshold is the $\hat{a}$ value associated with 50% POD, and the inspection threshold is always less than or equal to the decision threshold. To determine the decision threshold, analysis of the system noise needs to be conducted [8].

2.6.3.1 Noise Data
To determine the probability of a false positive for a specific NDT system and to ultimately determine the appropriate remedial action for eliminating possible false positives, analysis of the noise data needs to be conducted [8]. As the noise level in the system increases, detection of flaws may become increasingly more difficult. If the noise level is too high, remedial action such as re-inspection, repair and part replacement may be unnecessarily implemented, and an additional cost can be incurred. Acquiring noise data is the first step in the process in eliminating false positives and it can be obtained by evaluating the signal amplitude on an inspection site where no discontinuity is present. Although noise information can be obtained from the $\hat{a} \, vs \, a$ data or the **mh1823 POD** software, the preference is for actual noise measurement data to be obtained from the system signal response, as this is a direct representation of the noise present in the system [8].

2.6.4 Hit/Miss POD Modeling Using mh1823 POD Software

The hit/miss data for this thesis work was analyzed using the **mh1823 POD** software, the theory below provides specifics with respect to the modeling aspects that are used in this software.

2.6.4.1 Generalized Linear Models

In linear models, such as those involving ordinary least-squares regression and censored regression, the response, y, is related to the controlling variables as shown below:

$$y = f(X), \hspace{4cm} (2.62)$$

where X is the matrix of controlling variables. With ordinary linear regression, it is assumed that the response varies continuously and that it is unbounded. Binary or hit/miss data does not adhere to these assumptions as the observed responses are bounded and discrete, where the possible outcomes are only 0 or 1 values. Additionally, since the response of ordinary linear models is continuous, the error between the response and the model must also be continuous. This produces a Gaussian or normal distribution. Conversely, the resulting error between observed outcome and model prediction for binary data is binomial. Hence, assuming a Gaussian distribution for hit/miss data would result in inaccurate and unreliable parameter estimates.

To adequately address these inaccurate and unreliable parameter estimates, a Generalized Linear Model (GLM) is used. The GLM doesn't model the response, but the probability of the response, which can be a continuous function. A GLM links the binary response to the explanatory variables with link functions through the probability of either a 0 or 1 outcome [8]. The result is a transformed probability that can then be modeled as an ordinary polynomial function that is linear in the explanatory variables, and hence is considered a generalized linear model [8]. The variance of the transformed function is not constant. Therefore, iteratively reweighted least-squares regression and a special maximum likelihood method are necessary to estimate the GLM model parameters [8].

2.6.4.2 Link Functions

The algorithm of the **mh1823 POD** software utilizes one of four link functions to map the range (-∞,∞), required for the linear fitting, into [0,1], which covers the possible range of values for p, the probability of detection [8]. The link functions are ***logit, probit, cloglog, and loglog.*** The ***logit***, logistic or log-odds function is represented as follows:

$$f(X) = g(y) = \log\left(\frac{p}{1-p}\right),$$ (2.63)

where p is the probability of flaws being detected. When $p = 0$, $Y = -\infty$; when $p = 0.5$, $Y = 0$, and when $p = 1$, $Y = \infty$.

The ***probit*** or inverse normal function is

$$f(X) = g(y) = \Phi^{-1}(p).$$ (2.64)

The complimentary log-log or ***cloglog*** function is represented by

$$f(X) = g(y) = \log(-\log(1-p)),$$ (2.65)

and the ***loglog*** function is

$$f(X) = g(y) = -\log(-log(p)).$$ (2.66)

In equations (2.63) to (2.66), $f(X)$ represents any appropriate algebraic function that is linear in the parameters, and most often is a polynomial. The standard normal cumulative density function (cdf) is Φ ( ).

The probability of detection is defined as, $p_i = POD(a_i)$, which is a function linked to the *ith* crack, with size a_i. *Since $f(X) = g(y)$, there exists $g^{-1}(f(X))$ where $g^{-1}( )$ is the link.* To obtain $POD(a, ...)$, the **mh1823 POD** software utilizes the following links [8]:

$$\boldsymbol{probit\ link}\ \ POD(a, ...) = 1 - \Phi\left(f(X)\right)$$ (2.67)

$$\boldsymbol{logit\ link}\ \ POD(a, ...) = \frac{\exp\left(f(X)\right)}{1 + \exp\left(f(X)\right)}$$ (2.68)

$$\boldsymbol{cloglog\ link}\ \ POD(a, ...) = 1 - \exp\left(-\exp\left(f(X)\right)\right)$$ (2.69)

$$\textit{cloglog link} \quad POD(a, \dots) = -\exp\left(-\exp\left(-f(X)\right)\right). \tag{2.70}$$

It is important to note that although equation (2.67) has the form of the Gaussian probability density, it is not a distribution of crack sizes but an S-shaped function that describes the relationship between POD and size [8]. In equations (2.67) through (2.70), $POD(a, \dots)$ can be modelled such that

$$POD(a, \dots) = p\ (y = 1|X), \tag{2.71}$$

where $POD(a, \dots)$ determines the probability that a hit (*y = 1)* is observed, given other conditions, *X*. The term $POD(a)$ indicates that the probability of detection is a function of size a. The ellipsis (…) in $POD(a, \dots)$ identifies that the mathematical model that relates target size, a, with the POD, can include other parameters. These other parameters can represent target shape, density, and chemistry, depth of inspection, and system features such as type of probe and inspector [8].

2.6.4.3 The Logit Link

The logit link function has been proven to effectively and correctly model the dependence of p on a or $\log(a)$[8]. The logit link model, dating back to Berens and Hovey [65][35], is most commonly applied for its effectiveness in modelling hit/miss data in POD curves. Since it is used in analyzing samples in this work, additional background information on this model is provided below. As defined in equation (2.63), the log of the odds is the logit function.

The expression for the log-odds POD model is as follows [35]:

$$\log\left(\frac{POD(a)}{1 - POD(a)}\right) = \beta_0 + \beta_1 a, \tag{2.73}$$

or if size is transformed logarithmically,

$$\log\left(\frac{POD(a)}{1 - POD(a)}\right) = \beta_0 + \beta_1 \log(a), \tag{2.74}$$

where β_0 and β_1 are the intercept and slope respectively, and the maximum likelihood values in the Generalized Linear Model. Solving equation (2.74) for POD(a) results in [35]:

$$POD(a) = f(a, \boldsymbol{\theta}) = \frac{e^{\beta_0 + \beta_1 \cdot h(a)}}{1 + e^{\beta_0 + \beta_1 \cdot h(a)}}, \tag{2.75}$$

where $\theta = (\beta_0, \beta_1)^T$. There is no physical interpretation for $(\beta_0, \beta_1)^T$, hence with re-parameterization, equation (2.75) is rewritten as [35]:

$$POD(a) = f(a, \boldsymbol{\theta}) = \Phi_{link}^{-1} \left(\frac{x - \mu}{\sigma} \right), \tag{2.76}$$

where μ is the size or log(size) corresponding to the x value at which $POD = 0.5$ and σ is the inverse GLM regression slope [35]. The logit link can be written as [35]:

$$\Phi_{link}^{-1} \left(\frac{x - \mu}{\sigma} \right) = \log \left(\frac{POD(a)}{1 - POD(a)} \right), \tag{2.77}$$

where

$$\sigma = \frac{1}{\beta_1} \; and \; \mu = -\beta_0 \sigma. \tag{2.78}$$

The typical 'S' shaped POD(a) curve is transformed to a straight line on the logit grid as shown below in Figure 12, where the scale parameter σ shown in the left graph, is transformed to 1/slope in the right graph [35]. The Z scores are measures of standard deviation and are associated with the standard normal distribution.
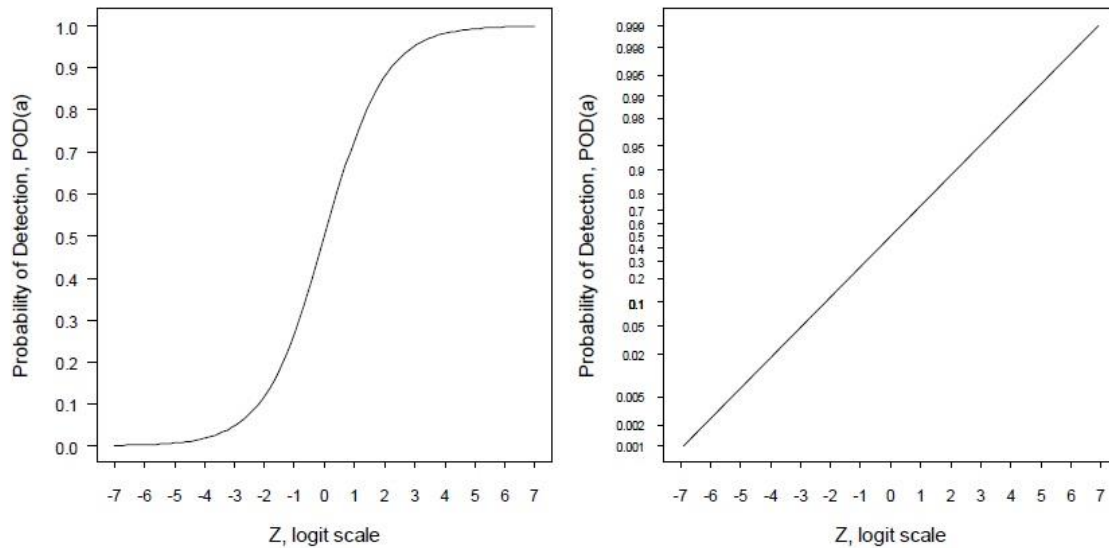
Figure 12: Typical POD(a) curve (left) plotted as straight line on logit grid (right) when scale parameter σ undergoes transformation to inverse of the slope, where z=(x-μ)/σ [35].

2.6.5 Analyzing Hit/Miss Data

To produce a *POD vs size* plot, the first step in the analysis process with the **mh1823 POD** software is to enter the hit/miss data from a single inspection with one inspector using a single probe. Although it has become standard practice to take the logarithm of size in producing POD models, and the **logit** link is usually the best overall model, it is important to assess all the *POD vs a* models available [8]. The **mh1823 POD** software presents eight possible *POD vs. a* models, two each for the four link functions presented in equations (2.67) through (2.70) as shown in Figure 13 [8].
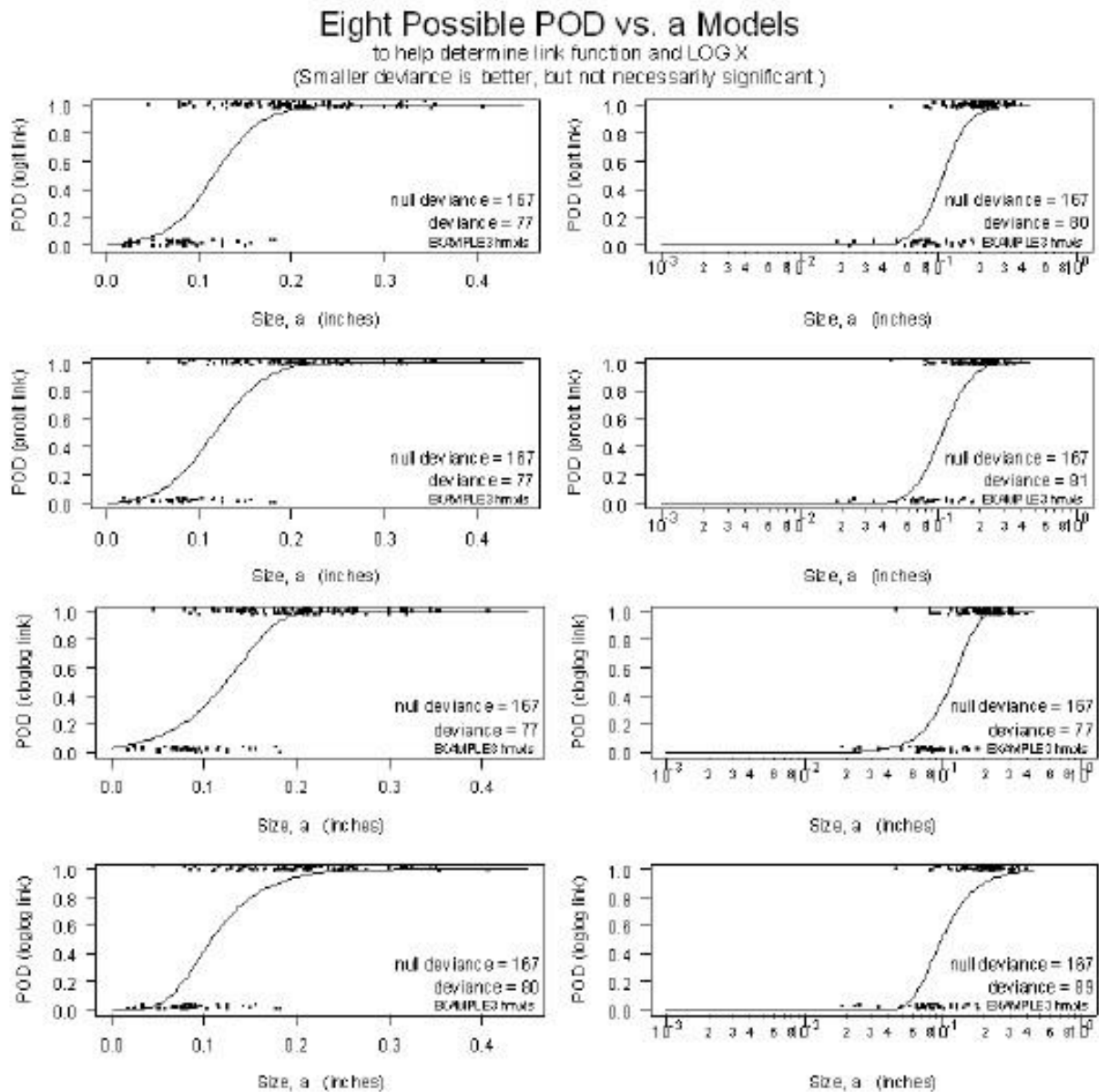
Figure 13: POD vs. a Models [8].

The deviance, computed from the loglikelihood ratio, represents a measure of the overall data scatter [8], and can be thought of as equivalent to the sum square error (SSE) in ordinary least squares regression. Hence, in selecting the POD vs. a model, one that has a smaller deviance is considered a better model as it provides a better fit [35]. The null deviance quantifies the scatter for a POD, where variances are close and

the POD is a constant 0.5 model [8]. Correspondingly, the model deviance reveals the improvement obtained when the model considers the influence of target size on POD [8]. These eight diagnostic POD curves aid in the decision with respect to selecting a link function and whether to use size or log of the size. Once the best model is selected for the data set, a *POD vs. a* plot can be constructed (Figure 14). It is important to note that the **mh1823 POD** software spreads the data points slightly above and below their actual 0.0 or 1.0 POD values, to provide a clear and distinguishable visual representation of the data, as opposed to an overlay of points forming a solid line.
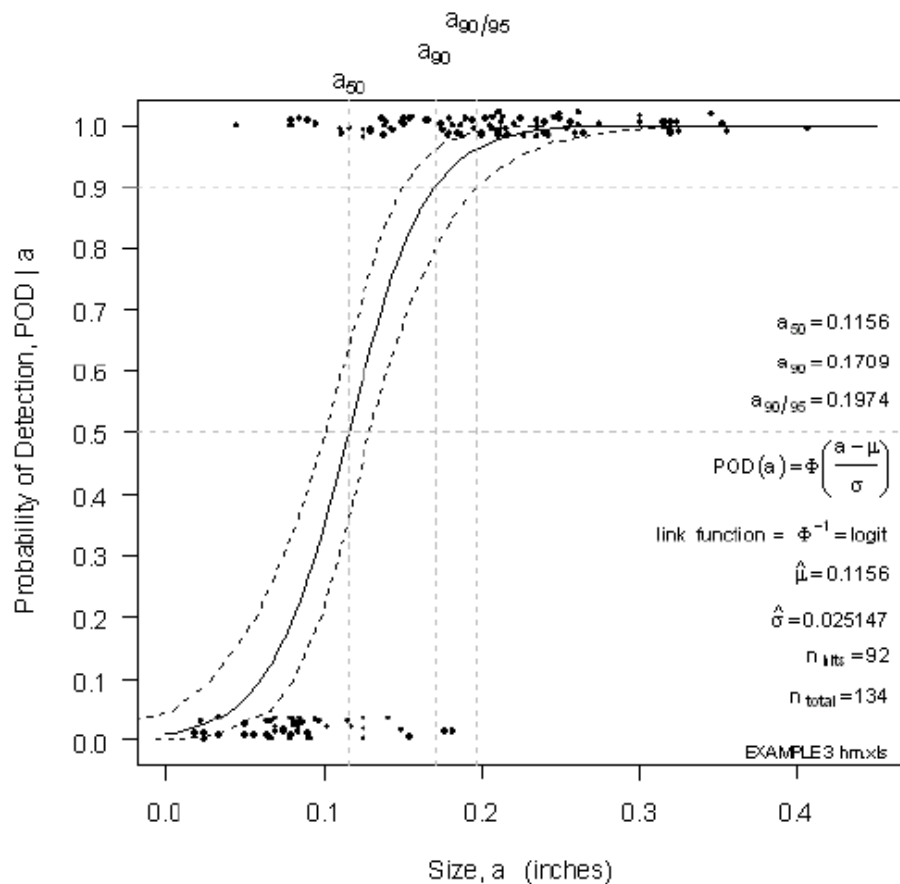
Figure 14: POD vs. size plot, displaying true POD vs size plot (solid line) and confidence bounds (dashed lines) [8].

Several important features of the model selected are shown on the plot, including name of the dataset used, the parameters of the model and the representative crack sizes -

a_{50} (the crack size having 50% POD)*, a_{90}* (the crack size having 90% POD)*, and $a_{90/95}$*
(the 95% confidence bound on the *a_{90}* estimate).

2.6.5.1 Hit/Miss Confidence Bounds
A confidence bound is usually associated with a POD function to indicate that the curve
was constructed using a sample population. With an increase in the sample population
to obtain the POD, the confidence level will also increase. In Figure 14, the significance
of the dashed confidence bounds is such that they enclose the true POD(a) curve (solid
line) in 95 of 100 similar experiments [35].

2.6.5.2 Determining Confidence Bounds
The term likelihood refers to the probability that the model fits the data, given specific
model parameters. To maximize the likelihood that specific POD model parameters
more accurately reflect the inspection outcomes, maximum likelihood parameter
estimates are used [8]. Choosing slightly different values will result in a diminished
likelihood. A means for constructing the likelihood ratio confidence bounds is based on
the Central Limit Theorem [8]. A consequence of this theorem is that the loglikelihood
ratio, Λ, which is the ratio of the logs of the new values to their maximum values, has an
asymptotic chi-square (χ^2) distribution [8]. The log-likelihood ratio has an asymptotic
chi-square distribution, with degrees of freedom equal to the number of parameters in
the model, when sample size is increased [35]. This relationship is detailed in the
following expression [35]:

$$-2log\left(\frac{L(\theta_0)}{L(\theta)}\right) \sim \chi^2_{1-\alpha;df},\tag{2.79}$$

where θ is maximum likelihood model, θ_0 represents other models for comparison, df is
the degrees of freedom (number of model parameters), and $1-\alpha$ is the confidence
probability. As the sample size increases, the log-likelihood ratio approaches χ^2_{df} more
rapidly than using Wald-type bounds which earlier methods used, and it produces
confidence bounds on the parameter estimates that are closer to nominal [35].

The idea is to move the POD(a) model parameters away from their maximum values
only until the criterion is reached [8]. Hence, parameters close to the best estimates are
plausible at describing the data, whereas those that are farther away from the best

estimates are not adequate at describing the data. The loglikelihood ratio determines this relative closeness [8].

The solid line in Figure 14 represents the POD (a) curve, with two model parameters μ and σ determining the line. The variable μ locates the curve horizontally and σ is the inverse of the POD curve's slope. For different values of μ and σ, a plot of the loglikelihood can be constructed, as shown in Figure 15.
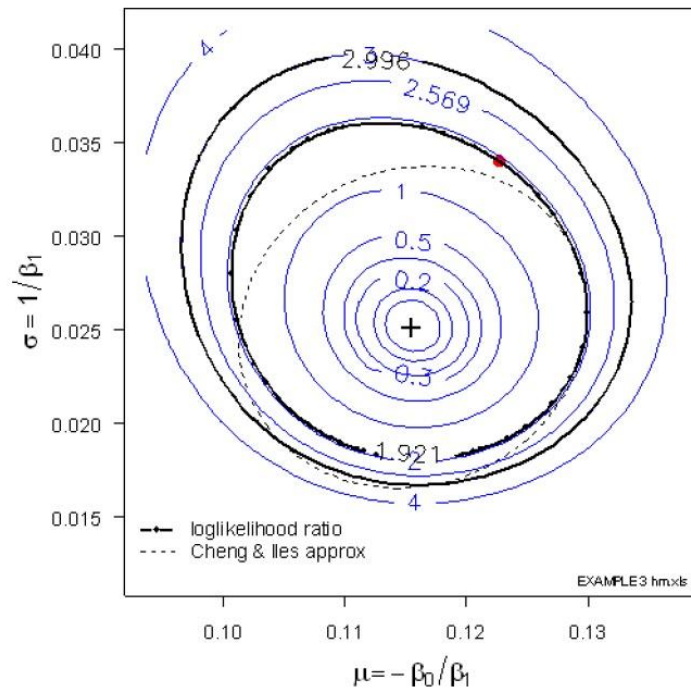
Figure 15: Plot of the loglikelihood ratio surface [8].

The MLE position representing the maximum likelihood estimate is shown with the '+' in the centre of Figure 15. Moving the (μ, σ) pair from the MLE position changes the loglikelihood and this is shown in the contour lines. The contour line, shown with alternating lines and dots that passes through the large red dot, is the 95% confidence bound for the parameter estimates based on the test data. Hence, in 95% of similar experiments, the true (μ, σ) pair is expected to be contained within such a confidence bounded ellipse. The large dot shows the (μ, σ) pair that produces $a_{90/95}$. All the (μ, σ) pairs on the contour that represent all the POD(a) curves create a family of POD(a)

curves. The envelope that contains this family of POD(a) curves represents the 95% confidence bounds on the original, maximum likelihood POD(a) curve [8].

It is interesting to note that the maximum likelihood estimates are not in the centre of the loglikelihood contours. When the sample size is small (as for the data analyzed in Figure 15) the contours are not symmetric, but as the sample size increases, the contours contract toward the MLEs and the contours tend to become symmetrically centred asymptotically [8]. Figure 15 also shows a dotted line ellipse that is centred on the MLE values, and this is the Cheng and Iles approximation to the confidence contour [64]. As revealed in Figure 15, for small sample sizes the Cheng and Iles approximation to the confidence contour shows poor results [8].

2.6.5.3 ImageJ for Multidimensional Analysis

ImageJ is an image processing program that was developed by the National Institutes of Health (NIH) for analyzing scientific multidimensional images [66]. ImageJ was developed to allow custom acquisition, analysis and processing using a built-in editor and Java compiler. The public domain software accommodates pixel value statistics of user-defined selections, such as measuring distances and can scale and sharpen images [66]. User-written plugins facilitate solutions to many image processing and analysis problems including multiple imaging system data comparisons [67], as will be implemented in this thesis.

3. Assessing Reliability and Capability of NDE System

The United States Department of Defense MIL-HDBK-1823A provides the guidance for assessing the capability and reliability of an NDE system. To ensure a valid assessment of an NDE system in terms of the POD as a function of target size, a, specific guidelines for planning, conducting, analyzing and reporting NDE reliability evaluations need to be followed [8]. Specifics with respect to experimental design, test variables and test specimens are outlined below.

3.1 Experimental Design

The objective of an NDE system reliability assessment is to determine the largest size crack the system can miss, and not the smallest crack the system can find [8]. To accomplish this, a relationship must be established between the POD and target size and the potential for false positives that defines the capability of the NDE system under specific application conditions [8]. A POD estimate is rendered ineffective and irrelevant if it is accompanied by an unacceptably high false positive rate. There are several variables that can contribute to a variation in NDE system response, and ultimately the uncertainty in flaw detection. These test variables that can affect the detection rate include the physical attributes of the target sites, NDE process variables, system settings and test protocol [8].

3.2 Test Variables

Each NDT system contains variables that should be considered when evaluating the specific inspection system in the application environment. Variables that could have a significant effect on POD are specimen pre-processing, inspectors, calibration, inspection process and imaging analysis.

3.3 Specimen Pre-processing

The variable associated with specimen pre-processing includes factors such as in-service environmental contamination and structural loading associated with maintenance and operations, and general surface condition [8]. Removal of the fastener in preparing the specimen for testing can also introduce additional damage to the test site, not only in the form of damage on the surface but also damage along the bore of the hole. A fastener hole that has significant surface damage can inadvertently affect the inspectors' ability to discern machine damage from the material discontinuity associated with a flaw. The PEC inspection method alleviates the need for fastener removal, hence damage associated with the removal process is minimized.

3.4 Inspector

Although some inspection systems have been shown to be inspector-independent, many of the NDE systems being analyzed reveal that the inspector is the most significant variable in the process [8]. In a manual inspection, such as BHEC where the inspector assesses the visual image and decides as to a hit/miss, the human factor can produce outcomes with large variation. It is also important to select inspectors with varying certification and training to ensure that the entire population of inspectors is represented and not just the most experienced inspectors. Although the PEC inspection method does not require an inspector to make an assessment as to a hit/miss decision, it is important that the inspector position the probe concentrically over the fastener head and minimize liftoff effects.

3.5 Calibration

To ensure that the inspection is conducted with the same sensitivity, independent of time or place, a calibration standard should be used prior to specimen inspection. This is particularly important in electronic inspection processes such as ECT. When there is more than one calibration standard available, the variation among standards should also be considered [8]. In the absence of a representative set of calibration standards and where the actual flaw population is sparse, outlier detection methods can be used [36].

3.6 Inspection Process

In designing the inspection process, specific controls on inspection parameters should be identified. In the case of ECT these inspection parameters could include signal frequency, gain and scan rates. With respect to the PEC inspection method, additional guidelines are followed to acquire specific components of the signal and to post-process the measurements data.

3.7 Imaging Factors

In ECT inspections, whether using manual inspections or automated C-Scan methods, an image is produced for evaluation by an inspector from which an assessment is made as to a hit/miss decision. Variables that should be considered when processing and evaluating images include maintaining consistency in hardware and software tools. Although there are significant human factors issues associated with inspectors manually evaluating images, controlling how the image is processed provides a standard basis from which all inspectors can then evaluate the image. Gain and frequency values can directly impact the BHEC signal response images that are presented to the inspector on the impedance plane display. These variables should be defined to ensure that the inspector conducts the analysis and makes a hit/miss determination based on an optimized image. With respect to automated C-Scan image processing, circuit design including the utilization of low and high-pass filters can be used to reduce noise and ultimately present the best image possible for inspector analysis. For the PEC inspection method investigated here, a hit/miss assessment is made, and analysis of the resulting image is performed using ImageJ software to assess the confidence of the $a_{90/95}$ value obtained through the **mh1823 POD** software.

3.8 Test Specimens

In evaluating the capability and reliability of a NDE system, the test specimens utilized in this assessment should reflect the structural types that would be encountered in the application being assessed. The specimen characteristics referred to include similarity with respect to geometry, material, surface condition, specimen processing and flaw

characteristics. Multiple specimen sets may be needed in a reliability assessment to not only account for varying structural types but also to minimize the possibility of the inspector becoming familiar with the specimen [8]. If the inspector gains familiarity with the test specimen the inspection does not represent an actual blind inspection and correspondingly, the POD assessment is not valid.

3.8 Physical Characteristics of Test Specimens

Test specimens should resemble the local geometries of those structural types that are to be assessed by a NDE system in a field application. This is important especially with ECT and UT as probe manipulation and features of the inspection process such as magnetic fields and sound waves are geometry dependent [8]. Bolt holes are one of the typical shapes that influence inspections and this is the application that will be assessed in this POD study.

Residual stresses resulting from raw material and manufacturing processes, part geometry and service history can significantly influence inspection results. This major influence on inspections has been shown with fluorescent penetrant testing (PT), UT and ECT.

Location and orientation of simulated flaws should be representative of the types of flaws encountered in an actual field environment. The inclusion of corner or surface flaws in the test specimens are significant geometric considerations with respect to the inspection techniques being assessed [8]. Flaws should be manufactured to a size typical of the component and application they represent.

The surface condition of the specimen can have a significant influence on the inspection signal-to-noise ratios [8]. Any machining of the specimen is to be consistent with machining that would be encountered with an actual part, to provide similar signal responses [8].

3.9 Target Size and Number of Inspection Sites

The statistical precision of the POD analysis depends on the number of inspection sites with flaws, the size of the flaws at these sites and the category of inspection result [8]. The data obtained through the inspector examination of the test specimens for the POD study can be grouped into two categories: those with only a binary output (hit/miss) or those that provide information as to the apparent flaw size ($\hat{a}\ vs\ a$).

It has been found that for a system that provides only a binary, hit/miss response, to obtain reasonable precision in the POD estimates, the specimen test set should contain at least 60 targeted sites [8]. This minimum value is an increase over Berens [7] recommendation of 30 targeted sites. For binary responses, an increase in the targeted sites will result in a significantly more precise estimate of a_{50} (target size at 50% POD), resulting in a smaller value for $a_{90/95}$. The reference to $a_{90/95}$ is important as it has become the de facto standard for design criterion [8]. It is the minimum detectable flaw size that a NDT method can detect 90% of the discontinuities of that size, 95% of the time [8]. For a system that provides a quantitative target response, $\hat{a}$, the specimen test should contain at least 40 targeted sites [8]. Additionally, the specimen set should contain at least three times as many unflawed inspection sites as flawed sites [8]. This ensures that an estimation can be obtained for the false positive rate.

Considering the importance of the $a_{90/95}$ criterion, there is increased emphasis on estimating the 90th percentile more precisely than other parts of the POD (a) curve [8]. This can be achieved by placing more targets in this region, while ensuring that there is a range of target sizes to enable an estimation of the entire POD curve [8]. There is a tendency to include an abundance of larger targets in the test set because smaller targets are difficult to produce with precision, it is imperative that there are smaller targets to gain a more accurate POD representation [8]. Flaws that are so large that they are always detected and those that are so small that they are always missed, result in limited information about the POD function [8].

3.10 Specimen Maintenance

It is imperative that test specimens be protected from mechanical damage and contamination that could alter the response acquired and therefore, an assessment of a specific NDE system. Mishandling of the specimen sets during evaluation and testing can result in the specimen being unrepresentative of actual inspection conditions, making the resultant POD analysis invalid. Damage can be minimized by packaging each test specimen in protective enclosures when not in use and carefully handling the specimens when they are being evaluated.

4. Experimental Technique

4.1 General

The purpose of this section is to provide an overview of the experimental setup utilized in the acquisition of data for the eventual probability of detection analysis. The specifications of the pulsed eddy current (PEC) probe utilized for this work are presented. An overview is provided of the data acquisition (DAQ) system and associated dual power operational amplifier (op-amp) that was used to control and measure the critical PEC system parameters. The specifications are then presented for the NAVAIR samples utilized in this analysis, that are representative of the wing structure of a CP-140 Aurora (Lockheed P-3 Orion) aircraft. Finally, the technique and associated equipment used to ensure concentric alignment of the probe on the fastener head is presented.

4.2 Pulsed Eddy Current Probe

All experimental data collection was conducted using a PEC probe with a central driving coil wound around a single 8 mm ferrite core. The ferromagnetic properties of the ferrite core contribute to the magnification of the magnetic flux produced by the driving coil and optimizes the resulting eddy currents encircling the ferrous fastener [13]. The 8 mm ferrite driving coil core being slightly larger in diameter than the head of the ferrous fasteners in the specimens being evaluated (7 mm), allows for more magnetic flux to be transferred to the fastener. In the work presented by Butt [12], four different PEC probes with varying specifications for driving coil diameter, resistance, differential pair spacing and number of turns, were assessed. It was determined that the 8mm ferrite core probe configuration (Figure 16) with the specifications outlined in Table 1 below provided the optimum results for second-layer crack detection in aircraft wing structures. The probe used in the work presented by Butt *et al.* [12][36][37] was the probe utilized for the POD analysis presented in this thesis.

Figure 16: Face of PEC probe utilized in experimental data collection showing 8 mm ferrite core surrounded by array of eight pick-up coils.

The probe used in this work has an array of eight pick-up coils positioned outside the driving coil, which are differentially paired and symmetrically oriented 180⁰ from each other.

The probe body was manufactured using a 3-D printer and the differential pair spacing was estimated based on the computer-aided design (CAD) rendering of the probe body [12].

Table 1: Detailed probe specifications for PEC probe used in experimental data collection [12].

Probe Parameter	Specifications
Driving Coil Inner Diameter	8 mm
Average Differential Pair Spacing	14.7 mm
# of Turns (pickup coils)	400
Approximate Driving Coil Resistance	15 Ω

4.3 Data Acquisition Equipment

The experimental setup with respect to the data acquisition equipment remained precisely as was presented by Butt [12]. The PEC probe detailed above is connected to

the amplifier circuit described in Section 4.4 below. A National Instruments (NI) USB-6361 data acquisition (DAQ) module, (specifications in Appendix K), digitizes the signal amplified by the op-amp with a sample rate of 100kHz/channel. The custom software interface designed in LabVIEW 2011 and used by Butt [12], controls specific acquisition parameters such as drive voltage, sample rate and signal post-processing. A 10-volt DC pulse excites the central driving coil, and the corresponding induced response of the differential pick-up coils is captured via an analog input on the NI DAQ module [12]. The signal responses are stored and then available for post-processing. A flowchart depicting the data acquisition process used for experimental data collection in [12] and in this work is presented in Figure 17.
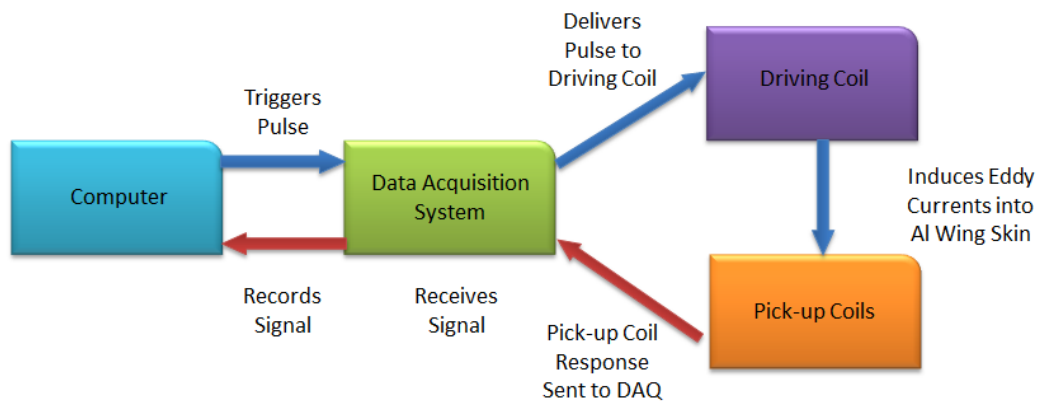
Figure 17: Flowchart depicting data acquisition process [12].

4.4 Driver Operational Amplifier Circuit

There was a requirement to amplify the driver signal produced by the NI analog output to improve the input signal and the resulting signal reproduction for MPCA analysis [12]. Butt [12] utilized a PA75 dual power op-amp wired in a noninverting configuration to amplify the driving coil excitation as shown below:
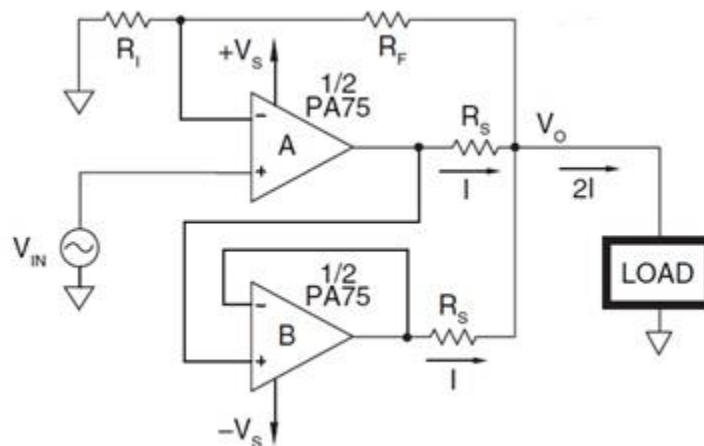
Figure 18: Circuit diagram for PA75 dual power op-amp wired in noninverting configuration [68].

The PA75 dual power op-amp used in the noninverting configuration provided a wide bandwidth, which contributed to a significant reduction in the level of noise in the first five eigenvectors resulting from the pick-up coil signal response signals [12]. The first five eigenvectors produced from MPCA of the pick-up coil pair response signal when utilizing the PA75 op-amp configuration is presented below in Figure 19.
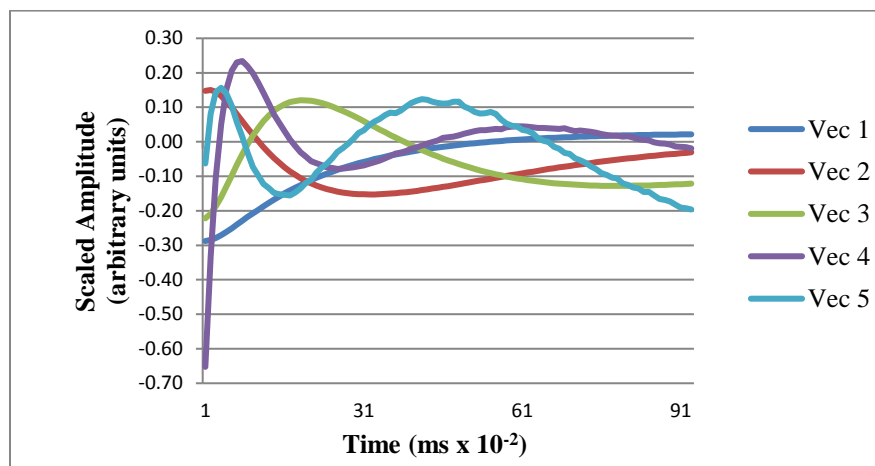
Figure 19: First five eigenvectors produced from MPCA of a pick-up coil pair response signal when the driver pulse was produced using an op-amp configured in a noninverting amplifier [12].

4.5 NAVAIR Sample Series Description

The series of samples acquired by the Aerospace and Telecommunications Engineering Support Squadron (ATESS) located at 8 Wing Trenton and used in the work presented by Butt *et al.* [12][36][37], were analyzed to complete the POD analysis presented in this thesis work. The samples were manufactured for the United States Naval Air Systems Command (NAVAIR) depot and are based on the wing lap-joint structure of the Lockheed P-3 Orion aircraft [12]. The P-3 Orion aircraft has the same airframe as the CP-140 Aurora aircraft, which is used by the Royal Canadian Air Force (RCAF). The samples are constructed of 2024-T3 aluminum plate with two sections joined together by a row of between 22 to 24 ferrous fasteners in a lap-joint configuration as shown in Figure 20.
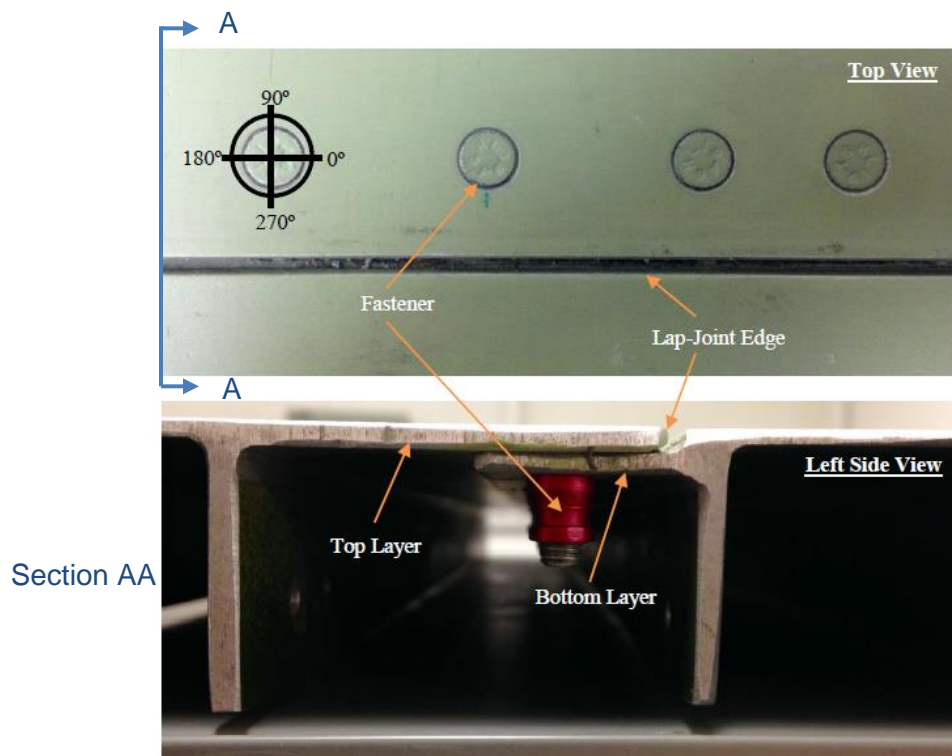
Figure 20: Top and side view of P-3 Orion sample showing the location of top and bottom layers along with fasteners and lap-joint edge [12].

The thickness of each of the two sections varies between 2.1 mm and 2.8 mm and the

interlayer gap between the top and bottom layers ranged from 0 mm to 0.12 mm [12]. A summary of the top and bottom layer thickness and interlayer gap for the complete NAVAIR series samples is provided in Appendix B. Fastener specifications for these samples include fastener length of 15 mm, head diameter of 7.0 mm and shaft diameter of 4.5 mm [69]. The distance from each individual fastener to the lap-joint edge was measured using a scaled image where the scale depicts precision down to 0.5 mm. The mean fastener-to-edge distance across all seven samples was 6.81 mm, with standard deviation ranging from 0.37 mm (Standard 25) to 0.78 mm (Standard 22).

Each of the seven NAVAIR samples contain both blank fastener sites where no notches are present and fastener sites where electric discharge machined (EDM) notches are present. In these samples, the EDM notches have been cut at a 45° angle to the edge of the bore hole such that they present a 1:1 aspect ratio. They have been cut into the bottom of the top layer, the top of the bottom layer, or both and range in size from 0.76 mm (0.030") to 6.10 mm (0.240") in length, with the width remaining constant at 0.254 mm (0.010"). The notch location, length, orientation and fastener number for each of the samples are provided in Appendices C and L. There are a total of 159 inspection locations, consisting of 101 notched and 58 blank fastener sites. The blank fastener locations for each of these samples is also provided in Appendix A. The NAVAIR samples contain a high percentage of notched fastener sites, which is ideal for assessing the PEC technique, although it is important to note that in field applications this percentage would be expected to be less than 1%.

4.6 Probe Alignment

Data acquisition of the pick-up coil signal response is obtained by positioning the PEC probe concentrically over each individual fastener head. An acrylic alignment guide is used to align the probe. The alignment guide includes a thin clear plastic sheet affixed to the base that has a center hole cut to the same size as the fastener head. Although the intent of the guide is to assist in aligning the probe concentrically over

the fastener head, it is a visual alignment technique and as such small, centering errors occur even when the inspection is performed by an experienced operator. To characterize the error due to misalignment, multiple measurements were taken at each fastener site for each sample data set to ensure proper representation of probe placement variation within the data set. Figure 21 depicts the data acquisition hardware components displaying the PEC probe, alignment guide and sample.
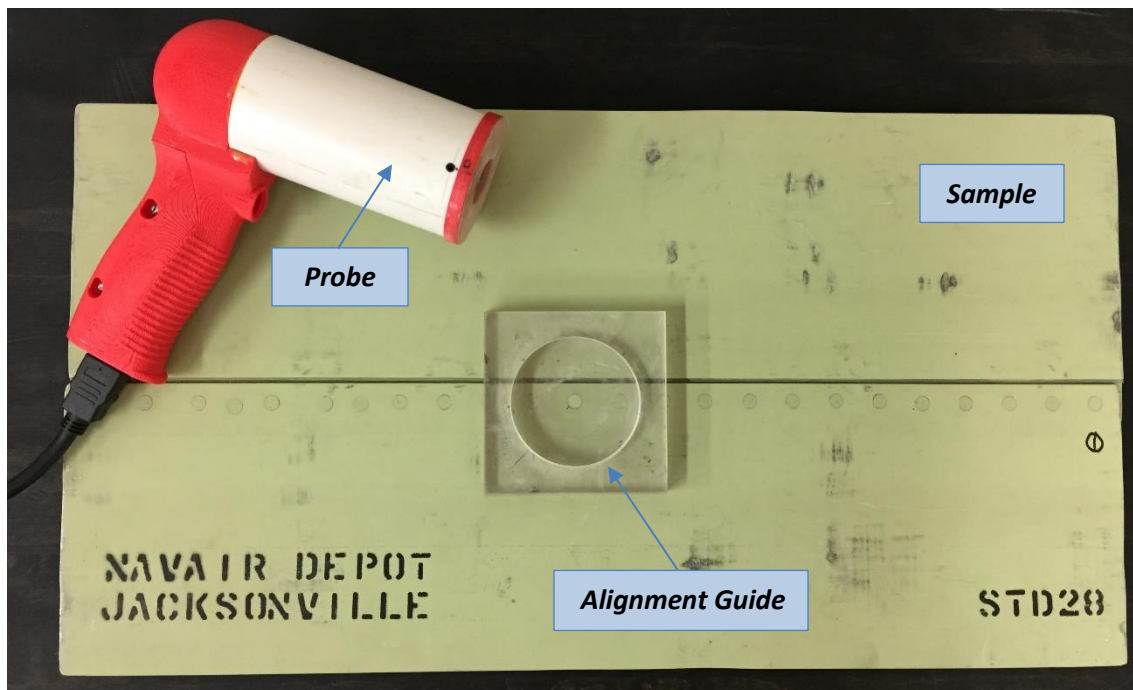
Figure 21: Experimental setup showing cut-out hole of alignment guide positioned concentrically on fastener head.

Figure 22 shows the actual positioning and orientation of the probe and the alignment guide during the process of data acquisition. Probe alignment over the fastener is one of the few operator dependent aspects of the PEC inspection technique and as such, experience of the inspector and operator fatigue can impact the results.
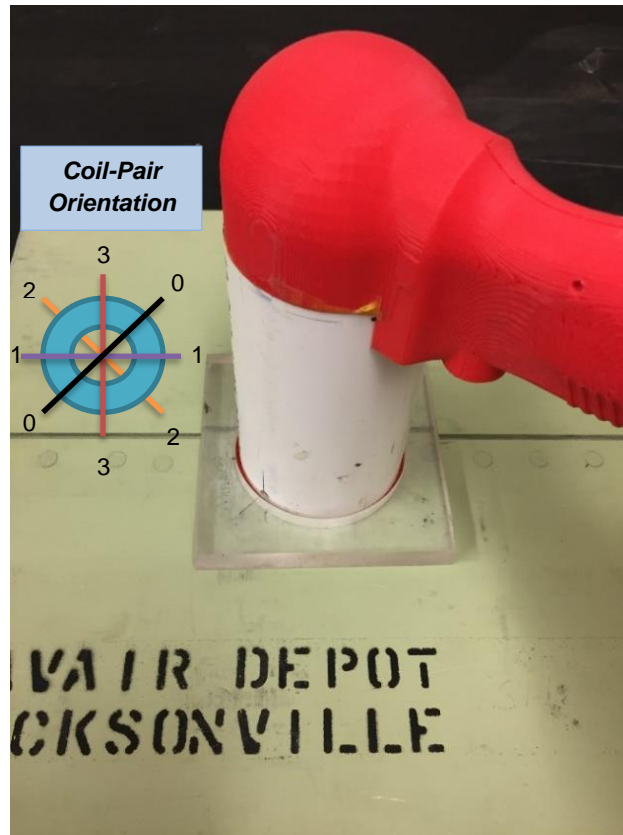
Figure 22: Actual positioning and orientation of the probe and the alignment guide on the sample during the data acquisition process.

4.7 Experimental Test Plan

The experimental test plan for this thesis research is presented in Table 2 below.

Table 2: Experimental Test Plan

Test Procedure	NAVAIR Sample Series						
	Std 22	Std 23	Std 24	Std 25	Std 26	Std 28	Std 32
Sample Detection - Measured Data							
Data measurement - 10 repeats	X	X	X	X	X	X	X
Cluster analysis - measured data	X	X	X	X	X	X	X
Fastener to lap-joint edge distance	X	X	X	X	X	X	X
Data measurement - revised edge distance	X	X	X	X	X	X	X
Cluster analysis - revised edge distance	X	X	X	X	X	X	X
Cluster / revised edge - top layer notches only	X	X	X	X	X	X	X

Test Procedure	NAVAIR Sample Series						
	Std 22	Std 23	Std 24	Std 25	Std 26	Std 28	Std 32
Cluster / revised edge - 2nd layer notches only	X	X	X	X	X	X	X
Probe temperature variation - 10 repeats	X		X			X	X
Probe off-centering - 10 repeats	X		X			X	
Cluster analysis - probe temperature variation	X		X			X	X
Cluster analysis - probe off-centering	X		X			X	
Sample Detection - Simulations							
Bootstrap / SHV / cluster analysis - 40 sample size	X	X	X	X	X	X	X
Bootstrap / SHV / cluster analysis - 80 sample size	X	X	X	X	X	X	X
Simulations - 5% False Call Rate							
Data Fraction 95% - 80 sample grouping size	X	X	X	X	X	X	X
Data Fraction 97.5% - 80 sample grouping size	X	X	X	X	X	X	X
Data Fraction 100% - 80 sample grouping size	X	X	X	X	X	X	X
Simulations - 10% False Call Rate							
Data Fraction 95% - 80 sample grouping size	X	X	X	X	X	X	X
Data Fraction 97.5% - 80 sample grouping size	X	X	X	X	X	X	X
Data Fraction 100% - 80 sample grouping size	X	X	X	X	X	X	X
Simulations - 15% False Call Rate							
Data Fraction 95% - 80 sample grouping size	X	X	X	X	X	X	X
Data Fraction 97.5% - 80 sample grouping size	X	X	X	X	X	X	X
Data Fraction 100% - 80 sample grouping size	X	X	X	X	X	X	X
Comparison of Measured and Simulation Results	X	X	X	X	X	X	X
POD Analysis							
POD - Measured Results							
Measured detection results - Link comparison	All samples combined						
Plot of even distribution binned notch sizes - Logit	X	X	X	X	X	X	X
Measured detection results - Logit link function	All samples combined						
Logit link function - Top layer notches only	All samples combined						
Logit link function - 2nd layer notches only	All samples combined						

Test Procedure	NAVAIR Sample Series						
	Std 22	Std 23	Std 24	Std 25	Std 26	Std 28	Std 32
POD on Simulations - 5% False Call Rate							
Data Fraction 95% - 40 sample grouping size	All samples combined						
Data Fraction 97.5% - 40 sample grouping size	All samples combined						
Data Fraction 100% - 40 sample grouping size	All samples combined						
POD on Simulations - 10% False Call Rate							
Data Fraction 95% - 40 sample grouping size	All samples combined						
Data Fraction 97.5% - 40 sample grouping size	All samples combined						
Data Fraction 100% - 40 sample grouping size	All samples combined						
POD on Simulations - 15% False Call Rate							
Data Fraction 95% - 40 sample grouping size	All samples combined						
Data Fraction 97.5% - 40 sample grouping size	All samples combined						
Data Fraction 100% - 40 sample grouping size	All samples combined						
POD on Simulations - 5% False Call Rate							
Data Fraction 95% - 80 sample grouping size	All samples combined						
Data Fraction 97.5% - 80 sample grouping size	All samples combined						
Data Fraction 100% - 80 sample grouping size	All samples combined						
POD Specimen Coverage	All samples combined						
Revised $a_{90/95}$ - ImageJ - all FC/data fractions	All samples combined						
POD on Simulations - 10% False Call Rate							
Data Fraction 95% - 80 sample grouping size	All samples combined						
Data Fraction 97.5% - 80 sample grouping size	All samples combined						
Data Fraction 100% - 80 sample grouping size	All samples combined						
POD Specimen Coverage	All samples combined						
Revised $a_{90/95}$ - ImageJ - all FC/data fractions	All samples combined						
POD on Simulations - 15% False Call Rate							
Data Fraction 95% - 80 sample grouping size	All samples combined						
Data Fraction 97.5% - 80 sample grouping size	All samples combined						

Test Procedure	NAVAIR Sample Series						
	Std 22	Std 23	Std 24	Std 25	Std 26	Std 28	Std 32
Data Fraction 100% - 80 sample grouping size	All samples combined						
POD Specimen Coverage	All samples combined						
Revised $a_{90/95}$ - ImageJ - all FC/data fractions	All samples combined						
POD Simulations- Optimum Data Fraction/FC Rate							
Comparison of Measured vs. Simulation POD	All samples combined						
Top layer notches only	All samples combined						
2nd layer notches only	All samples combined						

5. Signal Processing and Analysis

5.1 General

The purpose of this section is to detail the signal response post-processing methods utilized in this work to determine notch detection capability. The individual steps involved with post-processing of the pick-up coil signal response will be discussed below, commencing with signal gating. Signal gating selects only the portion of the time-domain signal that contains the desired information associated with the notch that is to be extracted for analysis. Modified Principle Components Analysis is then applied to the signal response at the gated area of interest, generating MPCA scores for each of the four coil pairs. The MPCA scores are then further processed using a cluster analysis approach, which accounts for environmental factors, and variability in repeat measurements while accounting for the distance from the fastener to the lap-joint edge. An overview will then be presented of the robust statistics that were used in conjunction with a blank fastener simulation algorithm to achieve Type 1 (blind) outlier detection.

5.2 Signal Gating

The signal response measurement data was collected and analyzed using MPCA [19] as described in Section 2.5.2. The differential pick-up coils have a distinct response to the DC pulse that powers the driving coil. A typical signal response obtained for a pair of differentially wired pick-up coils positioned 180º apart, as configured in the PEC probe used in this work, is shown in Figure 23. When the DC voltage is applied, the response of the pick-up coils due to induction produces the front end of the signal response, whereas the reverse induction produces the response at the back end of the signal when the DC voltage is switched off. A defined signal gate (Figure 23) was applied to each of the four differential coil pairs to capture the portion of the time-domain signal that includes the second layer notch information to be analyzed. MPCA is then used to analyze only the signal variations encountered within the gated bounds of the signal.
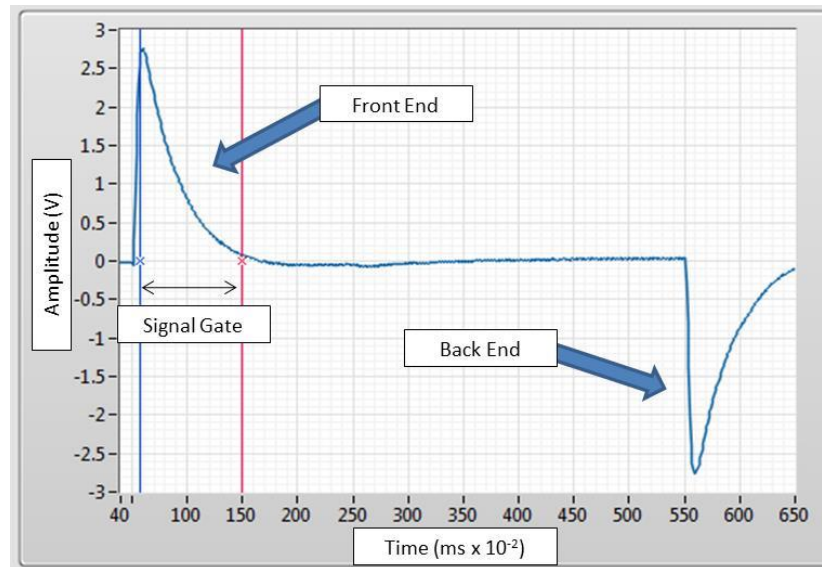
Figure 23: Raw PEC differential pick-up coil response showing the front and back end transient response, and the selected signal gate obtained from screenshot of the LabVIEW display [12].

Based on previous work by Stott [9][10] and Butt *et al.* [12][36][37], signals were gated from 0.58 ms to 1.5 ms (0.92 ms signal gate) as this portion of the transient response signal provided the vital second layer crack information needed for detection analysis. The gates were applied to each of the four differential coil pair signals using custom-built LabVIEW acquisition software used by Butt [12], producing eigenvectors and MPCA scores. Additional software output parameters utilized in the signal processing, to produce eigenvectors and MPCA scores, consisted of a setup with 1000 points/section with background removed, producing 45 points and 5 eigenvectors.

5.2.1 Differential Coil Pair Eigenvectors

In this thesis work, there was no change to the signal gate values identified by Stott and Butt. The signal gate values were applied to the unique transient response signal associated with the experimental measurements obtained in this work. Unique MPCA scores were calculated separately for each of the four differential coil pairs, using the same LabVIEW software setup as detailed in [12].

Through extensive analysis of the post-processed MPCA scores, for each of the four differential coil pairs, it was determined that the first five eigenvectors produced from

MPCA were useful. The eigenvectors obtained in this thesis work produced nearly equivalent, yet slightly reduced, detection results when compared with those obtained by Butt *et al.* [12][36][37]. It was evident that the eigenvectors were intricately linked to the specific 8 mm ferrite core PEC probe that was used. Since the eigenvectors for each of the four coil pairs determined by Butt [12] were optimized as part of his research and produced similar results to those obtained in this thesis work, these were used for all data post-processing of measurements obtained in this research. Consequently, the eigenvectors linked to the PEC probe [12] were used to process data measurements obtained in this thesis work, thus producing unique MPCA scores.

5.3 Factors Affecting the Measurements

After calculating the scores for each of the four coil pairs, the cluster analysis method detailed in Section 2.5.3 was applied to each separately. The covariance matrix for the known blanks for each coil pair was determined from the MPCA raw scores. The MPCA raw scores were further processed as detailed below to remove shifts in the data due to environmental factors, and effects due to repeat measurements and distance from the fastener to the lap-joint edge.

5.3.1 Removing Shifts Due to Temperature Differences

It was noted by Butt [12] that when collecting experimental data from different samples on different dates and times that the data obtained was influenced by slightly different environmental conditions. This resulted in a shift of the blank cluster from one sample to the next. Although it was suspected by Butt [12] that temperature could be responsible for the shift of the blank cluster, due to the results obtained by Buck *et al.* [23], Butt utilized mathematical methods to remove the shift. It was determined that the shift could be removed by calculating the mean blank scores for a particular sample data set and subtracting these means from all scores in that specific data set [12].

Although Butt [12] did not investigate the assumption by Buck *et al.* [23] that small variations in temperature produced notable effects on PEC inspection results, a small study was conducted to investigate this further in this work. Four NAVAIR samples were chosen (Standards 22, 24, 28 and 32) as they exhibited distinct differences associated

with edge distance, layer thickness and gap. Standards 22 and 24 were shown by Butt to produce near perfect (close to 100%) detection rates. For Standard 28 and Standard 32 from the NAVAIR sample series, Butt obtained substantially reduced detection rates, which was attributed to the distance to the lap-joint edge and larger nominal gap, respectively [12].

Previous work showed that only the temperature of the probe and not the temperature of the sample caused the data to shift for temperature ranges between 10ºC and 30ºC [70]. The electrical properties of the sample were only weakly dependent on temperature, but the probe was very sensitive to thermal expansion and resistivity changes. Consequently, a study was conducted to assess the effects of probe temperature variation.

The temperature study of the four NAVAIR samples consisted of obtaining experimental measurements of the samples utilizing the procedure as detailed in Section 4, with the only difference being the variation in the temperature of the probe. Three sets of experimental measurements were obtained for each sample with the temperature of the probe being 10ºC, 20ºC and 30ºC. The temperature of the probe was regulated by placing it in a temperature-controlled chamber. Once the probe reached the desired temperature, one measurement was taken of each fastener on that specific sample. The probe was then returned to the temperature chamber and the process was repeated until all 10 repeat measurements were obtained. Limiting the fastener measurements to one pass of the sample, with less than 8 seconds wait time between fasteners, minimized the probe cooling effects.

5.3.2 Probe Placement Effects

It has been recognized that repeat measurements of the same flaw can produce variation in the signal response outputs due to minute variations in probe positioning and individual inspector setup [35]. Previous work by both Horan *et al.* [19] and Babbar *et al.* [69] showed that probe off-centering can change the amplitude and shape of the signal response. Repeat measurements were taken of each fastener location to characterize the effects of probe placement variation. Butt *et al.* [12][36][37] determined that the variation due to repeat measurements produced a linear relationship with respect to the

first two scores, s_1 and s_2. Butt *et al.* [12][36][37] demonstrated that by applying a unitary rotation to all scores, the variation due to repeat measurements could be isolated to s_1 [12]. Since s_1 no longer contained useful information associated with notch detection, only the remaining blank scores (s_2 - s_5) were used for subsequent MD calculations.

Measurements were obtained for three NAVAIR samples (Standards 22, 24, and 28) to observe the results obtained when the probe was placed off-centre of the fastener head by up to 2 mm in distance. Previous work assessed the effects of probe off-centering in the direction parallel and perpendicular to the lap-joint edge, with the most significant impact on signal response occurring when the probe off-centering was parallel to the lap-joint edge [32]. In this thesis work, utilizing the PEC probe consisting of four pick-up coil pairs positioned at 45 degrees with respect to each other, the probe offset was randomly oriented in all directions with respect to angular position to assess the effect on the overall detection rate. Using an additional distance guide, the off-centering was confined to a range of between 0.5 mm to 2.0 mm. A single data set was obtained for each of the three samples, with each data set consisting of ten repeat measurements for each fastener. The off-centering of the probe was done for each fastener to ensure that a random set of notches and blanks had measurements taken where the probe was not concentrically placed over the fastener head.

5.3.3 Fastener Proximity to the Lap-Joint Edge

As demonstrated by Butt [12], the distance from each fastener to the lap-joint edge impacts the pick-up coil signal response. It was shown that the signal response changes as a function of distance from the edge, when each of the remaining blank scores (s_2 - s_5) was plotted with respect to edge distance [12]. Since there was a linear relationship observed, the resulting slopes for each score was used as a distance correction [12]. Plots presented by Butt [12] of blank scores plotted with no edge correction (Figure 24) and after edge correction (Figure 25) are provided below.

Although Butt *et al.* [12][36][37] obtained close to 100% detection for the three NAVAIR samples indicated in Figures 24 and 25 when using the edge correction method, there were four other NAVAIR samples that produced detection rates that were significantly

lower in comparison. Butt *et al.* [12][36][37] recognized that for some samples there was still spread in the blank cluster after the edge correction was applied.
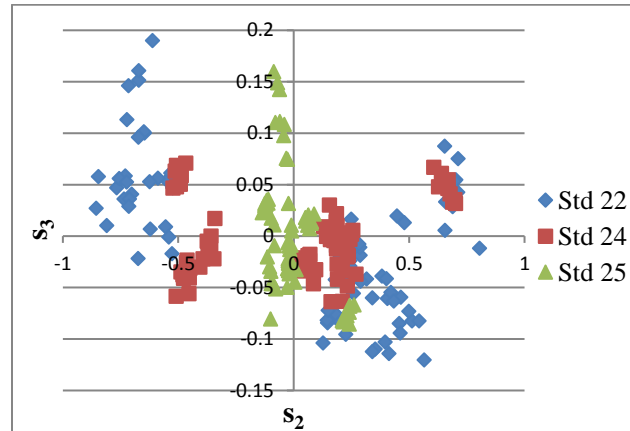
Figure 24: Plot of blank scores for three standards showing s_3 vs. s_2 without edge corrections [12].
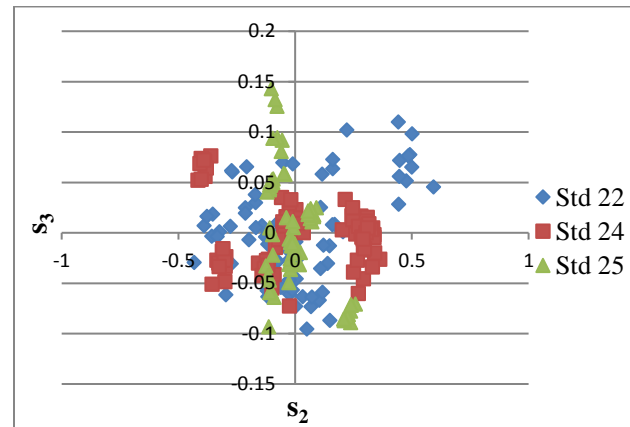
Figure 25: Plot of blank scores for three standards showing s_3 vs. s_2 after edge corrections [12].

The variation in the detection rate achieved for the four remaining NAVAIR sample series was investigated further in this thesis work. Using the results obtained by the robust statistics simulation method for all four coil pairs on each of the samples, the eigenvalues for the covariance matrices of the raw and zeroed scores were analyzed and compared. Additionally, correlation matrices were calculated for the raw and zeroed scores and those eigenvalues were analyzed across all samples. Although no pattern was recognized that would explain the variation in detection rate, additional analysis was conducted on the distance from the fastener to the lap-joint edge.

It was noticed that one of the lowest sample detection rates obtained by Butt [12] was for Standard 26. In determining fastener to lap-joint edge distance for all samples the images displayed a measurement scale with 1 mm increments. For Standard 26, Butt still used the same measurement scale, but utilized a longer exposure time and smaller aperture and focal length camera settings. This resulted in lower resolution images in comparison to the other samples Butt *et al.* [12][36][37] evaluated.

The results obtained for Standard 26 using the experimental measurements obtained in this thesis work resulted in the lowest detection rate in comparison to other samples. Higher quality images with increased resolution (480 dpi vs. 72 dpi) were obtained for each fastener in Standard 26 as shown below in Figure 26, utilizing a 7.1-megapixel Kodak EasyShare Z710 zoom digital camera. The image quality obtained for each individual fastener image resulted in an image size of approximately 1.3 MB (3072 pixels x 2304 pixels), in comparison to 440 kB (1280 pixels x 960 pixels) of those that were utilized in [12]. The high-resolution images display a measurement scale in 0.5 mm increments, and the fastener to lap-joint edge distances were measured in comparison to this scale. The high-resolution images were accessed in Microsoft Word, where the shape/line function was utilized on the image to obtain the distance measurement. The distance for each fastener was measured from the lap-joint edge to the edge of the fastener, with the average of five independent measurements.

Figure 26: High resolution image used for fastener to lap-joint edge distance measurements.

Using the fastener to lap-joint edge distance measurements obtained with higher resolution images with 0.5 mm scale increments, the detection rate for Standard 26 increased by 29.4% and the false call rate decreased by 60.8%. Based on the increased detection rate attained for this sample, higher resolution images were obtained for all NAVAIR sample series specimens and the fastener to lap-joint edge distance measurements were recalculated. Higher resolution sample images were not obtained for Standard 22 as the detection rate for this sample was already at 100%.

To confirm that the edge correction process used by Butt *et al.* [12][36][37] and detailed in Section 6.3.3 was still valid with the revised fastener to lap-joint edge distance measurements, the graph of s_3 vs. s_2 was replotted (Figure 27).
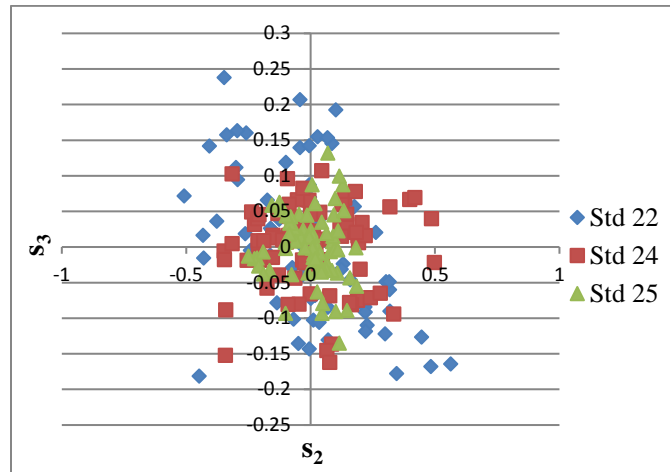
Figure 27: Plot of blank scores for three standards showing s_3 vs. s_2 after edge correction using revised fastener to lap-joint edge distance measurements.

Using the experimental measurements and revised fastener-to-lap-joint edge distances obtained in this thesis, it is evident from Figure 27 that the edge correction process presented by Butt *et al.* [12][36][37] still holds.

5.3.4 Decision Threshold

A decision threshold is determined based on a target false call rate and the MD is compared to this decision threshold to determine whether the fastener hole has a notch present [30]. As the MD has a chi-square distribution with n degrees of freedom correlating to the number of variables or scores utilized, the decision threshold was calculated using the chi-squared distribution function in Microsoft Excel with four degrees of freedom. The four degrees of freedom are correlated to the four remaining blank scores (s_2 - s_5) which were analyzed using the cluster analysis method.

5.4 Measured Results Analysis

Analysis of the experimental measurements obtained for the seven NAVAIR samples was conducted using the cluster analysis method. Each sample was analyzed individually, where the raw scores were rotated, edge corrected for distance and zeroed. The decision threshold was determined, so as to achieve a false call rate of 5%. The

MD calculated for each of the repeat measurements, for each of the fasteners, was compared to the decision threshold to register the result of whether the fastener hole has a notch or blank present. This process was repeated for the remaining NAVAIR samples. This utilized a Type 3 outlier detection method [17], as there was a requirement that the signal information associated with the blank fasteners be known in determining the MD.

5.5 Robust Statistics Simulations

As a consequence of the variations in wing skin thickness and interlayer gap in this application, a representative calibration standard may not be possible to achieve in field applications. A field deployable PEC inspection technique must be able to distinguish between cracks and blanks without the need for a calibration standard. It has been suggested that a potential solution to the in-situ calibration issue is a robust statistics method known as smallest half volume (SHV) [33]. The SHV method assumes that only a small fraction of fastener locations inspected will contain defects. Thus, this analysis technique can obtain blind detection results with no prior knowledge of blank fastener locations [20]. Although the original SHV method could accommodate up to 49% outliers, the inclusion of a data fraction as described in Section 5.5.1 below, allows for improvement in the blind detection results as the number of notches in the simulation increases [20].

The NAVAIR sample series used in this analysis were specifically engineered so that the samples contained a larger number of flawed than unflawed sites (flaw-centric). These NAVAIR standards were initially developed to test fast-scan methods [71], hence the intent in manufacturing these samples was to produce as comprehensive a set of defects as possible. Correspondingly, of the 22-24 fastener sites in each sample, the number of unflawed fastener sites ranged from seven to ten. Hence, the flaw-centric samples with fastener heads inserted, were also ideal for assessing PEC methods associated with this thesis work. To compensate for the inflated number of notched fastener sites in the samples, which do not represent in-service samples, a method was devised to increase the number of blank fastener measurements. This was accomplished by using the bootstrap technique described in Section 2.5.5.2 to randomly

select blank fastener data, and consequently produce covariance matrices, that have statistically similar properties to real experimental measurements. The randomly generated covariance matrices simulate real world detection scenarios where there are only a small percentage of flawed sites in the sample being inspected. Once the randomly generated covariance matrices are generated, the simulation technique utilizes the cluster analysis and robust statistics methods as outlined above to determine notch detection results.

5.5.1 Data Fraction

Simulations such as the one used in this analysis are invaluable in drawing statistical conclusions on the reliability of the NDT system being investigated, because not only can the number of flawed sites be predetermined, but other variables can be controlled.

To use the modified SHV algorithm, an appropriate data fraction [20] must be selected. The data fraction defines the percentage of data that will be used in determining the statistical center and the fraction of data used in the construction of the covariance matrix [20]. Hence, a data fraction of 95% indicates that the most tightly clustered 95% of the data is used in constructing the covariance matrix. Underhill *et al.* [20] concluded that the best detection performance for second layer crack detection with a small number of outliers was obtained when the data fraction was set to 97.5%. Figure 28 reveals the results of the study, which included a series of simulations where the sample included four randomly selected notches and 40 blank fastener sites, utilizing varying data fraction values. The plot produced by Underhill *et al.* [20] shows the detection results obtained with varying data fraction values ranging from 90 – 100%.
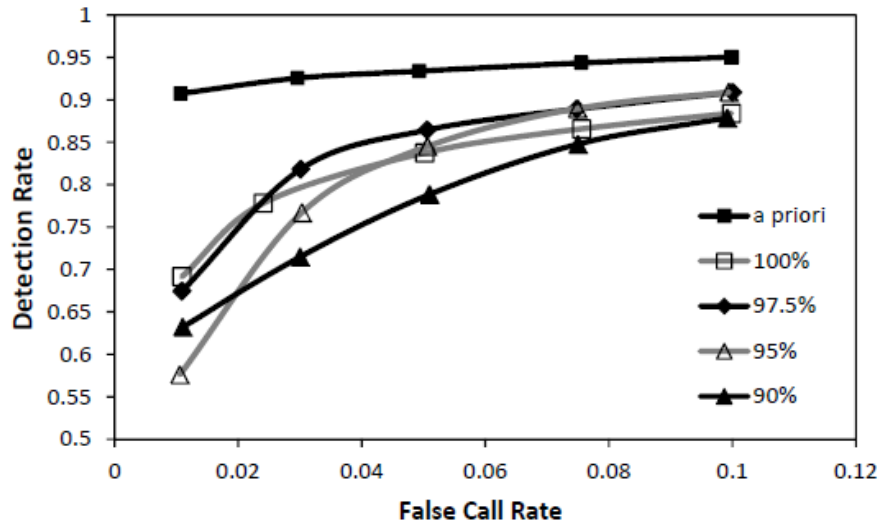
Figure 28: Detection rate vs. false call rate for simulations using the SHV algorithm with a data set containing 2 notches and 40 blanks [20].

The *a priori* results in Figure 28 represent the case when the blank fastener information is known prior to analysis, which would be equivalent to the "measured results" discussed in this thesis work. The results for the case involving 4 notches and 40 blank fastener sites represents 10% outlier contamination. The fact that the best results were obtained when the data fraction was set to 97.5%, and not 90% indicates that the SHV algorithm demonstrates robustness, when some outliers are included in the data set [12]

In this thesis work, the simulation technique and SHV algorithm described above were used to determine detection rates for each of the NAVAIR sample series. The simulations were repeated for data fraction values set to 95%, 97.5% and 100%.

5.5.2 Sample Grouping Size

Previous work by Underhill *et al.* [33] postulates that for the inspection technique that is being assessed in this work, a minimum of 40 blank fastener sites is required to obtain reliable detection results. Annis and Gandossi [34] recommend that for hit/miss POD vs. size modelling, the minimum number of targets is 60.

In this thesis work, simulations were conducted using the SHV algorithm for two cases, consisting of 40 samples and 80 samples, respectively. This allowed for an assessment to be made with respect to the influence of sample size on detection rate and POD vs. size results. It also allowed for an easy comparison of the simulation results with respect to the percentage of notches overall in the sample. In both the 40 and 80 sample set, simulations were conducted with 1, 2, 3 and 4 notches present. For the simulations with 80 samples, this resulted in a maximum of 5% notches, whereas in the simulations with 40 samples included a maximum of 10% notches. The simulations were set up to inspect each notch 100 times during the simulation, resulting in a maximum number of runs being 1600 for one complete simulation.

5.5.3 Determination of Decision Threshold for Specific False Call Rates

As in the analysis of the experimentally measured data, with simulations conducted using the SHV algorithm, there is a requirement to determine the decision threshold that will be compared to the MD to register whether there is a notch present at the fastener site.

Simulations were conducted using normally distributed random numbers for both the blanks and notches. A normal distribution with mean of zero and standard deviation of one is represented by a bell curve. On an interval, the normally distributed random numbers close to the mean are more likely to be selected than those far away from the mean. In determining the decision threshold for specific false call rates, when a small fraction of the data that displays the largest deviations is neglected, the data is no longer normally distributed and hence the chi-squared statistics do not apply. Once the simulation results were obtained, the cut-off value associated with specific false call rates was determined. This cut-off value was then used to calculate the decision threshold as described in Section 5.3.4.

In this thesis, simulations were conducted using the SHV algorithm for both 40 and 80 samples, with 95%, 97.5% and 100% data fraction, and decision thresholds linked to 5%, 10%, and 15% false call rates.

6. POD Analysis

6.1 General

A probability of detection (POD) analysis provides the metric needed to assess the reliability associated with a specific non-destructive testing technique. This section outlines the experimental technique used to conduct the POD analysis on PEC inspection technique for the detection of second layer cracks at ferrous fasteners in aircraft lap-joint structures. Factors that were significant in conducting this POD analysis, including sample size, POD model, distribution of notch sizes, Probability of False Positive (PFP), and $a_{90/95}$ determination will be discussed below.

6.2 Focus of Analysis on POD vs. Size Model

A POD analysis can be conducted using either binary response (hit or miss) or continuous response ($\hat{a}$ vs. a) data. Electromagnetic methods, such as conventional eddy current testing can produce $\hat{a} \, vs. \, a$ data. The PEC signal response in this thesis work was processed to produce hit/miss data as there were insufficient experimental samples containing smaller notch sizes, and the PEC technique and associated cluster analysis is inherently a hit/miss method. Correspondingly, it was not possible to group the notches to produce an accurate assessment of size detection rate with this method. Hence, the POD vs size (hit/miss) model was utilized in this analysis. It is important to note that once a crack is detected, conventional eddy current methods such as BHEC can be used to size the flaw after fastener removal [60].

Binary data contains less information than $\hat{a}$ data and problems can be encountered when all targets are detected in an $\hat{a} \, vs. \, a$ experiment [34]. Since binary data does contain less information, as indicated in [8] and [34], there is a requirement to have a greater number of specimens when analyzing hit/miss, as opposed to $\hat{a} \, vs. \, a,$ data. This will be discussed in greater detail in Section 6.5.

As Gandossi and Annis [35] indicate, the single most influential factor on POD is target size. This provides the foundation for further investigation of other influential factors, but

in the initial POD analysis it is critical that experimental specimens have characteristics, such as size, that are measurable [34]. For example, if a specific measure can be obtained for known corrosion in a sample, only then could corrosion be a measurable characteristic. In this case, NDE equipment would produce essentially the same output if the specific measure can be associated with the pre-defined corrosion measurement, and detection of this characteristic can be made [34].

There are specific limitations associated with the **mh1823 POD** software. The crack sizes in the experimental samples need to be true and correct for the software to provide accurate results [34]. The NAVAIR samples used in this work included notches that were manufactured with precision to within +/- 0.381mm (0.015") tolerance. Schematics provided with the specimens show the notch size and orientation on each fastener. Only one of the 101 notched fastener sites were unknown, although this was a fastener where there were notches on both the top and bottom layer. During analysis, four fasteners were removed from Standard 26 to confirm the presence, orientation and size of the notches at these fastener sites, as they displayed less than 100% detection rate. The specifications of these notched sites were validated and provided independent verification with the use of laser-based image capturing equipment, with results as shown below in Figure 29.

Figure 29: Laser image of hole 18 on Standard 26 displaying an 0.080" notch located at the bottom of the top layer, oriented at 225º CW from the lap-joint edge.

An additional limitation of the **mh1823 POD** software is that the model is to agree with

the data, such that on the POD (a) curve there are no misses at large sizes where the POD is near one or hits at small sizes where POD is near zero [34]. The POD(a) results obtained from the analysis of the NAVAIR specimens confirmed that there was sufficient precision in specimen notch size and the model agreed with the data.

6.3 Logit Function

As indicated in Section 2.6.4.3, it has been shown that when considering binary response NDE data, it is most effectively modelled using the generalized linear model logit link [34]. Historically, this model has not only been used almost exclusively with NDE binary data, but also in other scientific research areas involving medicine and pharmacology [34]. Both the logit and probit models are symmetric links, although there is a tendency for the probit link to be more sensitive to outliers at the POD extremes, resulting in unexpected behaviour [34]. In conducting the POD analysis of the measurement data obtained with the NAVAIR samples, although the results for the probit link provided similar deviance and Generalized Linear Model (GLM) parameter correlation as the logit link, the logit link provided the best fit.

To confirm that the logit link adequately represented the measurement data, the notch sizes for all NAVAIR samples were binned with an even distribution of notches (Table 3). As shown in Figure 30, using the experimental measured data obtained for the NAVAIR samples, the binned data follow the logit curve.

Table 3: Notch data binned for measured results of NAVAIR samples.

Notch Interval (inch)	No. of Notches in Bin	Average Detection Rate	Average Notch Size (inch)
0.03-0.035	10	0.68	0.032
0.04-0.055	10	0.81	0.047
0.06-0.075	13	0.908	0.068
0.08-0.095	10	1	0.086
0.1-0.11	13	0.985	0.106
0.115-0.125	11	1	0.122
0.13-0.135	10	1	0.132
0.14-0.21	14	1	0.192
0.215-0.24	10	1	0.221
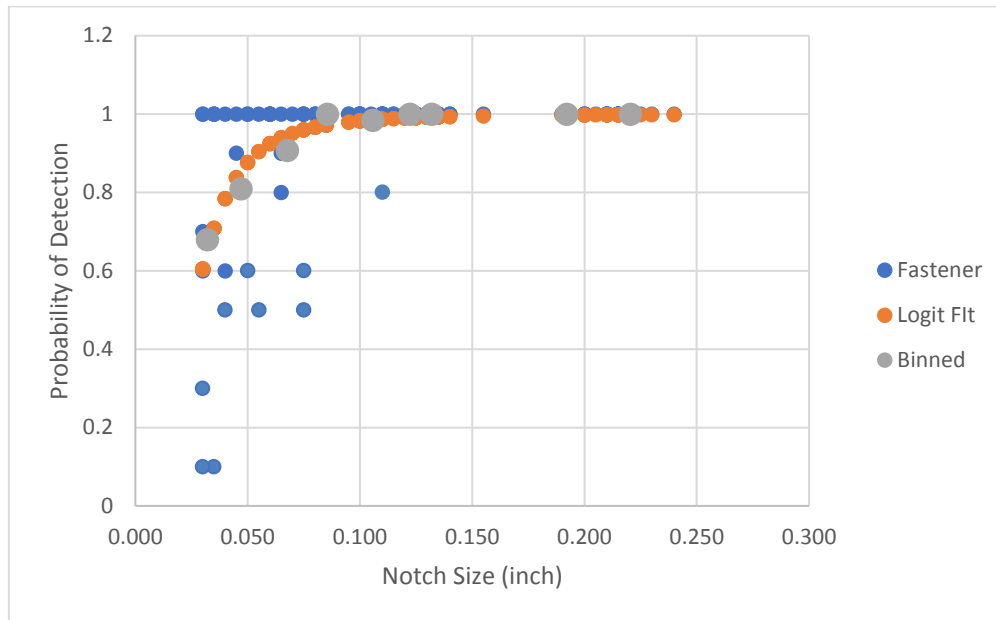
Figure 30: Measured results showing binned notch data following logit fit.

Similar plots were obtained using the simulated results for the NAVAIR samples when the notch data was binned. Hence, the POD analysis produced in this work contains results obtained utilizing the logit link exclusively.

6.4 Distribution of Notch Sizes and Impact on POD Results

The seven NAVAIR samples contained a total of 101 notched fastener sites. A summary of fastener sites by layer for each sample is provided in Table 4.

Table 4: Summary of fastener sites by layer for each NAVAIR sample.

Notch Location / Sample	Bottom of top	Top of Bottom	Both Layers
Std 22	2	10	3
Std 23	4	8	2
Std 24	1	9	4
Std 25	2	10	3
Std 26	4	8	3
Std 28	1	6	5
Std 32	4	10	2
Total	18	61	22

The notch sizes across all samples ranged in size from 0.76 – 6.10 mm (0.03 – 0.24 inches). Transforming the size logarithmically produced POD(a) plots that did not affect the results obtained for the generalized linear model, although the results were presented in this manner to clearly show the characteristic POD 'S' curve with its associated confidence bounds. It was found that models utilizing the log(size) transform, provided the best results when comparing the link models. It is important to note that scaling and transformation of size has no effect on the width of the POD confidence bounds [34].

6.4.1 Narrow Notch Size Range

Annis and Gandossi [34] discuss the influence of range and distribution of notch size on hit/miss POD curves. They revealed that a narrow notch size distribution will widen the confidence bounds that result [34]. An example of this, displaying narrow notch size range and the resulting wide confidence bounds on the POD(a) curve, is shown in Figure 31 [34].
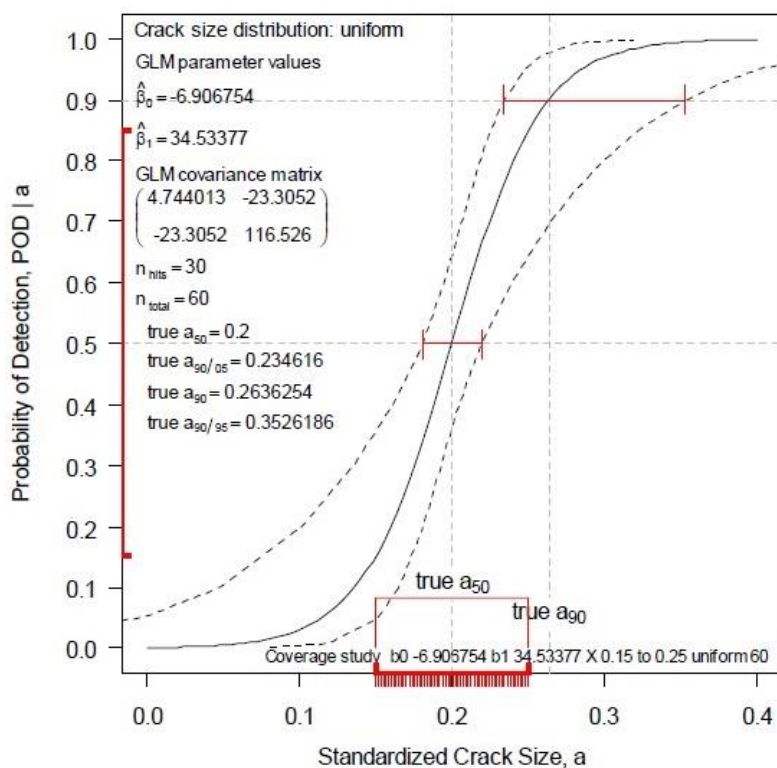
Figure 31: Wide confidence bounds on POD(a) plot due to narrow notch size range [34].

Increasing the width of the specimen range coverage will result in an increase in the range of POD coverage shown on the y-axis, as shown by the solid red vertical line. When the POD extremes, especially at POD > 0.9 are covered, the confidence bounds in those regions become more narrow [34]. Annis and Gandossi [34] determined that the optimum size distribution resulted in a POD coverage of between 0.3 and 0.97 [34].

In the POD analysis for this work, the notch size range and corresponding POD specimen coverage were considered to determine the effect on the confidence bounds and $a_{90/95}$ obtained for the experimental data.

6.4.2 Effect of Mis-located Targets

Another factor that Annis and Gandossi [34] discuss with respect to influencing confidence bounds is associated with the centre of the notch size range. They postulated that when the notch size range is not centred on the true a_{50} of the POD(a) curve, this could adversely affect the width and location of the confidence bounds [34].

When the size distribution is misplaced left, the POD specimen coverage is at the lower range of the POD axis, and results in wider confidence bounds at a_{90} [34]. Conversely, when notch size distribution is displaced right, the POD specimen coverage is at the higher range, resulting in minimal influence on the confidence bounds at a_{90} [34]. Figure 32 shows an example of a POD(a) plot with notch size distribution misplaced left, with the notch size range left of centre [34]. The impact of this irregular notch size distribution increases the width of the confidence bound at a_{90} and can produce a_{90} confidence bounds that are not symmetric.
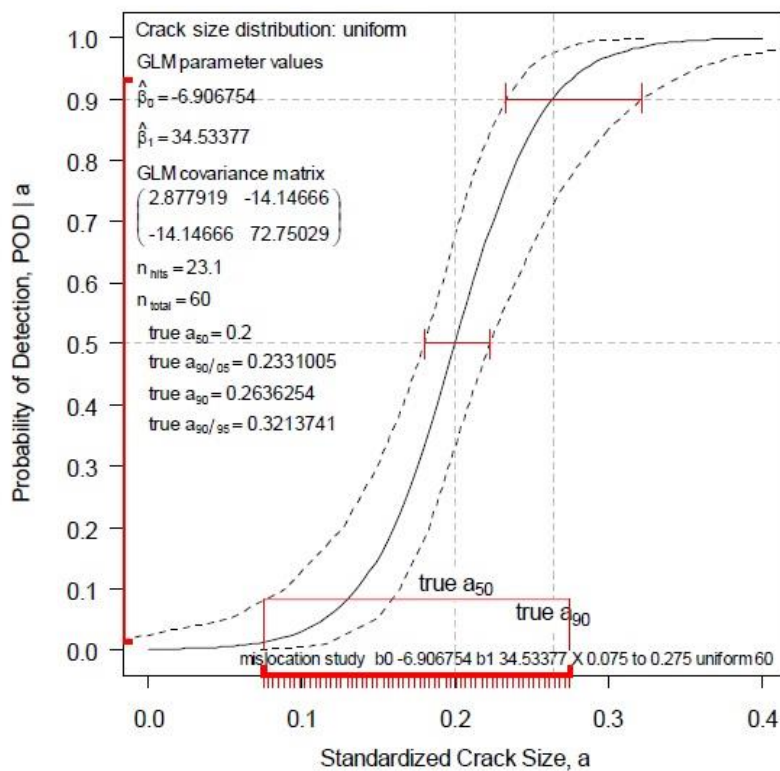
Figure 32: Notch size distribution misplaced left, centred slightly left of a_{50} [34].

The effect of mis-located targets such that the size distribution is not centred on the true a_{50} was considered in this POD analysis.

6.4.3 Non-uniform Size Distributions

Although transformations such as log(size) have no direct effect on the GLM of POD as a function of size, it can indirectly influence the distribution of target sizes with respect to

the true location and shape of the POD vs size plot, which is unknown a priori [34]. This results in smaller specimen POD coverage, as a transformed size can skew the distribution either left or right. When comparing differences in the target size distribution, as shown in Figure 33, a right-skewed distribution of target sizes results in the widest bounds at a_{90} [34].
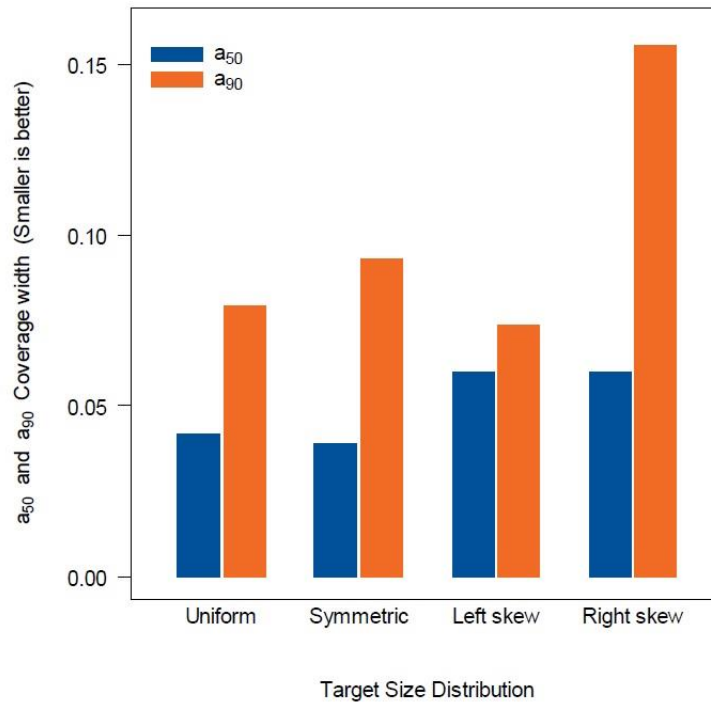
Figure 33: Right-skewed target distribution has widest bounds at a_{90} [34].

Annis and Gandossi [34] reveal that the optimum distribution is left-skewed such that the targets are grouped near a_{90} (Figure 34), as this produces the narrowest bounds at a_{90}, and correspondingly, the smallest $a_{90/95}$.
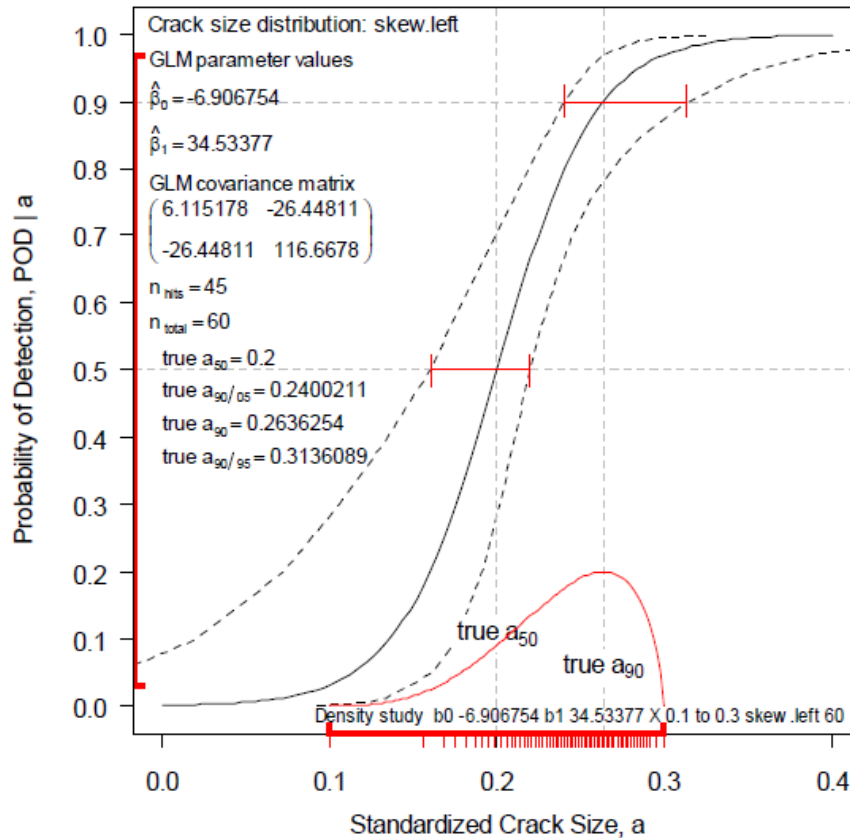
Figure 34: Optimum results obtained for a_{90} and $a_{90/95}$ with left-skewed distribution of target sizes [34].

The actual distribution of target sizes for the NAVAIR samples used in this work were not of uniform distribution, and when considering the results obtained from experimental measurements, were also not centred at a_{50}. As previously shown in Section 6.3, the notches were binned into groupings such that the bins included similar number of notches. This was to emphasize the logit fit to the data. Alternatively, to consider target distribution, Figure 35 displays the actual distribution of target sizes for the NAVAIR samples when the notches were plotted based on uniform crack size intervals.
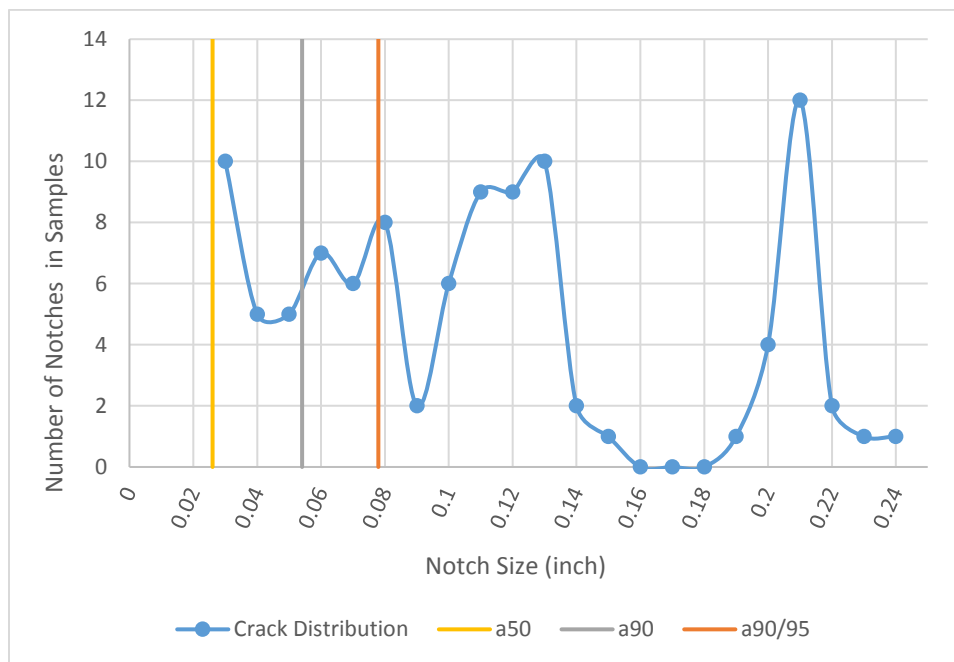
Figure 35: Actual distribution of target sizes for NAVAIR samples based on uniform notch size intervals.

As shown in Figure 35, the actual distribution of target sizes for the NAVAIR samples was not uniform, nor was it centred at a_{50}. This will be considered further in Section 6.5 below. The POD analysis was conducted using the logit model with log(size), as it naturally covers the range from $-\infty$ to $+\infty$. Section 8.5 provides the POD results obtained for both the measured and simulated results, and provides plots detailing the target size distribution and corresponding specimen POD coverage on the POD(a) curve.

6.5 Sample Size

Annis and Gandossi [34] conducted a formalized study on the influence of sample size (N) on hit/miss POD curves. As part of this study they produced guidelines not only for target size distribution and range, but also provided a recommendation with respect to the number of targets for hit/miss POD vs size modelling. They concluded that a decrease in sample size of flaws/discontinuities (targets) results in an increase in the uncertainty in the POD(a) curve with 60 being the recommended minimum number of targets [34]. The guidelines stressed that logistic POD vs size models with fewer than

60 samples are unstable, and can result in non-convergence of the GLM algorithm, issues with credible maximum likelihood parameter estimates and confidence bounds [34]. One of the conclusions in ENIQ report No 47 was that increasing the sample size to 90 improved confidence, but beyond $N = 90$ was not cost-effective [34].

6.5.1 Effect of sample size on confidence bound width

The study conducted at [34] revealed that with less than the recommended 60 hit/miss specimens, very broad confidence bounds resulted. Figure 36 provides a graphical representation of the effect of sample size on confidence bound width [34].
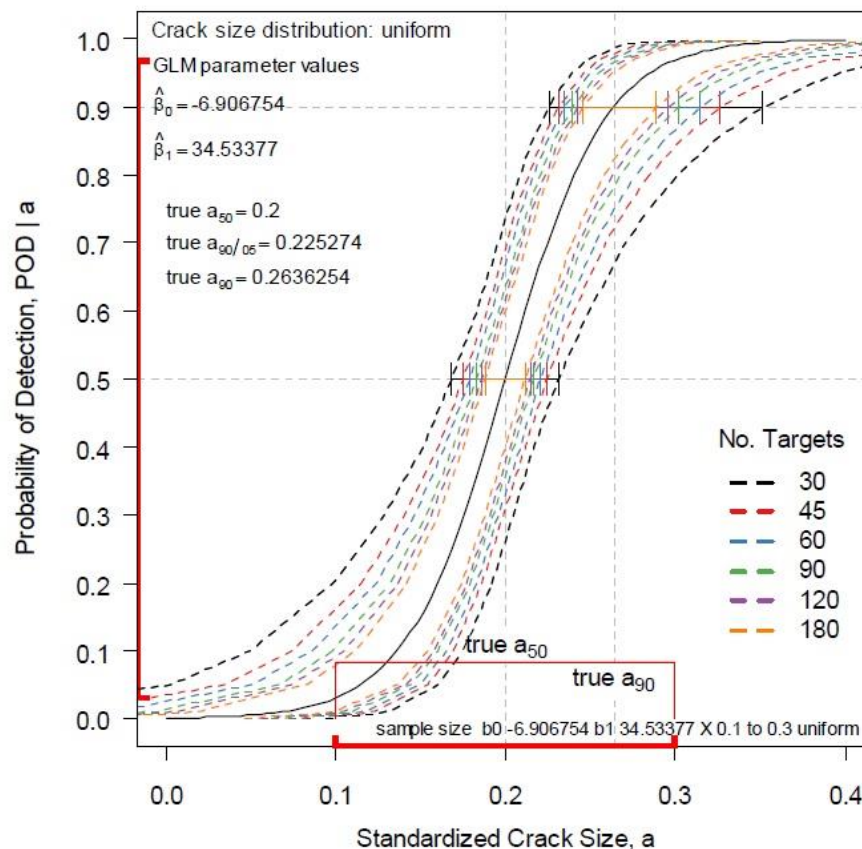
Figure 36: Confidence bounds on POD(a) curves decrease as sample size increases [34].

As seen in Figure 36, increasing the number of samples increases the precision of the confidence bounds. It was shown that the effect of doubling the sample size from $N = 30$ to $N = 60$, results in the confidence interval width at POD = 90% to be halved [34].

In comparing inspections, those that are superior have a higher probability of detection for targets larger than a_{90} and the confidence bound at that crack size is narrower [34].

6.6 Probability of False Positive

An assessment of the ability of an NDE system to discern signal from noise and the determination of the false call rate associated with an inspection is vital in evaluating the effectiveness of the NDE system. It is recommended that a Probability of False Positive (PFP) study be conducted as part of any POD analysis, as false positives and noise can affect the validity of POD results [8][35]. The **mh1823 POD** software provides the tools necessary to conduct the PFP study and the results can then be used to assess detection size ($a_{90/95}$) reliability. The expression for PFP, or false call, is as follows [8]:

$$PFP = 1 - specificity \qquad (7.1)$$

where *specificity* is the proportion of blank fastener sites where no false call is made. With respect to probability, this can be written as:

$$specificity = \frac{True\ blank\ fastener}{True\ blank\ fastener + false\ call} \qquad (7.2)$$

Hence, equation (7.1) can be expressed more generally in statistical form as

$$PFP = Prob(indications|no\ target\ present) \qquad (7.3)$$

Based on the equations shown above, if a sample with 150 blank fastener sites had no false calls during the inspection, the estimated PFP is not zero. The maximum likelihood estimate for the probability of zero false calls in 150 opportunities is zero, but it is not the best estimate for the outcome of the next inspection, as it is biased [8]. Conversely, the PFP is based on the outcome of the next inspection, and for zero false calls in 150 blank fastener sites, the actual PFP in 50 similar NDE tests (PFP$_{50}$) is 0.0046. The PFP$_{50}$ is based on Poisson statistics, and it is referred to as even-bet probability. It gives the expected value of the probability of the false call, such that 50% of the time there will be zero hits in 150 trials and 50% of the time there will be one or more hits. The **mh1823 POD** software provides results for PFP$_{50}$, PFP$_{90}$, and PFP$_{95}$,

that indicate increasing confidence, although recommends using PFP$_{50}$ when reporting results and using this information for component risk calculations [8].

The experimental data obtained for the blank fastener sites for all NAVAIR specimens was input into the **mh1823 POD** software to determine the hit/miss PFP. The results are provided in Section 7.2.5.

6.7 Refining $a_{90/95}$ Measurement

6.7.1 Repeated Measures POD(a) Curves for Simulations

The experimental signal response measurements obtained for the NAVAIR samples were processed using MPCA resulting in five scores for each of the four probe coil pairs. Ten repeat measurements were recorded for each fastener. These scores obtained for both the blank and notched fastener sites were then used in simulations to increase the sample size to a maximum of 80 to improve the precision of the POD confidence bounds.

The simulations accommodated a variation in the data fraction and the number of notches with the goal of encountering each notch approximately 100 times. The minimum number of repeats achieved, when the samples were combined, was 87. Each of these repeat measurements from the simulations were treated as independent inspections and were used in the POD analysis conducted with the **mh1823 POD** software. The software provides a feature that enables the POD(a) curves for all repeat measures to be plotted on the same graph with one set of confidence bounds. As shown in Figure 37 below and in Appendices F and G, in almost all cases for the simulations with 80 samples, all the combined POD(a) plots fit within the confidence bounds.
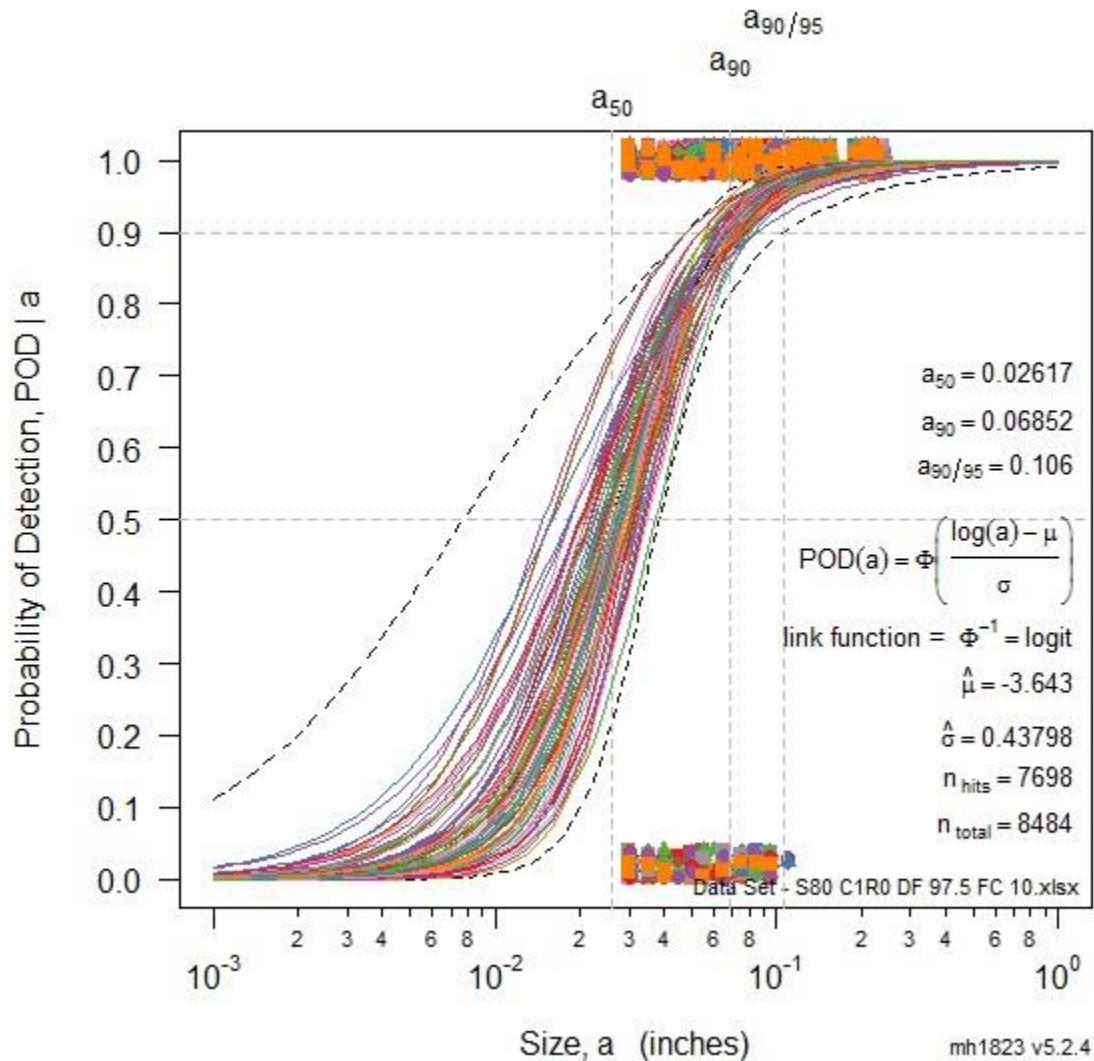
Figure 37: Repeat measures POD(a) plot – Simulation results using 80 samples, with independent POD(a) plots confined within confidence bounds.

This graphical result, obtained by the **mh1823 POD** software, reveals that it is reasonable to group the results and provides validity to the statistical outcomes based on their collective performance [8].

6.7.2 Rationale for Refining $a_{90/95}$ Measurement

Each POD(a) curve in the repeated measures plot of Figure 37 is treated as an individual inspection. The variability between inspections is expected and is evident in Figure 37, as there is a tight grouping of POD(a) curves but there are also a couple of

curves that are spread away from this grouping. The consequence of these curves being positioned away from the tight cluster of POD(a) curves impacts the confidence of the $a_{90/95}$ value.

As discussed previously in Sections 6.4.1 – 6.4.3, notch size range and distribution as well as mis-located targets can significantly affect the confidence bound widths. As discussed in Section 6.4.1 the notch size range for the NAVAIR samples is small, resulting in limited specimen POD coverage. As shown in Figure 37, the notch size distribution is not uniformly distributed nor is it centred on a_{50}. Figure 30 reveals that the data has a logit distribution when the notches are grouped into an evenly distributed number of fasteners, but the factors associated with notch size range and distribution spread the confidence bounds.

Additionally, when the POD results were compared between the experimental measurements and the simulations, the simulations produced slightly lower detection rates. The difference was noted even though the scores associated with the blank and notched fastener sites used in the analysis were identical. This indicates that the simulation results may display a marginal distribution, where unknown variables are influencing the simulation outcomes, or each measurement is not 100% independent of the others. It is expected that these factors can contribute to slightly lower detection rates as shown in the simulation results, when utilizing robust statistics methods.

Irregularities in the grouping of the POD(a) curves representing the repeat measures, the notch size range and distribution, and the slight reduction in detection rate when comparing POD results for both experimental measurements and simulations contributed to spreading the confidence bounds at a_{90}. Most importantly, it is expected that with 80 samples and 5% false call rate, there would be four POD(a) curves outside the confidence bounds. As all repeat measures POD(a) plots are within the confidence bounds, this indicates that there is an unexpected widening of the confidence bounds and an increase in the **mh1823 POD** calculated $a_{90/95}$ value. Consequently, these factors all contributed to, and resulted in an inflated $a_{90/95}$ value. Correspondingly, each of the repeat measures POD(a) plots for the simulations with 80 samples were analyzed with ImageJ software, as introduced in Section 2.6.5.3., to adjust for this spread at a_{90}.

The repeat measures POD(a) plots were loaded into ImageJ and scaled to 300%. If a POD(a) curve was spread away from the grouping at the $POD = 0.9$ line, distances were calculated from the edge of POD(a) grouping to the $a_{90/95}$ line. A distance measurement was also obtained using ImageJ for the distance between the a_{90} and $a_{90/95}$ lines on the POD(a) plot. The known values for a_{90} and $a_{90/95}$ calculated by the **mh1823 POD** software were used to interpolate the ImageJ measurements to determine a refined $a_{90/95}$ value. The complete procedure for $a_{90/95}$ revision utilizing ImageJ software is provided in Appendix M. The results of the refined $a_{90/95}$ values for all the simulations involving 80 samples are provided in a summary table in Section 7.5.5.

7. Results

7.1 General

This section presents the results obtained using the experimental measurements taken for the NAVAIR sample series described in Section 4.5, utilizing the signal analysis methods outlined in Section 5 and the POD analysis approach presented in Section 6.

First, the results obtained for Standard 26 with the revised fastener-to-lap-joint edge distance will be compared to the results obtained using previously defined measurements detailed in Section 5.3.3.

Next, single sample detection results are presented for all seven NAVAIR samples, which were obtained using the experimental measurements described in Section 4. This is followed by the results obtained for four samples, when varying the temperature of the probe for three cases, at 10ºC, 20ºC and 30ºC, respectively. Additional results are also presented on the experimental measurements for three NAVAIR samples, considering off-centering of the probe with respect to the fastener head.

This is followed by simulation detection results, which utilized the SHV algorithm and are presented for the two cases consisting of a sample size of 40 or 80. The results will compare variation in sample size, a data fraction of 95%, 97.5% and 100%, and decision thresholds that were linked to 5%, 10%, and 15% false call rates.

Finally, POD vs. size analysis results are presented for the combined NAVAIR sample series consisting of all seven samples. This analysis will include POD vs. size results for both 40 and 80 samples, the three data fraction percentages and false call rates indicated above, and the revision of the associated $a_{90/95}$ values.

7.2 Measurement Detection Results

7.2.1 Increased Precision of Fastener to Lap-Joint Edge Distance

Using the fastener to lap-joint edge distance measurements obtained with higher resolution images with 0.5 mm scale increments, the detection results for Standard 26,

after edge distance revision were determined. These results were compared to those where the initial edge distances were measured utilizing low resolution images. The analysis utilized the same MPCA scores obtained when single measurements were taken for each fastener with the measurements repeated a total of ten times. The only difference was the edge distance used in the cluster analysis process. The size of the notch with the largest miss, referred to in Table 5, is to within one standard deviation.

Table 5: Comparison of results for NAVAIR Standard 26 with initial and revised edge distances.

Standard 26	Results (Initial Edge Distance)	Results (Revised Edge Distance)
Detection Rate		
1st Layer	72.86%	95.71%
2nd Layer	73.64%	88.18%
Total	70.0%	90.67%
False Call Rate		
Total	7.14%	2.86%
Largest Miss (miss > 10% of Inspections)		
Size (mm)	2.16	1.27

Utilization of the revised fastener to lap-joint edge distances resulted in an increase of 29.4% in the detection rate and decrease of 60.8% in the false call rate. As indicated in Section 5.3.3., as a consequence of these results, higher resolution images were obtained for all NAVAIR sample series specimens (except Standard 22) and the fastener to lap-joint edge distance measurements were recalculated. Consequently, there was an increase in detection rate for all samples where edge distance measurements were obtained from high-resolution images. There was also a decrease in false call rate for all samples except for Standards 23 and 32, which revealed no change.

These results demonstrated the importance of precise edge distance measurements when conducting analysis on the NAVAIR samples using the techniques described in

Section 5, affecting not only detection rate, but ultimately the POD vs. size result. This will be discussed in greater detail in Section 8.1.

7.2.2 Single Sample Detection Results

For the series of seven NAVAIR samples evaluated in this work, each individual sample was evaluated separately. Experimental measurements were taken using the method outlined in Section 4, with one set of data collected for each sample, consisting of single measurements for each fastener repeated a total of ten times. Signal analysis was conducted as described in Section 5, consisting of cluster analysis, including score rotation and distance-to-edge correction. The scores for the blank fasteners from the individual sample being evaluated are used to compute the covariance matrix of the data matrix X, and thus the MD for all the experimental data, including both notches and blanks for the sample being evaluated. A summary of the results obtained for each individual sample in the NAVAIR sample series is provided below in Table 6. For each sample the detection rate is provided for notches located in the top layer, notches in the second layer, as well as the overall detection rate for that sample. The false call rate is also provided for each individual sample. The final column of Table 6 provides the largest notch size that was missed in greater than 10% of the inspections. In two of the samples the largest notch size missed may be slightly larger than the value indicated for that sample. These cases are highlighted with an asterisk as a miss occurred at another fastener site that included both top and second layer notches with a larger sized notch present than displayed in the final column of Table 6. The PEC technique used in this thesis work indicates a hit or a miss at a particular fastener site. When the fastener site contains notches in both layers of the sample, the PEC technique does not discriminate whether the notch hit was related to the top or second layer notch. Therefore, when both top and second layer notches were present at a fastener site location, this particular fastener site was not included when second layer detection rate was assessed. The averages across all samples are also presented to provide an overview of the effectiveness of the inspection technique for the NAVAIR sample series.

The results, as presented in Table 6, demonstrate reproducibility and reliability of the PEC inspection technique across all seven NAVAIR sample series, and will be discussed in greater detail in Section 8.1.

Table 6: Summary of NAVAIR individual sample detection results.

Sample ID	Detection Rate			False Call Rate	Largest Notch Size Miss (mm/inch)
	Top Layer	2nd Layer	Total		
Standard 22	100.00%	100.00%	100.00%	3.75%	N/A
Standard 23	85.0%	98.0%	92.1%	6.0%	0.89 / 0.035
Standard 24	100.00%	100.00%	100.00%	2.50%	N/A
Standard 25	100.00%	96.92%	97.33%	5.71%	0.76 / 0.03
Standard 26	95.71%	88.18%	90.67%	2.86%	1.27 / 0.05 *
Standard 28	83.33%	84.55%	85.83%	2.00%	1.02 / 0.04 *
Standard 32	91.67%	88.33%	88.13%	6.25%	1.91 / 0.075
Average	93.67%	93.71%	93.44%	4.15%	1.17 / 0.046
* A miss occurred at another fastener site with both top and second layer notches with a larger sized notch present. Unable to decipher which notch missed.					

7.2.3 Variation in Probe Temperature

The temperature study for the three NAVAIR samples consisted of obtaining experimental measurements of the samples utilizing the procedure detailed in Section 5.3.1. This technique, including probe positioning is the same as that used for all other experimental measurements in this thesis work, with the only difference being the variation in the temperature of the probe. Three sets of experimental measurements, with each set consisting of ten measurements per fastener, were obtained for each sample with the temperature of the probe being set at 10ºC, 20ºC and 30ºC, respectively. The detection rate obtained for each of the four samples at the different probe temperatures is displayed in Figure 38. The associated false call rates associated with these measurements are presented in Figure 39.
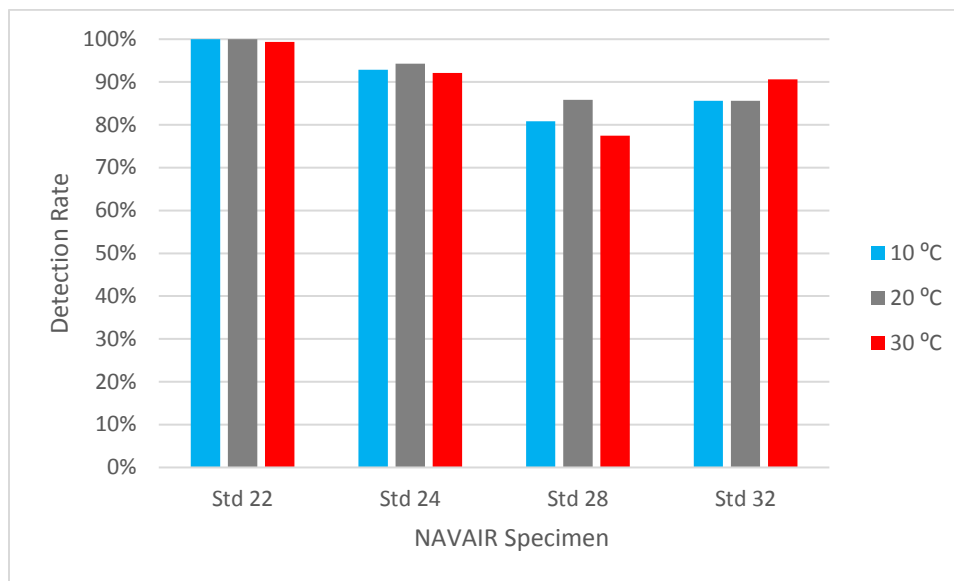
Figure 38: Effect of variation in probe temperature on detection rate for four NAVAIR samples.

The variation in detection rate was less than one standard deviation for all four samples when compared to results obtained when the probe was under normal temperature and environmental conditions. The variation ranged from 0.001% - 0.176%, with Standard 28 experiencing the largest reduction in detection rate at 30ºC.

In contrast with the effect of probe temperature on detection rate results, the false call rate increased significantly, not only across the three temperatures, but in comparison to results obtained under normal operating and environmental conditions. The variation in false call rates ranged from 0.016 % - 0.163 %, with Standard 28 again experiencing the largest variation in false call rate across all temperatures. When false call rates for Standard 28 were compared to those under normal operating conditions, the rate increased five-fold when measurements were taken with the probe temperature at 30ºC.
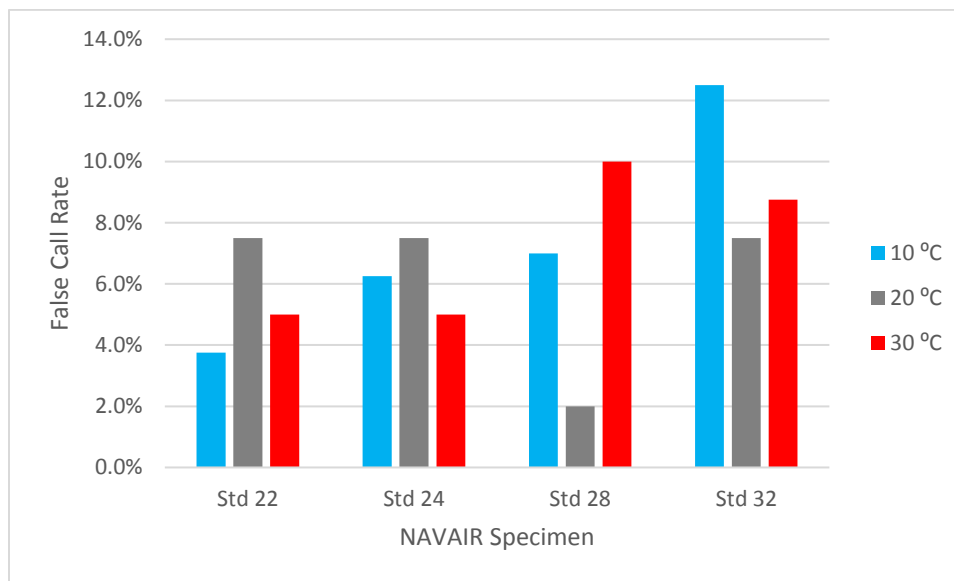
Figure 39: Effect of variation in probe temperature on false call rate for four NAVAIR samples.

The results presented in Figures 38 and 39 present the effects associated with variation in probe temperature on detection and false call rates and will be discussed in greater detail in Section 8.5.2.1.

7.2.4 Probe Off-Centering

As described in Section 5.3.2., measurements were taken of three NAVAIR samples (Standards 22, 24, and 28) to observe the results obtained when the probe was placed off-centre of the fastener head by up to 2 mm in distance. When processing the experimental data for the off-centered signal measurements, the reproduction of the input data in the MPCA analysis using the eigenvectors, resulted in mean square error values in the order 10^{-6} or 10^{-7}. Consequently, this revealed that the eigenvectors appropriately spanned the space being tested and the reproduction was adequate. The results including detection and false call rates as well as the largest notch size missed at a specific fastener in greater than 10% of inspections, are provided in Table 7 below. These results provide for comparison of the results obtained for the same NAVAIR standard where the probe is placed concentrically over the fastener head.

Table 7: Comparison of results obtained for three NAVAIR samples with and without probe off-centering technique.

Sample ID	Detection Rate			False Call Rate	Largest Notch Size Miss (mm/inch)
	Top Layer	2nd Layer	Total		
Standard 22	100.0%	100.0%	100.0%	3.75%	N/A
Std 22 - Off Centre	96.0%	93.8%	94.7%	11.3%	1.52 / 0.06
Standard 24	100.0%	100.0%	100.0%	2.50%	N/A
Std 24 - Off Centre	100.0%	79.2%	80.7%	10.0%	2.16 / 0.085
Standard 28	83.3%	84.5%	85.8%	2.00%	1.02 / 0.04 *
Std 28 - Off Centre	83.3%	81.8%	83.3%	8.00%	1.02 / 0.04 *
Average	94.44%	94.85%	95.28%	2.75%	0.339 / 0.013
Average - Off Centre	93.1%	85.0%	86.2%	9.8%	1.57 / 0.062

* Another miss occurred at another fastener site with both 1st and second layer notches with a larger notch size present. Unable to decipher which notch missed.

The results reveal that the detection rate for top layer notches remained relatively constant, while a significant reduction in detection rate was realized for second layer notches with respect to the overall detection rate for all samples. There was also a significant increase in the false call rate.

Scores s_1 and s_2 are most sensitive to centering of the probe, and since Standard 24 showed the largest variation with respect to detection and false call rate, plots were constructed of s_1 vs s_2 for the off-centred and centred results.
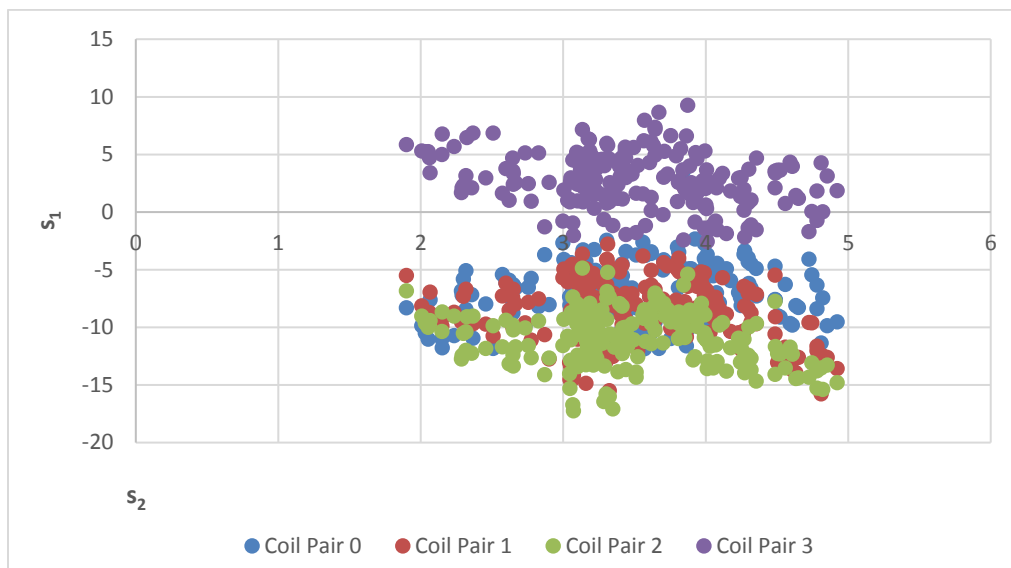
Figure 40: Plot of rotated scores s_1 vs. s_2 for NAVAIR Standard 24 with probe centered on fastener head.
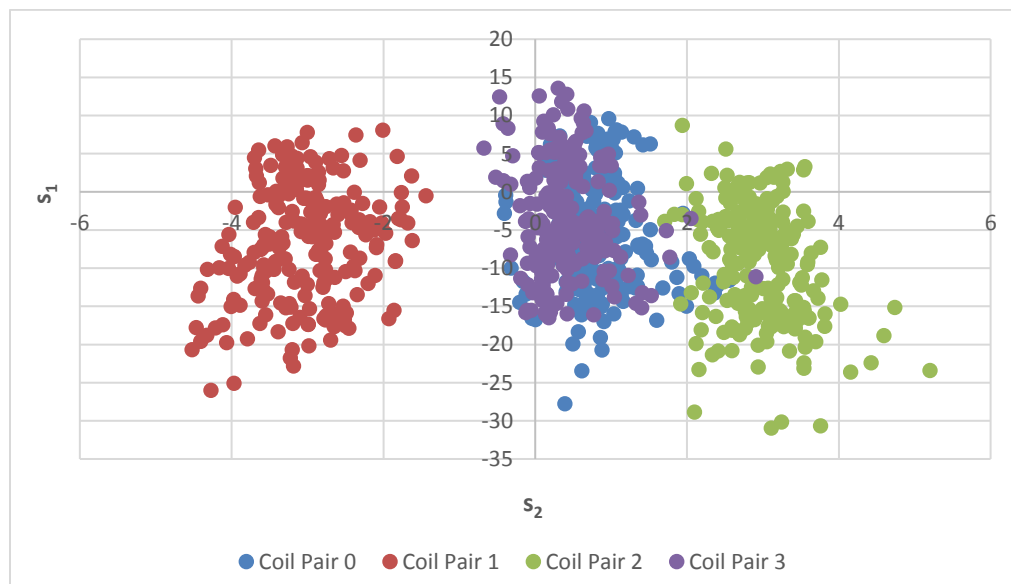
Figure 41: Plot of rotated scores s_1 vs. s_2 for NAVAIR Standard 24 with probe off-centered on fastener head.

The plots presented in Figures 40 and 41 reveal that when the probe is off-centered that most of the variation is along s_2 instead of s_1. The rotated scores of s_1 vs s_2 should

show the most variation along s_1, but there is more spread in the data along s_2 for the off-centered results.

The results presented in Table 7, and Figures 40 and 41 reveal the impact of probe off-centering on detection and false call rates and will be discussed in greater detail in Section 8.5.2.2.

7.2.5 Probability of False Positive

As indicated in Section 6.6, in evaluating the effectiveness of the NDE system, an assessment of the ability of an NDE system to discern signal from noise and the determination of the false call rate associated with an inspection is vital.   The experimental data obtained for the blank fastener sites for all individual NAVAIR specimens was input into the **mh1823 POD** software to determine the hit/miss PFP. The results are provided in Table 8, and although PFP_{50}, PFP_{90}, and PFP_{95}, in order of increasing confidence, are displayed, PFP_{50} is highlighted as the metric to assess results.   As previously indicated, the PFP is based on the outcome of the next inspection, and for even-bet probability, the actual PFP in 50 similar NDE tests ranged from 0.033 to 0.071.   These results can be used to assess detection size reliability ($a_{90/95}$) and will be discussed in greater detail in Section 8.5.

Table 8: False call and PFP results for NAVAIR sample series.

Sample ID	Standard 22	Standard 23	Standard 24	Standard 25	Standard 26	Standard 28	Standard 32
False Calls	3	6	2	4	2	2	5
Total Blanks	80	100	80	70	70	100	80
FC Rate	3.75%	6.00%	2.50%	5.71%	2.86%	2.00%	6.25%
Probability of False Positive (PFP)							
PFP (50)	0.0457	0.0665	0.0333	0.0664	0.038	0.0267	0.0706
PFP (90)	0.0816	0.1029	0.0651	0.111	0.0743	0.0523	0.1129
PFP (95)	0.0941	0.115	0.0766	0.126	0.0872	0.0616	0.1269

7.3 Simulation Detection Results

Randomly generated statistical data was generated using the simulation technique described by Underhill *et al.* [33][20], and outlined in Section 5.5, and numerous statistical simulations were conducted. As indicated in Section 5.5, the randomly generated data was derived from actual experimental data collected on the NAVAIR samples used in the analysis of the single sample detection case outlined above in Section 7.2.2. In these simulations, the data was processed in the same manner as described for the analysis of the samples using strictly experimental measurements. This post-processing of the data included removing the effects due to environmental conditions such as off-centering, repeat measurements, and fastener to lap-joint edge distance correction techniques as outlined in Section 5.3. This section will provide an overview of the simulation results obtained for single samples using known blanks, considering several cases consisting of variation in sample size, data fraction and false call rate.

7.3.1 Results Obtained for Single Sample with Known Blanks

Simulation results were obtained for each individual NAVAIR sample, where blank fasteners were randomly selected utilizing the bootstrap method, to produce the data matrix X, which was then used to determine the MD. As described in Section 2.5.5.2, blank fastener scores s_1 to s_5 were generated from real blank experimental data. Utilizing the bootstrap method with the SHV algorithm, randomly selected measurement data for both notches and blanks were used to generate test data sets that were blank fastener intensive and thus, representative of in-service samples.

In each simulation, a total of between 1 and 4 randomly selected notches from real experimental measurement data were used. The number of notches and blanks used in each simulation totalled 40, hence the number of randomly selected blanks ranged from 36-39 depending on the number of notches used. With the goal of seeing each notch 100 times on average, and each NAVAIR sample having between 12 and 16 notches, a simulation of one notch and 39 blanks, would require a maximum of 1600 runs. As detailed in Section 5.5.3, the specific cut-off value was set to generate a false call rate of 5%.

Simulations were conducted where the number of notches varied from one to four and the data fraction was set to 95%, 97.5% and 100%. The data fraction of 100% provided the best detection rate when only one notch was present. Results are shown in Table 9 and are compared to the measured experimental results for the individual NAVAIR sample.

Table 9: Comparison of experimental and simulation detection results for 40 samples, cut-off for 5% false call rate and data fraction set to 100%.

Sample ID		Measured Results	Simulation Results Data Fraction 100%
Standard 22	Detection	100.0%	96.8%
	False Call	3.8%	3.3%
Standard 23	Detection	92.1%	77.3%
	False Call	6.0%	5.4%
Standard 24	Detection	100.0%	82.6%
	False Call	2.5%	3.7%
Standard 25	Detection	97.3%	91.2%
	False Call	5.7%	4.1%
Standard 26	Detection	90.7%	68.7%
	False Call	2.9%	5.2%
Standard 28	Detection	85.8%	72.5%
	False Call	2.0%	3.4%
Standard 32	Detection	88.1%	78.6%
	False Call	6.3%	4.0%

The simulations conducted with a sample size of 40 produced detection results that were significantly lower for many of the NAVAIR specimens than those obtained using the experimental measurement data. It was shown through simulations that both the detection rate and false call rate decreased as the number of notches increased in the simulation from one to four.

7.3.2 Variation in Sample Grouping Size

The detection results obtained for simulations with a sample grouping size of 40 were not comparable to those obtained using experimental measurements. Hence simulations were conducted using the SHV algorithm with a sample size of 80, to assess the

influence of sample size on detection rate. A comparison of the simulation results obtained for grouping sizes of both 40 and 80 samples was conducted for all NAVAIR samples with 5% false call rate, 100% data fraction and the number of notches ranging from one to four. The comparison results for three of the NAVAIR samples showing the increase in detection rate as sample grouping size increases from 40 to 80 is shown in Figure 42.
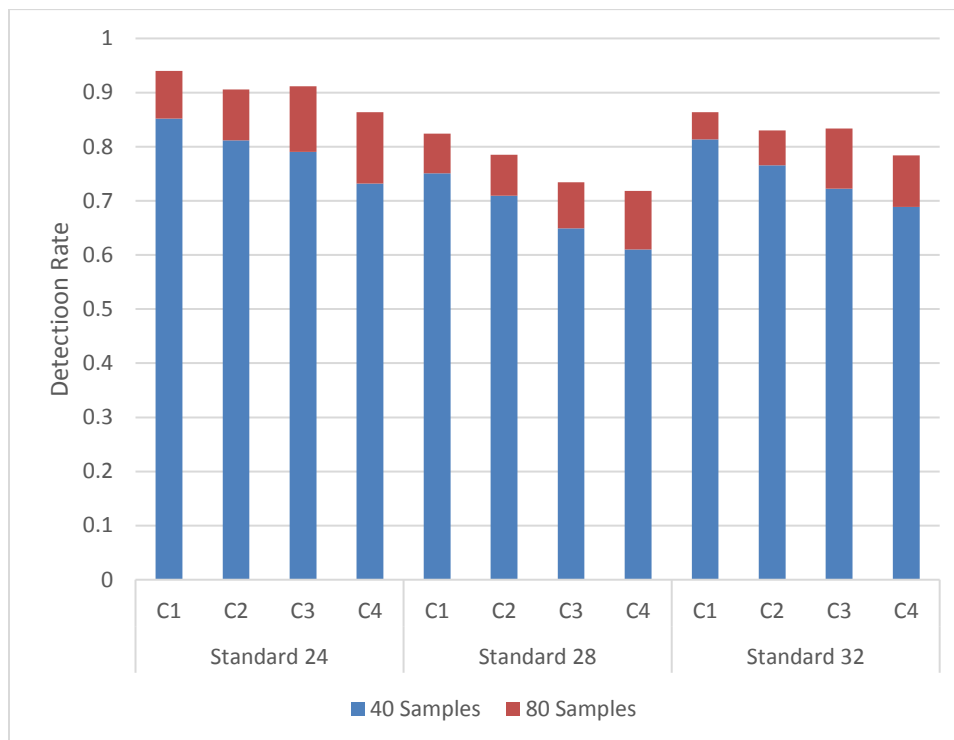
Figure 42: Simulation results for three NAVAIR samples showing increase in detection rate as sample grouping size is increased from 40 to 80 samples. Simulation specifications are constant at 5% false call rate and 100% data fraction across number of notches ranging from one to four.

The comparison of the results obtained when varying the sample size reveals that the detection rate increased when the sample size was increased, with an insignificant increase in false call rate.

For the simulations conducted with a sample size of 80 and the false call rate set to 5%, the best results were obtained when the data fraction was set to 100%. As shown in Table 10 below, for the simulations conducted with a sample grouping size of 80, as the

number of notches increased from one to four, the false call rate was reduced further below 5%. As more notches are added to the simulation the covariance matrix and standard deviation get larger and consequently, the false calls decrease.

Table 10: Summary of simulation detection results for notches and blanks totalling 80 samples, cut-off for 5% false call rate and data fraction set to 100%.

Sample ID		Number of Notches in Simulation			
		C1	C2	C3	C4
Standard 22	Detection	99.5%	97.2%	95.9%	92.7%
	False Call	3.9%	3.4%	3.1%	2.7%
Standard 23	Detection	84.1%	79.5%	77.8%	74.2%
	False Call	6.6%	6.1%	5.5%	5.3%
Standard 24	Detection	94.0%	90.6%	91.1%	86.4%
	False Call	4.0%	3.6%	3.2%	2.8%
Standard 25	Detection	96.7%	95.9%	92.7%	89.9%
	False Call	5.1%	4.5%	3.7%	3.1%
Standard 26	Detection	75.4%	72.3%	69.1%	66.2%
	False Call	6.5%	6.2%	5.7%	5.5%
Standard 28	Detection	82.4%	78.5%	73.4%	71.8%
	False Call	3.8%	3.4%	2.9%	2.6%
Standard 32	Detection	86.4%	83.0%	83.3%	78.4%
	False Call	4.9%	4.2%	3.6%	3.2%

Not only does the false call rate decrease as the number of notches in the simulation increases, but detection rate also decreases. The simulations conducted with a grouping size of 80 samples produced results that were comparable to the measured experimental results for the NAVAIR samples, thus validating the simulation methods.

The results presented in Tables 9 and 10 and Figure 42 reveal the effects of sample grouping size utilized in simulations involving cluster analysis, SHV, and bootstrap methodologies, with respect to detection rate. The influence of sample grouping size on detection rate will be discussed further in Section 8.2.

7.3.3 Variation in False Call Rate in Simulations

One of the advantages of using pulsed eddy current as opposed to conventional eddy current techniques for inspections is that fasteners do not need to be removed to conduct the inspection. Alternatively, conventional eddy current, or BHEC inspections, involves removing all the fasteners prior to inspection. With respect to BHEC, detection rates are generally associated with a false call rate of 5%.

The simulations described in Section 8.3.1 that were set up for a 5% false call rate, produced actual false call rates that ranged from 3.6 – 6.2% across all data fractions. Additionally, considering the PEC technique presented in this thesis work involves only removing those fasteners as having been inspected and identified as having a notch present, it is recommended that a higher false call rate be considered. Hence, simulations were conducted such that the specific cut-off value, using the technique detailed in Section 6.5.3, was set to generate a false call rate of 10%. The detection results obtained for simulations with a sample size of 80 and false call rate set to 10% are presented in Table 11.

Table 11: Summary of simulation detection results for notches and blanks totalling 80 samples, cut-off for 10% false call rate and data fraction set to 97.5%.

Sample ID		Number of Notches in Simulation			
		C1	C2	C3	C4
Standard 22	Detection	99.6%	99.4%	99.1%	99.0%
	False Call	8.5%	7.2%	6.2%	5.6%
Standard 23	Detection	85.9%	85.1%	85.1%	82.0%
	False Call	10.6%	10.0%	9.2%	8.5%
Standard 24	Detection	96.1%	95.7%	94.2%	93.5%
	False Call	7.7%	6.8%	6.1%	5.3%
Standard 25	Detection	97.9%	97.1%	97.3%	96.5%
	False Call	9.3%	8.2%	7.4%	6.5%
Standard 26	Detection	80.8%	80.3%	77.8%	74.7%
	False Call	9.4%	8.7%	8.2%	7.7%
Standard 28	Detection	84.6%	82.7%	79.7%	82.5%
	False Call	8.1%	7.0%	6.2%	5.6%
Standard 32	Detection	88.2%	87.4%	87.0%	89.1%
	False Call	9.3%	8.3%	7.3%	6.3%

The actual average false call rate obtained for simulations across all data fractions was 7.7%. The best detection rate achieved for simulations set up for 10% false call rate was obtained when utilizing a data fraction of 97.5%. The simulation results using this data fraction are presented in Table 11. The detection results obtained for this setup are consistent with those for the experimentally measured results presented in Table 9. The simulation results obtained for all NAVAIR samples with both 95% and 100% data fraction and a 10% false call rate are provided in Appendix E.

7.3.4 Variation in Data Fraction

All simulation results at 10% false call rate were compared to determine the data fraction that produced the best detection results. The average detection rate across all NAVAIR samples was calculated for results obtained for one, two, three and four notches introduced into the simulation. These average values, obtained for each number of notches present, were compared with regards to data fraction and are presented in Figure 43.
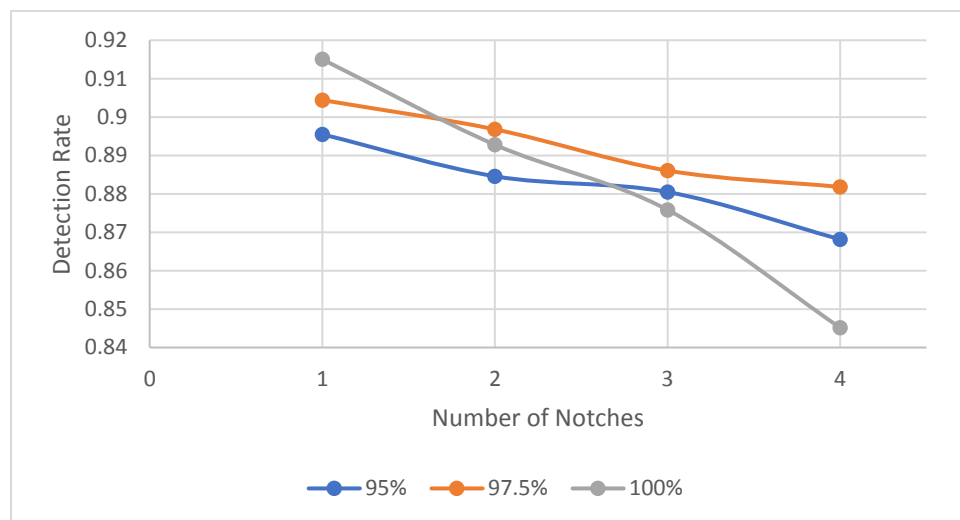
Figure 43: Comparison of simulation results obtained for 95%, 97.5% and 100% data fractions with 10% false call rate and one to four notches present.

Although simulations with 97.5% data fraction result in a slightly lower detection rate compared to 100%, when only one notch is selected for the simulation, this data fraction provides the best results overall, when considering both detection and false call rates. The results presented in Figure 43 provide confirmation that utilizing a data fraction of

97.5% in simulations involving cluster analysis, SHV, and bootstrap methodologies provides the best performance overall with respect to detection rate. The influence of data fraction will be discussed in greater detail with respect to POD vs. size results in Section 7.5.2.

7.4 Comparison of Measured and Simulated Results

When comparing the measured and simulation results obtained for each NAVAIR sample, five of the 101 notches displayed significantly lower detection rate, as shown in Table 12. The results for the simulations are presented for two data fractions, both at 10% false call rate.

Table 12: Comparison of measured and simulation results showing significant disparity in detection rate at specific fasteners.

Sample ID	Fastener Number	Bottom of Top	Top of Bottom	Size mm (+/- 0.38)	Orientation CW from Lap-Joint Edge	Distance from Lap-Joint Edge (mm)	Detection Rate Measured Result	Detection Rate Sim - Data Fraction 97.5%	Detection Rate Sim - Data Fraction 100%
Standard 23	24	X	X	1.14/2.41	315° / 135°	6.46	100%	18.8%	30.5%
Standard 24	17		X	1.65	270°	7.34	100%	61.3%	64.6%
Standard 26	2	X		1.65	135°	7.83	90%	14.7%	9.8%
Standard 26	9		X	1.02	270°	5.7	100%	57.4%	47.6%
Standard 26	10		X	2.16	45°	6.61	100%	57.8%	58.5%

The detection rates obtained for the remaining fasteners were all comparable to within less than 10%, when the results at all data fractions were assessed. The five fasteners identified in the summary table above revealed no obvious pattern with respect to notch orientation, size, location, or fastener to lap-joint edge distance. An image of fasteners nine and ten on Standard 26 is shown in Figure 44 below. Compared to almost all other samples typical of that shown in Figure 26, Standard 26 contained filler and other material extruding from the lap joint, which was also evident at fastener 2 of this sample, suggesting that gap variation may have been present.
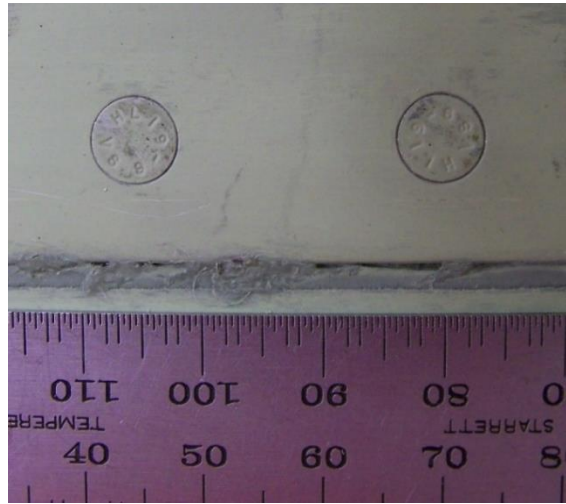
Figure 44: Image of fasteners nine (left) and ten (right) on NAVAIR Standard 26 showing material extruding from lap joint.

Surface damage was noted on and around fastener 24 of NAVAIR Standard 23, which is shown in Figure 45.

Figure 45: Image of fastener 24 on NAVAIR Standard 23 showing surface damage.

Considering that only 5 out of 101 total notched fastener sites showed inconsistencies between measured and simulation detection results, it proved difficult to identify a specific cause for this behaviour. This will be discussed in greater detail in Section 8.5.1.

7.5 POD Results

This section presents the POD vs. size results that were obtained for the seven NAVAIR samples, for both the measured and simulation scenarios. The influence of sample size on the $a_{90/95}$ value will be presented using results obtained for measured results, where sample size represents the number of notches present. Additionally, the influence of sample size grouping on the $a_{90/95}$ value will be discussed as it relates to simulations involving groupings of 40 and 80 samples. This will be followed by an overview of the variation in actual false call rate obtained during the analysis. Results will also be presented showing POD vs. size plots displaying the POD specimen coverage to display the effects of notch size range and distribution. Finally, the revised $a_{90/95}$ values will be presented for simulations conducted with 80 samples, with variation in data fraction and decision threshold linked to 5%, 10%, and 15% false call rate, respectively.

7.5.1 Measured Results

In conducting the POD analysis of the experimental measurements for all seven NAVAIR samples combined, the **mh1823** POD software produced eight possible POD vs. size models. There were slight differences (less than 1% variation) in the deviance for each of the models presented, and although smaller deviance is better, as indicated in Section 2.6.5, it was shown that this variation in deviance had no effect on the null deviance. After comparison of the eight models, it was apparent that the logit function had the best fit to the data, as predicted in Section 6.3. Transforming the target size logarithmically also provided the best results for all link models when modelling the data.

The POD vs. size results, obtained using the experimental measurements for all seven NAVAIR samples with ten repeat measurements on each fastener are shown in Table 13.

Table 13: POD vs. size results obtained for all NAVAIR samples combined using only ten repeat experimental measurements

Measured Results	All Samples	2nd Layer Notches Only
a_{50} (mm/inch)	0.66 / 0.026	0.55 / 0.22
a_{90} (mm/inch)	1.37 / 0.054	1.23 / 0.048
$a_{90/95}$ (mm/inch)	1.98 / 0.078	1.93 / 0.076

As shown in Table 13, the $a_{90/95}$ values were smaller when only considering the second layer notches. The detection rate achieved when considering all samples was 93.6%, whereas second layer detection was 94%.

The POD vs. size results presented in Table 13 demonstrate reliability and robustness of the PEC inspection technique across all seven NAVAIR sample series. The POD vs. size values obtained for the measured results will be discussed in greater detail in Section 8.4.1.

7.5.1.1 Influence of Sample Size on $a_{90/95}$

As identified in Section 5.5.2, Annis and Gandossi [34] recommend that for hit/miss POD vs. size modelling, the sample size or minimum number of targets is 60. The total number of notches present in all seven NAVAIR samples consisted of 101 top and 2nd layer notches. Table 13 above and Figure 46 below, give valid and acceptable $a_{90/95}$ values and stable POD vs. size plots, respectively.
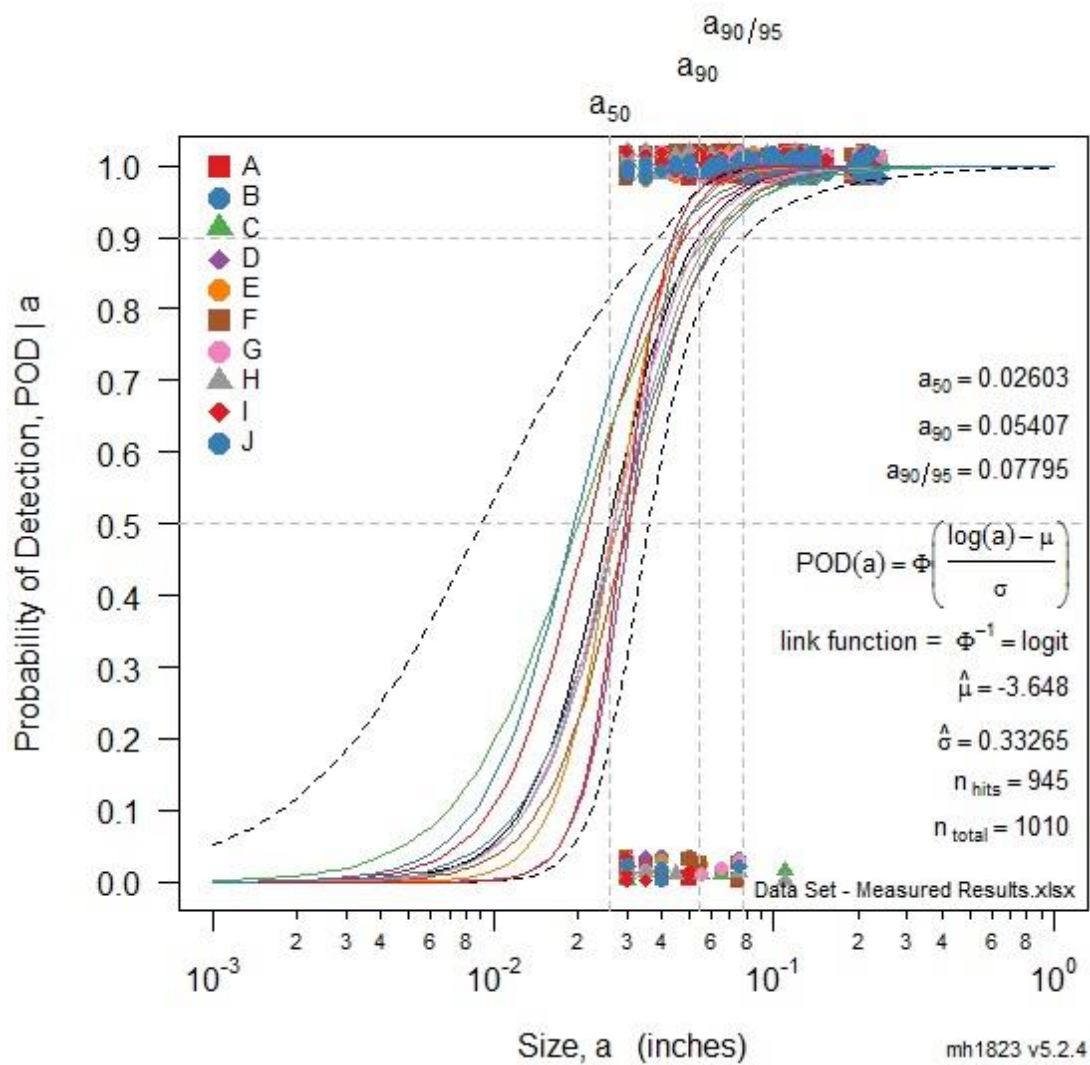
Figure 46: POD vs. size plot for all NAVAIR samples with ten repeat measurements.

Similarly, the results obtained when considering only the 83 second layer notches in all seven NAVAIR samples produced almost identical results as shown in Table 13, with a POD vs. size plot comparable to that of Figure 46. Conversely, when analyzing only the 40 first layer notches in all seven NAVAIR samples, a valid $a_{90/95}$ value was unattainable and the POD vs. size plot revealed instabilities and irregularities, as shown below in Figure 47.
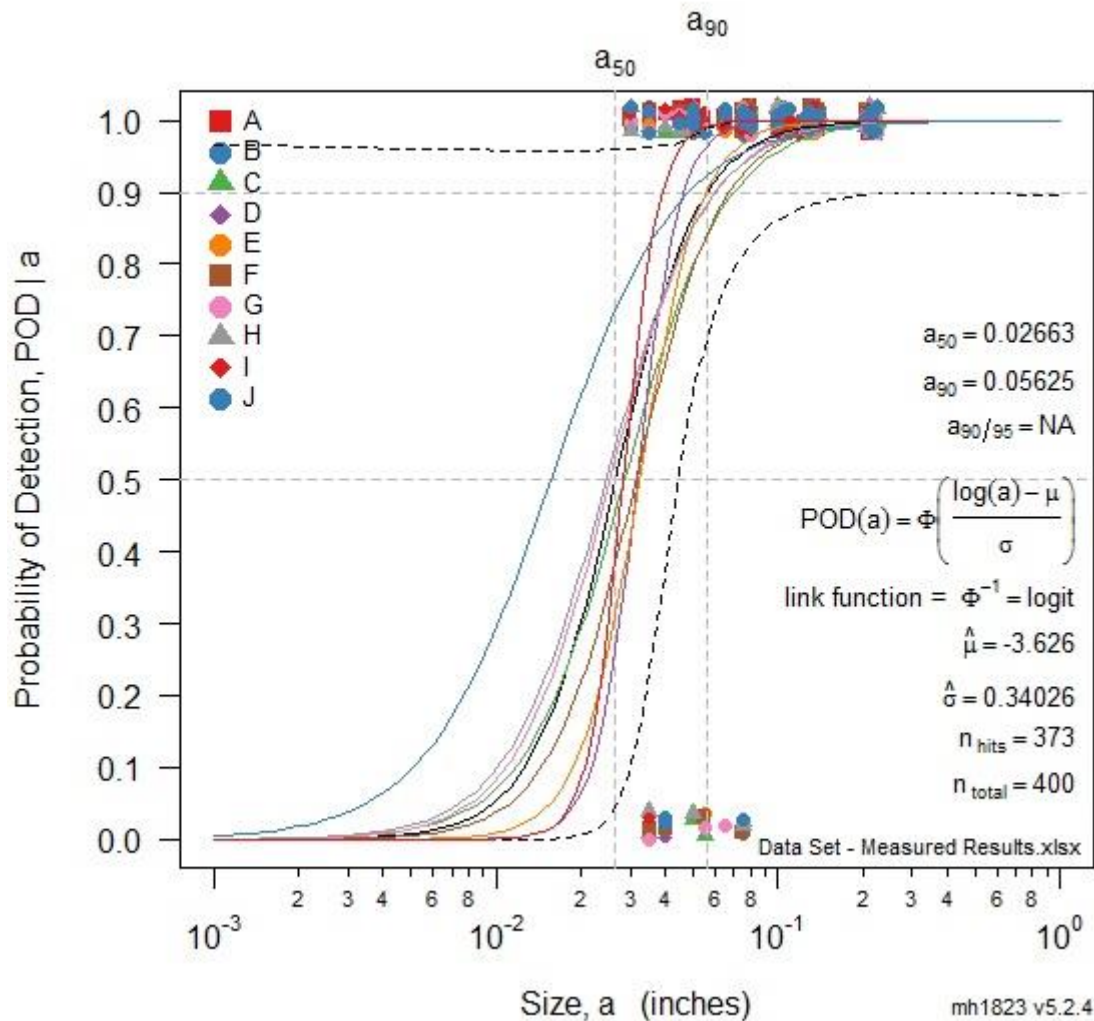
Figure 47: POD vs. size plot of measured results obtained for all seven NAVAIR samples when considering first layer notches only.

The results presented in Figures 46 and 47 reveal the effects of sample grouping size on the $a_{90/95}$ value when analyzing experimental measurement data. The influence of sample size on the $a_{90/95}$ value will be discussed in greater detail in Section 8.2.1.

7.5.2 Simulation Results

This section presents the POD vs. size results obtained for the simulations conducted using 40 or 80 samples, a data fraction of 95%, 97.5% and 100% and 5%, 10% and

15% false call rate, as presented in Section 7.3. The $a_{90/95}$ values will be presented for results obtained using the **mh1823 POD** software for all simulations conducted. Results comparing the influence of sample size will also be presented. Next, an overview of the impact of the variation of false call rate on POD vs size results will be provided. Following that, the impact of target range and notch size distribution will be examined through the specimen POD coverage. Finally, the $a_{90/95}$ values will be presented for results obtained using ImageJ software.

7.5.2.1 Influence of Sample Grouping Size on $a_{90/95}$

The $a_{90/95}$ values obtained for simulations with grouping sizes of 40 and 80 samples were assessed by comparing the same percentage of notches in the total samples used for the simulation. Figure 48 reveals the comparison results across all three data fractions. It shows that the simulations conducted with 80 samples produced $a_{90/95}$ values that were significantly lower than those obtained for simulations utilizing a sample size of 40. This was consistent across all data fractions. Not only were the $a_{90/95}$ values higher for the simulations with 40 samples, but when the false call rate and percentage of the notches in the total sample size was set to 5%, the $a_{90/95}$ value approached the largest notch present in the NAVAIR sample series. In simulations conducted with 40 samples, where the number of notches was between five and 10 percent of the total sample size, there where $a_{90/95}$ values larger than the largest notch size encountered in the specimens.
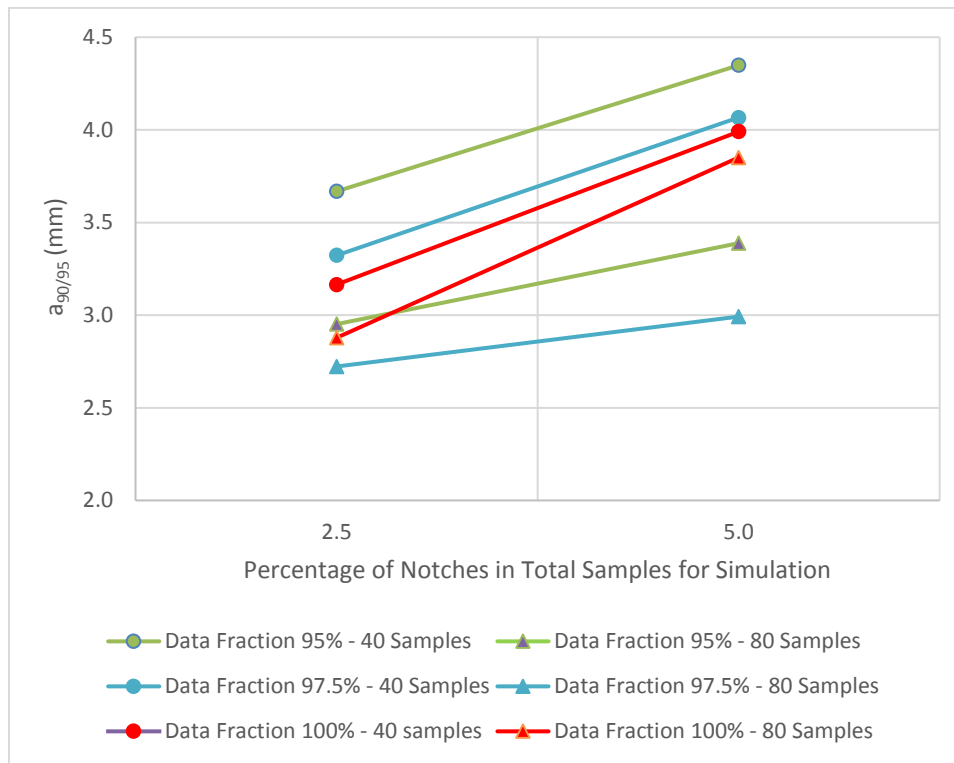
Figure 48: Summary of $a_{90/95}$ value obtained for simulations with grouping sizes of 40 and 80 samples comparing same percentage of notches in the total samples for the simulation by data fraction.

The POD vs. size plots showing repeat measurements were compared for simulations with both 40 and 80 sample grouping sizes. The plots for 40 samples revealed non-symmetric bounds at a_{90}, as shown in Figure 49. The repeat measurement plots were grouped such that they were skewed to the left of the confidence bounds at a_{90}, which is the most important reference line used in the calculation of $a_{90/95}$ [34]. In comparison, the symmetry of the confidence bounds at a_{90}, for the simulations conducted with a sample size of 80, showed greater symmetry and consequently, produced $a_{90/95}$ values with less variance.
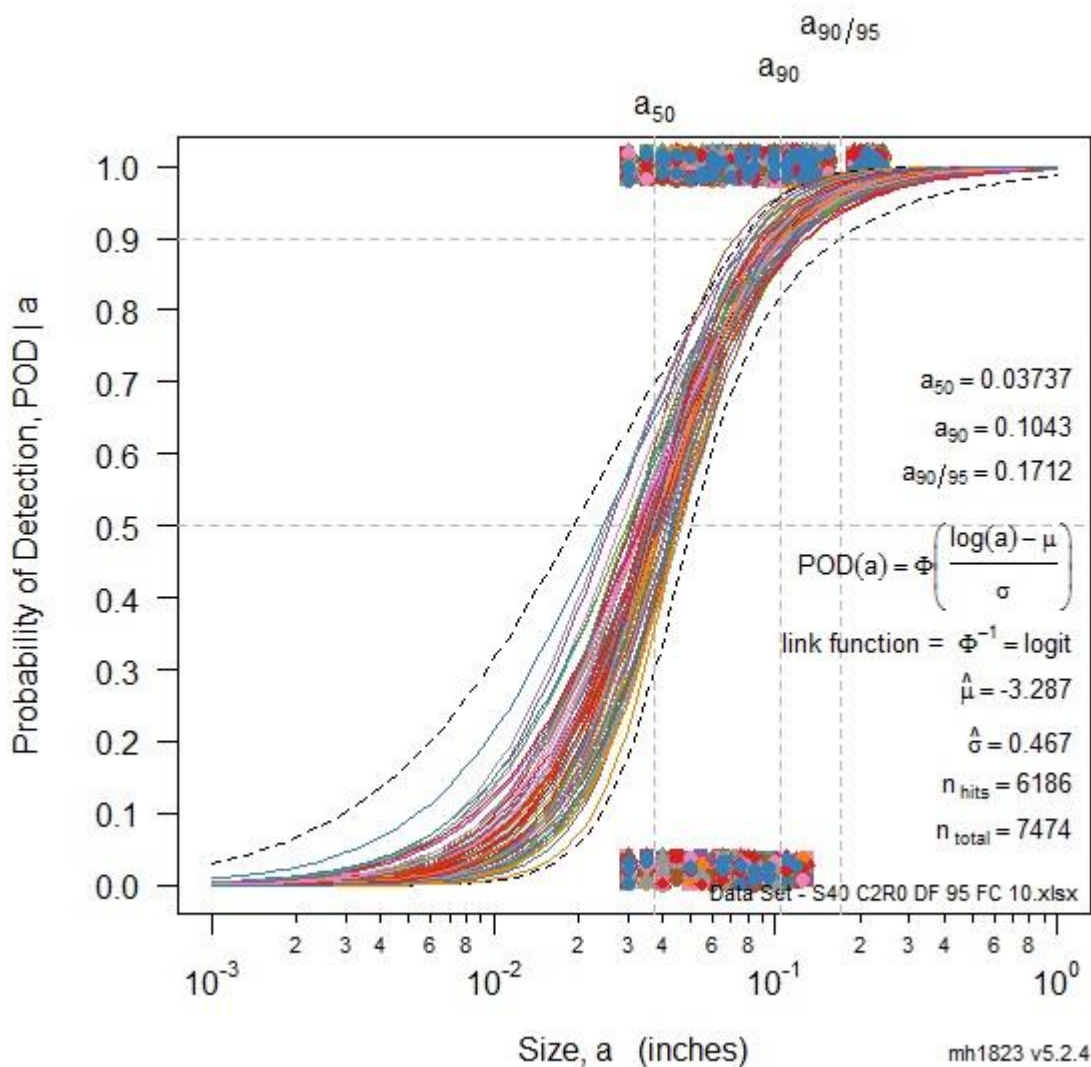
Figure 49: Non-symmetrical bounds at a_{90} for simulation with sample grouping size of 40, data fraction 95% and false call rate of 10%.

The simulations conducted with 80 samples produced POD vs. size curves that were more stable and $a_{90/95}$ values that showed less variance. Hence, the remainder of the POD analysis in this section focuses on the results obtained for simulations with a sample size of 80. A summary of the $a_{90/95}$ values obtained using the **mh1823 POD** software for simulations utilizing 80 samples, varying the false call and data fraction with the inclusion of up to four notches is presented in Table 14.

Table 14: Summary of $a_{90/95}$ values using mh1823 POD software for simulations utilizing grouping size of 80 samples and varying the false call and data fraction with the inclusion of up to four notches.

		Number of Notches in Simulation			
		C1	C2	C3	C4
False Call Rate	Data Fraction	$a_{90/95}$ (mm/inch) - **mh1823 POD** software			
5% False Call	95%	3.47 / 0.1367	3.64 / 0.144	3.83 / 0.151	4.29 / 0.169
	97.5%	3.38 / 0.133	3.55 / 0.14	3.74 / 0.147	4.07 / 0.160
	100%	3.04 / 0.120	3.56 / 0.140	4.02 / 0.158	4.87 / 0.192
10% False Call	95%	2.90 / 0.114	2.95 / 0.116	3.15 / 0.124	3.39 / 0.133
	97.5%	2.69 / 0.106	2.72 / 0.107	2.96 / 0.117	2.99 / 0.118
	100%	2.54 / 0.1	2.88 / 0.113	3.36 / 0.132	3.85 / 0.152
15% False Call	95%	2.45 / 0.096	2.51 / 0.099	2.65 / 0.104	2.79 / 0.110
	97.5%	2.34 / 0.092	2.32 / 0.091	2.57 / 0.101	2.62 / 0.103
	100%	2.24 / 0.088	2.53 / 0.10	2.95 / 0.116	3.43 / 0.135

The results presented in Figures 48 and 49, and Table 14, reveal the effects of sample grouping size utilized in simulations involving cluster analysis, SHV, and bootstrap methodologies, on the $a_{90/95}$ value. The influence of sample grouping size on the $a_{90/95}$ value will be discussed in greater detail in Section 8.2.

7.5.2.2 Influence of Data Fraction on $a_{90/95}$

When considering the results for 10% false call rate, amongst 95%, 97.5% and 100% data fractions, the lowest $a_{90/95}$ value (2.54 mm / 0.10 inch) was obtained for a simulation with one notch present and a data fraction of 100%. When comparing the $a_{90/95}$ results for simulations with 10% false call rate and up to four notches, a data fraction of 97.5% provided better results overall, as shown in Table 14 and Figure 50.
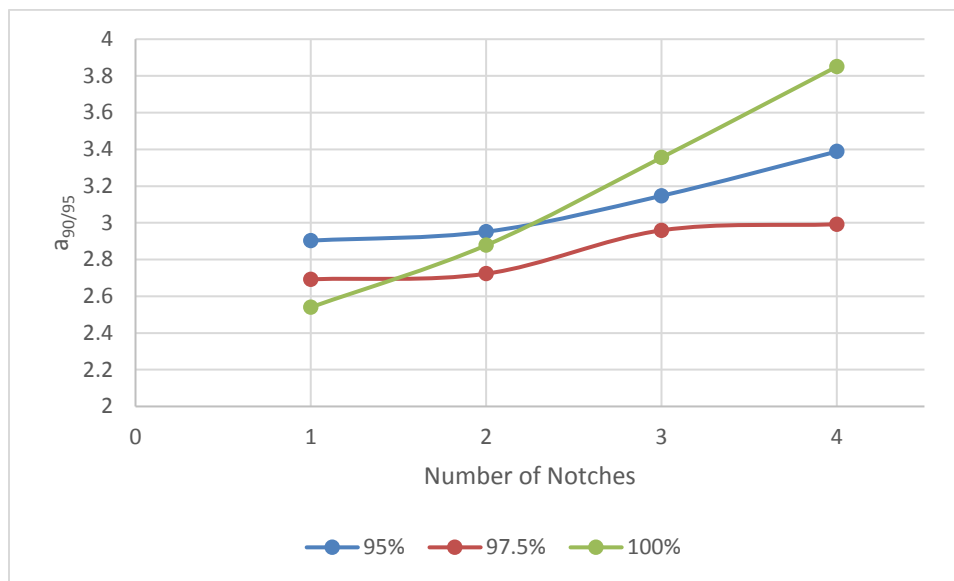
Figure 50: Summary of $a_{90/95}$ values represented by data fraction, using mh1823 POD software for simulations utilizing 80 samples, 10% false call rate, and the inclusion of one to four notches.

It is important to note that the data fraction of 97.5% provided the best results overall across all false call rates.

The detailed results for the simulation with 80 samples, 10% false call rate and a data fraction of 97.5% showing values for a_{50}, a_{90}, and $a_{90/95}$ are provided in Table 15 below.

Table 15: POD vs. size results for simulations with 80 samples, 10% false call rate and 97.5% data fraction

	Data Fraction 97.5%			
No. of Notches	C1	C2	C3	C4
a_{50} (mm/inch)	0.66 / 0.026	0.68 / 0.027	0.77 / 0.030	0.75 / 0.030
a_{90} (mm/inch)	1.74 / 0.069	1.76 / 0.069	1.93 / 0.076	1.94 / 0.076
$a_{90/95}$ (mm/inch)	2.69 / 0.106	2.72 / 0.107	2.96 / 0.117	2.99 / 0.118

The POD vs. size results for the remaining simulations using a sample grouping size of 80, false call rate variation between 5%, 10%, and 15%, and data fraction variation between 95%, 97.5% and 100% are provided in Appendix F. The POD vs. size repeat

measurement plots for simulations with a sample grouping size of 80 and false call rate set at 15% and 10% are provided in Appendices G and H respectively.

The results presented in Figure 50 and Table 15 reveal that a 97.5% data fraction utilized in simulations involving cluster analysis, SHV, and bootstrap methodologies, provides the best overall performance with respect to the $a_{90/95}$ value. The influence of data fraction on the $a_{90/95}$ value will be discussed in greater detail in Section 8.3.

7.5.3 Variation in False Call Rate

The simulations described in Section 7.3.1 that were set up for a 10% false call rate, produced actual average false call rates ranging from 6.0 – 9.6% across all data fractions (Figure 51). The maximum and minimum false call rates obtained for simulations with the number of notches ranging from one to four is also provided in Figure 51. As the number of notches increases from one to four, the false call rate decreases. For simulations set up with 80 samples and 5%, 10%, and 15% false call rate, the actual average false call rate achieved across data fractions was 4.6%, 7.7%, and 10.9%, respectively. A summary of actual false call rate results obtained across all data fractions where the number of notches increases from one to four is provided in Appendix I.
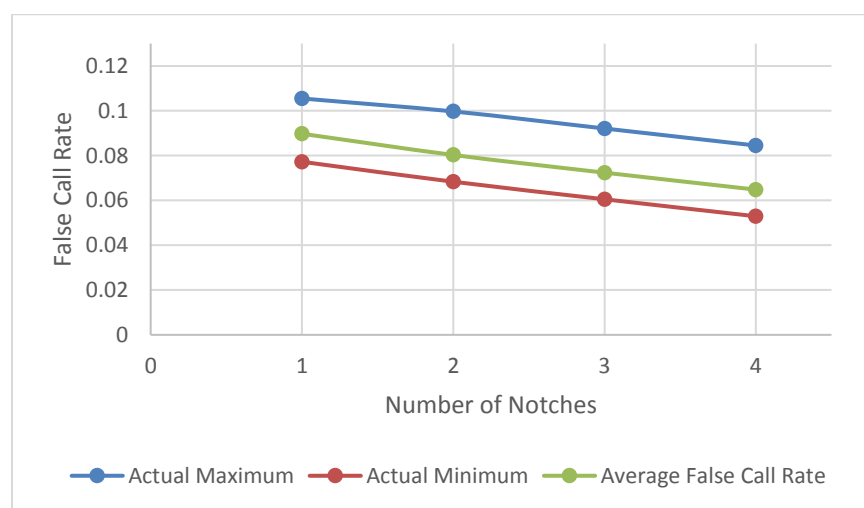
Figure 51: Actual maximum, minimum and average false call rates obtained for simulations set up with 80 samples, data fraction 97.5% and false call rate 10%, with the number of notches ranging from one to four.

The impact of false call rates on $a_{90/95}$ values is shown in Figure 52 for simulations with 80 samples, 10% false call rate, and the data fraction set to 95%, 97.5% and 100%. For each data fraction in the figure, the points from left to right represent notches increasing from one to four in the total sample set. As revealed in Figure 52, as the false call rate is decreased, the $a_{90/95}$ increases. The same outcome is realized with simulations where the false call rate is set to 5% and 15%. Additionally, it is apparent from Figure 52 that a data fraction of 97.5% provides the best overall results with respect to minimizing $a_{90/95}$ values and false call rates.
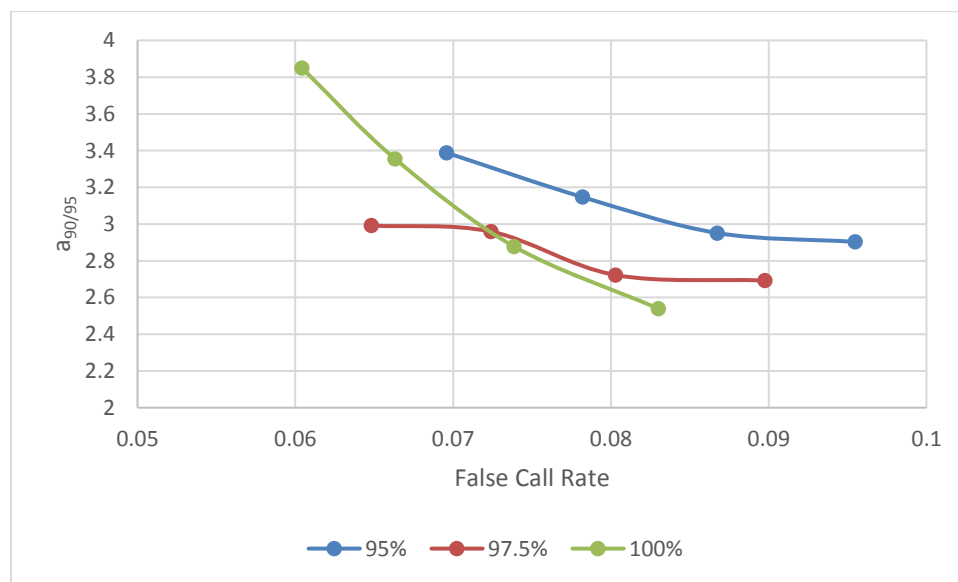
Figure 52: Plot of $a_{90/95}$ vs. false call rate with variation by data fraction, obtained for simulations set to 10% false call rate with sample grouping size of 80 and one to four notches in total samples.

The false call rate results will be discussed in greater detail in Sections 8.3 and 8.8.

7.5.4 Specimen POD Coverage

As discussed previously in Sections 6.4.1 – 6.4.3, notch size range and distribution, as well as mis-located targets can significantly affect the confidence bound widths. As shown in Figure 35 (Section 6.4.3), the actual distribution of target sizes for the NAVAIR samples was not uniform, nor was it centred at a_{50}.

Of the 101 notched fastener sites, 15% of the notches were below 1.27 mm (0.050"), 28% between 1.27 and 2.54 mm (0.050 – 0.10"), and the remainder at or above 2.54 mm (0.10"). There was also a gap in notch sizes where no cracks were present with sizes between 3.8 to 4.8 mm (0.15 – 0.19"). The actual distribution of target sizes produced what appeared to be two distributions of notch sizes, both left and right-skewed. The overall notch distribution was misplaced right of the a_{50} reference line.

For simulations conducted with a sample grouping size of 80, as the percentage of notches increased, the specimen POD coverage increased, and the notch distribution was centered more towards a_{50}. The data fraction of 95% produced the largest specimen POD coverage across all simulations. The notch size range for the NAVAIR samples is small, which resulted in limited specimen POD coverage for all the simulations conducted. A plot of POD vs. size for an 80-sample simulation with data fraction set to 97.5% and 10% false call rate that shows the limited specimen POD coverage is provided in Figure 53. Additionally, the notch size distribution is misplaced right of a_{50}, which has less impact on the confidence bound width at a_{90} than if the notch size distribution was misplaced left. The results associated with specimen POD coverage will be discussed in greater detail in Section 8.4.
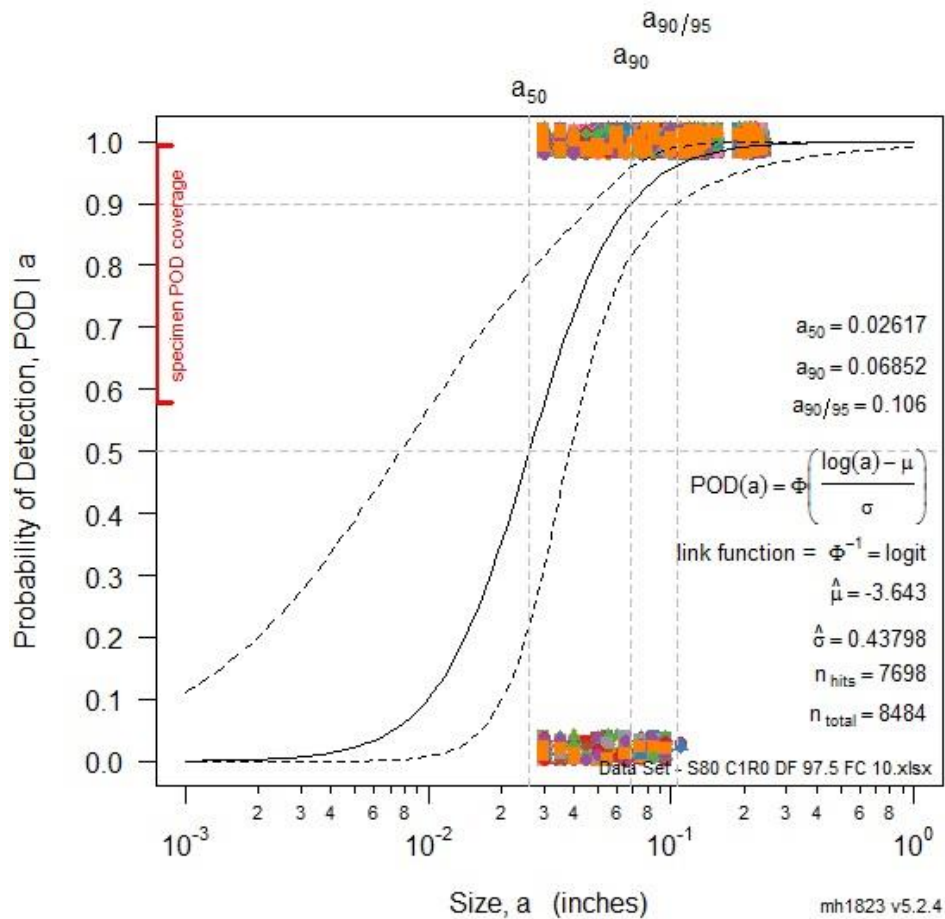
Figure 53: POD vs. size plot for an 80-sample simulation with 97.5% data fraction, 10% false call rate and one notch present, displaying limited specimen POD coverage and notch size distribution misplaced right of a_{50}.

The remaining POD vs. size plots showing specimen POD coverage for simulations conducted with 80 samples, a data fraction set to 97.5% and 10% false call rate with variation in number of notches within a given simulation are provided in Appendix J.

7.5.5 Revised $a_{90/95}$ using ImageJ Software

The actual distribution of target sizes and the POD vs. size results displayed in Sections 7.5.1 to 7.5.4 reveal various factors that contribute to increased width of confidence bounds at a_{90}. The decreased notch size range and non-symmetric notch distribution as well as the limited specimen POD coverage adversely impacts the width of the confidence bounds, and ultimately the $a_{90/95}$ value.

The POD(a) curves in the repeated measures plot of Figure 49 show the impact of these factors where non-symmetric bounds at a_{90} are evident, which then influences the $a_{90/95}$ result. Each POD(a) curve in the repeated measures plot is treated as an individual inspection and variability between inspections is expected. As shown in Figure 37, there is a distinct tight grouping of POD(a) curves, but there are also a couple of curves that are spread away from this grouping. The consequence of these curves being positioned away from the tight cluster of POD(a) curves also impacts the confidence associated with the $a_{90/95}$ value.

Additionally, when experimental measurement and simulations POD results were compared, the simulations produced lower detection rates and larger $a_{90/95}$ values across all false call rates. As shown in Table 16, the $a_{90/95}$ obtained for simulations at all data fractions was larger not only when analyzing all samples, but also when assessing second layer notches only. Although the measured results are for 5% false call rate, the $a_{90/95}$ is still less than the results achieved with simulations at 5%, 10% and 15% false call rate.

Table 16: Comparison of $a_{90/95}$ values obtained for measured results and simulation results at 10% false call rate when analyzing all samples and second layer notches only.

		All Samples	2nd Layer Notches Only
a$_{90/95}$ (mm/inch)	Measured Results	1.98 / 0.078	1.93 / 0.076
	Simulation - Data Fraction 95%	2.90 / 0.114	3.29 / 0.129
	Simulation - Data Fraction 97.5%	2.69 / 0.106	2.88 / 0.113
	Simulation - Data Fraction 100%	2.54 / 0.1	2.61 / 0.103

The difference was noted even though the scores associated with the blank and notched fastener sites used in the analysis were identical. This indicates that the simulation results may display a marginal distribution, suggesting that there are unknown variables that are influencing the simulation outcomes.

Irregularities in the grouping of the POD(a) curves represent repeat measures, the notch size range and distribution and slight reduction in detection rate, in both experimental and simulation PODs, have contributed to spreading of the confidence bounds at a_{90}. Consequently, these factors have all contributed to a gap in the grouping of the POD(a) curves, increasing the $a_{90/95}$ value, and ultimately its precision and confidence. At 5% false call rate, it is expected that 4 out of 80 POD(a) curves should exceed the $a_{90/95}$ value at 90% detection rate. The irregular spread in the confidence bounds at a_{90} and gap detected in the curve grouping can be accounted for by measuring the empirical $a_{90/95}$ directly from the graph. Correspondingly, each of the repeat measured POD(a) plots, for the simulations with 80 samples, were analyzed with ImageJ to reduce this spread at a_{90}. A summary of $a_{90/95}$ values obtained using ImageJ software on the simulation results with a sample size of 80, for all variations in false call rates and data fractions is shown in Table 17.

Table 17: Summary of $a_{90/95}$ values using ImageJ software for simulations varying the false call and data fraction with the inclusion of up to four notches.

| | | Number of Notches in Simulation | | | |
		C1	C2	C3	C4
False Call Rate	Data Fraction	$a_{90/95}$ (mm/inch) - Using ImageJ			
	95%	2.72 / 0.107	2.74 / 0.108	3.00 / 0.118	3.10 / 0.122
5% False Call	97.5%	2.60 / 0.102	2.67 / 0.105	2.90 / 0.114	2.98 / 0.117
	100%	2.43 / 0.095	2.75 / 0.108	3.24 / 0.128	3.53 / 0.139
	95%	2.22 / 0.088	2.33 / 0.092	2.35 / 0.093	2.66 / 0.105
10% False Call	97.5%	2.07 / 0.082	2.26 / 0.089	2.33 / 0.092	2.29 / 0.090
	100%	2.04 / 0.080	2.28 / 0.090	2.59 / 0.102	2.90 / 0.114
	95%	1.93 / 0.076	2.05 / 0.081	2.10 / 0.083	2.11 / 0.083
15% False Call	97.5%	1.91 / 0.075	2.02 / 0.080	1.96 / 0.077	2.08 / 0.082
	100%	1.71 / 0.067	1.93 / 0.076	2.28 / 0.090	2.56 / 0.101

Additional rationale for using ImageJ software and the corresponding $a_{90/95}$ results that are obtained will be discussed in greater detail in Section 8.9.

8. Discussion

The purpose of this section is to discuss the significance of the results obtained from evaluation of the seven NAVAIR sample series specimens. The impact of precision with respect to fastener to lap-joint edge distance measurements on detection rate is presented. This is followed by a discussion on the impact of sample size, notch range and distribution, and environmental factors on the POD vs. size and $a_{90/95}$ results obtained for simulations utilizing cluster analysis and SHV methodologies. Following this, the variation observed between measured and simulation results as well as specimen characteristics are discussed to emphasize that there are material parameters and simulation algorithm behaviours that require further study. Then, a discussion of the reliability of the PEC inspection technique for this application is presented, followed by the rationale for considering a 10% false call rate to be associated with this technique. Finally, the $a_{90/95}$ results obtained using both the **mh1823 POD** and ImageJ software packages are compared, and a discussion with respect to determining an $a_{90/95}$ value that is to be associated with this inspection technique is presented.

8.1 Distance to Lap-Joint Edge

From the results presented in Section 7.2.1, when the fastener to lap-joint edge distance measurements were revised to reflect greater precision, the results improved significantly for Standard 26. The results obtained for all NAVAIR samples, showing before and after edge distance revision are provided in Appendix D.   The increased precision in fastener to lap-joint edge distance resulted in an increase in detection rate for all samples and a significant decrease in false call rate except for Standards 23 and 32. The detection rate for top layer notches increased from 87.7 to 93.7%, second layer detection rate increased from 86.8 to 93.7% and overall the average detection rate for the NAVAIR sample series increased from 86.0 to 93.4%. The false call rate also showed improvement, as it dropped from 5.5 to 4.2%. It is apparent from these results of the importance of precise fastener to lap-joint edge distances, when using the PEC methodologies presented in this work.

These results demonstrate that the fastener to lap-joint edge distance has a direct relationship with detection and false call rates. Previous analysis by Butt *et al.* [12][36][37] showed that desirable detection results from Standards 22, 24, and 25 could be obtained, but that apparent detection problems with the remaining NAVAIR Standards were present. The fastener-to-edge distance observed by Butt [12] varied by +/- 1.8 mm across all seven samples, whereas in this research the value was +/- 1.94 mm. A comparison of the results obtained in this and previous research [12] for all NAVAIR samples is presented in Table 18.

Table 18: Comparison of single sample detection results from previous research [12] utilizing original fastener to lap-joint edge distance and detection results obtained in this thesis work with edge distance revision.

Sample ID	Detection Rate		False Call Rate
	Results Previous Research [12] at 5% False Call Rate	Results Edge Distance Revision	
Standard 22	99.0%	100.0%	3.8%
Standard 23	81.0%	92.1%	6.0%
Standard 24	95.0%	100.0%	2.5%
Standard 25	95.0%	97.3%	5.7%
Standard 26	87.0%	90.7%	2.9%
Standard 28	81.0%	85.8%	2.0%
Standard 32	84.0%	88.1%	6.3%
Average	89.0%	93.4%	4.2%

As identified by Butt *et al.* [12][36][37], the loss of detection for Standard 28 was attributed to the fact that the majority of the fasteners were placed closer to the lap-joint edge than other samples by an average of 1 mm. With the revision of the edge distances in this thesis work, the fasteners were still closer to the lap-joint edge than all other samples by an average of 0.96 mm. The closeness of the fastener to the lap-joint edge intensified the edge effect, which could impact detection rates. Additionally, as also noted by Butt [12], the Standard 28 sample had the largest differences in the fastener-to-edge distance for a single sample, which remained the case for the revised edge distances measured as 3.00 mm in this thesis work. Consequently, these

differences contributed to the blank fastener cluster being more spread out, resulting in more notches being contained within the blank cluster boundary and lower detection rates. Although the fastener-to-lap-joint edge distance revision completed in this thesis work improved the results from 81.0% detection rate and 5% false call rate to 85.8% and 2%, respectively, this sample still produced the lowest detection results.

Increasing the precision on the fastener-to-lap-joint edge distances not only increased the detection rate to above 90%, while reducing the false call rate below 5%, but seemed to improve previously identified inconsistencies in detectability [12]. The results demonstrated the importance of precise edge distance measurements when using the methodologies inherent in this PEC technique, as it affected not only detection rate but also the $a_{90/95}$ value. The results presented in Table 18 also demonstrate reproducibility and reliability of the PEC inspection technique across all seven NAVAIR samples.

8.2 Sample Size and Sample Grouping

8.2.1 Influence of Sample Size on $a_{90/95}$

As identified in Section 5.5.2, Annis and Gandossi [34] recommend that for hit/miss POD vs. size modelling, the sample size or minimum number of targets/notches is 60.

The $a_{90/95}$ values obtained in both the analysis of the 101 top and 2nd layer notches in all NAVAIR samples and when considering only the 83 2nd layer notches in these samples, produced almost identical results. The $a_{90/95}$ values obtained were 1.98 mm (0.078 inch) and 1.93 mm (0.076 inch) respectively. Similarly, the POD vs. size plots for both cases showed comparable robustness and confidence bounds. Conversely, when considering only the 40 first layer notches in all seven NAVAIR samples, a valid $a_{90/95}$ value was unattainable and the POD vs. size plot revealed instabilities and irregularities, as was shown above in Figure 47. It is apparent that a sample size of 40 notches is insufficient to produce a valid logit fit on the measured data. Hence, the POD analysis on the measured results, presented in Section 7.5.1.1, confirm the sample size guidelines of 60 set by Annis and Gandossi [34].

8.2.2 Influence of Sample Grouping Size on Detection Rate

With respect to detection results obtained with simulations with a sample grouping size of 40 and one notch in the total number of samples, the detection rate ranged from 80.8% to 91.0%. The 91.0% detection rate was achieved when the false call rate was set to 15%. At a 5% false call rate, the detection rate ranges from 85.9% to 88.5% across all three data fractions. As the detection rate for the measured results is 93.6%, and the cut-off is set to achieve a 5% false call rate, it is apparent that the results obtained for simulations with a sample grouping size of 40 were not as good.

Hence, simulations were conducted using the SHV algorithm with a sample grouping size of 80, to assess the influence of sample size on detection rate. Since detection rate decreases when increasing the number of notches in the simulation, the comparison between groupings of 40 and 80 samples was completed, keeping the percentage of notches the same with respect to the total sample size. In comparing simulation results obtained for groupings of both 40 and 80 samples the detection rate increased, with a negligible (0.9% to 1.1% average) increase in the false call rate, when the sample grouping size increased. The detection results obtained for simulations with a grouping size of 80 samples ranged between 85.2% to 92.8%. The simulations conducted with a sample grouping size of 80 produced results that were comparable to the measured experimental results for the NAVAIR samples. Additionally, as the number of notches in the simulation was increased, the false call rate was further reduced below 5%. Hence, the simulations conducted with 80 samples produced significantly better detection rates when compared to simulations with 40 samples. This confirms that as the number of samples in the simulation increases, the limits on the chi-squared distribution improve, and consequently the detection rate increases.

8.2.3 Influence of Sample Grouping Size on $a_{90/95}$

The $a_{90/95}$ values obtained for simulations with grouping sizes of 40 and 80 samples were compared in the same manner as that for detection rates, where the same percentage of notches in the total samples used for each simulation was compared. The simulations conducted with 80 samples produced $a_{90/95}$ values that were lower than for those simulations utilizing a grouping size of 40 samples, which was consistent

across all data fractions. As the sample grouping size increased, the skewing of the chi-squared distribution decreased, resulting in decreased asymmetry and improved results. For simulations using 97.5% data fraction and 5% of the total samples being notches, the $a_{90/95}$ value decreased from 4.07 mm to 2.99 mm (0.160 to 0.118 inch) when increasing the sample grouping size from 40 to 80. Not only were the $a_{90/95}$ values higher for the simulations with 40 samples, but when the false call rate and percentage of the notches in the total sample size was set to 5%, the $a_{90/95}$ value approached the largest notch present in the NAVAIR sample series.

A comparison of the simulations with sample grouping sizes of 40 and 80 samples was also conducted with respect to POD vs. size plots showing repeat measurements. The plots for the results obtained for simulations with 40 samples revealed wide and non-symmetric bounds at a_{90}. Additionally, the repeat measurement plots showed irregularities and were also grouped such that they were skewed to the left of the confidence bounds at a_{90}, revealing reduced stability and a lack of confidence associated with the $a_{90/95}$. In comparison, the symmetry of the confidence bounds at a_{90} for the simulations conducted with a grouping size of 80 samples showed greater symmetry and stability, and consequently produced $a_{90/95}$ values with increased precision.

For simulations with a sample grouping size of 80 with the inclusion of two notches, the smallest $a_{90/95}$ realized was 2.72 mm (0.107 inch) at 10% false call rate and 3.56 mm (0.140 inch) at 5% false call rate. For comparing results of simulations with sample grouping sizes of 80 and 40, the percentage of notches in the total sample size was held constant. Hence, the results for simulations with a grouping size of 40 samples and the inclusion of one notches were compared at the same false call rates, resulting in $a_{90/95}$ values of 3.16 mm (0.125 inch) and 3.73 mm (0.147 inch), respectively. When increasing the percentage of notches to 5% of the total sample grouping size, the resulting average increase in $a_{90/95}$ value across all data fractions was 0.750 mm (0.030 inch) for 40 samples and 0.560 mm (0.022 inch) for 80 samples.

The simulation results with a grouping size of 40 samples produced larger $a_{90/95}$ values, greater variation with increasing number of notches, and irregularities in the POD vs.

size plots that confirmed the lack of confidence in $a_{90/95}$ value. The analysis confirmed that 40 samples were insufficient for hit/miss POD analysis, whereas simulations with 80 samples produced results with greater robustness and confidence.

8.3 Data Fraction

8.3.1 Influence of Data Fraction on Detection Rate

When considering simulations with a sample size of 80 and cut-off set to produce a 10% false call rate, the best overall detection rate achieved was 91.5%, when averaging results for all NAVAIR samples. This result was obtained when the data fraction was set to 100%, with the inclusion of only one notch in the total sample size of 80. This result was consistent with false call rates set to both 5% and 15%. Although, the best detection rate was achieved when the data fraction was set to 100%, as the number of notches increased in the simulations, the detection rate dropped significantly. When the percentage of notches in the simulation was increased to 5% of the total sample size (four notches), the detection rate became the lowest when compared to data fractions of 95 and 97.5%. With the false call rate set to 10%, when comparing the variation in detection rate with the inclusion of one to four notches, the 100% data fraction showed the largest drop at 6.6%. It is apparent that when the data fraction is set to 100%, it is more sensitive to outliers, and data fractions of 95% and 97.5% produce better detection performance. This reveals the effectiveness of the robust statistics to reject outliers, in comparison to the results obtained with a data fraction of 100% which utilizes strictly normal statistical analysis. The smallest variation in detection rate (1.9%), when the number of notches increased from one to four, occurred when the data fraction was set to 97.5%. This was consistent across all false call rates. When comparing the detection rate at this data fraction to that of the best overall detection rate achieved (inclusion of one notch only), the simulation produced only a slightly lower detection rate at 90.4%.

This result, in concert with the small variation in detection rate across the inclusion of one to four notches in the simulation, confirms that robust statistics are optimum when the data fraction is set to 97.5%. The results for the case involving four notches and 76 blank fastener sites represents 5.0% outlier contamination. The best detection results

are achieved when the data fraction is set to 97.5% and not 95%. The improved detection, over 100% data fraction, indicates that the SHV algorithm demonstrates robustness when some outliers are included in the data set. When considering only second layer notches, the best detection performance achieved is still with a 97.5% data fraction, as it provides the best overall results with the lowest variation. Hence, these results confirm the conclusions made by Underhill *et al.* [20] with respect to data fraction and the PEC technique for second layer crack detection.

8.3.2 Influence of Data Fraction on $a_{90/95}$

Blind outlier detection methods were utilized in simulations involving cluster analysis and SHV methodologies, in order to demonstrate effective detection capability without prior knowledge of unflawed behaviour. The smallest $a_{90/95}$ value obtained for all simulations conducted with a sample size of 80 and the inclusion of one notch was 2.54 mm (0.10 inch) at 10% false call rate and 3.04 mm (0.120 in) at 5% false call rate. As with the best detection rate performance above, the lowest $a_{90/95}$ was achieved with the inclusion of one notch and data fraction set to 100%. This outcome was consistent across all false call rates. With the inclusion of one to four additional notches in the total sample size, the variation in the $a_{90/95}$ value (1.31 mm / 0.052 inch) was once again largest when the data fraction was set to 100%. When increasing the percentage of notches up to 5% of the total sample size, the least amount of variation (0.30 mm / 0.012 inch) with respect to $a_{90/95}$ values occurred when the data fraction was set to 97.5%. This pattern was evident with all simulated false call rates.

For simulations with 10% false call rate and the inclusion of one notch, the $a_{90/95}$ result with 97.5% data fraction was 2.69 mm / 0.106 inch, which was only slightly higher than the lowest $a_{90/95}$ recorded, at 100% data fraction. This result, in concert with the small variation in $a_{90/95}$ across one to four notches included in the simulation, confirms that robust statistics are optimum when the data fraction is set to 97.5% [12]. It provides the best performance overall with respect to minimizing $a_{90/95}$ values and false call rates. As with detection performance, the best POD vs. size performance is achieved when the data fraction is set to 97.5% and not 95%, which demonstrates the SHV algorithm robustness, when some outliers are included in the data set. When considering only

second layer notches, the best $a_{90/95}$ performance achieved is still with a 97.5% data fraction, as it provides the best overall results with the lowest variation. Hence, the conclusions made by Underhill *et al.* [20] with respect to data fraction and the PEC technique for second layer notch detection can be extended to $a_{90/95}$ performance.

8.4 Distribution of Notch Sizes in Samples

As concluded in [34], notch size range and distribution, as well as mis-located targets can significantly affect the confidence bound widths on POD vs. size plots, and ultimately adversely affect the $a_{90/95}$ value.

The 101 notched fastener sites in the seven NAVAIR samples ranged in size from 0.76 – 6.1 mm (0.03 - 0.24 inches). The samples had no notch sizes below 0.76 mm (0.03 inch) and only 15 notches below 1.27 mm (0.05 inch), and consequently, there was no POD specimen coverage on the POD vs size curves below POD = 0.4. The simulation results revealed that as the percentage of notches increased, the specimen POD coverage increased, and the notch distribution was centered more towards a_{50}. The small notch size range for these specimens contributed to limited specimen POD coverage for all simulations conducted. Although Annis and Gandossi [34] provide guidelines recommending a specimen POD coverage of 0.03 to 0.97, the impact of the lower bound of the coverage not reaching POD = 0.03 is not significant. The lower bound of the specimen POD coverage only directly affects the confidence bounds at a_0 and a_{50}, and has less of an impact on the upper bound of the POD vs. size curve where a_{90} and $a_{90/95}$ are found.

The distribution of target sizes for the NAVAIR samples produced what appeared to be two distributions, both left and right-skewed, due to the gaps in notch sizes indicated above. The overall notch distribution was also misplaced right of the a_{50}. Although having the notch size distribution misplaced to the right of a_{50} has less of an impact on the confidence bound width at a_{90} than if the notch size distribution was misplaced to the left, there is still a slight impact on the $a_{90/95}$.

The actual distribution of target sizes for the NAVAIR samples reveal a non-symmetric notch distribution, which significantly impacts the width of the confidence bounds at a_{90}, and ultimately the $a_{90/95}$ value. The POD vs. size curves in the repeated measurement plots (Figures 49 and 53) confirm the impact of the notch size distribution, as the non-symmetric bounds at a_{90} are evident.

The data fraction of 95% produced the largest specimen POD coverage across all simulations, although as indicated in [34], this factor alone does not produce the lowest $a_{90/95}$. The non-symmetric notch distribution impacts the $a_{90/95}$ value most significantly, with the limited notch size range and misplaced notch distribution all contributing to increasing the width of the confidence bounds at a_{90}, and ultimately increasing the $a_{90/95}$ value. The effects of these factors adversely influencing the width of the confidence bounds and decrease the precision on the $a_{90/95}$ value were evident in this thesis work, and thus confirm the influence of notch size range and distribution on POD vs. size values as noted by Annis and Gandossi [34].

8.5 Factors Other Than Notch Size Impacting Results

8.5.1 Specimen Characteristics

When comparing the measured and simulation results obtained for each NAVAIR sample, five of the 101 notches displayed significantly lower detection rate. These five fasteners revealed no obvious pattern with respect to notch orientation, size, location, or fastener to lap-joint edge distance, hence it proved difficult to identify a specific cause for this behaviour. Three of the five fasteners, all from one sample, contained filler and other material extruding from the lap joint and surface damage was evident on one of the other fasteners identified.

As Gandossi and Annis state [35], although notch size is the most influential factor that affects POD, other characteristics such as orientation, density, chemical, acoustic or electrical properties, component shape or radii of curvature can also influence it.

Although it has already been discussed that distance to lap-joint edge, and notch size range and distribution can influence POD results, fastener damage (Figure 54) can also

contribute to irregularities in the transient response signal, and ultimately affect the $a_{90/95}$.

Figure 54: Visible fastener damage encountered on NAVAIR sample.

Specimen thickness, gap, range of permeability and spacing of fasteners (Figure 55) can also influence POD results [34]. Although previous work by Pan *et al* [29] and Giguère *et al.* [72] suggest lift-off effects due to variation in gap when using PEC inspection techniques should be able to be accounted for, this was not investigated in this thesis as this information was not accessible.

Figure 55: Irregular spacing between fasteners on NAVAIR Standard 23.

Butt [12] posited that the lower detection rate for Standard 26 could be attributed to the fact that both the top and bottom layers are nominally thicker when compared to other NAVAIR standards, which implied that a greater depth of penetration is required in order

to detect the second layer notches. In this thesis work it was shown that detection for Standard 26 was increased to above 90% and false call rate was decreased to below 3% just by increasing the precision on the fastener to lap-joint edge distance. However, sheet thickness and interlayer gap (Appendix B) may still contribute to lost notch sensitivity when combining samples [12], since a combination of factors may result in outlying scores not being representative of fastener sites with defects present.

8.5.2 Environmental Factors

8.5.2.1 Temperature of Probe
When varying probe temperatures between 10ºC and 30ºC and comparing the results under normal temperature and environmental conditions, there was minimal effect on detection rate, whereas the false call rate produced greater variation.

As already discussed in Section 8.1, the variability with respect to the fastener-to-lap-joint-edge distance for Standard 28 affects its detection rate. With the probe temperature at 30ºC, the detection rate for the sample (77.5%) showed the greatest decrease when compared to other samples.

The false call rate increased significantly for all samples, not only across the three temperatures, but in comparison with results obtained under normal operating and environmental conditions. The largest variation in false call rate was observed for Standard 28, which demonstrated a five-fold increase when comparing normal operating conditions and when measurements were taken with the probe temperature at 30ºC. The increase in false call rate may have been attributed to the increased mechanical noise and vibration at inspection location. The increased noise in the signal would affect the blank fastener cluster, and the system would experience increased difficulty in differentiating signal noise from that associated with a notch.

8.5.2.2 Probe Off-centering
It has been shown in previous works by Horan *et al.* [19] and Babbar *et al.* [32] that the amplitude and shape of the resultant PEC signal can be altered due to slight off-centering of the probe with respect to the ferrous fastener head. Babbar *et al.* [32] observed that probe shifts in the direction of the lap-joint edge produced signals which were different than when the probe was centered over the fastener head. It was also

noted by Babbar *et al.* [32] that the resultant change in signal appeared to be negligible when the probe shift was perpendicular to the lap-joint edge [32] with Horan *et al.* [1] observing very little effect for displacements up to 0.5 mm [1]. Correspondingly, the work presented by Butt *et al.* [12][36][37] addressed the variation in repeat measurements due to horizontal off-centering (shifts in the direction of the lap-joint edge). As presented, the unitary rotation applied to the scores was used to isolate the uncontrolled variation arising due to repeat measurements of s_1 [12]. This provided a solution when the probe shift was confined to horizontal off-centering.

The results from the off-centering in this thesis work, where the probe shift was not confined to horizontal off-centering, reveal that most of the variation is along s_2 instead of s_1. The detection rate for top layer notches remained relatively constant, while a significant reduction in detection rate was realized for second layer and in the overall detection rate for all samples. There was also a significant increase in the false call rate. The detection methodologies used in this analysis process involves the removal of s_1, but if there is an issue with s_1 and s_2, then this may affect the results. If most of the variation is along s_2 instead of s_1, then the covariance matrix constructed from $s_2 \ to \ s_5$ may not be able to facilitate identification of outlier scores arising in the presence of notches, if the covariance matrix used to calculate the MD is flawed. With most of the variation along s_2 instead of s_1 and the significant reduction in detection rate for second layer notches, it is apparent that if the probe shift is not horizontal then it is no longer viable to disregard s_1 from subsequent MD calculations. The results confirm the importance of precise positioning of the probe and the associated direct impact of operator skill level on detectability.

8.6 Sources of Error

Airworthiness assurance is obtained through adequate and appropriate inspection intervals. Due to the nature of NDE inspection techniques, human inspection performance has a direct impact on inspection reliability. Various human factors can impact NDE reliability, including those related to the task, individual inspector characteristics, physical environment, and organization factors to name a few [73]. Human factors are an important aspect of NDE reliability, as inspection techniques such

as PEC require a human to conduct the inspection. As indicated in [34], human error 'sets an upper limit on POD' as the technique is capable of a certain minimum error given the best possible handling of the inspection. The field of human factors research is beyond the scope of this thesis work, but it is important to recognize specific factors that contributed to sources of error, which were inherent in this research. One source of human error could be attributed to probe lift-off and off-centering, which would have effects on the experimental measurements obtained, and ultimately detection rates and the $a_{90/95}$.

There were also irregularities observed in the gap at the lap-joint edge, as some samples contained material wedged in this gap. This could have resulted in measurement errors associated with the fastener to lap-joint edge distance, which has been identified as a critical factor with respect to notch detection results.

Environmental factors such as mechanical vibrations and machine noise from nearby apparatus was observed when obtaining measurements for the probe temperature analysis. This seemed to have an impact on the results, as the false call rate increased for every NAVAIR sample analyzed as part of this study.

8.7 Inspection Reliability

The focus of this thesis work is to assess the detection reliability of the PEC inspection technique in the application associated with detecting second layer cracks at ferrous fasteners in aircraft lap-joint structures. A probability of detection (POD) analysis was conducted to provide the metric needed to assess this reliability. As indicated in Ref. [35], the importance of obtaining a quantitative measure such as $a_{90/95}$, is essential to not only contribute to risk reduction but actually quantify the failure probability associated with in-service structures.

This thesis work provided evidence of reproducibility and repeatability with respect to effective notch detectability of the three NAVAIR samples presented by Butt *et al.* [12][36][37], but also extended his work to include four additional NAVAIR samples for a total of seven samples. The measured results obtained in this work provide an average

detection rate of 93.4% and false call rate of 4.2%, across all NAVAIR samples. The results obtained using cluster analysis and the SHV algorithm in simulations produced similar detection results at 88.5%, 91.6% and 92.8% with 5%, 10% and 15% false call rate, respectively.

The POD analysis conducted using the **mh1823 POD** software on all combined samples utilizing the measured results, produced an $a_{90/95}$ of 1.98 mm (0.078 inch) overall, and an $a_{90/95}$ of 1.93 mm (0.076 inch) for second layer notches only. For the simulations conducted with a sample grouping size of 80, the POD vs. size results using the **mh1823 POD** software produced an $a_{90/95}$ of 2.24 mm (0.088), 2.54 mm (0.1 inch) and 3.04 mm (0.12 inch) with the cut-offs set to 15%, 10% and 5% false call rate, respectively. The revised POD vs. size results obtained using ImageJ software produced an $a_{90/95}$ of 1.70 mm (0.067 inch), 2.04 mm (0.08 inch) and 2.43 mm (0.095 inch) at 15%, 10% and 5% false call rate, respectively.

Consequently, this thesis work provided the initial evidence and confidence that this PEC system is capable of detecting flaws in second layer lap-joint aircraft structures with high confidence in reliability with respect to detection. Additionally, the POD analysis confirmed confidence in the PEC system and it being used for in-service inspections, with $a_{90/95}$ values approaching that recognized for BHEC inspection.

8.8 Rationale for Increased False Call Rate for PEC Technique

With the cluster analysis approach and SHV methodologies, a decision threshold or cut-off is selected to produce a specific false call rate. As indicated in Section 8.7 above, when this decision threshold is decreased, the false call and detection rates increase. The NDE technique currently being used on CP-140 Aurora aircraft consists of conventional or BHEC inspection methods that require removal of every fastener on the aircraft structure being inspected. Conversely, with the PEC inspection technique, only those fasteners that indicate a flaw is present would be removed. As indicated above in Section 8.7, the $a_{90/95}$ ranged from 2.43 mm (0.095 inch) with ImageJ revision to 3.04 mm (0.12 inch) using the **mh1823 POD** software, when the decision threshold was set

to 5% false call rate. When the cut-off was set to produce a 10% false call rate, the $a_{90/95}$ values ranged from 2.04 mm (0.08 inch) to 2.54 mm (0.1 inch), when using ImageJ and **mh1823 POD** software packages respectively. At 15% false call rate, the $a_{90/95}$ reduced lower to 1.70 mm (0.067 inch) with ImageJ and 2.24 mm (0.088 inch) with **mh1823 POD** software. The actual average false call rate obtained with these latest simulations was 10.8%. The consequence of a false call rate higher than 5% with the PEC inspection technique is not as significant or impactful as with BHEC. The $a_{90/95}$ value for PEC approaches that for BHEC, whereas only 15% of the fasteners would have to be removed, as opposed to removal of 100% of the fasteners as with BHEC.

The PEC inspection technique could be utilized to identify fasteners that indicate irregular behaviour, and consequently highlight those fasteners that require further inspection. The fastener could then be removed, and the hole scanned using the conventional method of BHEC to gather more information about the potential defect. The advantage of using the PEC system is that the decision threshold that sets the false call rate is subjective, and a false call that is confirmed via BHEC has no safety implications. Increasing the false call rate increases the detection rate and lowers the $a_{90/95}$, and would ultimately increase the percentage of fasteners requiring removal, but it would never require 100% fastener removal as with BHEC inspection methods.

There are numerous advantages to the PEC inspection method over that of BHEC. The inspection does not require fastener removal, hence the risk of collateral damage on the surface structure is mitigated. Additionally, this inspection method minimizes the cost associated with labour and human resources as well as aircraft maintenance downtime. This can be advantageous, especially for specific structural areas on aircraft that are susceptible to fatigue damage, as this is an easy and quick way to check fasteners more regularly with less impact on operations and resources.

The results obtained in this thesis work using strictly experimental measurements resulted in a detection rate of 93.6% and an $a_{90/95}$ value of 1.98 mm (0.078 inch) at 5% false call rate. Considering these measured results and a 91.6% detection rate for simulations and corresponding $a_{90/95}$ values of 2.04 mm (0.08 inch) with ImageJ

revision at 10% false call rate, the recommendation is to set the standard for acceptable false call rate for PEC to be 10%. With the actual average false call rate obtained with simulations at 15% false call rate, being 10.8% and with $a_{90/95}$ being 1.70 mm (0.067 inch) with ImageJ, and 2.24 mm (0.088 inch) with **mh1823 POD** software, setting the false call rate at 10% is valid.

8.9 Determination of Detectable Flaw Size

The minimum detectable flaw size, $a_{90/95}$, must be determined prior to validating a NDE technique for use in field applications, as there are intrinsic airworthiness implications and requirements that must be met. This thesis work provided the reliability metric needed to conduct the initial assessment of the PEC system.

The results obtained using strictly experimental measurements produced the lowest $a_{90/95}$ at 1.98 mm (0.078 inch) for all samples and even better results when considering only second layer notches, where the $a_{90/95}$ was 1.93 mm (0.076 inch). These results provide assurance that the PEC technique is capable of detecting second layer cracks in aircraft lap-joint structures with high confidence, reliability, and detectability.

The simulations utilized the cluster analysis approach and SHV methodologies including the bootstrap method, and although they utilized the MPCA scores from actual experimental measurements, produced higher $a_{90/95}$ values. It is important to recognize that there is a requirement for these simulations, as there are difficulties associated with obtaining large numbers of samples, yet it is vital that the minimum number of targets (notches) and sample grouping sizes be maintained to increase the confidence associated with the $a_{90/95}$ value. Most importantly, simulations and methodologies such as SHV are critical because in any real inspection the blanks are unknown, thus the "measured" approach cannot be implemented. Hence, these simulations are vital in determining the reliability metric for this NDE technique, and correspondingly the combination of both experimental and simulation results were used to determine the $a_{90/95}$ value associated with this technique. The disadvantage of the simulations is that there are essential parameters that are adversely influencing the results as some disparity exists between measured results and those from simulations. These additional

parameters that are not related to the size of the flaw must be identified and accounted for. With only five fasteners out of 101 that showed disparity between the measured and simulation results, additional samples will be required to determine the specific parameter or variable that is impacting the results. Although temperature of the probe and fastener-to-lap-joint-edge distance was investigated, the factors of top and bottom layer thickness and interlayer gap were not considered in this thesis work.

The non-symmetric flaw size distribution contributed to irregular behaviour of the POD vs. size plots associated with repeat measures, which impacted the resulting $a_{90/95}$ values. Correspondingly, ImageJ software was used to revise the $a_{90/95}$ values. The utilization of this image software to revise the $a_{90/95}$ value introduces a level of subjectivity associated with this value. The inclusion of additional samples, which would result in a symmetric flaw size distribution, is necessary to produce POD vs. size curves that do not require revision and improve the confidence of the determined $a_{90/95}$ value.

For simulations with a sample grouping size of 80, 10% false call rate, the lowest $a_{90/95}$ achieved was 2.54 mm (0.1 inch). Increasing the number of notches in the simulation to 5% of the total sample grouping size produced an $a_{90/95}$ of 2.99 mm (0.118 inch). In field applications, the number of flaws, in comparison to unflawed fastener sites, would be small. Hence, it may be assumed that the $a_{90/95}$ is at the lower end of this range. ImageJ results at 10% false call rate produced $a_{90/95}$ values ranging from 2.03 to 2.22 mm (0.080 to 0.087 inch). The ImageJ revision on the simulation results are in between the $a_{90/95}$ values obtained for the measured results and the $a_{90/95}$ values using **mh1823 POD** software on the simulation results. The actual average false call rate obtained with simulations at 15% false call rate being 10.8% resulted in $a_{90/95}$ of 1.70 mm (0.067 inch) with ImageJ and 2.24 mm (0.088 inch) with **mh1823 POD** software. Based on both the measured and simulation results indicated above, it is determined that with regards to this PEC inspection technique the associated $a_{90/95}$ value be recognized as 2.0 mm (0.08 inch). The effective range is then between the measured result and the lower end of the results obtained for the simulation results using the **mh1823 POD** software. Further research work involving additional samples needs to be conducted to increase the precision on the actual $a_{90/95}$ value that should be associated with this technique.

9. Summary and Future Work

9.1 Summary

The motivation for this thesis work was to conduct a POD analysis, which provided the metric needed, in the form of an $a_{90/95}$ value, to assess the reliability associated with the developed PEC inspection technique for the detection of second layer cracks in aircraft lap-joint structures. Prior to the POD analysis, the robustness of the PEC inspection system was also analyzed by increasing parameter variability with respect to off-centering and temperature of the probe, and fastener to lap-joint edge distance. Effective detection capability without prior knowledge of unflawed behaviour is advantageous. Therefore, blind outlier detection methods were utilized in simulations involving cluster analysis and SHV methodologies. The capability of the PEC technique that was assessed, was shown to approach that of conventional BHEC ($a_{90/95}$ of 0.05 inch), with the advantage of there not being a requirement for fastener removal.

This thesis work utilized the PEC probe with an 8 mm ferrite core and eight differentially paired pick-up coils, highlighted by Butt *et al.* [12][36][37] to be the most effective at detecting fatigue cracks in second layer aluminum structures. The ferrous fastener was used as a conduit for the magnetic flux generated by the driving coil, which allowed for deeper penetration of eddy currents and the pick-up coils in the PEC probe to receive the resultant transient responses [9]. The resulting PEC signals were analyzed using the statistical analysis method known as modified principal components analysis (MPCA), described by Horan *et al* [19], which reduced the transient PEC signals to a series of eigenvectors and scores. The MPCA scores were analyzed using Type 3 outlier detection methods consisting of cluster analysis and determination of a Mahalanobis distance (MD), followed by a smallest half volume (SHV) analysis utilizing a bootstrap method, which is a Type 1 clustering analysis approach [17]. These approaches, in concert with the boundary threshold, distinguished between signals associated with electrical discharge machined (EDM) notches and those of blank fastener locations to identify locations with notches present [10] [20].

Previous PEC work completed by Butt *et al.* [12][36][37] demonstrated effective results for three NAVAIR samples using both cluster analysis and SHV analysis methodologies, with simulations producing blind outlier detection capability. Parameters implicit in the cluster analysis and robust statistics methods, such as eigenvalues, eigenvectors, covariance and correlation matrices, were analyzed to confirm that no patterns or specific variables that may adversely influence detection results were present. Additional parameter variability was introduced into the PEC inspection system with respect to off-centering and temperature of the probe, and fastener to lap-joint edge distance.

It has been shown in previous works by Horan *et al.* [19] and Babbar *et al.* [32] that the amplitude and shape of the resultant PEC signal can be altered due to slight off-centering of the probe with respect to the ferrous fastener head. It was also noted by Babbar *et al.* [32] that the resultant change in signal appeared negligible when the probe shift was perpendicular to the lap-joint edge [32], while Horan *et al.* [19] observed very little effect for displacements of up to 0.5 mm [19]. Correspondingly, the work presented by Butt *et al.* [12][36][37] addressed the variation in repeat measurements due to horizontal off-centering (shifts in the direction of the lap-joint edge) with a unitary rotation applied to the scores isolating the uncontrolled variation due to repeat measurements of s_1 [12]. The results from the off-centering in this thesis work, where the probe shift was not confined to horizontal off-centering and the displacement was up to 2 mm from the centre of the fastener, reveal that most of the variation was along s_2 instead of s_1. The detection rate for top layer notches remained relatively constant, while a significant reduction in detection rate was realized for second layer and in the overall detection rate for all samples. There was also a significant increase in the false call rate. The reduction in detection rate was attributed to most of the variation being along s_2 instead of s_1. This resulted in the covariance matrix, which was constructed from $s_2 \, to \, s_5$ and which was ultimately used to calculate the MD, being flawed. Consequently, some difficulties were experienced in detecting notches. It is apparent that if the probe shift is not horizontal then it is no longer viable to disregard s_1 from subsequent MD calculations.

The variation in temperature of the probe between 10ºC, 20ºC and 30ºC had minimal impact on detection rate for the four samples assessed. In contrast, the false call rate increased significantly, not only across the three temperatures, but in comparison to results obtained under normal operating and environmental conditions. It is hypothesized that the increase in false call rate was attributed to the increased mechanical noise and vibration at the inspection location. The increased noise in the signal would affect the blank fastener cluster and the system would experience increased difficulty in differentiating signal noise from that associated with a notch.

Fastener-to-lap-joint-edge distance measurements were obtained from higher resolution images, utilizing 0.5 mm scale increments, resulting in an increase in image size from 440 kB (1280 pixels x 960 pixels) utilized in [12] to approximately 1.3 MB (3072 pixels x 2304 pixels). The intent was to assess if this parameter contributed to the poor detection results obtained by Butt [12] with four of the seven NAVAIR samples. With increased precision of the fastener-to-lap-joint-edge distances, an increase in detection rate was obtained. The detection rate for top layer notches increased from 87.7% to 93.7%, second layer detection rate increased from 86.8% to 93.7% and overall the average detection rate for the NAVAIR sample series increased from 86.0% to 93.4%. The false call rate also showed improvement, as it dropped from 5.5% to 4.2%. These results emphasize the importance of precise fastener-to-lap-joint-edge distance measurements when using the PEC methodologies presented in this work. Correspondingly, effective detection results were realized with the sample set expanded to all seven NAVAIR samples. This revealed robustness of the PEC inspection system, and with an average detection rate for all samples over 93%, simulations were then conducted to assess the blind detection capability.

When comparing the measured and simulation results obtained for each NAVAIR sample, five of the 101 notches displayed significantly lower detection rate. It proved difficult to identify a specific cause for this behaviour, as the five fasteners revealed no obvious pattern with respect to notch orientation, size, location, or fastener to lap-joint edge distance. Although notch size is the most influential factor that affects POD, specimen thickness, gap, fastener spacing, range of permeability and specimen damage can also influence detection rate. As three of the five fasteners locations exhibited

material extruding from the lap joint and surface damage was evident on one of the other fasteners identified, the discrepancy between measured and simulation detection rates was attributed to these factors. Overall, the simulations utilizing cluster analysis, bootstrap and SHV methodologies, proved blind outlier detection capability for this PEC inspection system.

The POD analysis of the PEC inspection technique was then conducted, including the metric needed, in the form of an $a_{90/95}$ value, to assess its reliability in the detection of second layer cracks at ferrous fasteners in aircraft lap-joint structures. The POD analysis of the NAVAIR samples using the **mh1823 POD** software confirmed that the logistic model provided the best results and reinforced the superiority of this model, dating back to Berens and Hovey [65], for effectiveness in modelling hit/miss data in POD curves. Sample size, range and distribution of notch sizes, mis-located targets, simulation grouping size, and data fraction all influenced detection rates, and ultimately the $a_{90/95}$. This thesis work confirmed the work of Underhill *et al.* [20] with respect to data fraction and the PEC technique for second layer crack detection, as the best detection rate achieved was with a 97.5% data fraction. This work also confirmed that this can be extended to include $a_{90/95}$ performance.

The POD analysis conducted using the **mh1823 POD** software on all samples and the measured results produced an $a_{90/95}$ of 1.98 mm (0.078 inch) overall, and an $a_{90/95}$ of 1.93 mm (0.076 inch) for second layer notches only. For hit/miss POD analysis, Annis and Gandossi [34] recommended specimens that include a notch size range corresponding to a specimen POD coverage of 0.03 to 0.97, and a uniform notch size distribution centred at a_{50}. This work confirmed the effects of analyzing specimens with notch sizes outside of these guidelines, adversely influencing the width of the confidence bounds and decreasing the precision of the $a_{90/95}$ value. Additionally, all repeat POD(a) plots were confined within the confidence bounds, signifying a widening of the confidence bounds and a corresponding increase in the **mh1823 POD** calculated $a_{90/95}$ value. To account for the increased width of confidence bounds at a_{90} and the apparent inflation of the $a_{90/95}$ value, ImageJ software was used to increase the precision of $a_{90/95}$ values obtained from simulation results. For simulations set with a

15% false call rate, the $a_{90/95}$ obtained was 2.24 mm (0.088 inch) with **mh1823 POD** software, which was then reduced to 1.70 mm (0.067 inch) with ImageJ software. The actual average false call rate obtained for these simulations was 10.8%.

It is important to note that the consequence of a false call rate higher than 5% with the PEC inspection technique is not as significant or detrimental as with BHEC. The advantage of using the PEC system is that the decision threshold that sets the false call rate is subjective, and a false call that is confirmed via BHEC has no safety implications. Increasing the false call rate increases the detection rate and lowers the $a_{90/95}$. This research confirms an $a_{90/95}$ value for PEC that approaches that for BHEC, whereas only 15% of the fasteners may have to be removed, as opposed to removal of 100% of the fasteners as with BHEC. There are numerous advantages to the PEC inspection method. The inspection does not require fastener removal, hence there is reduced risk of collateral damage on the surface structure, the cost associated with labour and human resources is reduced and aircraft maintenance downtime is minimized. There is also the added advantage of the opportunity for increased inspection intervals for specific aircraft structural areas that are susceptible to fatigue damage, with less impact on operations and resources.

Based on both the measured and simulation results obtained in this thesis work, it is determined that with respect to this PEC inspection technique, that the associated $a_{90/95}$ value be recognized as being 2.03 mm (0.080 inch). These results provide assurance that the PEC technique is capable of detecting second layer cracks in aircraft lap-joint structures with high confidence, reliability, and detectability. The result of this work also has significant implications associated with the potential for this inspection technique to be recognized as an approved NDT technique for detection of cracks at bolt holes in lap-joint wing structures without the requirement of fastener removal.

9.2 Future Work

Based on the findings of this thesis work, the following research work is recommended to increase the confidence in the precision of the actual $a_{90/95}$ value that should be associated with this technique:

❖ Obtain additional samples, comparable to NAVAIR sample series, that would result in symmetric notch size distribution. Sections 5.4 and 6.3 of [34] should be used as a guide to determine appropriate sizes for additional NDE lab specimens.

❖ Identify additional parameters other than notch size that account for disparity between measured and simulation results.

❖ Develop improved methods of comparing and combining samples, accounting for differences in layer thickness and interlayer gap.

❖ Determine impact of thickness variation and interlayer gap and corresponding limitations with respect to POD vs. size curves.

❖ Enhance PEC analysis methods to extend beyond hit/miss results that provide a measure of crack size and orientation.

❖ Obtain samples where original fasteners have been replaced with oversize fasteners to assess impact on detection when fastener size is same size or larger than 8 mm ferrite core in PEC probe.

❖ Obtain and analyze in-service specimens using PEC inspection method to assess detection reliability and effectiveness under conditions with increased signal noise, associated with non-lab specimens that include surface defects and corrosion.

❖ Conduct field trial of PEC inspection system prior to BHEC inspection to validate effectiveness of system.

❖ Develop user-friendly software interface capable of accurately measuring fastener-to-lap-joint-edge distance that could be deployable and effective in field environment.

❖ Develop a deployable PEC system that includes a robust probe design and user-friendly software interface capable of post-processing measurement data efficiently and effectively.

References

[1]   N. E. Dowling, *Mechanical Behavior of Materials*, 4th ed. Upper Saddle River: Pearson Education Inc., 2013.

[2]   D. E. Bray and R. K. Stanley, *Nondestructive Evaluation: A Tool in Design, Manufacturing, and Service (Revised Edition)*, 1st ed. Boca Raton: CRC Press, 1997.

[3]   A. Fahr, *Aeronautical Applications of Non-destructive Testing*, 1st ed. Lancaster: DEStech Publications Inc., 2014.

[4]   Department of Defense, "Nondestructive Inspection Methods, Basic Theory." Department of Defense, 2007.

[5]   European Commission - Joint Research Centre, "ENIQ Recommended Practice 1 Influential / Essential Parameters," *ENIQ Rep. nr. 24*, no. 2, 2005.

[6]   H. Lemire, P. R. Underhill, T. W. Krause, M. Bunn, and D. J. Butcher, "Improving Probability of Detection of Bolt Hole Eddy Current Inspection," *Res. Nondestruct. Eval.*, vol. 21, pp. 1–16, 2010.

[7]   A. P. Berens, "NDE Reliability Data Analysis," *Nondestruct. Eval. Qual. Control. ASM Int. Met. Handb.*, vol. 17, pp. 689–701, 1989.

[8]   Department of Defense, "Nondestructive Evaluation System Reliability Assessment, MIL-HDBK-1823A," Dayton, 2010.

[9]   C. A. Stott, P. R. Underhill, and T. W. Krause, "Pulsed Eddy Current Detection of Cracks in Multilayer Aluminum Lap Joints," *IEEE Sens. J.*, vol. 15, pp. 956–962, 2015.

[10]   C. A. Stott, "Pulsed Eddy Current Inspection of Second Layer Wing Structure," Royal Military College of Canada, Kingston, 2014.

[11]   Technical Engineering Services Directorate, "In-Service Inspection Flaw Assumptions for Metallic Structures, EN-SB-08-012, Revision C," Dayton, 2013.

[12]   D. M. Butt, "Optimizing Pulsed Eddy Current for Inspection of Second Layer Wing Structure," Royal Military College of Canada, 2016.

[13]   P. P. Whalen, "Transient Eddy Current Inspection in the Presence of Ferrous Fasteners in Multi-Layered Aluminum Structures," Royal Military College of Canada, Kingston, 2010.

[14]   T. W. Krause, C. Mandache, and J. H. Lefebvre, "Diffusion of Pulsed Eddy Currents in Thin Conducting Plates," *Am. Inst. Phys.*, vol. 975, pp. 368–375, 2008.

[15]   V. Barnett and T. Lewis, *Outliers in Statistical Data*. New York: John Wiley & Sons, 1994.

[16]   V. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artif. Intell. Rev.*, vol. 22, no. 2, pp. 85–126, 2004.

[17]   W. J. Egan and S. L. Morgan, "Outlier Detection in Multivariate Analytical Chemical Data," *Anal. Chem.*, vol. 70, no. 11, pp. 2372–2379, 1998.

[18]   J. Lattin, J. D. Carroll, and P. E. Green, *Analyzing Multivariate Data*. Pacific Grove: Brooks/Cole, 2003.

[19]   P. F. Horan, P. R. Underhill, and T. W. Krause, "Pulsed eddy current detection of cracks in F/A-18 inner wing spar without wing skin removal using Modified Principal Components Analysis," *NDT&E Int.*, vol. 55, pp. 21–27, 2013.

[20]   P. R. Underhill, D. M. Butt, and T. W. Krause, "Blind Detection of Cracks at Fasteners in CP-140 Aircraft Using Pulsed Eddy Current," in *19th World Conference of Non-Destructive Testing*, 2016.

[21]   A. Sophian, G. Y. Tian, D. Taylor, and J. Rudlin, "Design of a pulsed eddy current sensor for detection of defects in aircraft lap joints," *Sensors and Actuators*, vol. 101, pp. 92–98, 2002.

[22]   Y. Y. He *et al.*, "Pulsed eddy current technique for defect detection in aircraft riveted structures," *NDT&E Int.*, vol. 43, pp. 176–181, 2009.

[23]   J. A. Buck, P. R. Underhill, J. E. Morelli, and T. W. Krause, "Simultaneous Multiparameter Measurement in Pulsed Eddy Current Steam Generator Data Using Artificial Neural Networks," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 3, pp. 119–129, 2001.

[24]   V. B. Wwedensky, "Concerning the eddy currents generated by a spontaneous change of magnetization," *Ann. Phys.*, vol. 64, pp. 609–620, 1921.

[25]   D. Desjardins, "Analytical Modeling for Transient Probe Response in Eddy Current Testing," Royal Military College of Canada, Kingston, 2011.

[26]   C. V. Dodd and W. E. Deeds, "Analytical solutions to eddy current coil probe problems," *J. Appl. Phys.*, vol. 39, no. 6, pp. 2829–2838, 1968.

[27]   R. C. Callaroti, P. Schmidt, and H. Arqué, "Theory of the measurement of thickness and conductivity of cylindrical shells by an inductive method," *J. Appl. Phys.*, vol. 43, no. 10, pp. 3952–3958, 1972.

[28]   Y. He, M. Pan, D. Chen, and F. Luo, "PEC defect automated classification in aircraft multi-ply structures with interlayer gaps and lift-offs," *NDT&E Int.*, vol. 53, pp. 39–46, 2013.

[29]   M. Pan, Y. He, G. Tian, D. Chen, and F. Luo, "PEC Frequency Band Selection for Locating Defects in Two-Layer Aircraft Structures With Air Gap Variations," *Trans. Instrum. Meas.*, vol. 62, no. 10, pp. 2849–2856, 2013.

[30]   Y. He, M. Pan, F. Luo, D. Chen, and X. Hu, "Support vector machine and

optimised feature extraction in integrated eddy current instrument," *Measurement*, vol. 46, pp. 764–774, 2013.

[31]   D. R. Desjardins, G. Vallieres, P. P. Whalen, and T. W. Krause, "Advances in Transient (Pulsed) Eddy Current For Inspection of Multi-Layer Aluminum Structures in the Presence of Ferrous Fasteners," *Rev. Prog. Quant. Nondestruct. Eval.*, vol. 1430, pp. 400–407, 2012.

[32]   V. K. Babbar, P. R. Underhill, C. A. Stott, and T. W. Krause, "Finite Element Modeling of Second Layer Crack Detection in Aircraft Bolt Holes with Ferrous Fasteners Present," *NDT&E Int.*, vol. 65, pp. 64–71, 2014.

[33]   P. R. Underhill, C. A. Stott, and T. W. Krause, "In-Situ Calibration of Pulsed Eddy Current Detection of Cracks at Fasteners in CP-140 Aircraft," *Rev. Prog. Quant. Nondestruct. Eval.*, vol. 1706, pp. 90004-1-8, 2016.

[34]   C. Annis and L. Gandossi, "ENIQ TGR Technical Document: Influence of Sample Size and Other Factors on Hit/Miss Probability of Detection Curves (ENIQ report No 47)," Luxembourg, 2012.

[35]   L. Gandossi and C. Annis, "ENIQ TGR Technical Document: Probability of Detection Curves: Statistical Best-Practices (ENIQ report No 41)," Luxembourg, 2010.

[36]   D. M. Butt, P. R. Underhill, and T. W. Krause, "Examination of Pulsed Eddy Current for Inspection of Second Layer Aircraft Wing Lap-Joint Structures Using Outlier Detection Methods," *CINDE J.*, vol. 37, no. Sept/Oct, pp. 6–10, 2016.

[37]   D. M. Butt, P. R. Underhill, and T. W. Krause, "Enhancing Pulsed Eddy Current for Inspection of P-3 Orion Lap-Joint Structures," in *AIP Conf. Proc. 1706*, 2016, pp. 90017-1-7.

[38]   D. J. Griffiths, *Introduction to Electrodynamics*, 3rd ed. Upper Saddle River: Prentice-Hall, 1998.

[39]   H. Ohanian and J. Markert, *Physics for Scientists and Engineers, Volume 2*, 3rd ed. New York: W.W. Norton and Company, Inc., 2007.

[40]   D. Desjardins, "Analytical Modeling for Transient Probe Response in Eddy Current Testing," Queen's University, Kingston, 2015.

[41]   R. M. Bozorth, *Ferromagnetism*, 1st ed. Ann Arbor: Wiley-IEEE Press, 1993.

[42]   G. Dietzmann and M. Schaefer, "Rayleigh hysteresis with sinusoidal wave form of magnetic induction," *J. Magn. Magn. Mater.*, vol. 110, no. 1–2, pp. 151–160, 1992.

[43]   D. R. Desjardins, T. W. Krause, and L. Clapham, "Transient response of a driver coil in transient eddy current testing," *NDT E Int.*, vol. 73, pp. 22–27, 2015.

[44]   H. Ohanian, "On the approach to electro- and magneto-static equilibrium," *Am. J.*

Phys., no. 51, p. 1020, 1983.

[45] Olympus Corporation, "Introduction to Eddy Current Testing." [Online]. Available: http://www.olympus-ims.com/en/eddycurrenttesting/. [Accessed: 30-Jul-2017].

[46] P. Hammond, "The calculation of the magnetic field of rotating machines. Part 3: Eddy currents induced in a solid slab by a circular current loop," *Proc. IEE - Part C Monogr.*, vol. 109, no. 16, pp. 508–515, 1962.

[47] V. S. Cecco, G. Van Drunen, and F. L. Sharp, "Eddy Current Testing Manual on Eddy Current Method," no. 11, pp. 1–206, 1981.

[48] M. A. Heald and B. Marion, Jerry, *Classical Electromagnetic Radiation*, 2nd ed. Orlando: Academic Press, 1980.

[49] W. R. Smythe, *Static and Dynamic Electricity*, 3rd ed. New York: McGraw-Hill Inc., 1968.

[50] NDT Resource Center, "Depth of Penetration & Current Density," 2016. [Online]. Available: https://www.nde-ed.org/EducationResources/CommunityCollege/EddyCurrents/Physics/depthcurrentdensity.htm. [Accessed: 30-Jul-2017].

[51] H. D. Young and R. A. Freedman, *University Physics*. San Francisco: Pearson Education Inc., 2004.

[52] G. Rizzoni, *Principles and Applications of Electrical Engineering*. New York: McGraw-Hill Inc., 2007.

[53] S. Goldman, *Transformation Calculus and Electrical Transients*. New York: Prentice-Hall, 1949.

[54] D. Desjardins, T. W. Krause, and L. Clapham, "Transient response of a driver-pickup coil probe in transient eddy current testing," *NDT E Int.*, vol. 75, pp. 8–14, 2015.

[55] K. I. Kim, K. Jung, and H. J. Kim, "Face recognition using kernel Principal Components Analysis," *IEEE Signal Process. Lett.*, vol. 9, no. 2, pp. 40–42, 2002.

[56] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. Hoboken: Wiley, 2003.

[57] P. J. Rousseeuw and M. Hubert, "Robust statistics for outlier detection," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 1, no. 1, pp. 73–79, 2011.

[58] S. C. Chapra and R. P. Canale, *Numerical Methods for Engineers*. Montreal: McGraw-Hill Inc., 1988.

[59] H. M. Stephens Jr., "NDE Reliability – Human Factors – Basic Considerations," in *15th World Conference on Non-Destructive Testing*, 2000.

[60] N. Nakagawa and J. C. Moulder, "A model of bolt hole inspection via eddy currents," *Rev. Prog. Quant. Nondestr. Eval.*, vol. 12, pp. 259–263, 1993.

[61] A. P. Berens and P. W. Hovey, "Statistical Methods for Estimating Crack Detection Probabilities," in *Probabilistic Fracture Mechanics and Fatigue Methods: Applications for Structural Design and Maintenance*, West Conshohocken: ASTM International, 1983, pp. 79–94.

[62] A. P. Berens and P. W. Hovey, "Characterization of NDE Reliability," *Rev. Prog. Quant. NDE*, vol. 1, pp. 579–585, 1982.

[63] J. C. Brewer, "Estimate of Probability of Crack Detection from Service Difficulty," Cambridge, 1994.

[64] R. C. H. Cheng and T. C. Iles, "Confidence Bands for Cumulative Distribution Functions of Continuous Random Variables," *Technometrics*, vol. 25, no. 1, pp. 77–86, 1983.

[65] A. P. Berens and P. W. Hovey, "Evaluation of NDE reliability characterisation, AFWAL-TR-81-4160," Wright-Patterson Air Force Base, 1981.

[66] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "NIH Image to ImageJ: 25 Years of Image Analysis," *Nat. Methods*, vol. 9, no. 7, pp. 671–675, 2012.

[67] B. Raiwa, H. McNally, P. Varadharaian, J. Sturgis, and J. Robinson, "AFM/CLSM Data Visualization and Comparison Using an Open-source Toolkit," *Microsc Res Tech.*, vol. 64, no. 2, pp. 176–184, 2004.

[68] APEX Microtechnology Inc., "Dual Power Operational Amplifers - PA75." Tucson, 2012.

[69] V. K. Babbar, P. R. Underhill, C. A. Stott, and T. W. Krause, "Finite Element Modeling of Second Layer Crack Detection in Aircraft Bolt Holes with Ferrous Fasteners Present," *NDT&E Int.*, vol. 65, pp. 64–71, 2014.

[70] P. R. Underhill, private communication, May 13, 2016.

[71] M. Bunn, private communication, Aerospace and Telecommunications Engineering Support Squadron, Trenton, 2015.

[72] S. Giguère, B. A. Lepine, and J. M. S. Dubois, "Pulsed Eddy Current Technology: Characterizing Material Loss with Gap and Lift-off Variations," *Res. Nondestruct. Eval.*, vol. 13, no. 3, pp. 119–129, 2001.

[73] A. D'Agostino, S. Morrow, C. Franklin, and N. Hughes, "Review of Human Factors Research in Nondestructive Examination," Washington, 2017.

[74] N. Instruments, "National Instruments USB-6361 Block Diagram," 2016. [Online]. Available: http://www.ni.com/documentation/en/multifunction-io-device/latest/usb-6361/block-diagram/. [Accessed: 01-Aug-2017].

Appendix A

Table 18: NAVAIR sample series blank fastener locations.

Sample ID	Blank Fasteners
Standard 22	8, 15, 16, 17, 18, 20, 22, 23
Standard 23	2, 5, 7, 8, 11, 12, 14, 17, 21, 23
Standard 24	2, 10, 14, 16, 18, 20, 21, 22
Standard 25	2, 4, 6, 8, 12, 16, 18
Standard 26	1, 4, 6, 11, 15, 17, 19
Standard 28	1, 4, 5, 10, 11, 16, 18, 20, 21, 22
Standard 32	2, 6, 10, 17, 19, 21, 23, 24

Appendix B

Table 19: P-3 Orion sample series top and bottom layer thickness with interlayer gap (uncertainties are estimated as being to within stated significant figures) [12].

Sample ID	Top Layer Thickness (mm)	Bottom Layer Thickness (mm)	Interlayer Gap Thickness (mm)
Standard 22	2.5	2.1	0.05
Standard 23	2.6	2.1	0.12
Standard 24	2.4	2.1	0.00
Standard 25	2.5	2.1	0.03
Standard 26	2.8	2.5	0.03
Standard 28	2.5	2.2	0.06
Standard 32	2.5	2.4	0.08

Appendix C

Table 20: NAVAIR Standard 22 – Fastener distance, notch size and orientation.

Sample ID	Fastener Number	Bottom of Top	Top of Bottom	Size in (+/- 0.015)	Size mm (+/- 0.38)	Orientation CW from Lap-Joint Edge	Distance from Lap-Joint Edge (mm)
Standard 22	1		X	0.13	3.3	135°	8.520
	2	X		0.075	1.91	270°	8.520
	3		X	0.07	1.78	225°	8.248
	4		X	0.035	0.89	90°	8.338
	5		X	0.215	5.46	45°	8.338
	6	X	X	.130/.035	3.30/0.89	90°/315°	8.248
	7		X	0.08	2.03	270°	6.627
	8						6.985
	9		X	0.06	1.52	225°	7.075
	10		X	0.11	2.79	135°	6.896
	11		X	0.035	0.89	90°	6.627
	12	X	X	.215/.130	5.46/3.30	45°/225°	6.716
	13		X	0.06	1.52	270°	6.429
	14		X	0.2	5.08	315°	6.071
	15						7.679
	16						6.607
	17						6.696
	18						6.696
	19	X		0.11	2.79	225°	6.607
	20						7.143
	21	X	X	.035/.110	0.89/2.79	45°	6.875
	22						8.214
	23						7.857

Table 21: NAVAIR Standard 23 – Fastener distance, notch size and orientation.

Sample ID	Fastener Number	Bottom of Top	Top of Bottom	Size in (+/- 0.015)	Size mm (+/- 0.38)	Orientation CW from Lap-Joint Edge	Distance from Lap-Joint Edge (mm)
Standard 23	1	X	X	.075/.125	1.91/3.18	270° / 90°	7.646
	2						7.418
	3		X	0.21	5.33	270°	6.508
	4	X		0.125	3.18	45°	6.608
	5						6.521
	6		X	0.065	1.65	90°	7.040
	7						6.708
	8						6.671
	9		X	0.085	2.16	45°	7.631
	10	X		0.21	5.33	90°	7.123
	11						7.176
	12						6.787
	13		X	0.11	2.79	225°	6.970
	14						5.845
	15	X		0.035	0.89	270°	6.681
	16		X	0.2	5.08	135°	6.123
	17						6.172
	18		X	0.06	1.52	315°	7.221
	19	X		0.1	2.54	90°	6.774
	20		X	0.05	1.27	45°	6.842
	21						7.419
	22		X	0.125	3.18	270°	6.234
	23						6.341
	24	X	X	.045 /.095	1.14/2.41	315° / 135°	6.463

Table 22: NAVAIR Standard 24 – Fastener distance, notch size and orientation.

Sample ID	Fastener Number	Bottom of Top	Top of Bottom	Size in (+/- 0.015)	Size mm (+/- 0.38)	Orientation CW from Lap-Joint Edge	Distance from Lap-Joint Edge (mm)
Standard 24	1		X	0.14	3.56	225°	5.738
	2						5.836
	3	X	X	0.08/0.215	2.03/5.46	225° / 90°	5.928
	4		X	0.045	1.14	270°	6.241
	5	X		0.135	3.43	270°	6.004
	6		X	0.06	1.52	45°	6.247
	7		X	0.035	0.89	135°	6.010
	8		X	0.085	2.16	270°	6.320
	9	X	X	0.205/Unkn	5.21/Unkn	90° / 225°	7.009
	10						6.637
	11		X	0.115	2.92	90°	6.844
	12		X	0.085	2.16	270°	7.557
	13		X	0.2	5.08	45°	6.912
	14						6.732
	15	X	X	0.05/0.135	1.27/3.43	270° / 135°	6.890
	16						6.848
	17		X	0.065	1.65	270°	7.339
	18						6.264
	19	X	X	0.105/0.11	2.67/2.79	270° / 90°	6.694
	20						6.621
	21						6.675
	22						6.616

Table 23: NAVAIR Standard 25 – Fastener distance, notch size and orientation.

Sample ID	Fastener Number	Bottom of Top	Top of Bottom	Size in (+/- 0.015)	Size mm (+/- 0.38)	Orientation CW from Lap-Joint Edge	Distance from Lap-Joint Edge (mm)
Standard 25	1		X	0.135	3.43	45°	6.430
	2						6.423
	3	X	X	.075/.215	1.91/5.46	225° / 90°	6.673
	4						6.889
	5		X	0.03	0.76	315°	6.880
	6						6.781
	7	X	X	.135/.045	3.43/1.14	90° / 315°	6.650
	8						6.508
	9		X	0.03	0.76	90°	6.522
	10	X		0.22	5.59	135°	6.879
	11		X	0.085	2.16	270°	6.419
	12						6.551
	13		X	0.115	2.92	225°	6.609
	14		X	0.23	5.84	270°	6.292
	15	X		0.05	1.27	45°	6.848
	16						6.232
	17		X	0.08	2.03	135°	6.319
	18						6.234
	19		X	0.055	1.4	270°	6.332
	20	X	X	.1/.035	2.54/0.89	135° / 270°	6.793
	21		X	0.12	3.05	90°	7.998
	22		X	0.105	2.67	225°	6.563

Table 24: NAVAIR Standard 26 – Fastener distance, notch size and orientation.

Sample ID	Fastener Number	Bottom of Top	Top of Bottom	Size in (+/- 0.015)	Size mm (+/- 0.38)	Orientation CW from Lap-Joint Edge	Distance from Lap-Joint Edge (mm)
Standard 26	1						7.634
	2	X		0.065	1.65	135°	7.832
	3		X	0.24	6.1	270°	6.845
	4						7.467
	5		X	0.125	3.18	45°	8.269
	6						7.150
	7		X	0.05	1.27	225°	7.048
	8	X		0.13	3.3	225°	5.647
	9		X	0.04	1.02	270°	5.704
	10		X	0.085	2.16	45°	6.613
	11						6.617
	12	X		0.215	5.46	315°	7.200
	13		X	0.12	3.05	225°	7.330
	14	X	X	.030/.210	0.76/5.33	270°/90°	6.644
	15						6.890
	16		X	0.03	0.76	270°	7.591
	17						7.673
	18	X		0.08	2.03	225°	6.748
	19						6.839
	20	X	X	.1/.06	2.54/1.52	90°/225°	6.915
	21		X	0.155	3.94	315°	6.779
	22	X	X	.05/.110	1.27/2.79	45°/45°	7.260

Table 25: NAVAIR Standard 28 – Fastener distance, notch size and orientation.

Sample ID	Fastener Number	Bottom of Top	Top of Bottom	Size in (+/- 0.015)	Size mm (+/- 0.38)	Orientation CW from Lap-Joint Edge	Distance from Lap-Joint Edge (mm)
Standard 28	1						5.885
	2		X	0.125	3.18	135°	5.960
	3		X	0.21	5.33	315°	5.998
	4						6.020
	5						4.814
	6		X	0.04	1.02	90°	5.450
	7		X	0.03	0.76	45°	5.122
	8	X	X	.075/.055	1.91/1.40	90°/270°	6.409
	9	X	X	.130/.085	3.40/2.16	270°/135°	5.111
	10						5.352
	11						5.303
	12	X		0.215	5.46	45°	5.435
	13		X	0.11	2.79	225°	5.469
	14	X	X	.080/.210	2.03/5.33	225°/90°	5.538
	15		X	0.13	3.3	270°	5.552
	16						5.715
	17	X	X	.055/.05	1.40/1.27	135°/225°	7.821
	18						6.351
	19	X	X	.130/.110	3.30/2.79	315°/135°	5.949
	20						7.455
	21						6.052
	22						5.850

Table 26: NAVAIR Standard 32 – Fastener distance, notch size and orientation.

Sample ID	Fastener Number	Bottom of Top	Top of Bottom	Size in (+/- 0.015)	Size mm (+/- 0.38)	Orientation CW from Lap-Joint Edge	Distance from Lap-Joint Edge (mm)
Standard 32	1	X	X	.05/.125	1.27/3.18	90°/315°	7.646
	2						7.418
	3		X	0.03	0.76	270°	6.508
	4	X		0.075	1.91	45°	6.608
	5		X	0.19	4.83	225°	6.521
	6						7.040
	7		X	0.075	1.91	45°	6.708
	8	X		0.1	2.54	225°	6.671
	9		X	0.1	2.54	45°	7.631
	10						7.123
	11		X	0.03	0.76	270°	7.176
	12		X	0.21	5.33	90°	6.787
	13	X		0.04	1.02	135°	6.970
	14	X		0.14	3.56	315°	5.845
	15		X	0.045	1.14	225°	6.681
	16		X	0.075	1.91	45°	6.123
	17						6.172
	18		X	0.095	2.41	90°	7.221
	19						6.774
	20		X	0.125	3.18	315°	6.842
	21						7.419
	22	X	X	.225/.05	5.71/1.27	270°/225°	6.234
	23						6.341
	24						6.463

Appendix D

Table 19: Measured results - before and after fastener to lap-joint edge distance revision.

		Detection Rate			False Call Rate
	Sample ID	Top Layer	2nd Layer	Total	
Standard 22		100.0%	100.0%	100.0%	3.8%
Standard 23		83.3%	91.0%	86.4%	6.0%
	Distance Revision	85.0%	98.0%	92.1%	6.0%
Standard 24		100.0%	87.7%	88.6%	5.0%
	Distance Revision	100.0%	100.0%	100.0%	2.5%
Standard 25		98.0%	96.2%	96.0%	7.1%
	Distance Revision	100.0%	96.9%	97.3%	5.7%
Standard 26		72.9%	73.6%	70.0%	7.1%
	Distance Revision	95.7%	88.2%	90.7%	2.9%
Standard 28		71.7%	76.4%	78.3%	3.0%
	Distance Revision	83.3%	84.5%	85.8%	2.0%
Standard 32		88.3%	82.5%	82.5%	6.3%
	Distance Revision	91.7%	88.3%	88.1%	6.3%
Average		87.7%	86.8%	86.0%	5.5%
	Distance Revision	93.7%	93.7%	93.4%	4.2%
Green highlight indicates increase in detection rate or decrease in false call rate					

Appendix E

Table 28: Simulation results for NAVAIR samples with number of notches ranging from one to four, 10% false call, and data fraction 95%.

Sample ID		Number of Notches in Simulation			
		C1	C2	C3	C4
Standard 22	Detection	99.73%	99.13%	99.07%	99.00%
	False Call	9.21%	8.29%	7.31%	6.26%
Standard 23	Detection	85.57%	84.36%	83.94%	82.29%
	False Call	11.39%	10.56%	9.62%	9.02%
Standard 24	Detection	93.43%	93.86%	92.79%	91.07%
	False Call	8.36%	7.48%	6.84%	5.91%
Standard 25	Detection	97.20%	97.00%	96.00%	94.93%
	False Call	9.85%	8.82%	8.00%	6.71%
Standard 26	Detection	81.07%	78.53%	76.80%	75.13%
	False Call	9.66%	8.94%	8.66%	8.10%
Standard 28	Detection	82.92%	80.08%	80.92%	78.92%
	False Call	8.21%	7.40%	6.29%	5.78%
Standard 32	Detection	86.94%	86.25%	86.83%	86.38%
	False Call	10.17%	9.22%	8.02%	6.94%

Table 20: Simulation results for NAVAIR samples with number of notches ranging from one to four, 10% false call, and data fraction 100%.

Sample ID		Number of Notches in Simulation			
		C1	C2	C3	C4
Standard 22	Detection	99.5%	97.2%	95.9%	92.7%
	False Call	3.9%	3.4%	3.1%	2.7%
Standard 23	Detection	84.1%	79.5%	77.8%	74.2%
	False Call	6.6%	6.1%	5.5%	5.3%
Standard 24	Detection	94.0%	90.6%	91.1%	86.4%
	False Call	4.0%	3.6%	3.2%	2.8%
Standard 25	Detection	96.7%	95.9%	92.7%	89.9%
	False Call	5.1%	4.5%	3.7%	3.1%
Standard 26	Detection	75.4%	72.3%	69.1%	66.2%
	False Call	6.5%	6.2%	5.7%	5.5%
Standard 28	Detection	82.4%	78.5%	73.4%	71.8%
	False Call	3.8%	3.4%	2.9%	2.6%
Standard 32	Detection	86.4%	83.0%	83.3%	78.4%
	False Call	4.9%	4.2%	3.6%	3.2%

Appendix F

Table 30: POD vs. size results for simulations with a sample size of 80 and 5% false call rate.

No. of Notches	Data Fraction 95%				Data Fraction 97.5%				Data Fraction 100%			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
a_{50} (inch)	0.034	0.035	0.372	0.040	0.033	0.034	0.037	0.038	0.030	0.034	0.038	0.041
a_{90} (inch)	0.088	0.091	0.096	0.106	0.085	0.089	0.095	0.101	0.077	0.089	0.100	0.116
$a_{90/95}$ (inch)	0.137	0.144	0.151	0.169	0.133	0.140	0.147	0.160	0.120	0.140	0.158	0.192

Table 31: POD vs. size results for simulations with a sample size of 80 and 10% false call rate.

No. of Notches	Data Fraction 95%				Data Fraction 97.5%				Data Fraction 100%			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
a_{50} (inch)	0.028	0.029	0.031	0.033	0.026	0.027	0.030	0.030	0.025	0.027	0.030	0.034
a_{90} (inch)	0.074	0.075	0.080	0.085	0.069	0.069	0.076	0.076	0.065	0.073	0.083	0.094
$a_{90/95}$ (inch)	0.114	0.116	0.124	0.133	0.106	0.107	0.117	0.118	0.100	0.113	0.132	0.152

Table 32: POD vs. size results for simulations with a sample size of 80 and 15% false call rate.

No. of Notches	Data Fraction 95%				Data Fraction 97.5%				Data Fraction 100%			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
a_{50} (inch)	0.024	0.024	0.026	0.027	0.022	0.023	0.026	0.025	0.021	0.024	0.027	0.029
a_{90} (inch)	0.062	0.064	0.067	0.071	0.059	0.059	0.066	0.067	0.057	0.064	0.073	0.083
$a_{90/95}$ (inch)	0.096	0.099	0.104	0.110	0.092	0.091	0.101	0.103	0.088	0.100	0.116	0.135

Table 33: POD vs. size results for simulations with a sample size of 40 and 5% false call rate.

No. of Notches	Data Fraction 95%				Data Fraction 97.5%				Data Fraction 100%			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
a_{50} (inch)	0.042	0.046	0.055	0.066	0.040	0.047	0.058	0.067	0.035	0.041	0.050	0.058
a_{90} (inch)	0.109	0.129	0.155	0.179	0.106	0.131	0.156	0.188	0.093	0.118	0.148	0.183
$a_{90/95}$ (inch)	0.173	0.216	0.271	0.313	0.169	0.220	0.262	0.340	0.147	0.201	0.268	0.364

Table 34: POD vs. size results for simulations with a sample size of 40 and 10% false call rate.

No. of Notches	Data Fraction 95%				Data Fraction 97.5%				Data Fraction 100%			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
a_{50} (inch)	0.035	0.037	0.043	0.051	0.030	0.034	0.041	0.045	0.029	0.034	0.039	0.046
a_{90} (inch)	0.092	0.104	0.123	0.141	0.082	0.097	0.111	0.133	0.079	0.096	0.120	0.145
$a_{90/95}$ (inch)	0.144	0.171	0.210	0.239	0.131	0.160	0.181	0.234	0.125	0.157	0.212	0.275

Table 35: POD vs. size results for simulations with a sample size of 40 and 15% false call rate.

No. of Notches	Data Fraction 95%				Data Fraction 97.5%				Data Fraction 100%			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
a_{50} (inch)	0.028	0.030	0.033	0.040	0.025	0.028	0.033	0.036	0.024	0.029	0.033	0.038
a_{90} (inch)	0.077	0.086	0.095	0.112	0.070	0.081	0.090	0.104	0.069	0.082	0.102	0.124
$a_{90/95}$ (inch)	0.121	0.139	0.157	0.185	0.111	0.131	0.145	0.177	0.110	0.132	0.180	0.232

Appendix G

The plot shows Probability of Detection (POD | a) vs. Size, a (inches) on a logarithmic x-axis. Annotations on the figure include:

$a_{50} = 0.02388$

$a_{90} = 0.06248$

$a_{90/95} = 0.09635$

$$POD(a) = \Phi\left(\frac{\log(a) - \mu}{\sigma}\right)$$

link function $= \Phi^{-1} = $ logit

$\hat{\mu} = -3.735$

$\hat{\sigma} = 0.43767$

$n_{hits} = 7258$

$n_{total} = 7878$

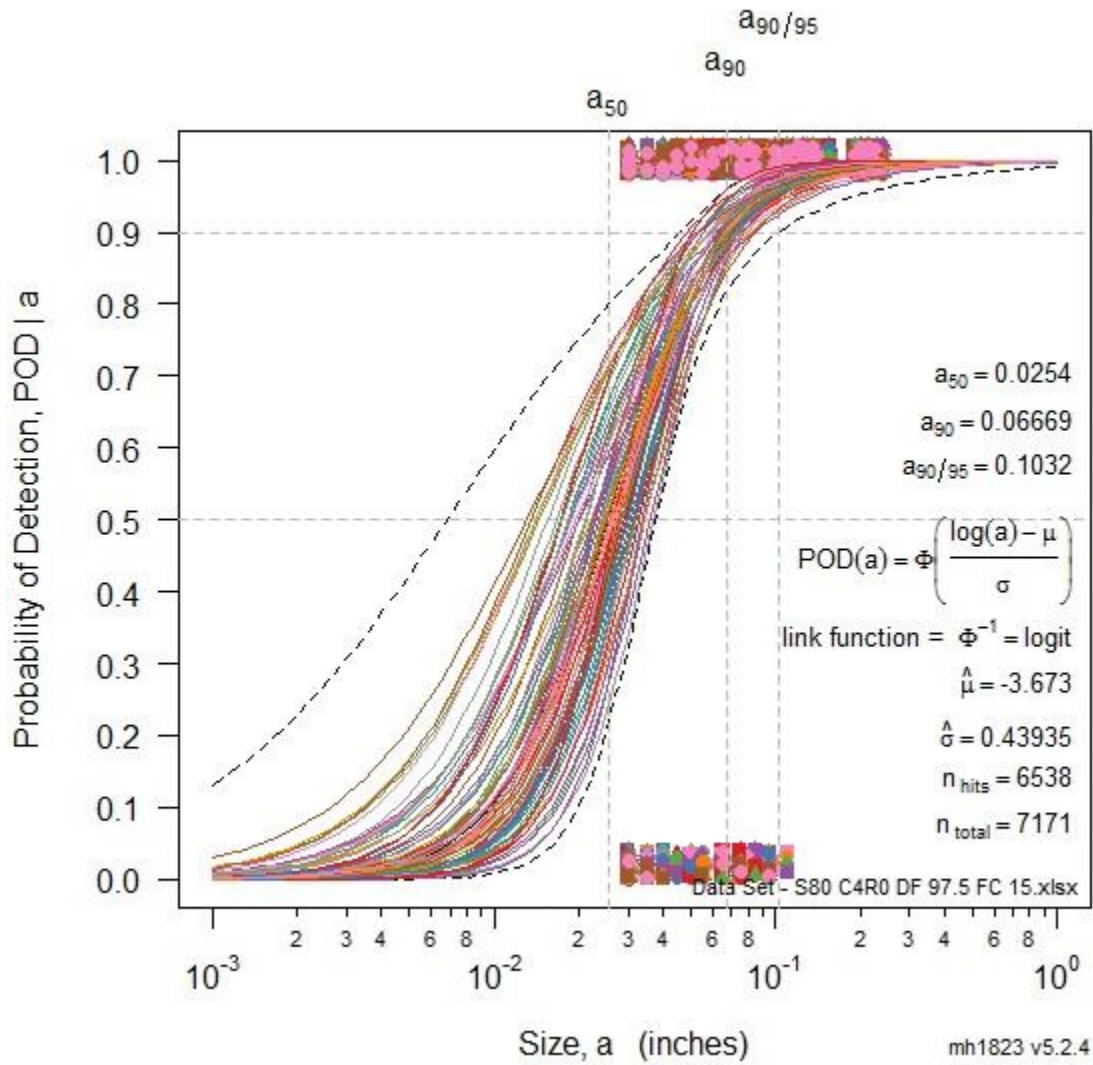Data Set - S80 C1R0 DF 95 FC 15.xlsx

mh1823 v5.2.4
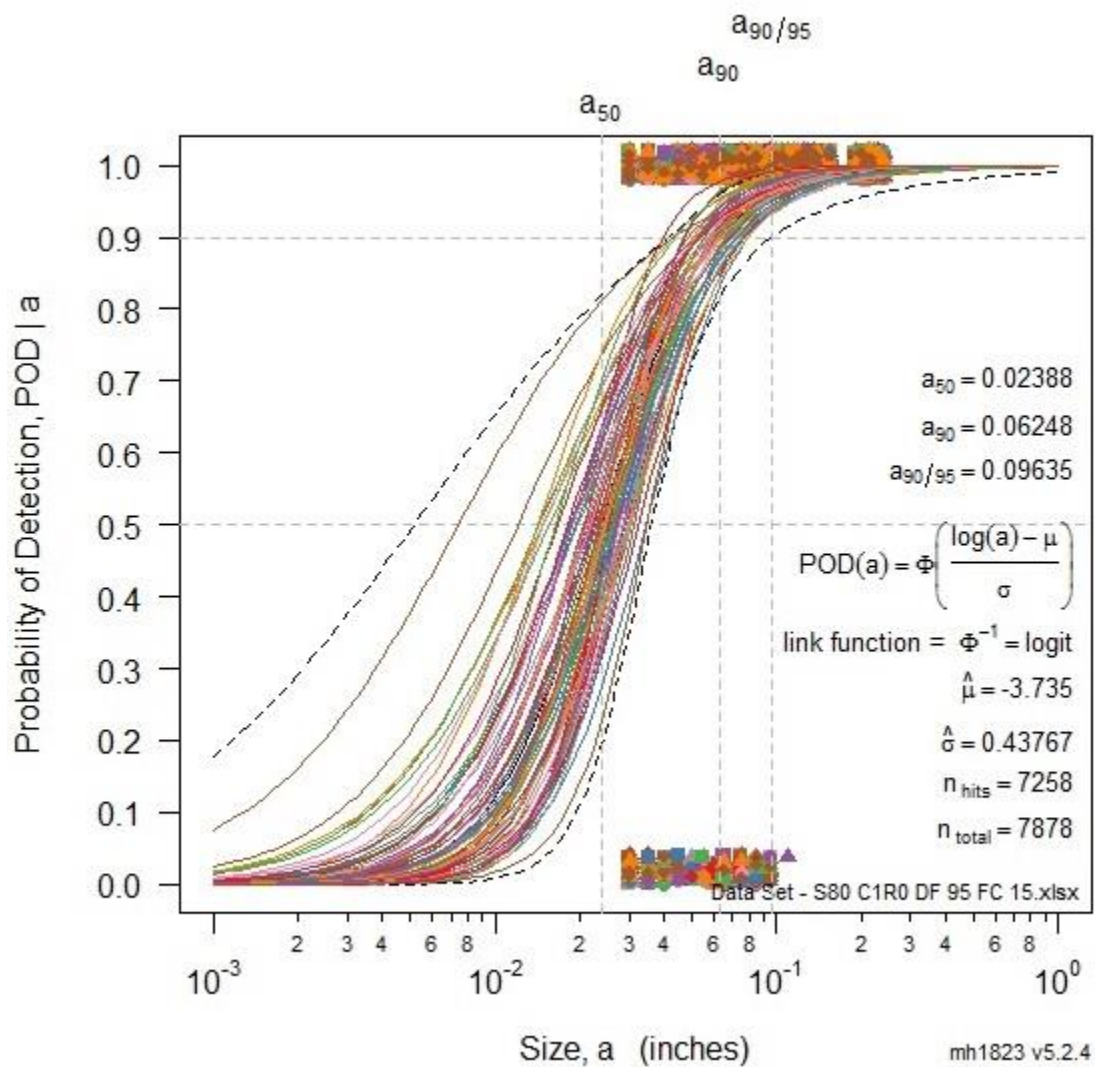
Figure 56: POD vs. size repeat measurements plot - 80 samples, 1 notch, Data Fraction 95%, False Call 15%.
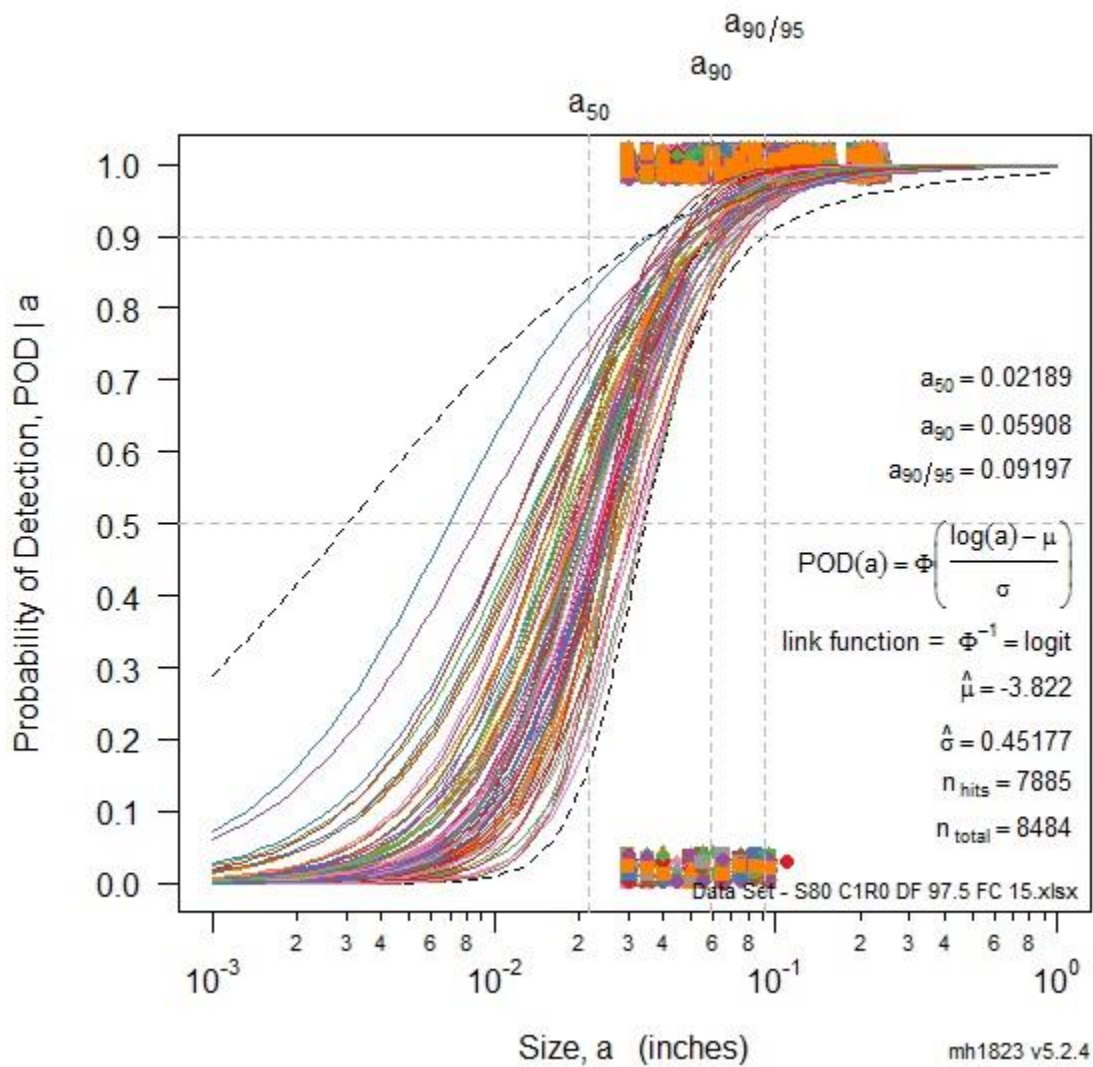
Figure 57: POD vs. size repeat measurements plot - 80 samples, 1 notch, Data Fraction 97.5%, False Call 15%.

$a_{90/95}$

a_{90}

a_{50}

$a_{50} = 0.02115$

$a_{90} = 0.0568$

$a_{90/95} = 0.08823$

$$POD(a) = \Phi\left(\frac{\log(a) - \mu}{\sigma}\right)$$

link function $= \Phi^{-1} = $ logit

$\hat{\mu} = -3.856$

$\hat{\sigma} = 0.4497$

$n_{hits} = 7737$

$n_{total} = 8282$

Data Set - S80 C1R0 DF 100 FC 15.xlsx

Size, a (inches)

mh1823 v5.2.4
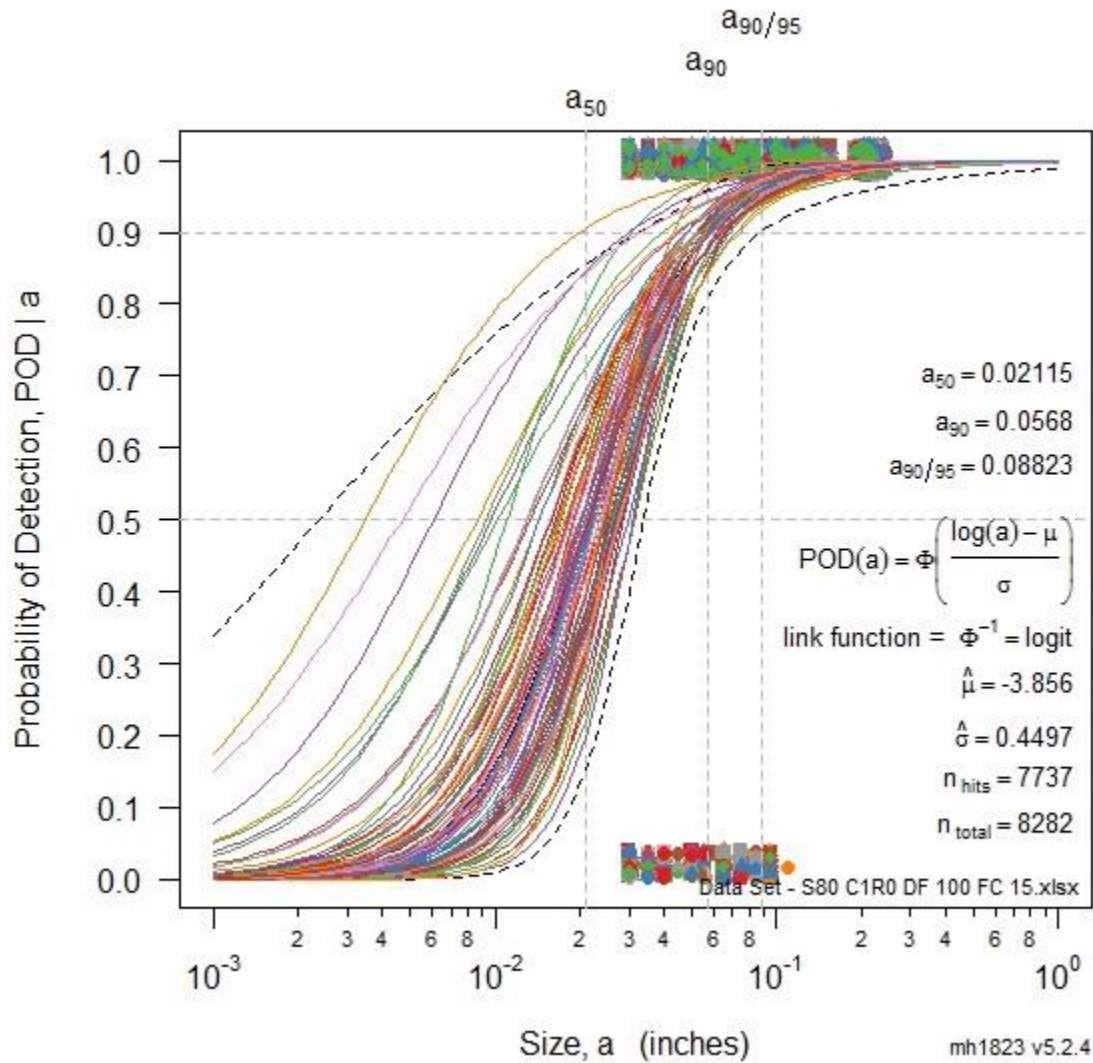
Figure 58: POD vs. size repeat measurements plot - 80 samples, 1 notch, Data Fraction 100%, False Call 15%.

$a_{90/95}$
a_{90}
a_{50}

$a_{50} = 0.02385$
$a_{90} = 0.06366$
$a_{90/95} = 0.09898$

$$POD(a) = \Phi\left(\frac{\log(a) - \mu}{\sigma}\right)$$

link function $= \Phi^{-1} = $ logit
$\hat{\mu} = -3.736$
$\hat{\sigma} = 0.44688$
$n_{hits} = 7148$
$n_{total} = 7777$

Data Set - S80 C2R0 DF 95 FC 15.xlsx

mh1823 v5.2.4
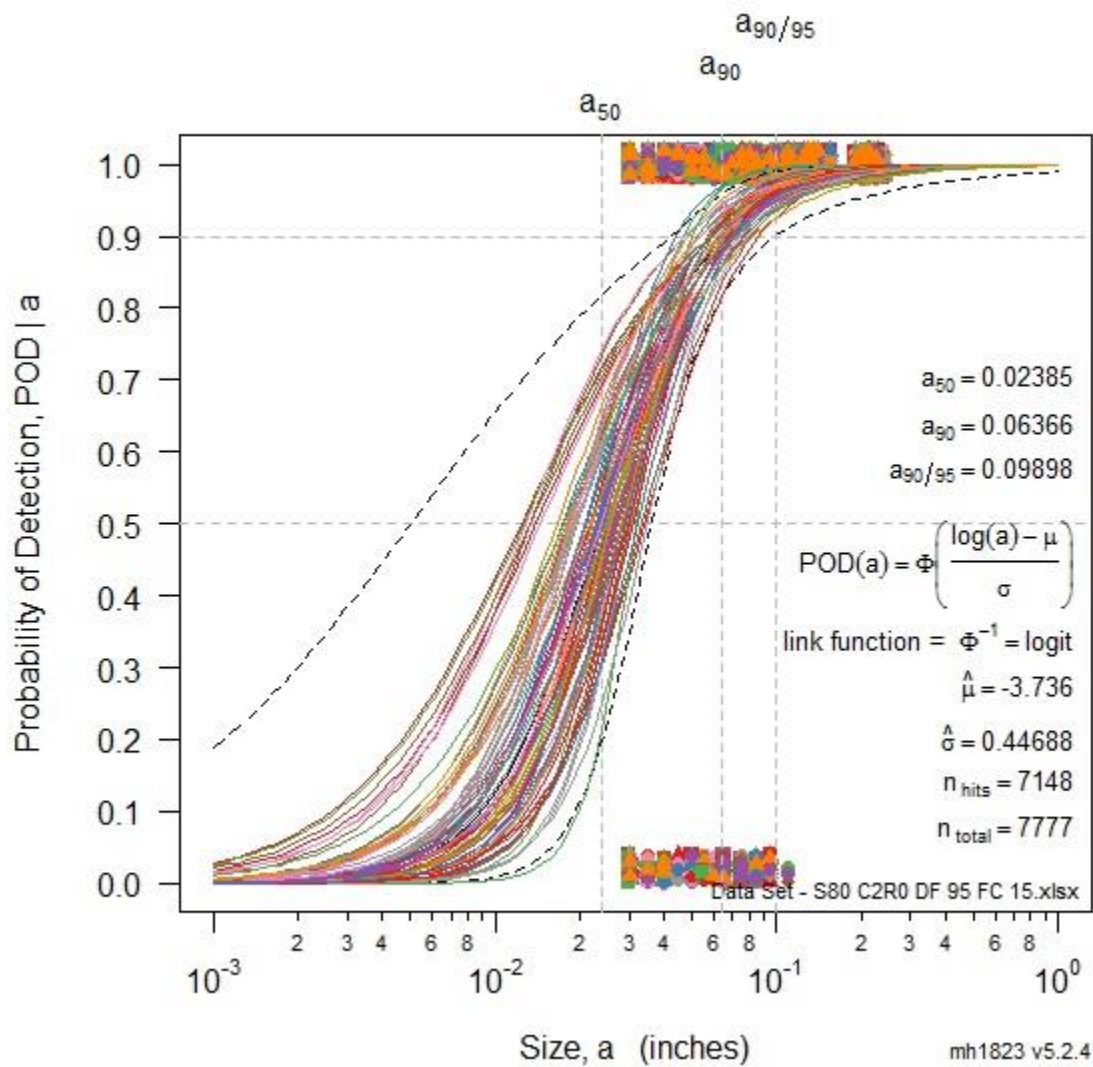
Figure 59: POD vs. size repeat measurements plot - 80 samples, 2 notches, Data Fraction 95%, False Call 15%.
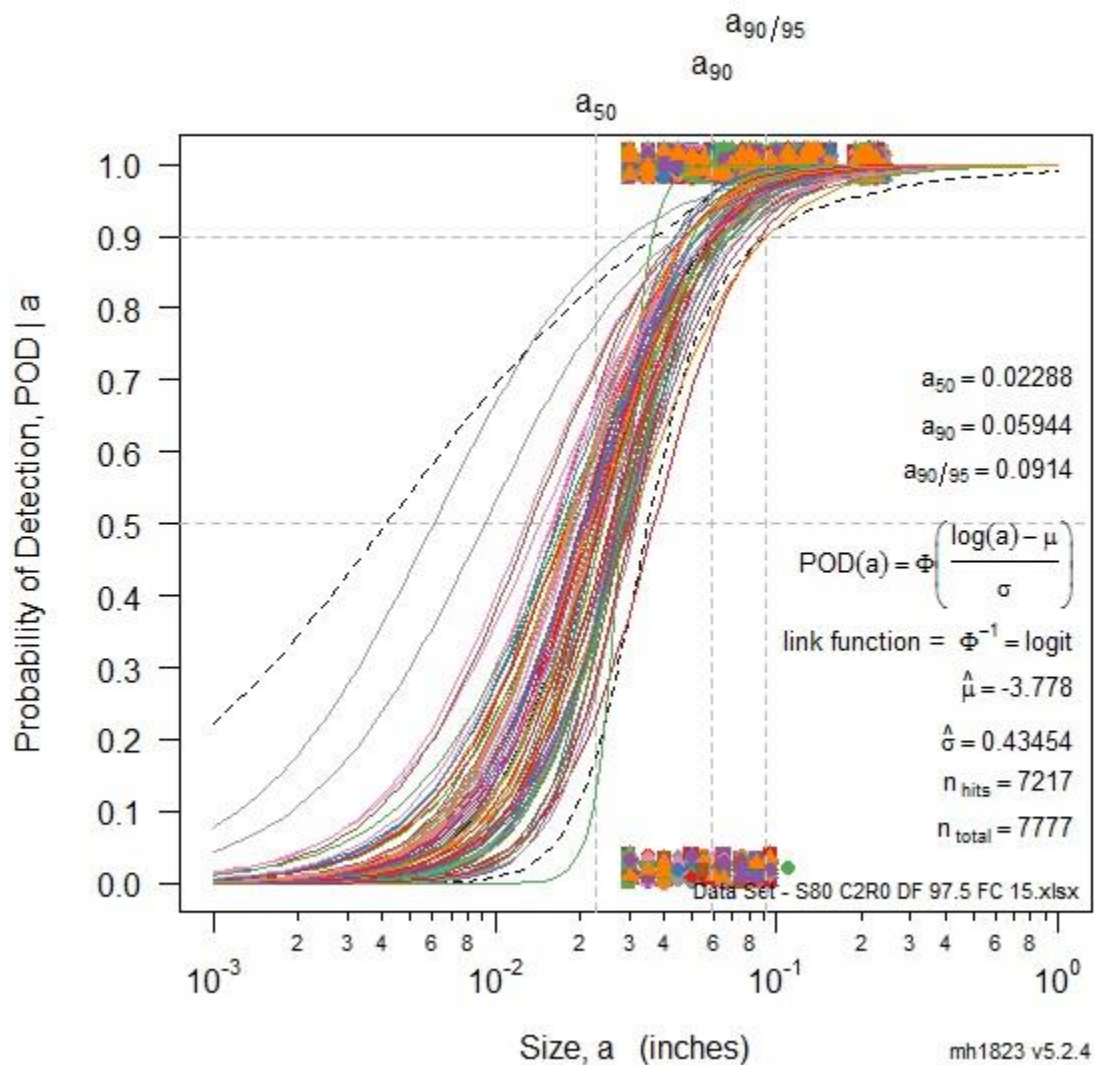
Figure 60: POD vs. size repeat measurements plot - 80 samples, 2 notches, Data Fraction 97.5%, False Call 15%.
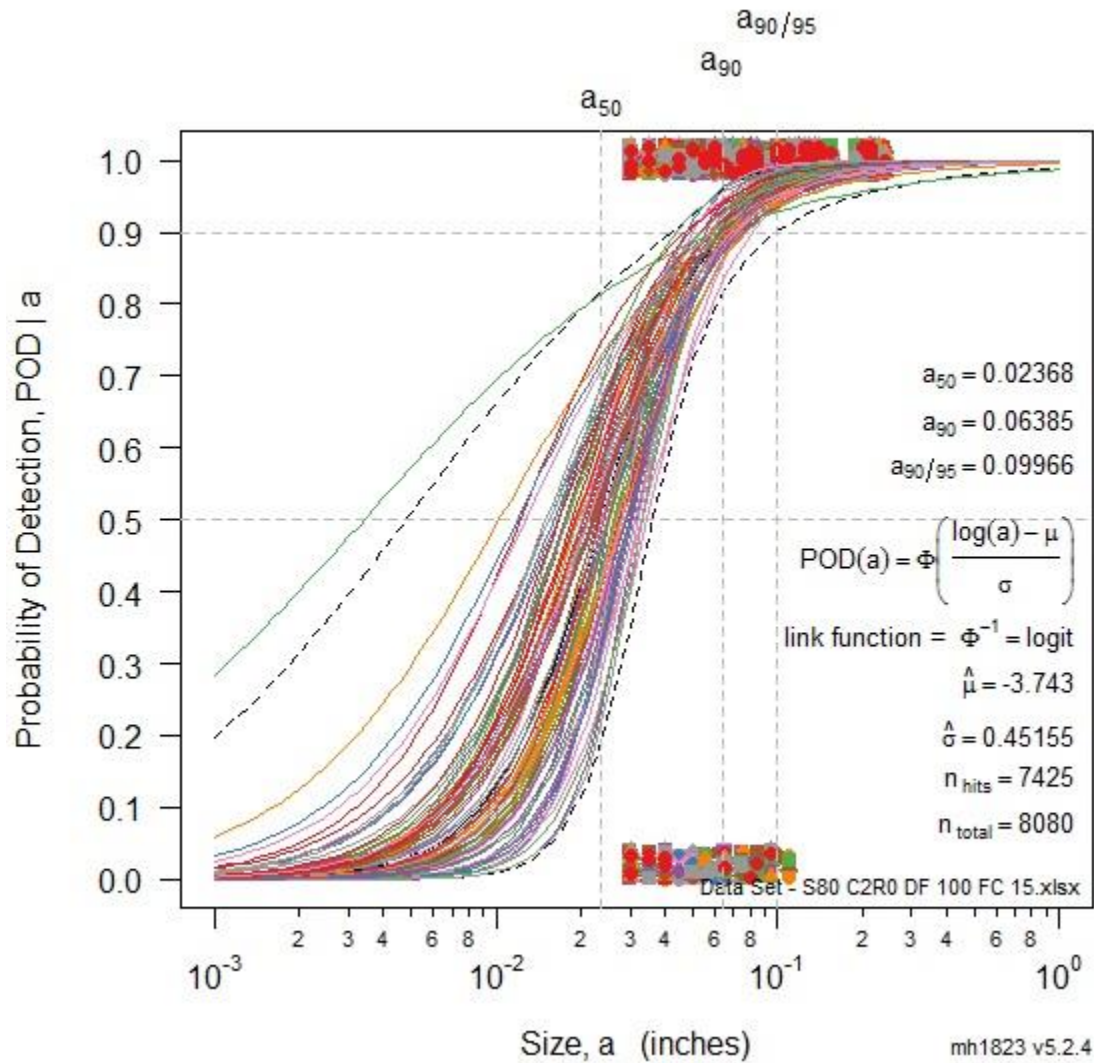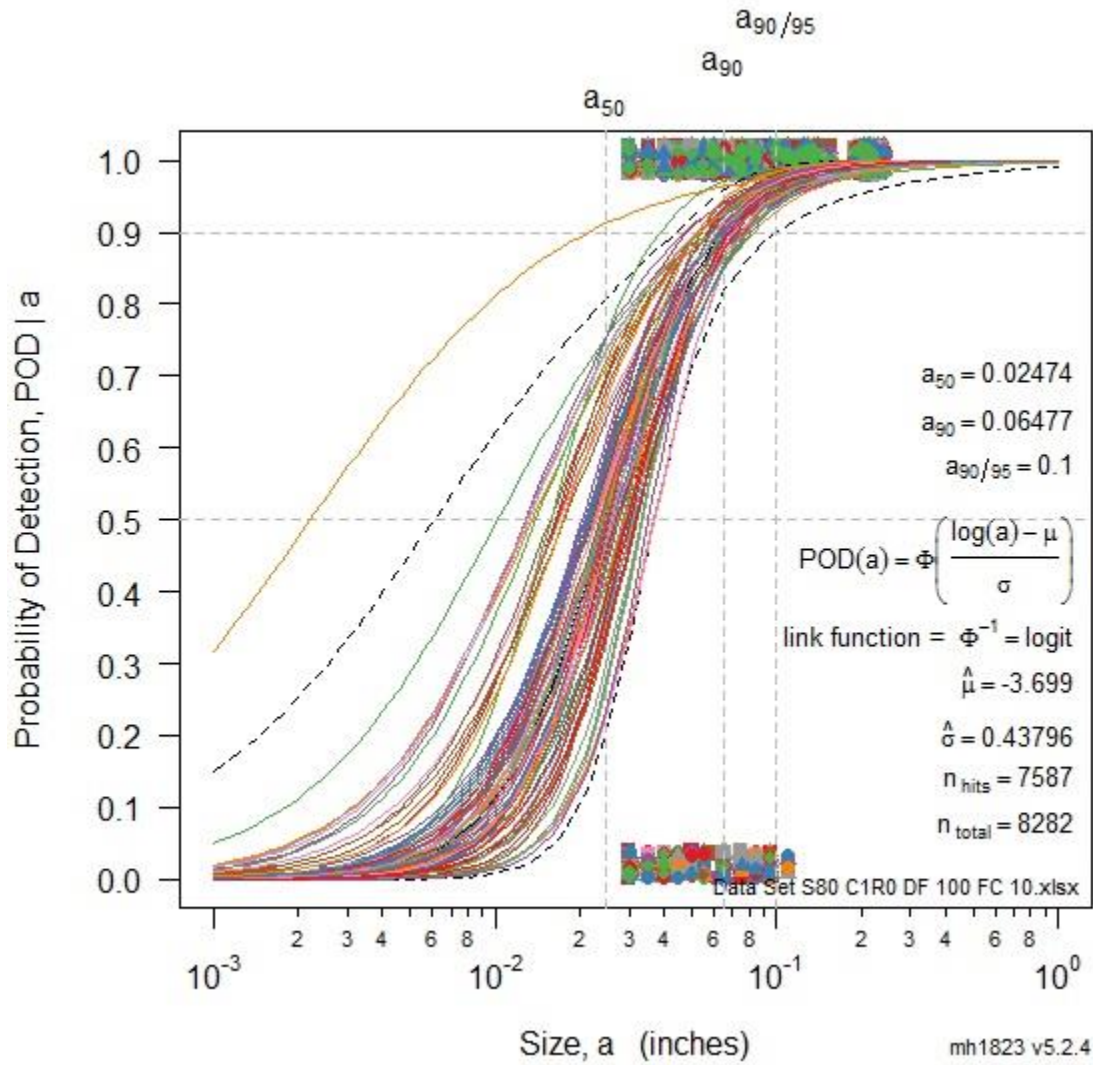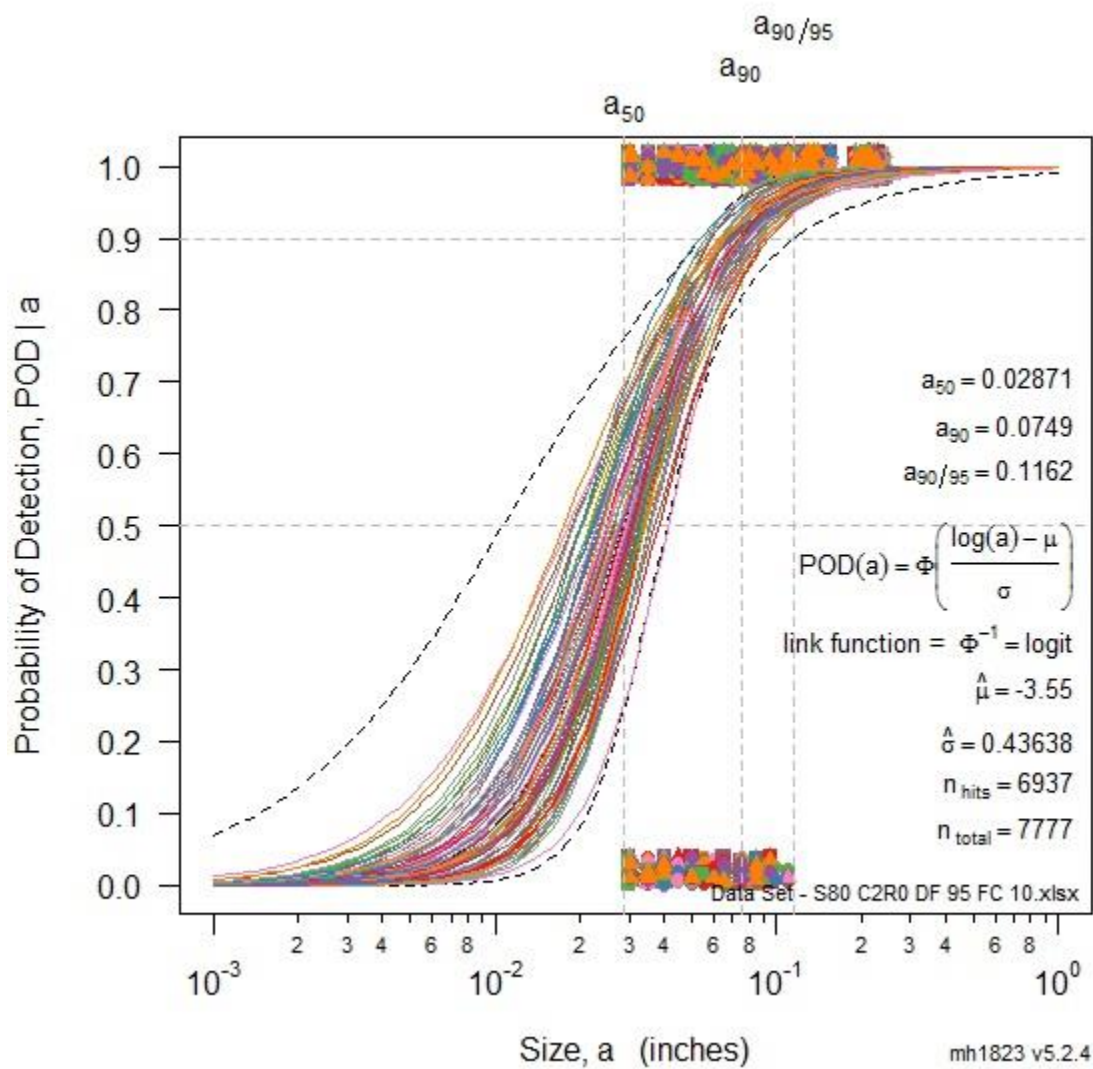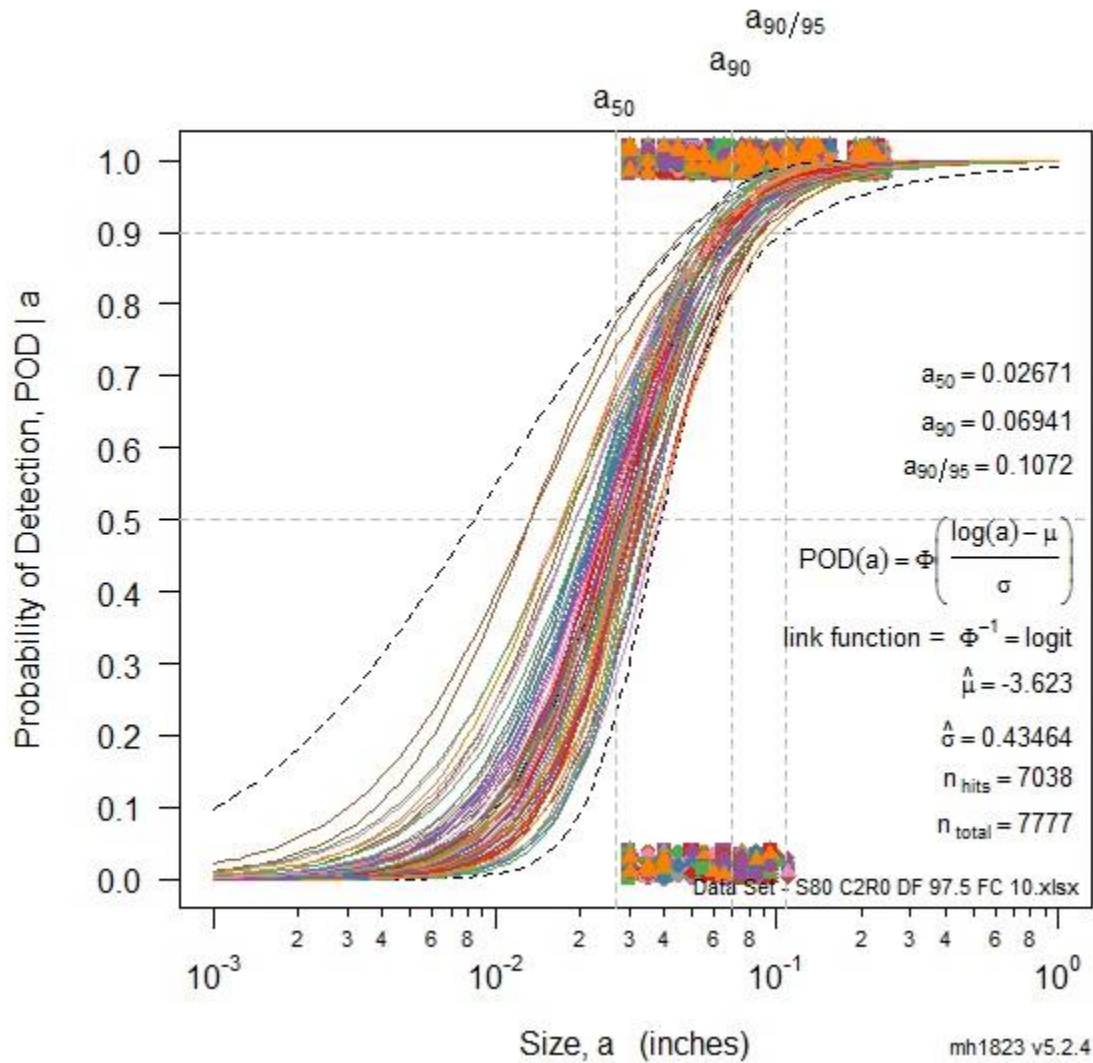
Figure 61: POD vs. size repeat measurements plot - 80 samples, 2 notches, Data Fraction 100%, False Call 15%.

The figure shows a plot of Probability of Detection, POD | a, versus Size, a (inches).

Labels on plot: a_{50}, a_{90}, $a_{90/95}$

$a_{50} = 0.02587$
$a_{90} = 0.06747$
$a_{90/95} = 0.1042$

$$POD(a) = \Phi\left(\frac{\log(a) - \mu}{\sigma}\right)$$

link function $= \Phi^{-1} =$ logit
$\hat{\mu} = -3.654$
$\hat{\sigma} = 0.43616$
$n_{hits} = 7534$
$n_{total} = 8282$

Data Set - S80 C3R0 DF 95 FC 15.xlsx

mh1823 v5.2.4
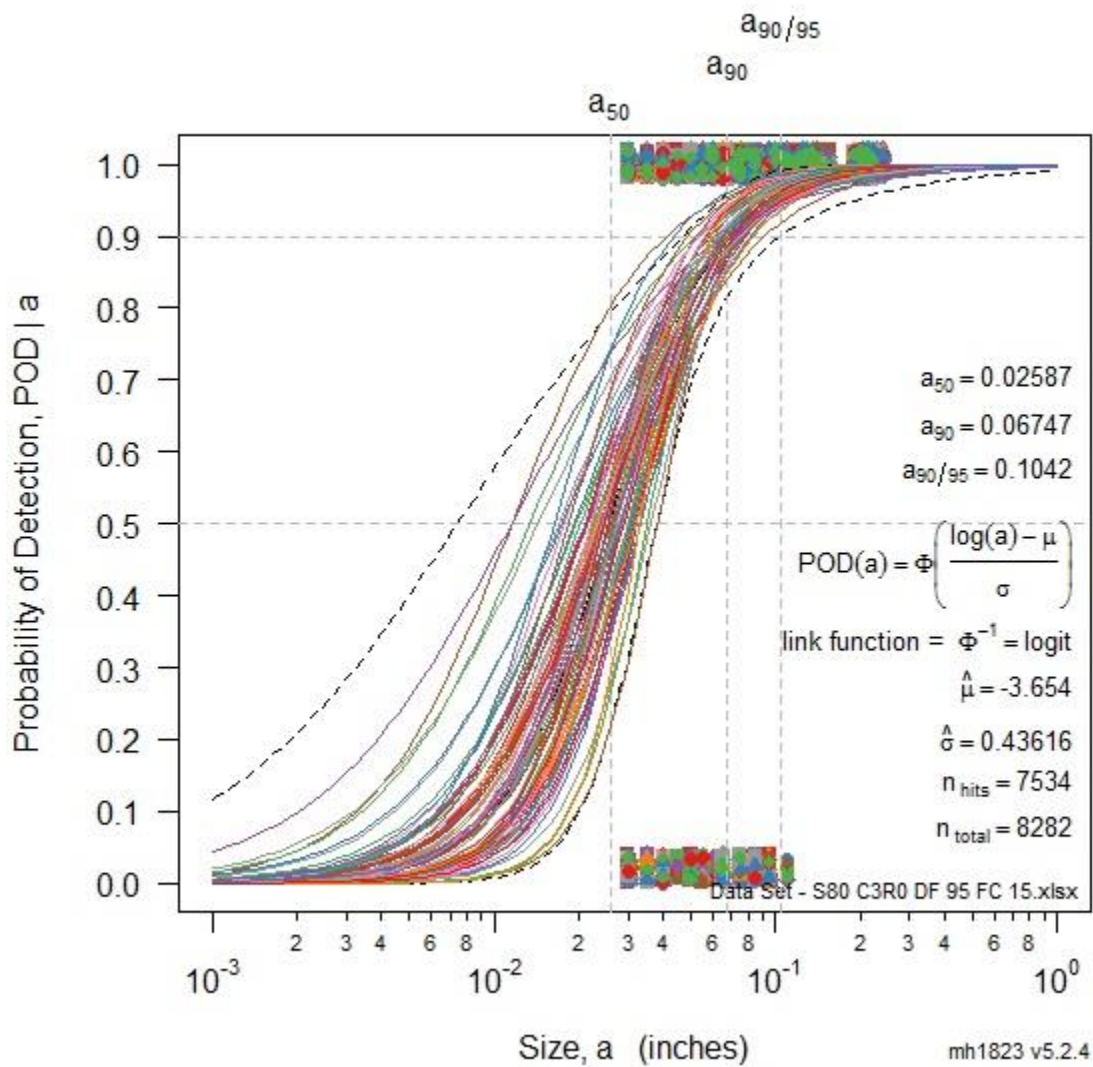
Figure 62: POD vs. size repeat measurements plot - 80 samples, 3 notches, Data Fraction 95%, False Call 15%.
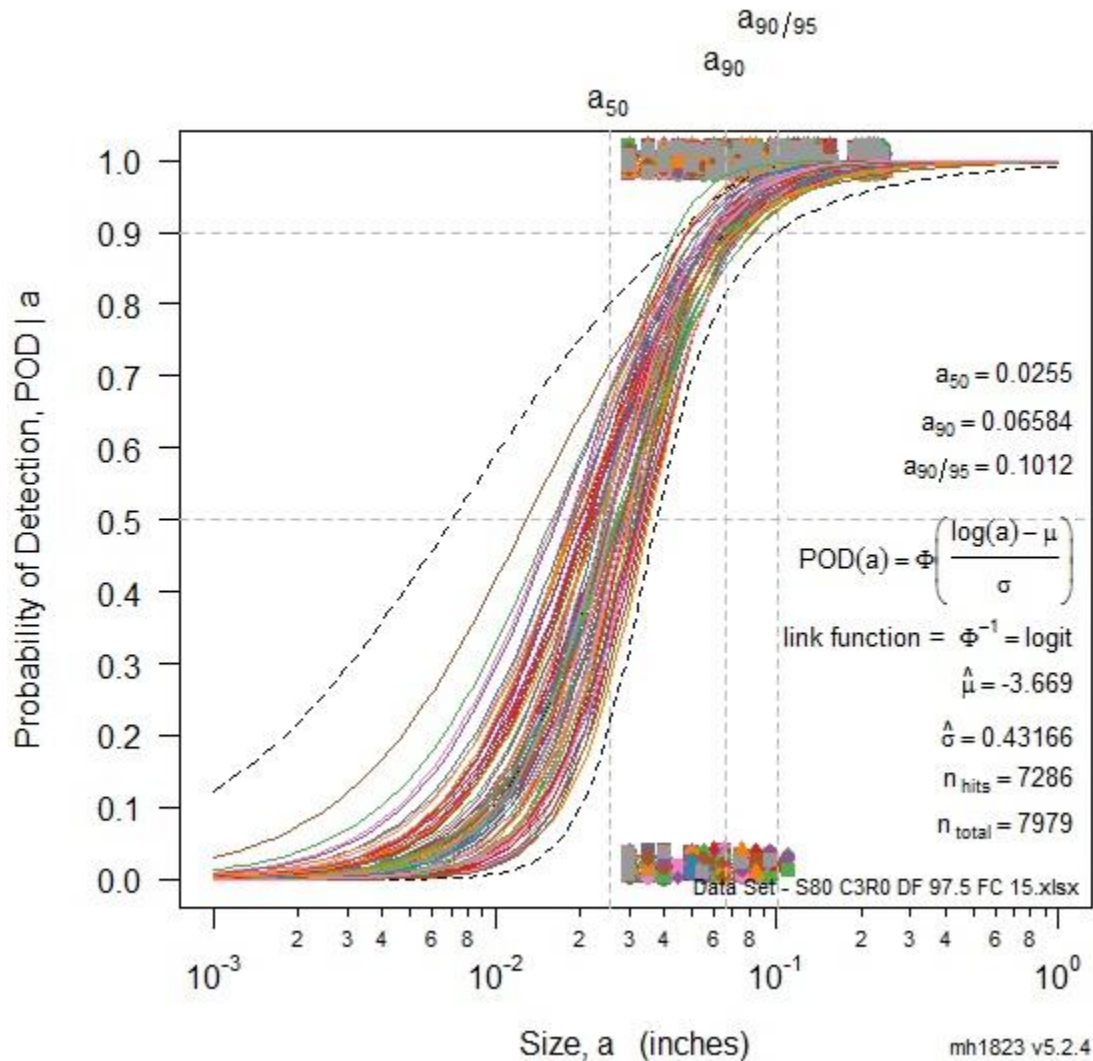
Figure 63: POD vs. size repeat measurements plot - 80 samples, 3 notches, Data Fraction 97.5%, False Call 15%.
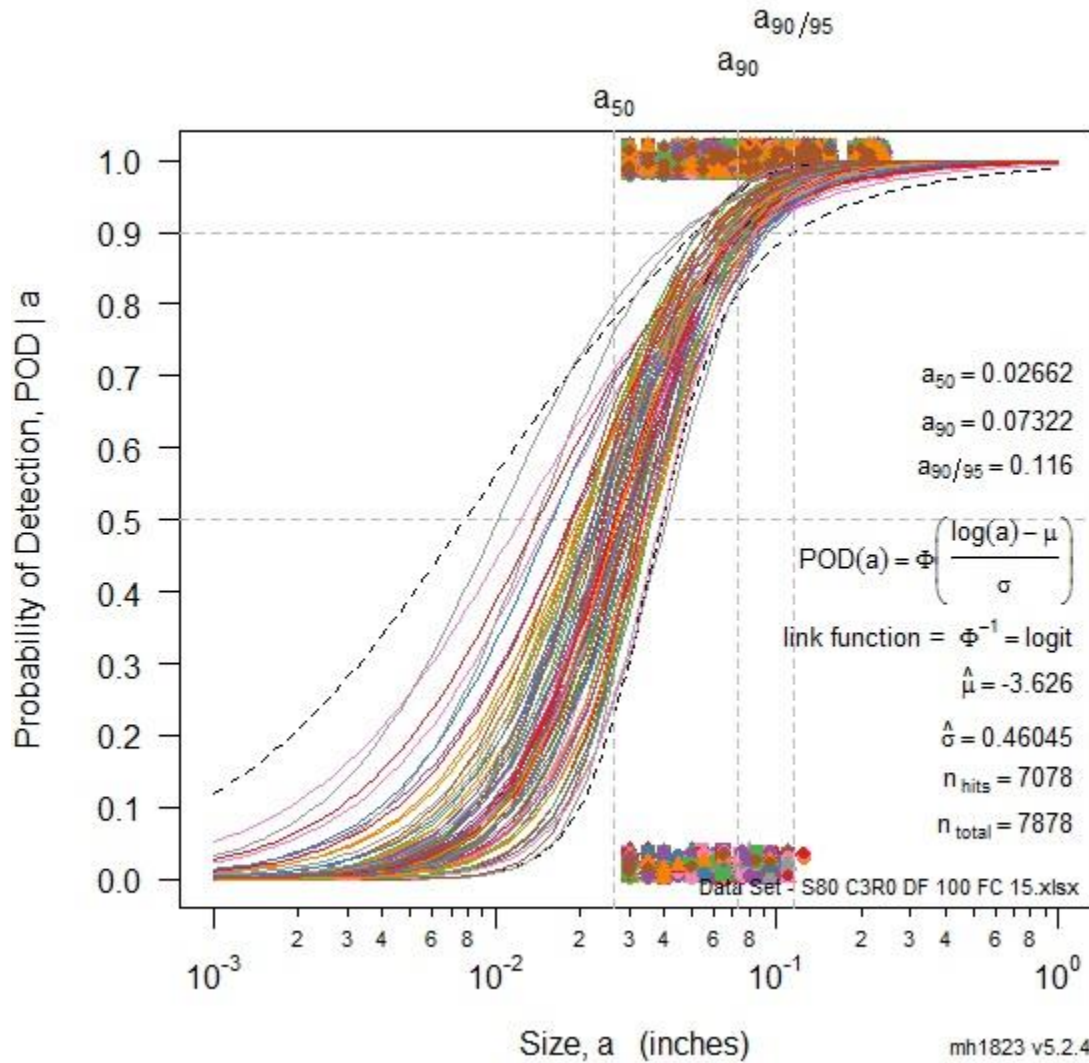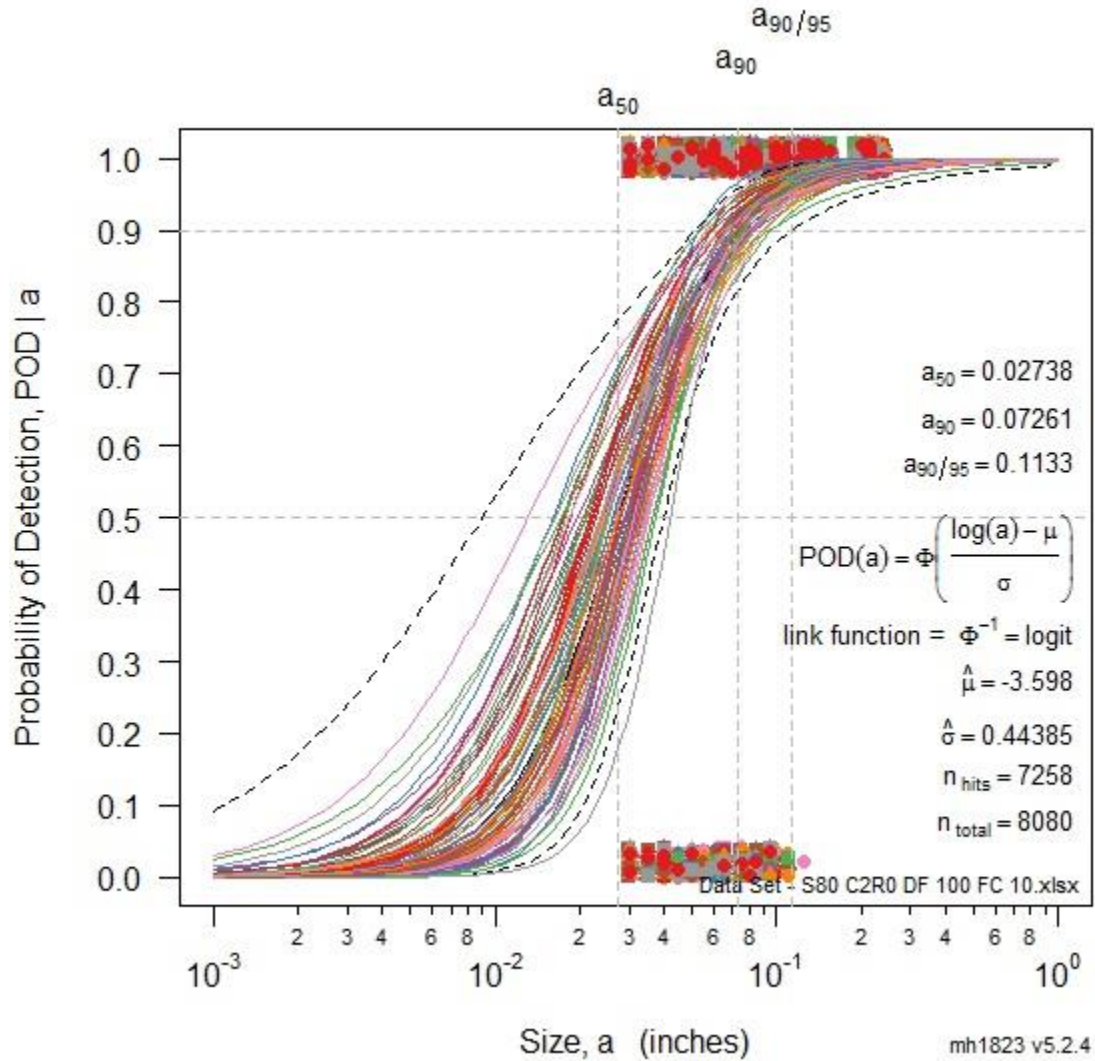
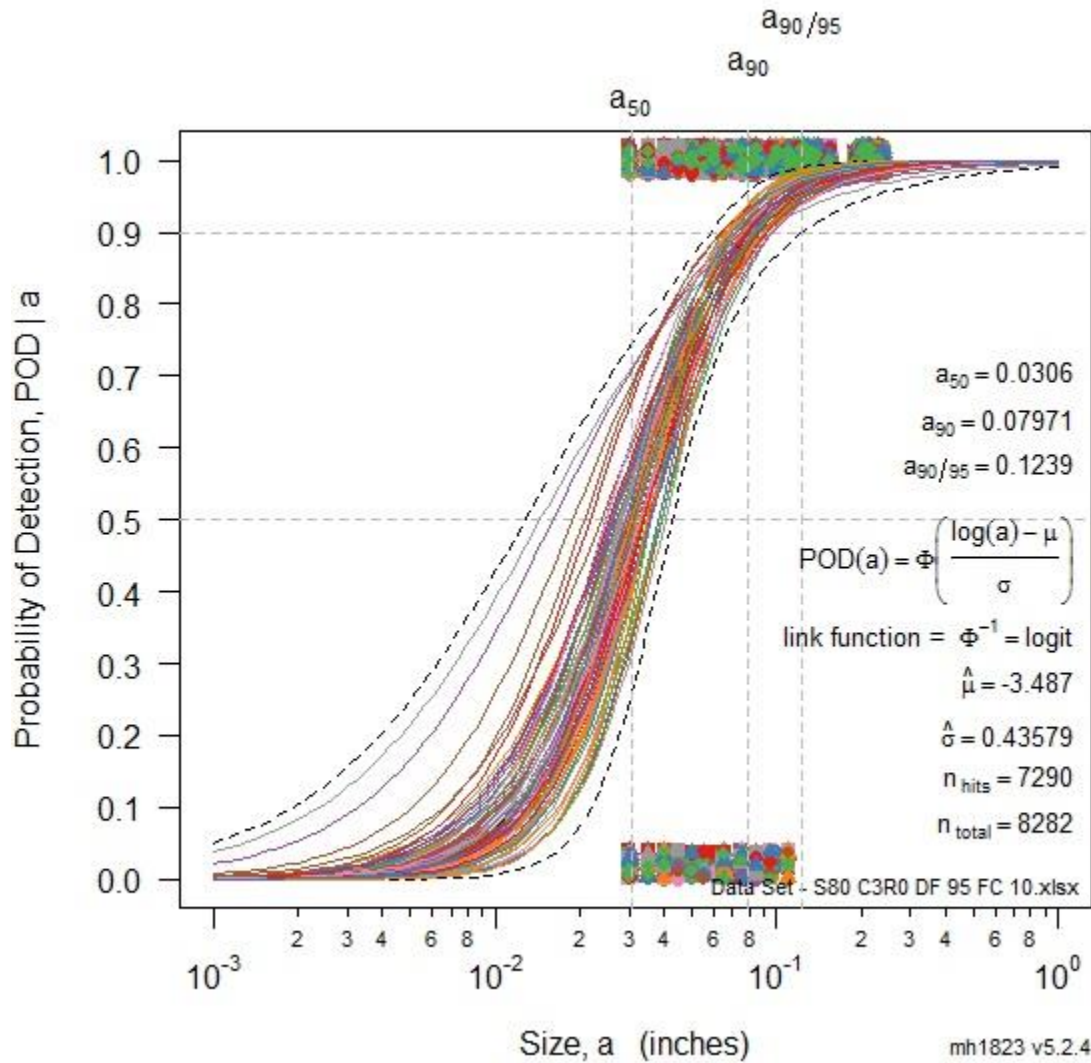**Figure 64: POD vs. size repeat measurements plot - 80 samples, 3 notches, Data Fraction 100%, False Call 15%.**
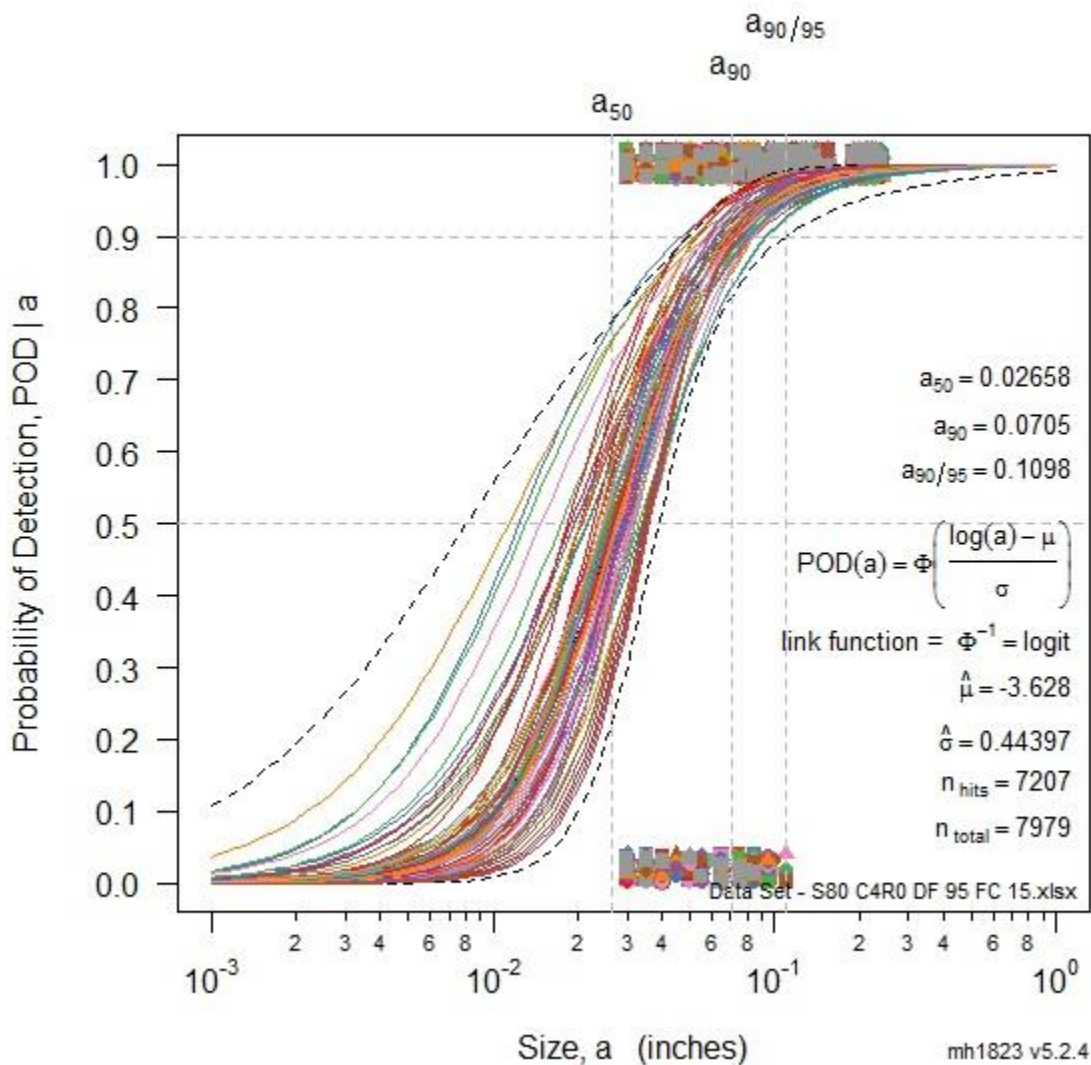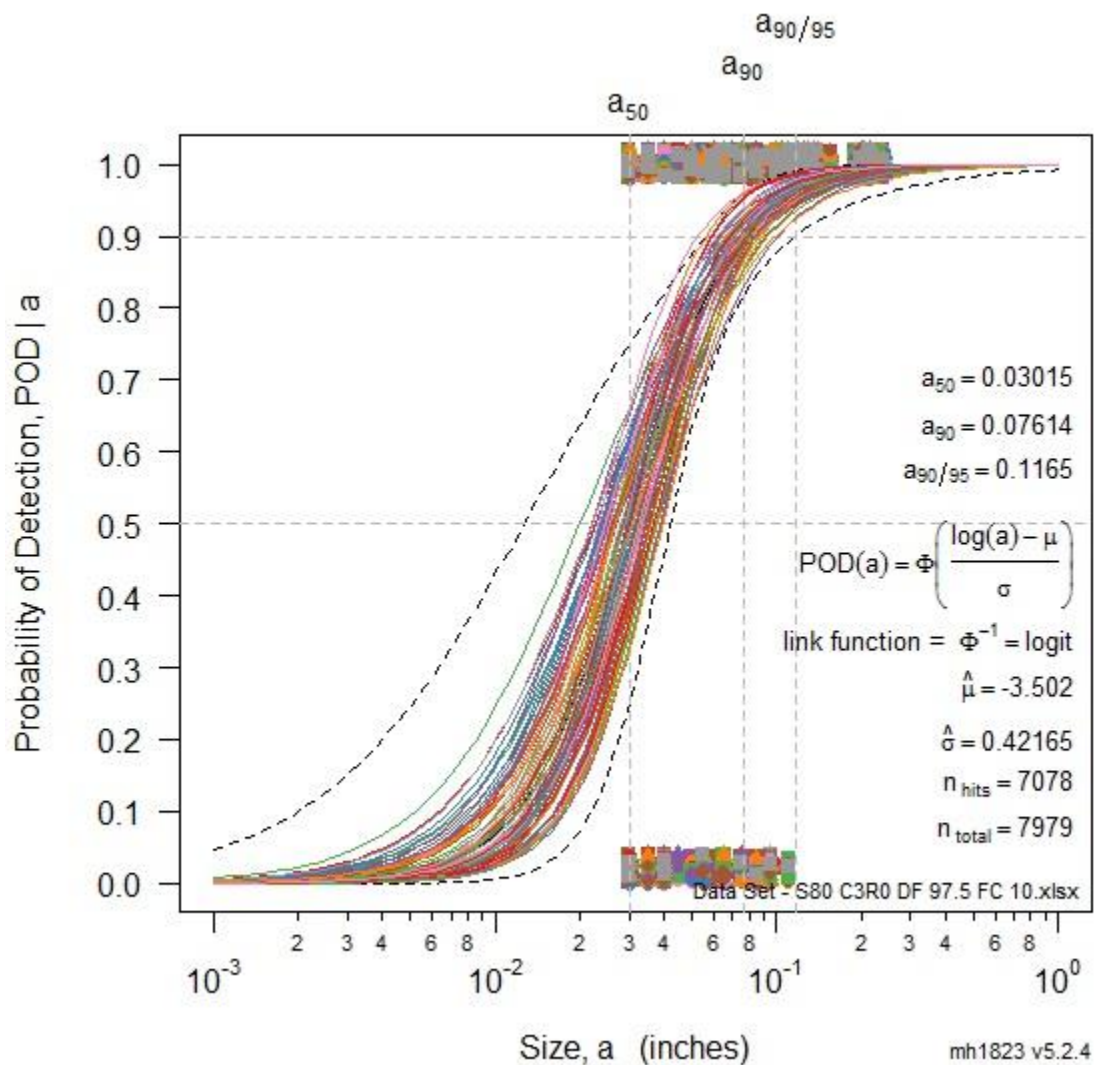
Figure 65: POD vs. size repeat measurements plot - 80 samples, 4 notches, Data Fraction 95%, False Call 15%.

Figure 66: POD vs. size repeat measurements plot - 80 samples, 4 notches, Data Fraction 97.5%, False Call 15%.
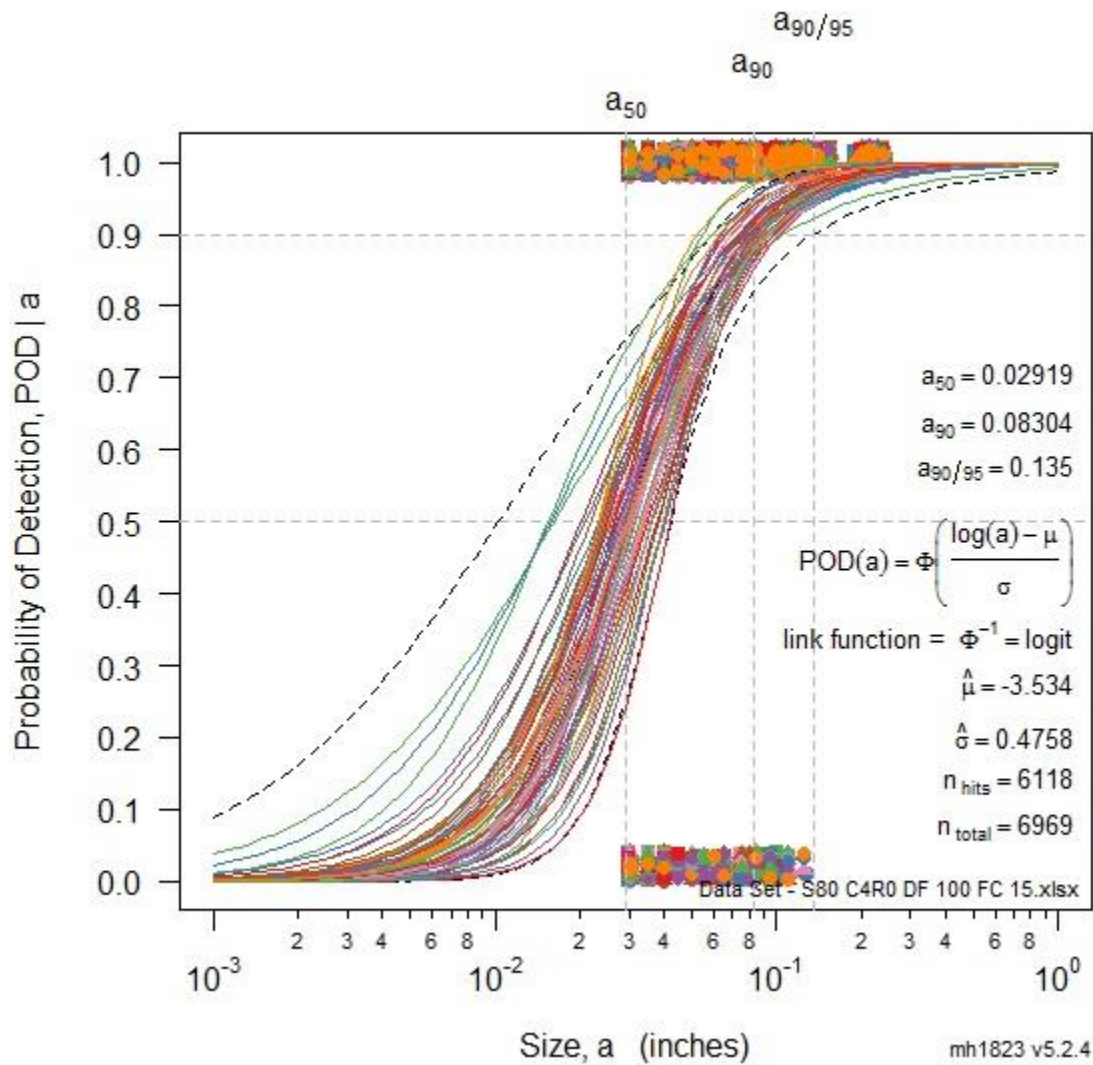
Figure 67: POD vs. size repeat measurements plot - 80 samples, 4 notches, Data Fraction 100%, False Call 15%.

Appendix H
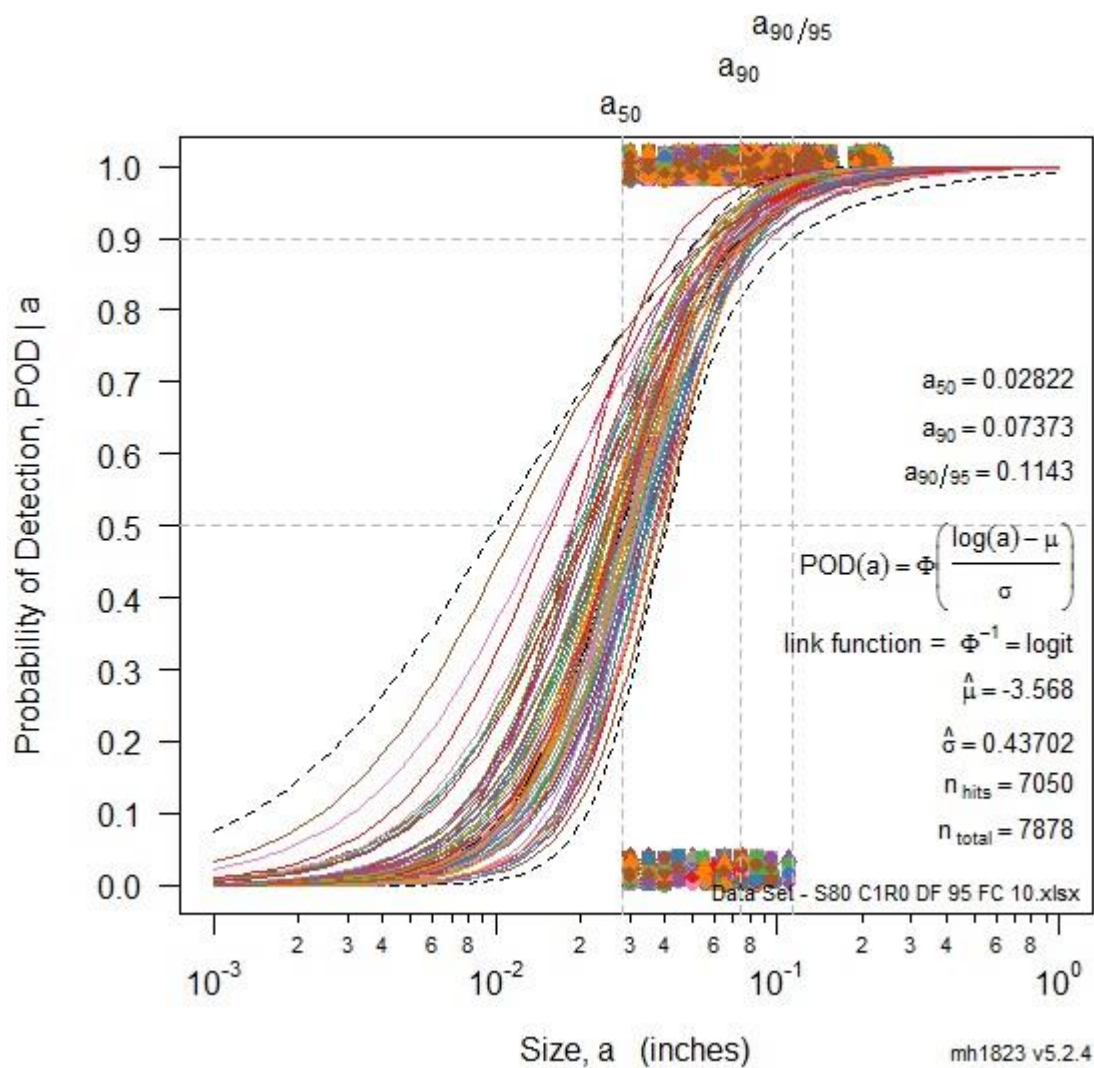
Figure 68: POD vs. size repeat measurements plot - 80 samples, 1 notch, Data Fraction 95%, False Call 10%.
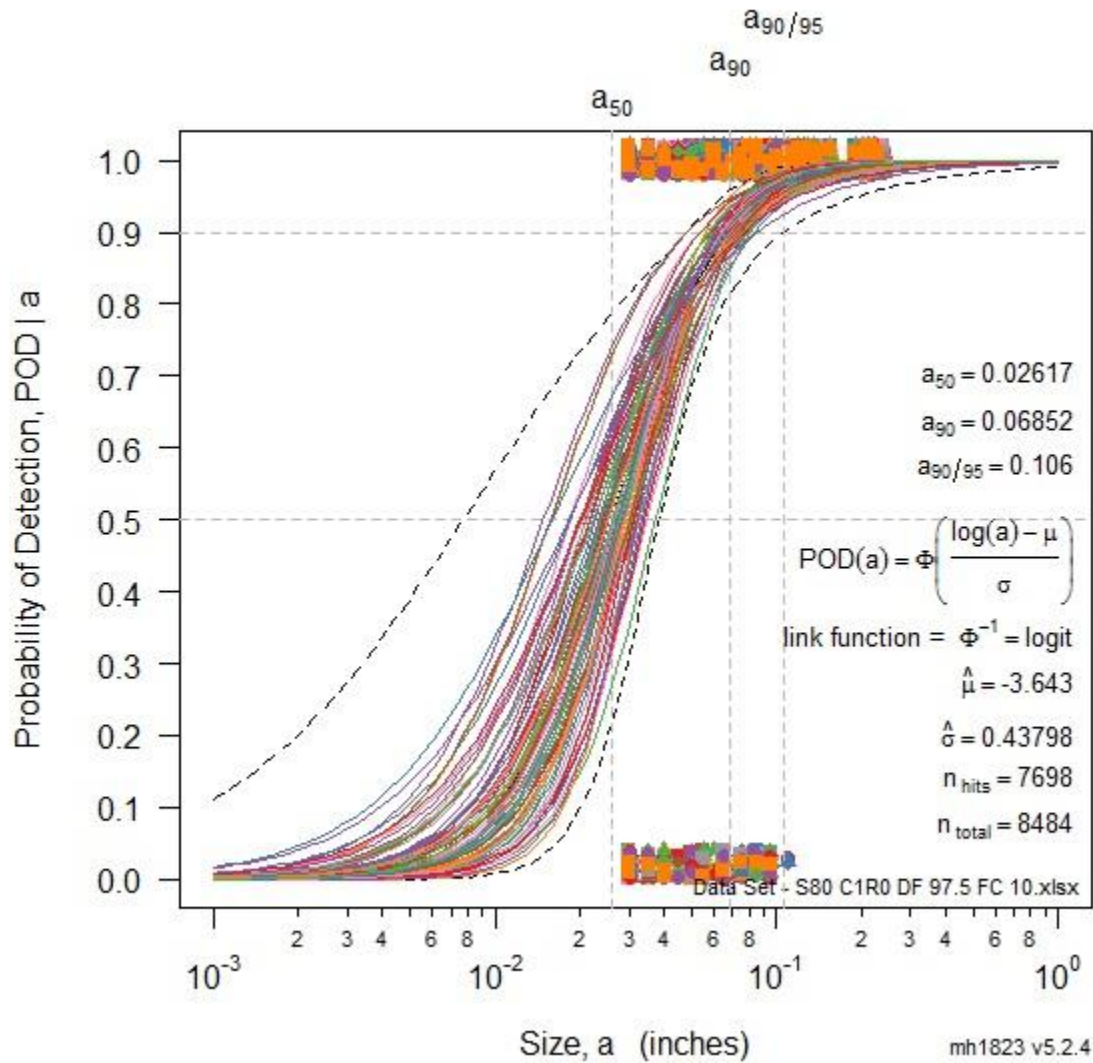
Figure 69: POD vs. size repeat measurements plot - 80 samples, 1 notch, Data Fraction 97.5%, False Call 10%.
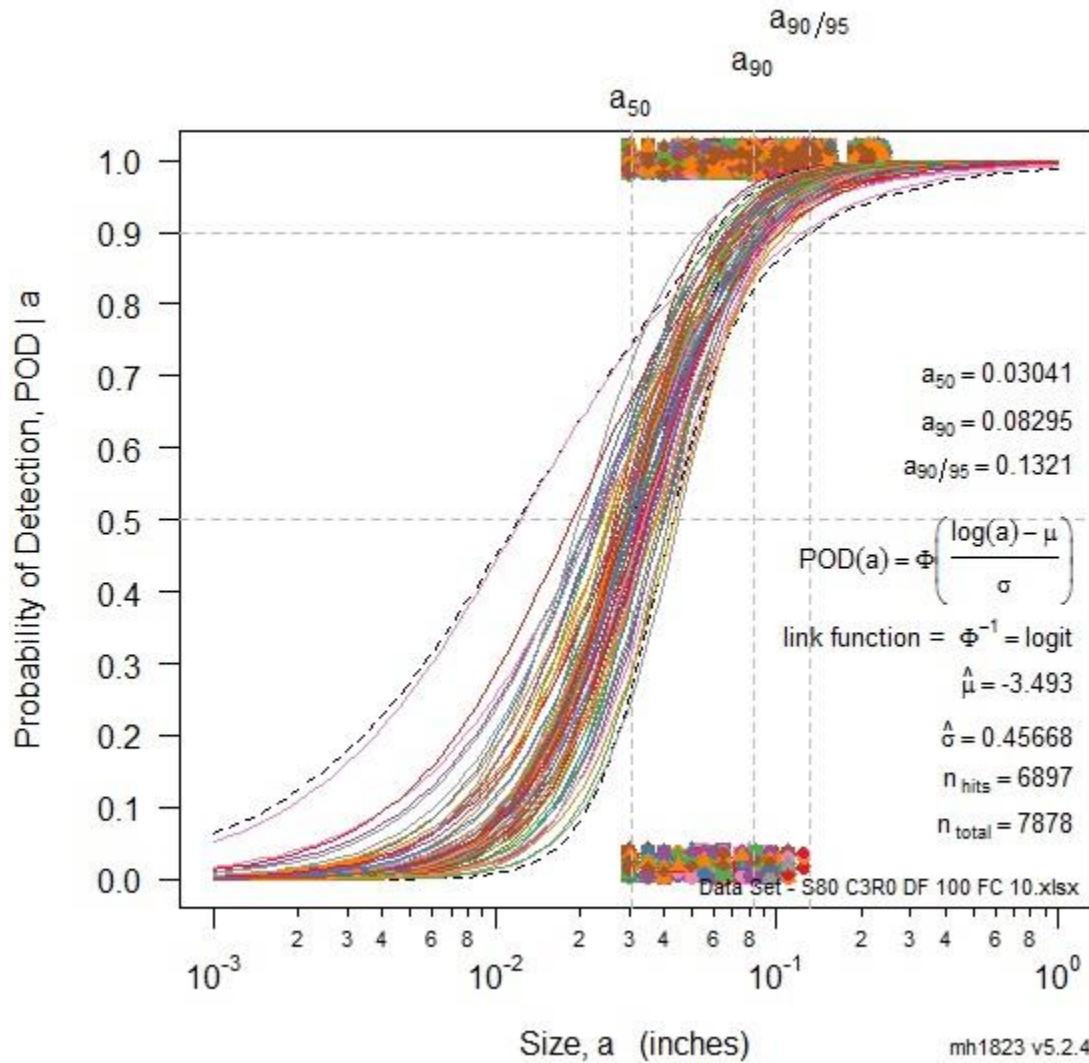
Figure 70: POD vs. size repeat measurements plot - 80 samples, 1 notch, Data Fraction 100%, False Call 10%.
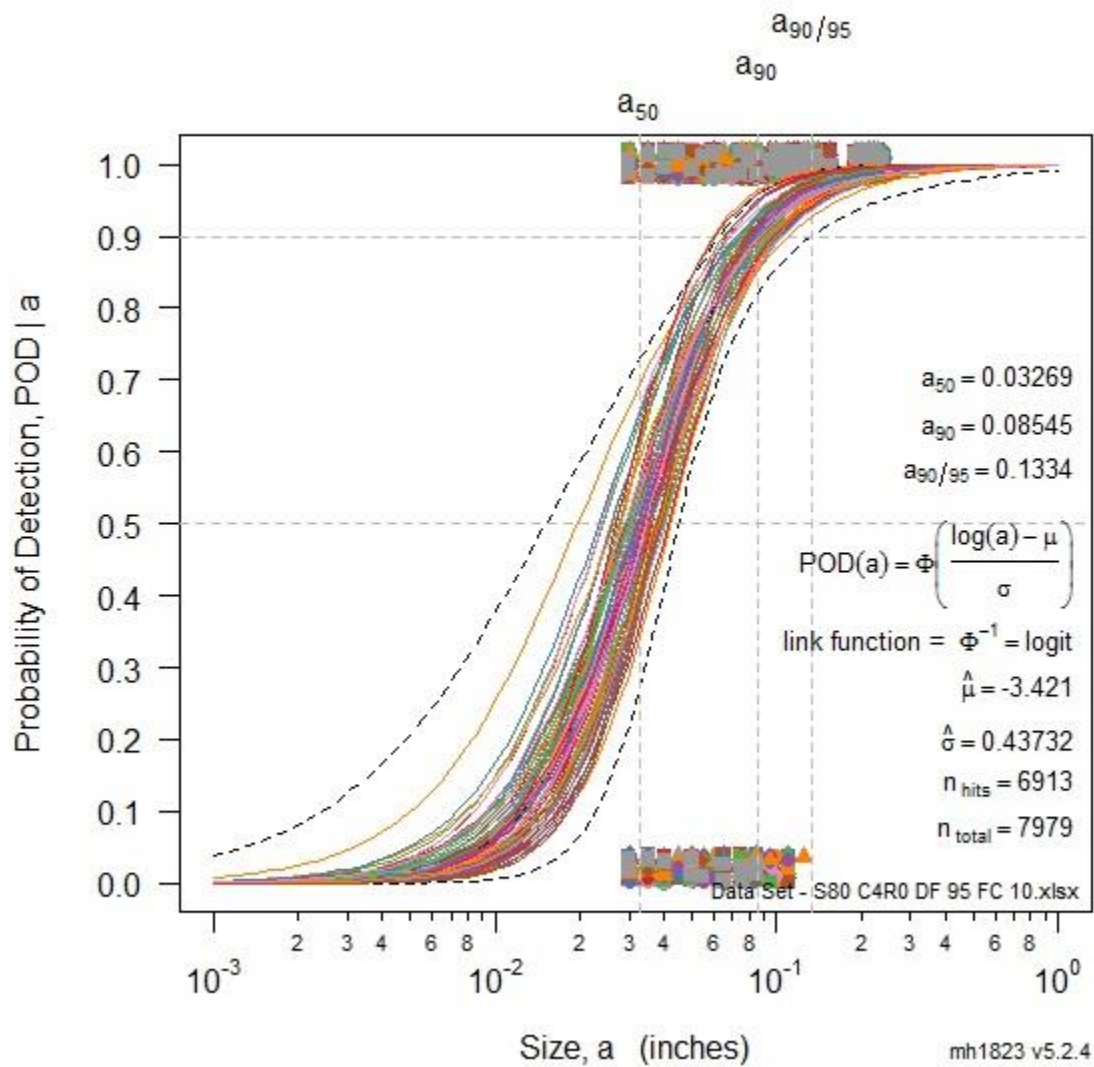
Figure 71: POD vs. size repeat measurements plot - 80 samples, 2 notches, Data Fraction 95%, False Call 10%.

Figure 72: POD vs. size repeat measurements plot - 80 samples, 2 notches, Data Fraction 97.5%, False Call 10%.

Figure 73: POD vs. size repeat measurements plot - 80 samples, 2 notches, Data Fraction 100%, False Call 10%.

Figure 74: POD vs. size repeat measurements plot - 80 samples, 3 notches, Data Fraction 95%, False Call 10%.
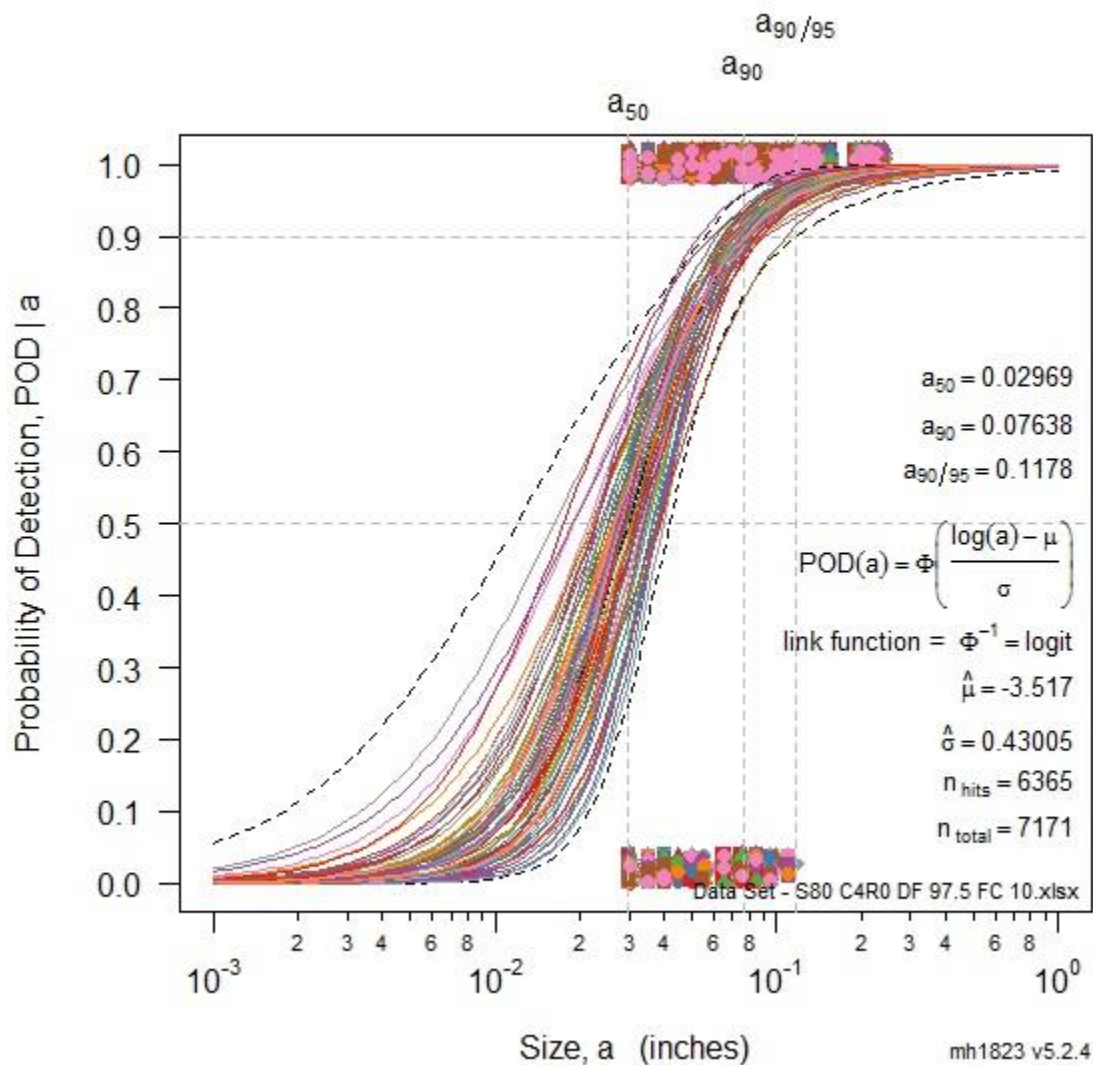
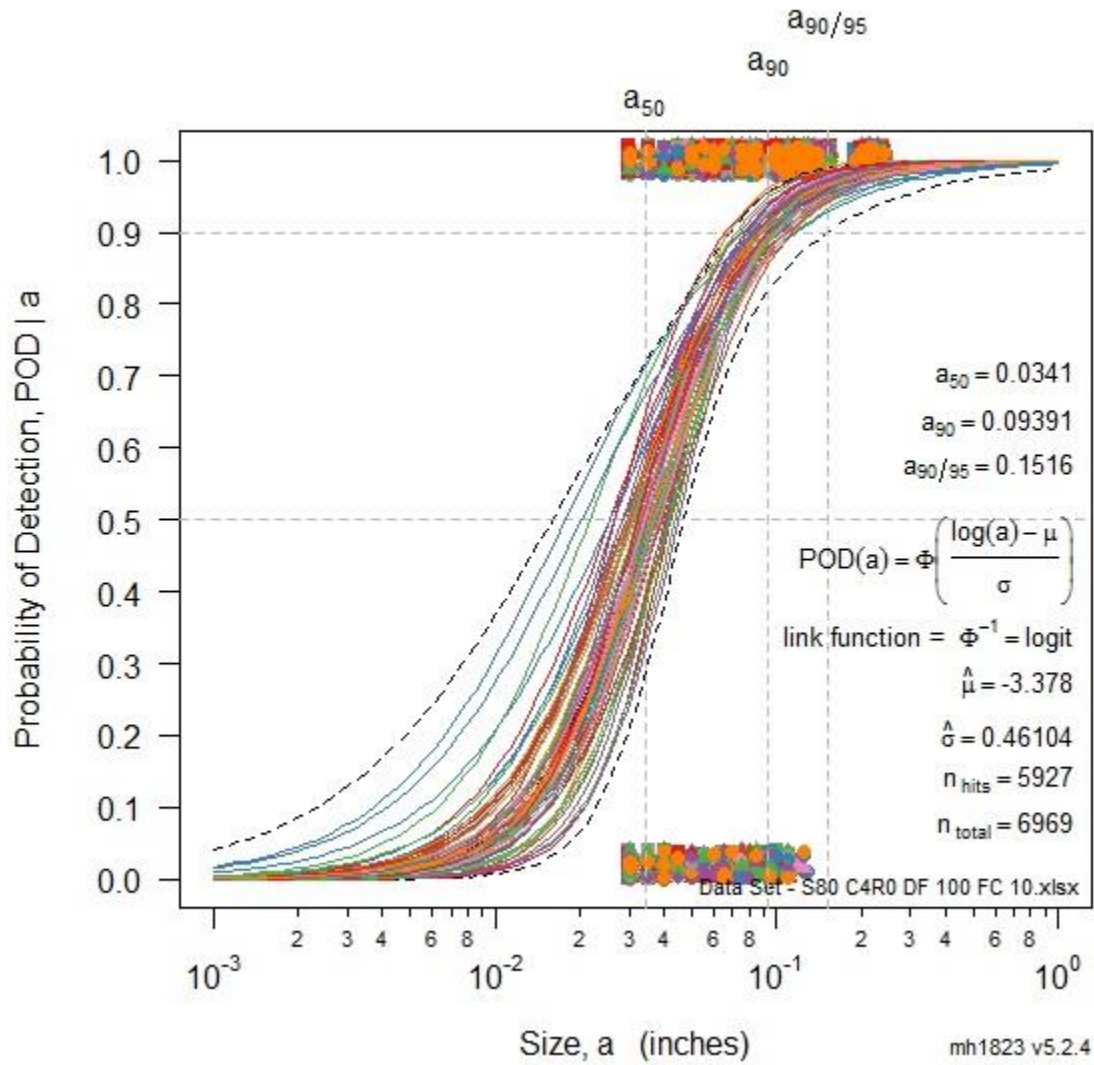**Figure 75: POD vs. size repeat measurements plot - 80 samples, 3 notches, Data Fraction 97.5%, False Call 10%.**

Figure 76: POD vs. size repeat measurements plot - 80 samples, 3 notches, Data Fraction 100%, False Call 10%.

The plot displays the following annotations:

$a_{50} = 0.03269$

$a_{90} = 0.08545$

$a_{90/95} = 0.1334$

$$POD(a) = \Phi\left(\frac{\log(a) - \mu}{\sigma}\right)$$

link function = Φ^{-1} = logit

$\hat{\mu} = -3.421$

$\hat{\sigma} = 0.43732$

$n_{hits} = 6913$

$n_{total} = 7979$

Data Set - S80 C4R0 DF 95 FC 10.xlsx

mh1823 v5.2.4

Figure 77: POD vs. size repeat measurements plot - 80 samples, 4 notches, Data Fraction 95%, False Call 10%.

Figure 78: POD vs. size repeat measurements plot - 80 samples, 4 notches, Data Fraction 97.5%, False Call 10%.

Figure 79: POD vs. size repeat measurements plot - 80 samples, 4 notches, Data Fraction 100%, False Call 10%.

Appendix I

Table 36: Summary of actual false call rates for simulations set with sample size of 80 and false call rate of 5%.

False Call Actual	Data Fraction 95%				Data Fraction 97.5%				Data Fraction 100%			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
	Average False Call - Range 3.58% - 6.16%											
Max	7.72%	7.17%	6.58%	6.07%	6.55%	5.94%	5.82%	5.17%	6.56%	6.18%	5.69%	5.50%
Min	4.24%	3.65%	2.90%	2.52%	3.78%	3.11%	2.57%	2.21%	3.84%	3.38%	2.91%	2.61%
Average	6.16%	5.54%	4.88%	4.28%	5.18%	4.45%	3.90%	3.36%	4.99%	4.49%	3.95%	3.58%
	Average False Call Rate for Sims with 5% False Call Cut-off - 4.56%											

Table 37: Summary of actual false call rates for simulations set with sample size of 80 and false call rate of 10%.

False Call Actual	Data Fraction 95%				Data Fraction 97.5%				Data Fraction 100%			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
	Average False Call - Range 6.04% - 9.55%											
Max	11.39%	10.56%	9.62%	9.02%	10.55%	9.97%	9.21%	8.45%	10.29%	9.40%	8.60%	8.42%
Min	8.21%	7.40%	6.29%	5.78%	7.73%	6.84%	6.05%	5.29%	7.02%	6.14%	5.54%	4.82%
Average	9.55%	8.67%	7.82%	6.96%	8.98%	8.03%	7.24%	6.48%	8.30%	7.39%	6.63%	6.04%
	Average False Call Rate for Sims with 10% False Call Cut-off - 7.67%											

Table 38: Summary of actual false call rates for simulations set with sample size of 80 and false call rate of 15%.

False Call Actual	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
	Average False Call - Range 8.37% - 13.1%											
Max	15.16%	14.35%	13.01%	12.17%	14.45%	13.80%	12.84%	12.28%	13.63%	12.49%	11.65%	10.94%
Min	11.62%	10.45%	9.74%	8.72%	11.16%	9.92%	9.20%	8.23%	9.91%	8.50%	7.78%	6.98%
Average	13.11%	12.10%	11.11%	10.15%	12.63%	11.58%	10.64%	9.72%	11.42%	10.15%	9.18%	8.37%
	Average False Call Rate for Sims with 15% False Call Cut-off - 10.85%											

Appendix J
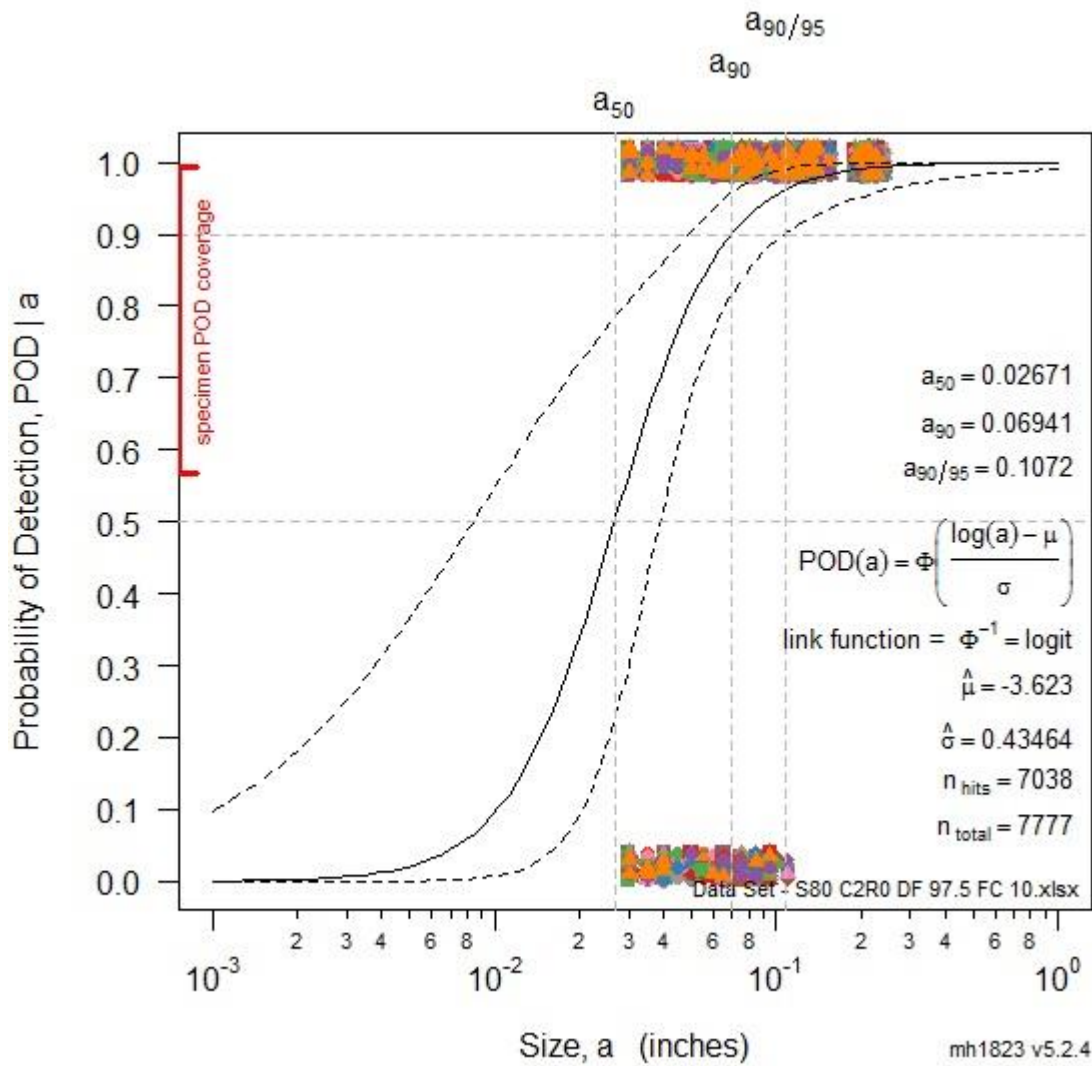
Figure 80: POD vs. size plot for an 80-sample simulation with 97.5% data fraction, 10% false call rate and two notches present, displaying specimen POD coverage.
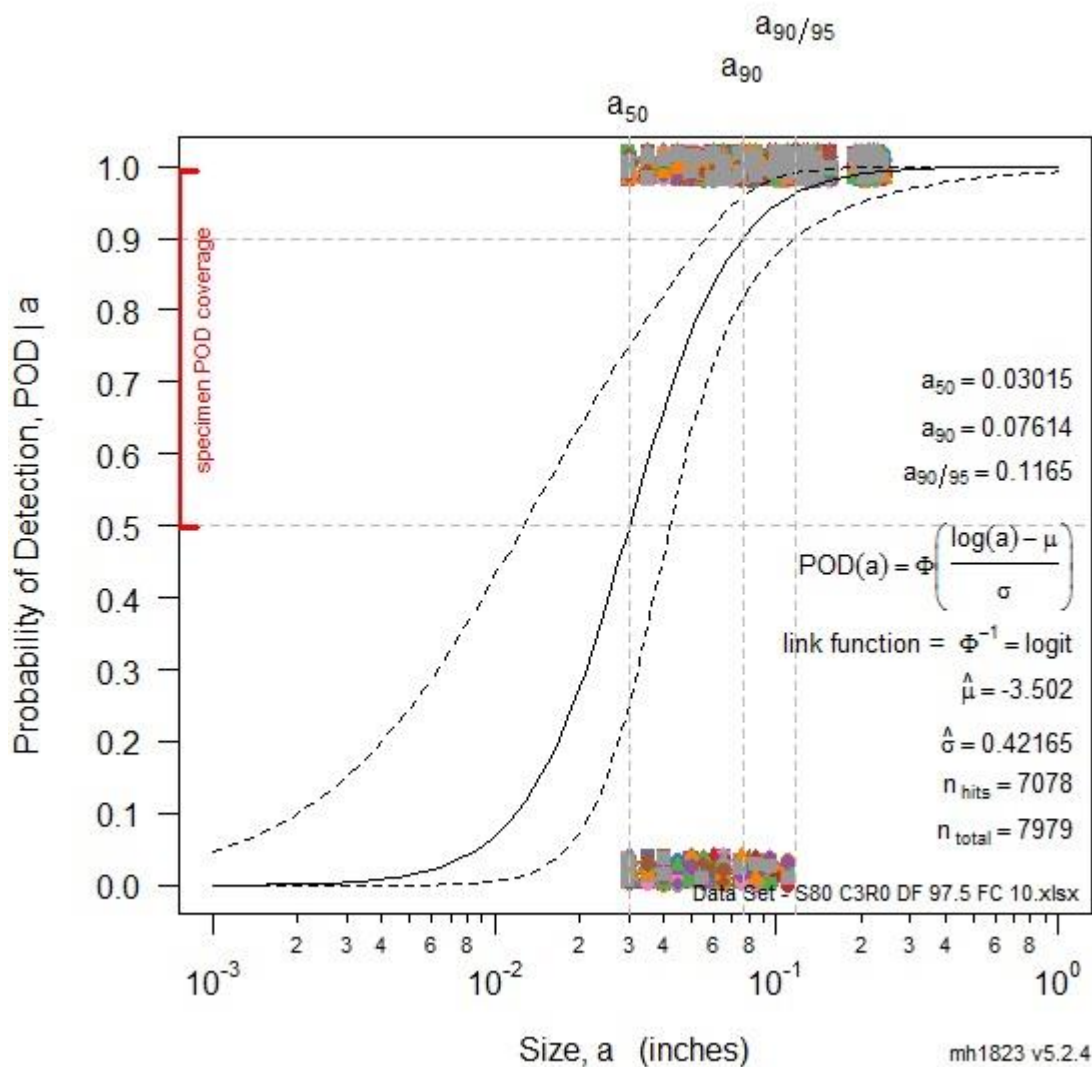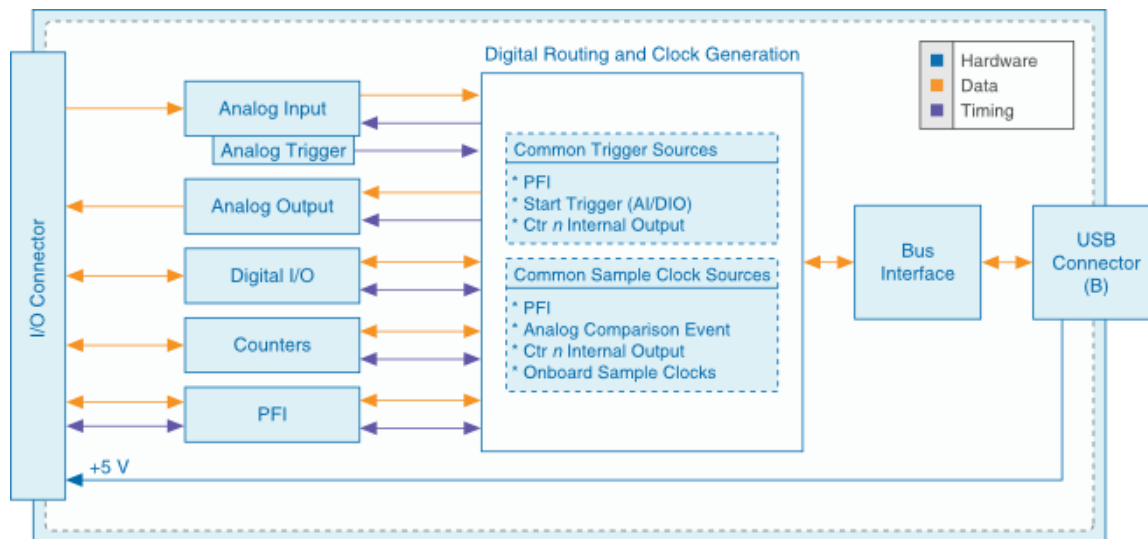
Figure 81: POD vs. size plot for an 80-sample simulation with 97.5% data fraction, 10% false call rate and three notches present, displaying specimen POD coverage.
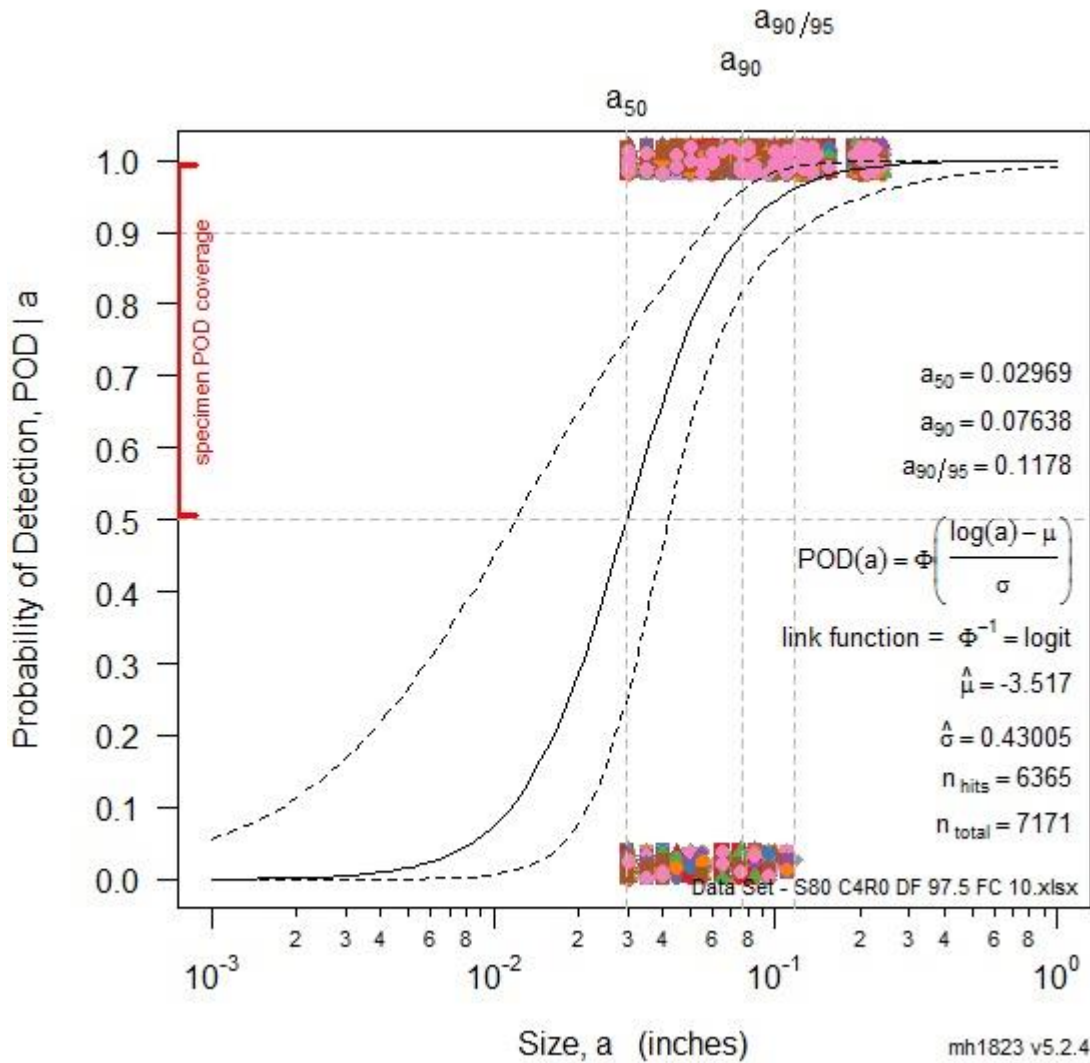
Figure 82: POD vs. size plot for an 80-sample simulation with 97.5% data fraction, 10% false call rate and four notches present, displaying specimen POD coverage.

Appendix K

Figure 83: NI USB-6361 Block Diagram [74].

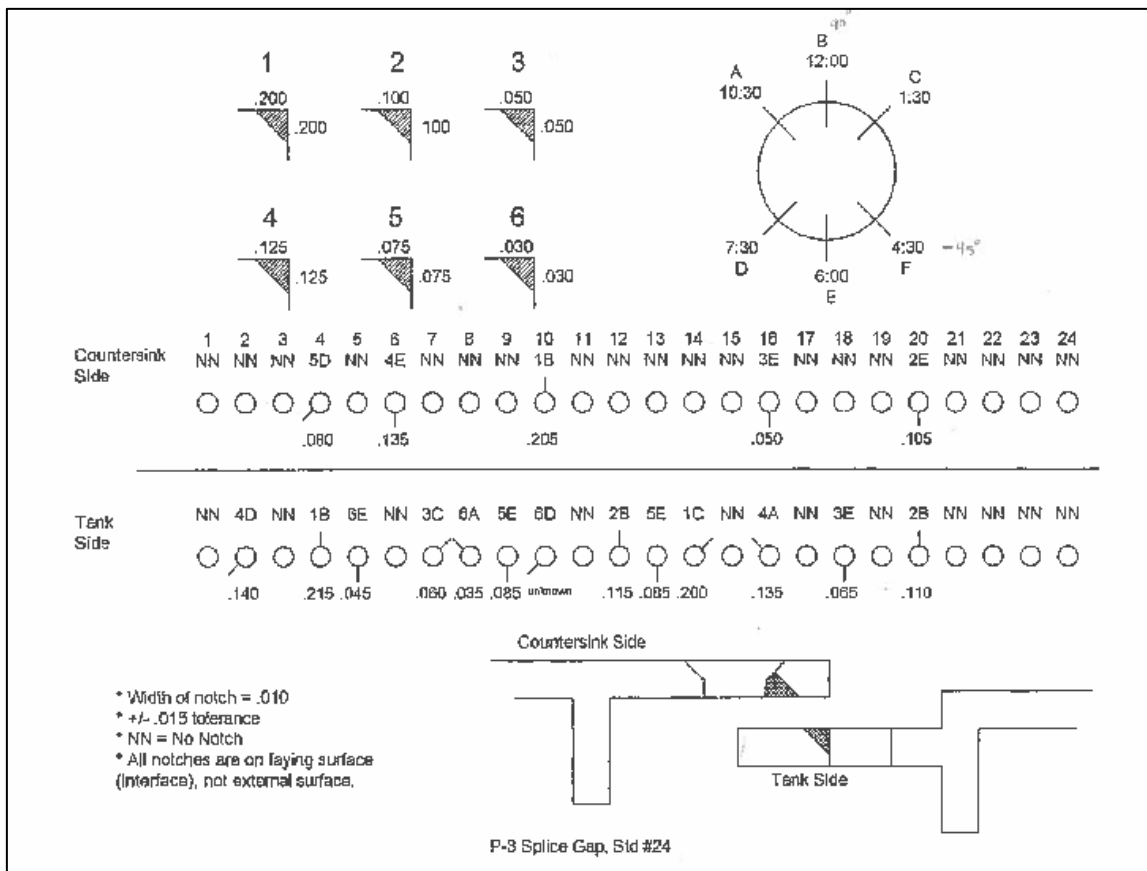# Appendix L
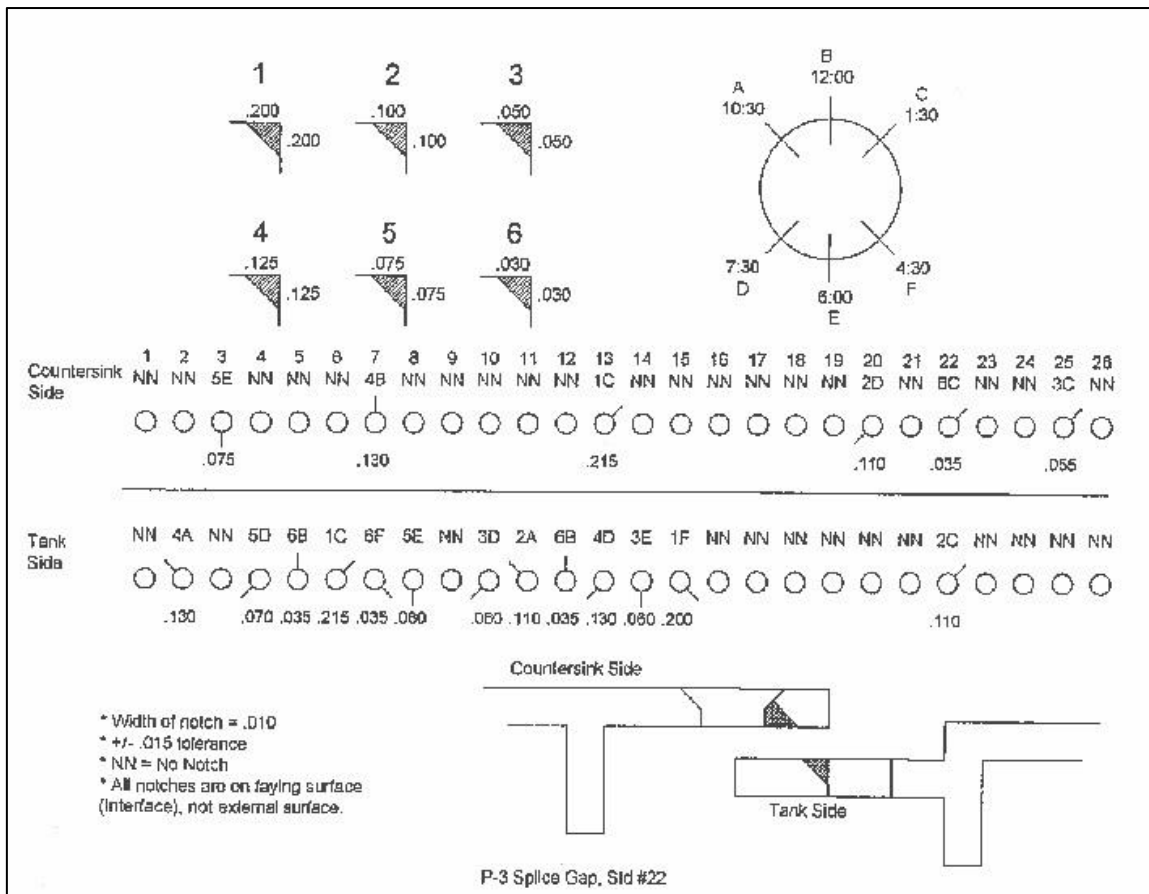
Figure 56: NAVAIR Standard 22 - Notch Schematic Diagram
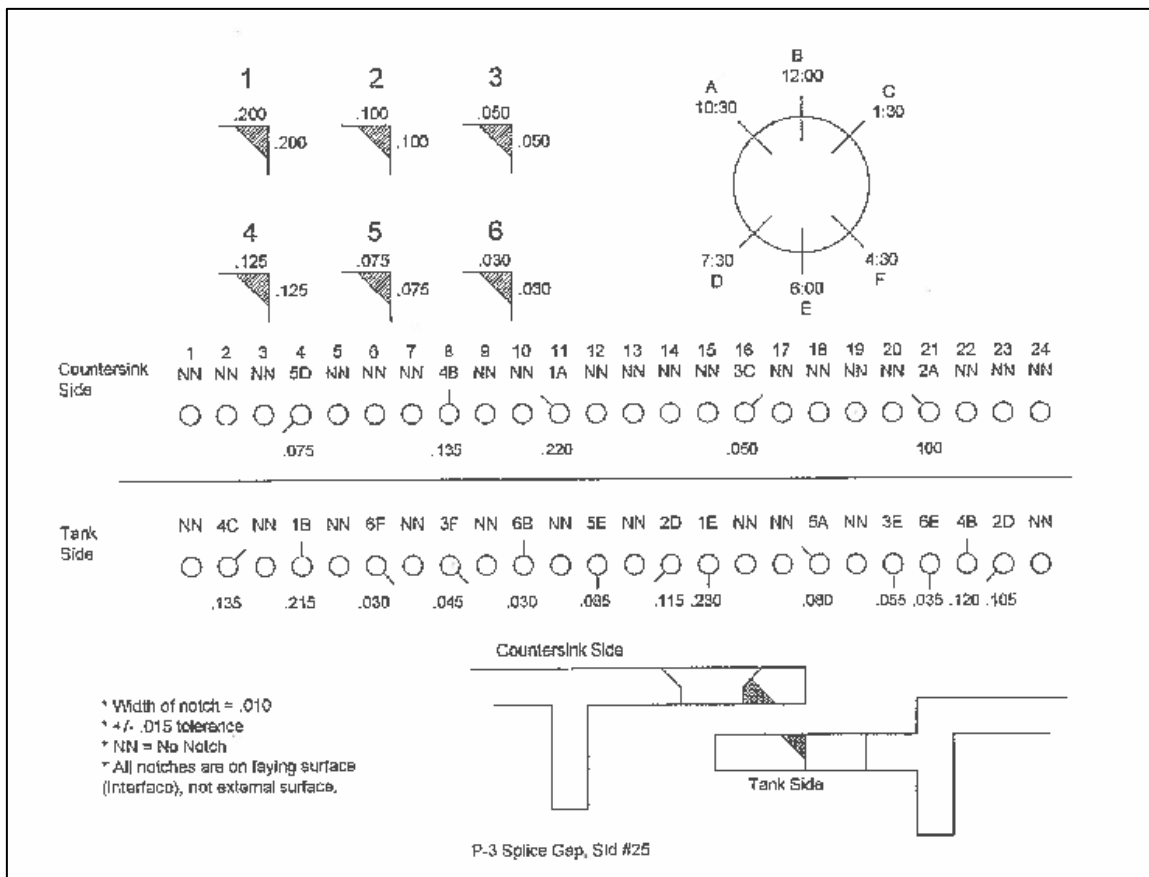
Figure 57: NAVAIR Standard 24 - Notch Schematic Diagram

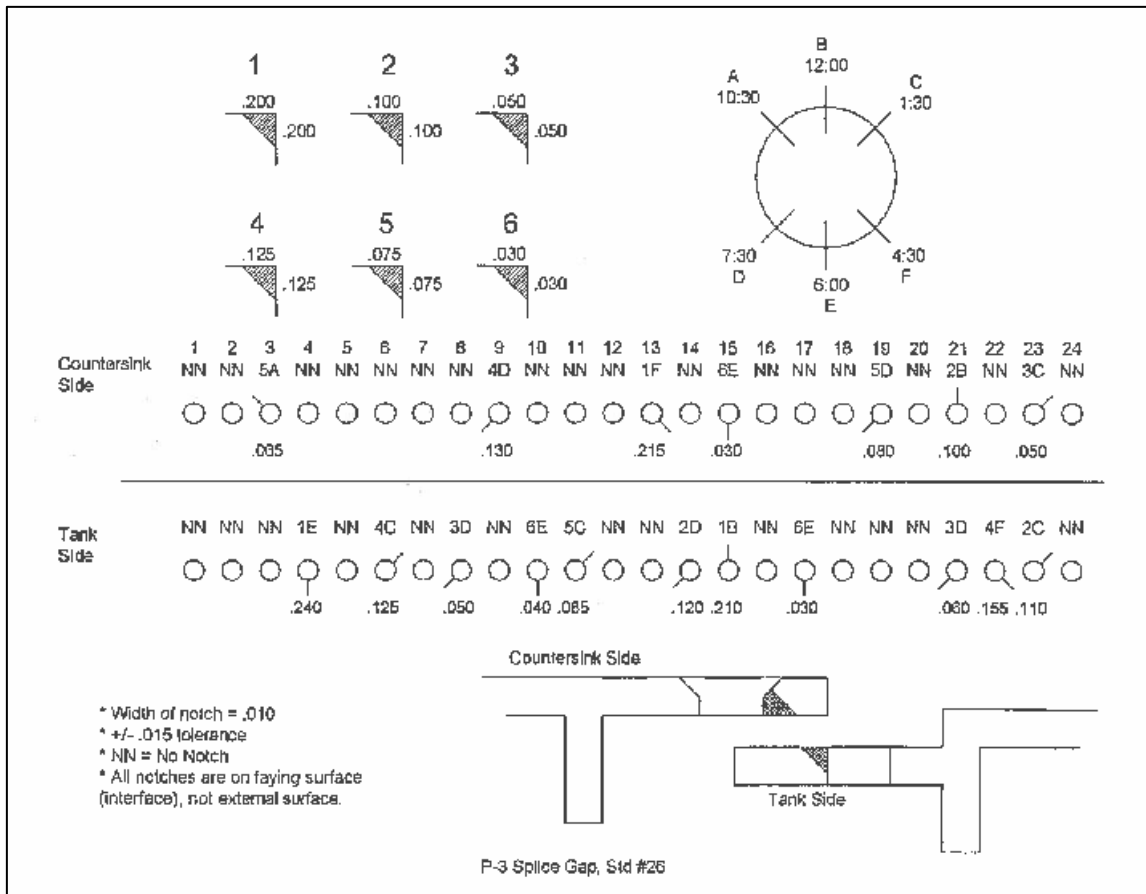**Figure 58: NAVAIR Standard 25 - Notch Schematic Diagram**

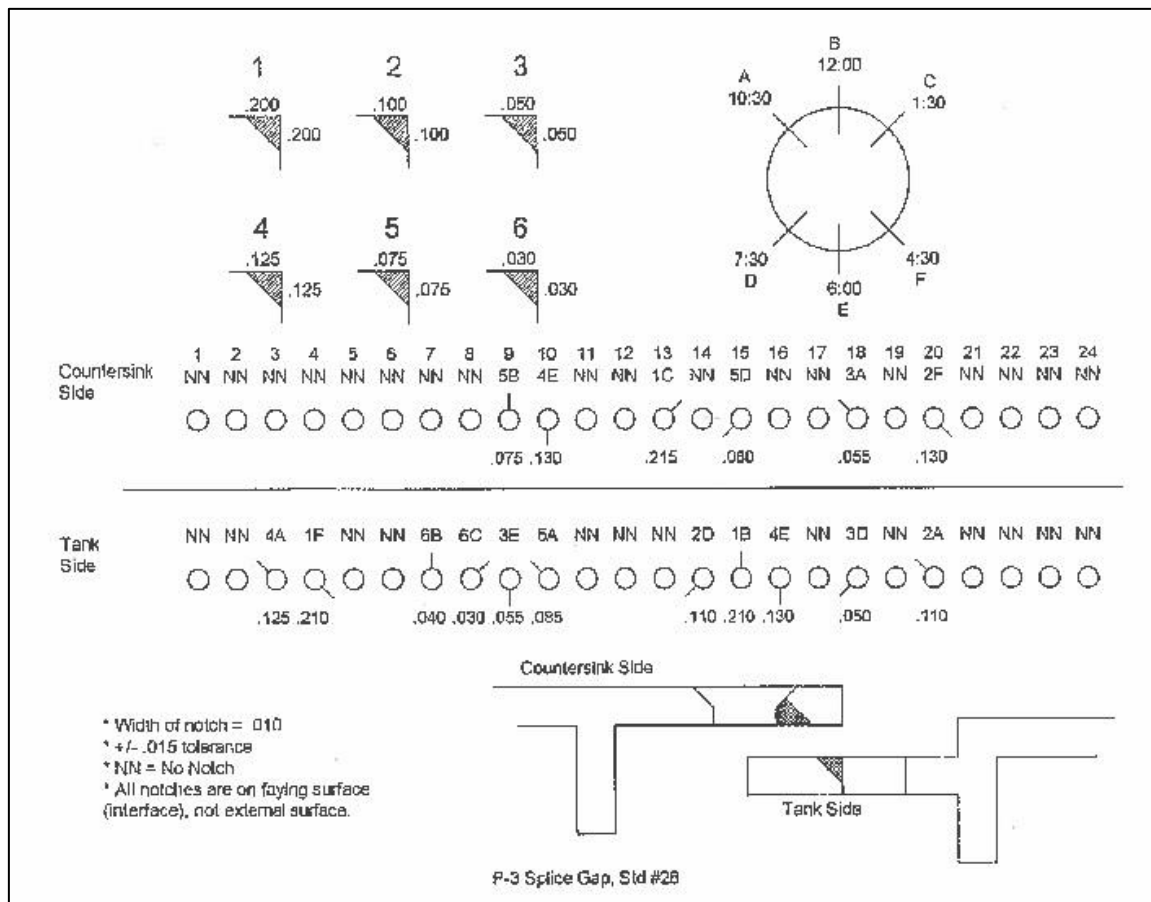**Figure 59: NAVAIR Standard 26 - Notch Schematic Diagram**

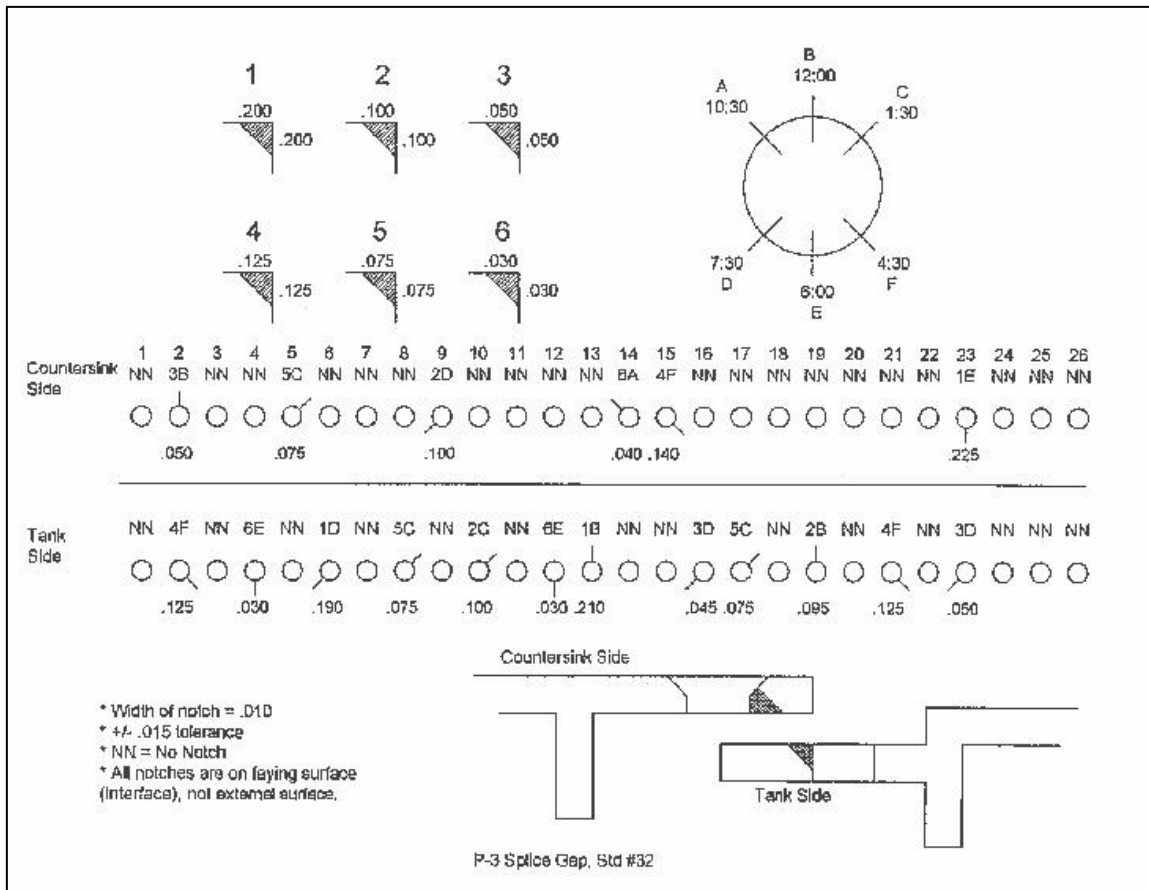**Figure 60: NAVAIR Standard 28 - Notch Schematic Diagram**

Figure 61: NAVAIR Standard 32 - Notch Schematic Diagram

Appendix M

Procedure for $a_{90/95}$ Revision Utilizing ImageJ Software

To adjust for the spread at a_{90}, repeat measures POD(a) plots for the simulations with 80 samples were analyzed with ImageJ software as follows:

- Load repeat measures POD(a) plot into ImageJ.
- Scale image to 300%.
- Open Microsoft Excel spreadsheet.
- Select line function in ImageJ.
- Construct line from edge of POD(a) grouping at the $POD = 0.9$ line to the a_{90} line on repeat measures plot. This length represented as LEnv and entered in Microsoft Excel spreadsheet.
- Measure length between a_{90} and $a_{90/95}$ lines on POD(a) plot. This length is represented as LDelta and entered in Microsoft Excel spreadsheet.
- The known values for a_{90} and $a_{90/95}$ calculated by the **mh1823 POD** software were entered as D1 and D2 respectively in Microsoft Excel spreadsheet.
- To interpolate the ImageJ measurements to determine a refined $a_{90/95}$ value, the following mathematical calculation was performed:

$$a_{90/95} = 10^{\frac{\log(D2) - \log(D1)}{(LDelta * LEnv) + \log(D1)}}$$

Curriculum Vitae

Cole Omega Uemura was born in Windsor, Ontario. She obtained an Electro-Mechanical Engineering diploma from the Marine Institute, Memorial University of Newfoundland. She attained the highest average in the program history and was awarded the Marine Institute of Memorial University Bronze Medallion. Upon graduation, Cole was selected as the recipient of the Herbert Lott Naval Trust Fund Award for her professionalism and academic accomplishments. This award is presented to one recipient annually from all Commonwealth countries. She was subsequently employed as a Naval Weapons Technician on the Royal Canadian Navy destroyer, Her Majesty's Canadian Ship (HMCS) Algonquin, in the Pacific Fleet.

Upon completion of her posting to Victoria, British Columbia, Cole was selected to attend Queen's University in engineering and was awarded the Science 48½ Scholarship, which is the largest award available at Queen's. She obtained a Bachelor of Applied Science in Engineering (Mathematics and Engineering – Applied Mechanics) and was awarded the Boyd Lemna Graduation Award.

Upon graduation from Queen's, Cole attended the Aerospace Engineering Officer Basic Course (AOBC) in Borden, Ontario. She was subsequently posted to the Directorate of Airworthiness and Engineering Support in Ottawa, Ontario, in the position of Air Force Test and Evaluation Coordinator. There, she was responsible for coordination of Air Force flight test and evaluation activities required in support of projects involving aircraft modifications with airworthiness implications.

In 2015, Cole was posted to Kingston, Ontario to complete a Master of Science in Physics at the Royal Military College of Canada.