DOT/FAA/AR-97/73

Office of Aviation Research
Washington, D.C. 20591

Detection Reliability for Small Cracks Beneath Rivet Heads Using Eddy-Current Nondestructive Inspection Techniques

PB99-126864

December 1998

Final Report

This document is available to the U.S. public through the National Technical Information Service (NTIS), Springfield, Virginia 22161.

U.S. Department of Transportation
Federal Aviation Administration

REPRODUCED BY: NTIS
U.S. Department of Commerce
National Technical Information Service
Springfield, Virginia 22161

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof. The United States Government does not endorse products or manufacturers. Trade or manufacturer's names appear herein solely because they are considered essential to the objective of this report.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AR-97/73		

4. Title and Subtitle	5. Report Date
DETECTION RELIABILITY FOR SMALL CRACKS BENEATH RIVET HEADS USING EDDY-CURRENT NONDESTRUCTIVE INSPECTION TECHNIQUES	December 1998
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Floyd W. Spencer	
9. Performing Organization Name and Address	PB99-126864 AIS)
FAA Airworthiness Assurance NDI Validation Center Sandia National Laboratories Albuquerque, NM 87185-0829	11. Contract or Grant No.
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
U.S. Department of Transportation Federal Aviation Administration Office of Aviation Research Washington, DC 20591	Final Report
	14. Sponsoring Agency Code
	AFS-300

15. Supplementary Notes

Federal Aviation Administration William J. Hughes Technical Center COTR is Christopher Smith

16. Abstract

Advanced inspection technology that is emerging from the laboratory is generally far superior to the less capable systems around which aircraft inspections are designed. In light of the very conservative nature of these inspection designs, it is apparent that today's advanced technology is being employed at only a fraction of its full potential. In order to assess the full potential of advanced eddy-current inspection technology on representative aircraft applications, the FAA's Airworthiness Assurance Nondestructive Inspection Validation Center (AANC) was tasked to assess the full capability of several advanced systems. The task involved inspections of several rivet skin splices—representative of actual aircraft structure—containing cracks ranging from 0.040-, 0.060-, and 0.080-inch standards, and thresholds were set to the lowest reasonable level for the particular system. The results demonstrated that some of the systems were able to reliably detect cracks as small as 0.040 inch with false call rates remained less than 1%.

17. Key Words	18. Distribution Statement
Nondestructive Inspection, Eddy current, Inspection reliability, Probability of Detection	This document is available to the public through the National Technical Information Service (NTIS), Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	55	N/A

Form DOT F1700.7 (8-72) Reproduction of completed page authorized

ACKNOWLEDGEMENTS

Many people were involved in the experiments that are reported on in this document. There is a genuine interest in the NDI community for techniques to be subjected to blind tests. This has been demonstrated by the willingness of manufacturers, developers, and users to subject themselves and their equipment to testing.

The following people participated willingly in the experiments reported here, often returning for multiple visits to the Federal Aviation Administration's (FAA) Airworthiness Assurance NDI Validation Center (AANC): Tom Reep, Staveley Instruments, Dave Jankowski, Krautkramer Branson, John Calvert, Hocking NDT Limited, Owen Manning, late of Northrop Corporation, Bill Sheppard, Northrop Corporation, Buzz Wincheski and Min Namkung, NASA Langley Research, John Simpson, Lockheed Engineering and Sciences Co., Jim Fulton, late of Analytical Services and Materials, Inc.

Also providing support and gathering data at the AANC were David Moore, Phil Walkington, Darin Graf, and Kyle Thompson.

PROTECTED UNDER INTERNATIONAL COPYRIGHT
ALL RIGHTS RESERVED.
NATIONAL TECHNICAL INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE

TABLE OF CONTENTS

	Page
EXECUTIVE SUMMARY	ix
1. INTRODUCTION	1
2. SPECIMEN DESCRIPTION AND ANALYSIS TECHNIQUES	1
2.1 Specimen Overview	2
2.2 Analysis Overview	2
3. EDDY-CURRENT EQUIPMENT AND TECHNIQUES	3
3.1 Nortec 30 Eddyscan	3
3.1.1 Nortec Experiment Background	4
3.1.2 Nortec PoD Analysis	4
3.2 Krautkramer Branson Crackfinder	7
3.2.1 Crackfinder Experiment 1 Background	7
3.2.2 Crackfinder Experiment 1 Analysis	8
3.2.3 Crackfinder Experiment 2 Background	12
3.2.4 Crackfinder Experiment 2 Analysis	12
3.3 Hocking FastScan	14
3.3.1 Hocking FastScan Background	14
3.3.2 Hocking FastScan Analysis	16
3.4 Nortec 19e With Sliding Probe	19
3.4.1 Nortec 19e Background	19
3.4.2 Nortec 19e Analysis	20
3.4.2.1 Analysis of Signals From Noncracked Rivets	21
3.4.2.2 Analysis ($\hat{a}$ Versus a) of Signals From Cracked Rivets	22
3.4.2.3 Use of Inverse Regression in Analysis of Signals	26
3.5 Pencil Probe with Zetec MIZ-22	29
3.5.1 Zetec MIZ-22 Background	29
3.5.2 Zetec MIZ-22 Analysis	29
3.6 Low-Frequency Eddy-Current Array (LFECA)	32

3.6.1 LFECA Inspection Background 32
3.6.2 LFECA Analysis 32

3.7 NASA Self-Nulling Rotating Probe 34

3.7.1 NASA Self-Nulling Probe Inspection Background 34
3.7.2 NASA Self-Nulling Probe Data Analysis 37

3.8 McDonnell Douglas Aerospace/GK Engineering Surface Scanning Probe 39

3.8.1 GK Engineering Probe Background 39
3.8.2 GK Engineering Probe Analysis 41

4. SUMMARY AND DISCUSSION 42

5. REFERENCES 43

APPENDIX A—Using Inverse Regression Fits to Establish PoD Functions

LIST OF FIGURES

Figure		Page
1	Nortec Eddyscan Log Signal Strength Versus Log Crack Length	5
2	Nortec Eddyscan PoD Curves	6
3	Crackfinder Data for 0.25-Inch Probe-Unmodified and Reduced Sensitivity Units	8
4	Crackfinder Data for 0.50-Inch Probe-Unmodified and Reduced Sensitivity Units	9
5	Crackfinder PoD Curves for Unmodified Unit With 0.25-Inch Probe-Combined Inspection Surface	9
6	Crackfinder PoD Curves for Reduced Sensitivity Unit With 0.25-Inch Probe	10
7	Crackfinder PoD Curves for Unmodified Unit With 0.50-Inch Probe	10
8	Crackfinder PoD Curves for Reduced Sensitivity Unit With 0.50-Inch Probe	11
9	Crackfinder Signals-Template and Transverse Procedures	13
10	Second Visit Crackfinder PoD-Template Procedure	13
11	Second Visit Crackfinder PoD-Transverse Procedure	14
12	Hocking FastScan Signals on Small Panels	16
13	Hocking FastScan PoD Curves-Setup 1	18
14	Hocking FastScan PoD Curves-Setups 2 and 3	18
15	Nortec 19e Sliding Probe Setup Signals	20
16	Nortec 19e Sliding Probe Signal	20
17	Nortec 19e—No Flaw Signals by Rivet Within Panel	21
18	Nortec 19e—No Flaw Signals by Panel	21
19	PoD Estimates for Nortec 19e Sliding Probe 20-kHz Inspections on Bare Surface	23
20	PoD Estimates for Nortec 19e Sliding Probe 20-kHz Inspections on Painted Surface	24
21	PoD Estimates for Nortec 19e Sliding Probe 30-kHz Inspections on Bare Surface	24
22	PoD Estimates for Nortec 19e Sliding Probe 30-kHz Inspections on Painted Surface	25

23	Nortec 19e ROC Curves	25
24	Nortec 19e Predicted Crack Length Distribution at Nonflawed Rivets	28
25	Zetec MIZ-22 Signals	29
26	Signal Data from Zetec MIZ-22 Pencil Probe Inspection	30
27	Miz-22 Pencil Probe Estimated PoD Curves	31
28	LFECA Estimated PoD Curves—Inspections by Crack	33
29	LFECA Estimated PoD Curves—Inspections by Rivet for Inspector 2	33
30	NASA Self-Nulling Rotating Probe Raw Signal	35
31	NASA Self-Nulling Rotating Probe Processed Signal	36
32	NASA Self-Nulling Rotating Probe Processed Signal in Polar Coordinates	36
33	NASA Self-Nulling Probe Response Versus Crack Length—Small Panels	38
34	Estimated Probability of Detection Curves—NASA Rotating Probe	38
35	GK Engineering Probe—Signal Data	41
36	GK Engineering Probe—Estimated PoD Curves	41

LIST OF TABLES

Table		Page
1	Test Specimen Factors and Levels Included in Analysis	3
2	Nortec Eddyscan Inspection Parameters	4
3	Nortec Eddyscan False Calls	5
4	Crackfinder False Call Rates With 95% Upper Confidence Bounds	11
5	Hocking FastScan Inspection Parameters	15
6	Hocking FastScan False Call Rates Versus Signal Thresholds	17
7	Hocking FastScan False Call Rates	18
8	Nortec 19e Instrument Settings	19
9	Nortec 19e—No Flaw Variations and False Call Levels	22
10	Nortec 19e PoDs From Predictive Fits	27
11	Zetec MIZ-22 Pencil Probe PoD Parameters	31
12	NASA Self-Nulling Rotating Probe—Summary of Responses	37
13	GK Probe/Elotest B1 Instrument Settings	40
14	Summary PoD Values	42

EXECUTIVE SUMMARY

As part of its validation efforts the Federal Aviation Administration (FAA) Airworthiness Assurance Nondestructive Inspection Validation Center maintains a set of lap splice specimens with cracks growing out from beneath flush-head fasteners. These specimens are used in blind experiments where the inspector does not have knowledge concerning the specific location and size of cracks. Manufacturers, developers, and users of various eddy-current equipment have performed experiments to characterize reliability using these test specimens. Results on eight such experiments are reported here. Major conclusions are summarized below.

Commonly employed eddy-current procedures are capable of reliably detecting cracks as small as 0.050 inch while maintaining false calls below 1 percent. However, to achieve such detection rates requires careful settings of threshold levels and appropriate setup standards.

Currently, there are newly developed techniques and instruments that are capable of doing better than 0.050-inch detection. The National Aeronautics and Space Administration's (NASA) self-nulling rotating probe demonstrated that it could reliably be used to find cracks as small as 0.032 inch and Northrop's low frequency eddy-current array (LFECA) has demonstrated a capability of reliably detecting cracks as small as 0.040 inch. These rates are achievable without an increase in false calls.

It has been previously reported that the Hocking FastScan, Nortec-30 Eddyscan, Northrop LFECA, and GK Engineering/Elotest are capable of detecting surface cracks 1.0 mm (0.040 inch) in length under flush-head aluminum rivets. Our results indicate that although capable of detecting this size crack, the probability associated with routinely detecting them (at false call rates < 0.01) are approximately 0.23 (FastScan), 0.74 (Eddyscan), 0.88 (LFECA), and 0.67 (GK).

The effect of inspecting through 0.003 to 0.005 inch of paint is often a decrease in the probability of detection. However, this effect is due primarily to the difficulty in properly centering the probe over the rivets. Techniques that give the operator signal feedback that can be used to assure proper centering are effective in removing thin layers of paint as a deterrent to the reliability factor.

We have also proposed a new analysis that extends the a-hat versus a analysis to include multiple dimensions of a signal. The methodology proposes to treat the flaw characteristic as the dependent variable and the (possibly many) signal characteristics as independent variables. The fits are then treated as the dependent variable with the flaw characteristics as the independent variables in an a-hat versus a analysis to estimate the probability of detection curve. This proposed methodology is shown to be equivalent to the a-hat versus a analysis when there is a single-dimensional signal.

1. INTRODUCTION.

As part of its validation efforts, the FAA Airworthiness Assurance Nondestructive Inspection (NDI) Validation Center (AANC) maintains a set of lap splice specimens with cracks from beneath flush-head fasteners. These specimens are used in blind experiments where the inspector does not have knowledge concerning the specific location or size of the cracks. Manufacturers, developers, and users of various eddy-current equipment have performed experiments to characterize reliability using these test specimens. This document presents results of various experiments performed over the past 3 years.

The discussed equipment ranges from the newly developed to those that are right off the shelf. All of the inspections employed eddy currents and all were done on the same set of test specimens. Where possible, signal strength data were gathered in addition to detection versus nondetection data. This allows for a strengthened analysis since the criteria for making calls can be altered after the inspection and the trade-off between detections and false calls can be characterized.

A major source of variation in inspections is inspector-to-inspector difference [1]. Inspector differences that can affect results include decision making as well as skill aspects of manipulating a probe. In general, a single individual using a single piece of equipment performed each of the inspections. Thus, most of the detection curves derived here are particular to an instrument-inspector combination. The inspectors included equipment developers, company sales representatives, and experienced NDI technicians. All inspectors were experienced or well versed in the capabilities of the inspection technique they were using. Nevertheless, care should be taken in interpreting the detection and false call data presented in this report. In transferring the results of this report to use-conditions additional consideration should be given to the possible effect of conditions of the inspection that may or may not have been addressed in this report.

Equipment used in this survey include the following off-the-shelf equipment: Nortec 30 Eddyscan, Krautkramer Branson Crackfinder, Hocking FastScan, Nortec 19e with sliding probe procedure, and the Zetec MIZ-22 with pencil probe and template procedure. The last two are instruments and techniques commonly employed in aircraft maintenance facilities [1]. For these two instruments we study procedure variables in order to establish capability for detection. The capability can then be compared to field achieved reliability as reported in reference 1.

Also included are instruments or probes recently developed that are not commercially available. These include the Northrop developed low-frequency eddy-current array (LFECA), the NASA developed rotating self-nulling probe, and a McDonnell Douglas Aerospace/GK Engineering surface scanning probe used with the Elotest B1 mini rotor. The test results for each of these pieces of equipment will be presented in subsequent sections.

2. SPECIMEN DESCRIPTION AND ANALYSIS TECHNIQUES.

The inspections reported here were not performed solely for the purpose of deriving probability of detection (PoD) curves. In many cases, developers were looking for firsthand experience on the type of specimens provided. They used lessons learned from the inspections to alter

equipment or procedures. Meeting several goals in doing the inspections means that there are some variations in the number of inspections performed for each of the techniques discussed. Multiple inspections occurred using different procedures and equipment changes with some of the techniques. The commonality of all the inspections was that they were performed in the AANC hangar in Albuquerque and were all monitored by AANC personnel and all used test specimens from the same basic set.

2.1 SPECIMEN OVERVIEW.

The test specimens included 43 panels, each 20 by 20 inches and configured in a lap splice. The lap splice is a 3-inch overlap of 0.040-inch-thick 2024-T3 clad aluminum. The lap splice consists of three rows of rivets with 1-inch spacing between rows as well as between the rivets within a row. The rivets are countersunk flush 5/32 rivets. All flaws are fatigue flaws in the top skin with lengths measured from the rivet shank ranging from 0.014 to 0.8 inch. In this configuration, a crack of 0.050 inch would extend slightly from beneath the countersunk rivet head. The flaws were grown in the top skin prior to riveting to the bottom skin. (Details of crack length distribution and specimen fabrication are given in reference 1.)

All of the panels were painted. Paint thickness was nominally 0.003 to 0.005 inch thick. On 18 panels the paint was removed from the lap splice leaving a bare aluminum inspection surface around the rivet sites.

In addition, two large panels (8.5 feet long by 4 foot tall) each containing 102 rivet sites were used in some of the inspections. These panels also contained a lap splice but in addition contained stringer and frame substructures that are representative of those found in a Boeing 737 aircraft. Flaws in these structures are fatigue flaws that were grown after the structure had been built. One panel was painted and the other panel was a bare aluminum surface.

Also available as a test specimen was the AANC's Boeing 737 test bed. Inspections were not performed on the 737 in all cases. Specific crack length data are not known for cracks that may be underneath rivets in the Boeing 737 test bed. Thus, information specific to deriving PoD curves would not be available. However, in some cases signals from the aircraft are compared to the signals taken on the fabricated test specimens to judge commonality. The section of the Boeing 737 used in these experiments consisted of 41 rivet sites from body stations 430 to 470 and on left stringer 14.

2.2 ANALYSIS OVERVIEW.

The analysis of inspection data depended on the form of the data. The models used as a foundation for the analysis are those discussed in reference 2 and implemented in the Probability of Detection analysis software developed at University of Dayton Research Institute for the US Air Force. They are briefly explained below.

PoD curves were derived for hit/miss (detect/no detect) data using binary regression with a probit link function [2,3]. When the data consisted of the signal strength on which a hit/miss call could be made, the PoD curves were fit using regressions or analysis of variance models based on the signal strength as the dependent variable. This latter method is referred to as an a-hat (or $\hat{a}$)

versus a analysis, where a is the crack length [2,3,4,5]. In this analysis some aspect of the instrument response, $\hat{a}$, is treated as being directly related to the crack length. Specifically, the regression model is most often taken to be $\ln(\hat{a}) = \beta_0 + \beta_1 \cdot \ln(a) + \varepsilon$, where a is the crack length and ε is the residual error and is assumed to have a Gaussian distribution with mean 0 and variance δ^2. The parameters β_0 and β_1, defining the relationship, and the amount of variation (δ) remaining after fitting the model are estimated by maximum likelihood estimation. The original quantities, $\hat{a}$ or a, either singly or together, can be used in the analysis instead of their logarithms. The choices are made according to which best satisfies the linear assumption.

In both the hit/miss and signal data models, the basic PoD model is extended to include explanatory factors other than crack length. Primary variables included in the analysis are the surface condition of the test specimen, angle of crack, and the number and location of cracks at each rivet inspection. The levels considered in the analysis presented here are given in table 1.

TABLE 1. TEST SPECIMEN FACTORS AND LEVELS INCLUDED IN ANALYSIS

Factors	Levels
Surface condition	paint, bare
Crack angle	0 deg (horizontal), 11 deg, 22 deg
Crack location	-1 (left only), 1 (right only), 2 (both sides)

The crack lengths in the specimens are measured from the rivet shank. In some of the inspections, signals were produced that reflected overall conditions at the rivet site and were not specific to each side of the rivet where a crack might be. In these cases, analyses were performed for each of the explanatory variables, *sum of crack lengths* and *maximum crack length*. In those cases where the rivet contains two cracks, the *sum of crack lengths* does not represent a crack-tip-to-crack-tip length because the rivet hole diameter is not included.

All the figures showing estimated PoD curves also show the 95 percent upper confidence bounds for the crack length estimated to have a detection rate of 50 percent and the crack length estimated to have a detection rate of 90 percent. This was thought to be adequate to indicate the strength of the data for making the PoD curve estimations without cluttering the plots. The method for calculating the confidence bounds for probability of detection crack lengths is given in references 2 and 3. In tables addressing false calls, the false call rate is denoted by FCR and the 95 percent upper confidence bound is denoted by FCR_{95}.

3. EDDY-CURRENT EQUIPMENT AND TECHNIQUES.

3.1 NORTEC 30 EDDYSCAN.

The inspection reliability of the Nortec 30 Eddyscan was reported on earlier [6]. A follow-up to the earlier characterization in which signal strength and different setup standards are used is reported on here.

3.1.1 Nortec Experiment Background.

In a follow-up to an earlier reported experiment [6], signal data were acquired on 12 of the test specimens. A Nortec representative performed all inspections over a period of 2 days. The inspections included a setup on a 0.080-inch standard and a setup on a 0.060-inch standard. Limited data were also taken from a setup on a 0.040-inch standard. The test specimens included an equal number of painted and bare panels.

For each of the inspections, the equipment was set up for the signal to read approximately 80 percent full scale on the size of the electro-discharge machined (edm) notch of interest. For the 0.080-inch set up, the 0.060-inch standard produced a signal of 40 percent full scale, and the 0.040-inch standard produced a signal of 16 percent full scale. In the 0.060-inch setup, the 0.080-inch standard produced a saturated signal and the 0.040-inch standard produced a signal of 30 percent full scale. Table 2 contains the resulting setup parameters. Also shown for comparison are the parameters from reference 6. Note: The alarm level was set during the inspections at a level slightly less than one-half of the level obtained for the setup notch. However, the analysis addresses the impact of changing that value.

TABLE 2. NORTEC EDDYSCAN INSPECTION PARAMETERS

Parameter	Set up on 0.080-Inch Standard	Set up on 0.060-Inch Standard	Reference 6 Setup
Gain	28.0 dB	33.0 dB	28.5 dB
Gate start	250 micro seconds	100 micro seconds	100 micro seconds
Gate width	50 micro seconds	50 micro seconds	30 micro seconds
Alarm level	35 percent	35 percent	not used
Rotation	0 degrees	0 degrees	354 degrees

During the inspections, the signals for any rivet inspections that were noticeably above background were stored in the Nortec 30. This included all signals above 10 percent full scale and most between 5 and 10 percent full scale. Upon reaching storage capacity, all signals were printed and labeled. For analysis, the percent of full scale was measured directly off the printed signals.

3.1.2 Nortec PoD Analysis.

The initial analysis of the signal data was to determine the call level, as a percent of full scale, for which false calls were produced. There are two types of false calls made. The first type is to make a call that a crack exists at a rivet when there is no crack there. The other type of false call is to make a call of multiple cracks at a rivet when there is only a single crack. Table 3 shows the number of false calls, the rate, and the 95 percent upper confidence bound for three different call levels.

TABLE 3. NORTEC EDDYSCAN FALSE CALLS

Call Level (Percent Full Scale)	Setup on 0.080-Inch Standard		Setup on 0.060- Inch Standard	
	Crack-Free Rivets Identified, FCR, FCR_{95}	False Signals at Cracked Sites, FCR, FCR_{95}	Crack-Free Rivets Identified, FCR, FCR_{95}	False Signals at Cracked Sites, FCR, FCR_{95}
30	0, 0, 0.018	0, 0, 0.069	0, 0, 0.018	0, 0,0.069
25	0, 0, 0.018	1, 0.024, 0.108	2, 0.012, 0.039	1, 0.024, 0.108
15	1, 0.006, 0.029	1, 0.024, 0.108	2, 0.012, 0.039	3, 0.071, 0.174

An $\hat{a}$ versus a analysis was performed on the data shown in figure 1 for each of the setups. In the graphs of figure 1, different symbols are used for the cases of the signal being at a rivet site containing only a left crack, only a right crack, or a crack both left and right. Two cracks at the same rivet would be diametrically opposed from each other. Under these conditions, there was more crack signal variation due to an apparent increase in the uncertainty of the centering of the probe. That is, the inspector was able to shift the signal from one side to the other as he moved the probe.

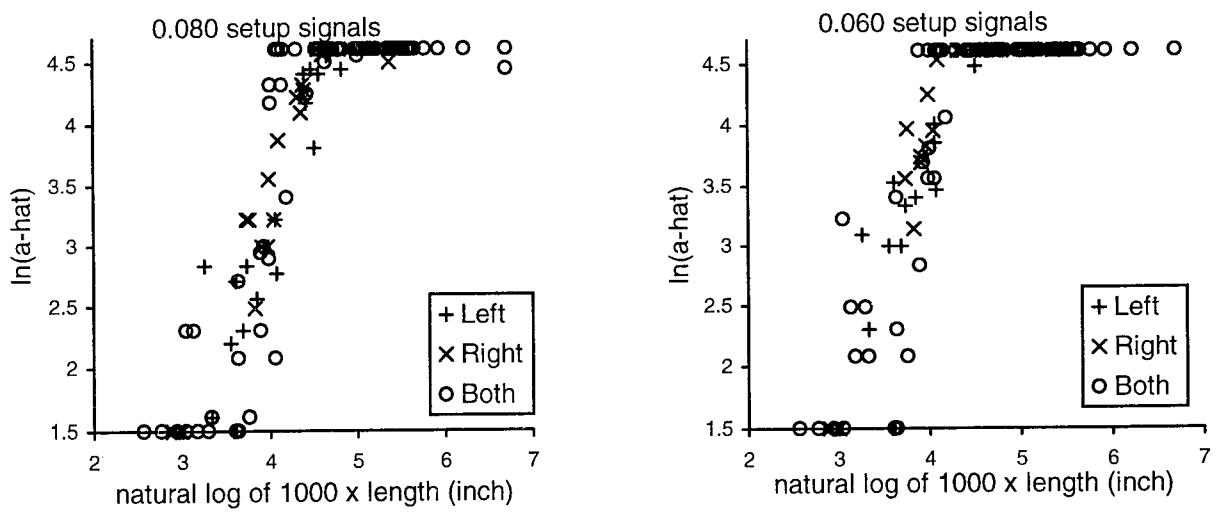

FIGURE 1. NORTEC EDDYSCAN LOG SIGNAL STRENGTH VERSUS LOG CRACK LENGTH

A similar phenomenon accounts for the false signals for a second crack when only one crack was present. In this case, if the single crack was moderately large, it had the effect of pulling the center of the probe toward it. In doing so, the probe was more likely to pick up the rivet edge on the opposing side and display a signal that was interpreted as a second crack.

The one point just under the saturation line (x = 6.7 = ln(812 mils), y = 4.6 = ln(100 percent)) in the 0.080-inch standard setup was considered to be a saturation point in the analysis. Although the signal was not saturated for this point, it occurred at a rivet containing two cracks. The other side of the signal was saturated and corresponded to a crack that was large and visually detectable. The phenomenon of signal transfer from side to side as previously discussed was

believed to be in effect. This, combined with the fact that this point exerted substantial influence on the results, led to considering it as saturated.

PoD curves estimated from the log signal strength versus log length analysis are given in figure 2 for three threshold levels and the two setups. An estimated 90 percent detection rate was achieved at both setups for cracks the length of the setup notch, using 30 percent of full scale for making calls. This was with no false calls. At a threshold of 15 percent full scale, false call rates for a rivet site were in the 0.5 to 1.0 percent rate. The 90 percent detectable crack lengths were 0.061 and 0.047 inch. These are consistent with previously reported values [6] but indicate that detection reliability can be improved without substantial changes in the false call rate for rivets by setup on a smaller standard.

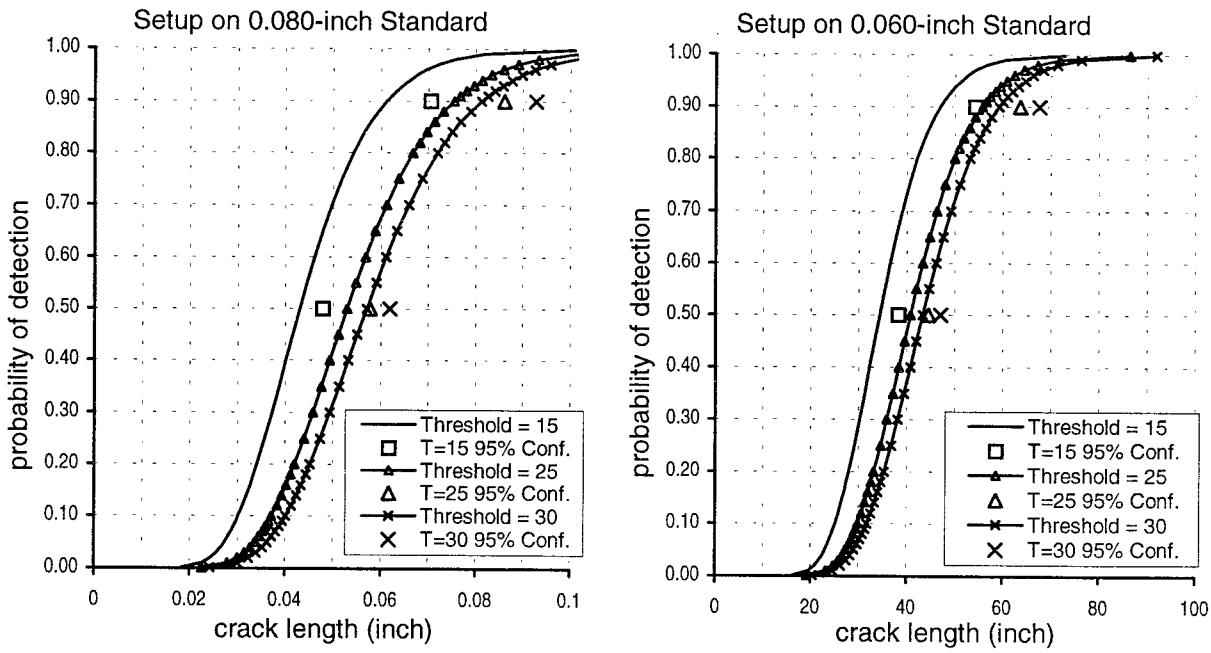

FIGURE 2. NORTEC EDDYSCAN PoD CURVES

Table 3 indicates that the false call rate for calling two cracks at sites that contain only a single crack is higher than the rate for calling a noncracked rivet as cracked. With the 15 percent full-scale threshold and using the 0.060 setup, the rate for adding a crack to an already cracked rivet was around 7 percent (3/42). Making such a call is likely to be inconsequential, as repair actions will likely assure further investigation.

The use of a standard electro-discharge machined 0.040-inch notch as a reference setup was pursued briefly. One of the painted and one of the bare panels were each inspected using a 0.040-inch edm notch as a reference standard and setting the signal to 80 percent of full scale. The parameters for this inspection were the same as for the 0.060 setup in table 2 except that the gain was set at 41 decibels. There were not enough data to estimate PoD curves properly. However, it was clear from this limited inspection that the inspector spent more time centering the probe and attributed more of the signals to background noise from the rivets.

3.2 KRAUTKRAMER BRANSON CRACKFINDER.

McDonnell Douglas nondestructive testing personnel performed inspections using the KrautKramer Bransons Crackfinder. The experimental results presented here are only a part of their independent evaluation of the instrument. Experiments using the Crackfinder were carried out in two different time periods. The first set of experiments was performed to characterize two different units combined with two different probes. The second set of experiments was performed to answer questions concerning procedures.

3.2.1 Crackfinder Experiment 1 Background.

Four configurations of the Krautkramer Branson Crackfinder were tested. Configuration 1 consisted of a base unit (commercially available) with a 0.25-inch probe operating at 500 kHz. Configuration 2 was the same base unit with a 0.50-inch probe operating at 320 kHz. Configuration 3 used a modified unit developed by McDonnell Douglas, in which the sensitivity had been reduced, with the same probe used in configuration 1. Configuration 4 was the reduced sensitivity unit with the 0.050-inch probe of configuration 2. In reporting results the following codes will be used: U25 (unmodified unit with 0.25-inch probe), U50 (unmodified unit with 0.50-inch probe), RS25 (reduced sensitivity unit with 0.25-inch probe), and RS50 (reduced sensitivity unit with 0.50-inch probe) for configurations 1 to 4, respectively.

Inspections of the small test specimens occurred in two different modes. In the first mode, inspections were performed with the test specimens mounted on frames to simulate a lap splice on an aircraft. In the second mode, the inspections were performed on a laboratory bench top. All inspections of the two larger panels were performed with the panels in a vertical orientation, simulating the side of an aircraft.

The inspections of the large panels and the frame mounted smaller panels required the inspector to handle the test equipment just as he would when inspecting on the side of an aircraft. The test surface was vertical with a row of rivets at about eye level. In this case the inspector used a neck strap and rested the bottom of the instrument on the lower part of the chest. The neck strap held the upper portion of the instrument at an angle that made it easy to look down from the inspection site to view the instrument.

Prior to starting the inspection the inspector used standards with edm notches of known 0.150-, 0.100-, 0.080-, and 0.050-inch lengths in setting up the inspection task. The unit was adjusted to obtain full scale (39-40 light bars) for the 0.150-inch notch. Setup readings of 29, 19, and 4 light bars were obtained for the 0.100-, 0.080-, and 0.050-inch notches, respectively. All subsequent setups nominally matched these values.

All the inspections for the 0.25-inch probe were performed using a 21/32-inch circular guide for the probe from a draftsmen template. Those for the 0.50-inch probe used a 7/8-inch circular guide. The inspector visually centered each rivet within the guide and then ran the probe several times around the guide. The position of the probe at the time of a signal enabled the inspector to characterize the direction of the crack. For all detects, the direction (to the resolution of top, bottom, left, or right) and the signal strength (number of lights out of 40 possible) were recorded.

3.2.2 Crackfinder Experiment 1 Analysis.

Figures 3 and 4 show the data that were analyzed. For the Crackfinder, 40 light bars represent saturation. No calls were made that were less than 5 light bars. Thus, the analysis was performed using 4 as a recording threshold. This is equivalent to assuming that all signals for which no call was made, were 4 or fewer light bars. In the $\hat{a}$ versus a analysis, saturated signals were treated as right censored values and the no calls were treated as left censored [7].

An initial analysis indicated that the bench top versus frame factor made no difference in the results. Thus, this factor is not presented here. The paint and bare surface inspection conditions were statistically different for most of the inspections. Estimated PoD curves are given for each of the unit and probe combinations in figures 5-8. False call rates are given in table 4 along with 95 percent upper confidence bounds. The thresholds of 5, 10, and 20 light bars for making calls are represented in each of the figures.

In all cases the false call rates were higher for the painted surface than they were for the bare surface. The detection rates were less for the painted surface than for the bare surface (as reflected in larger crack lengths to achieve the same detection rate) in all cases except the reduced sensitivity unit with the 0.25-inch probe. However, the false call rates were substantially higher for all three thresholds on the painted surface (0.19, 0.14, 0.03) than for the bare surface (0.01, 0.01, 0). The paint condition had a greater impact when the 0.50-inch probe was used than when the 0.25-inch probe was used. (Note the scale change between the bare and painted surface graphs of figures 7 and 8.)

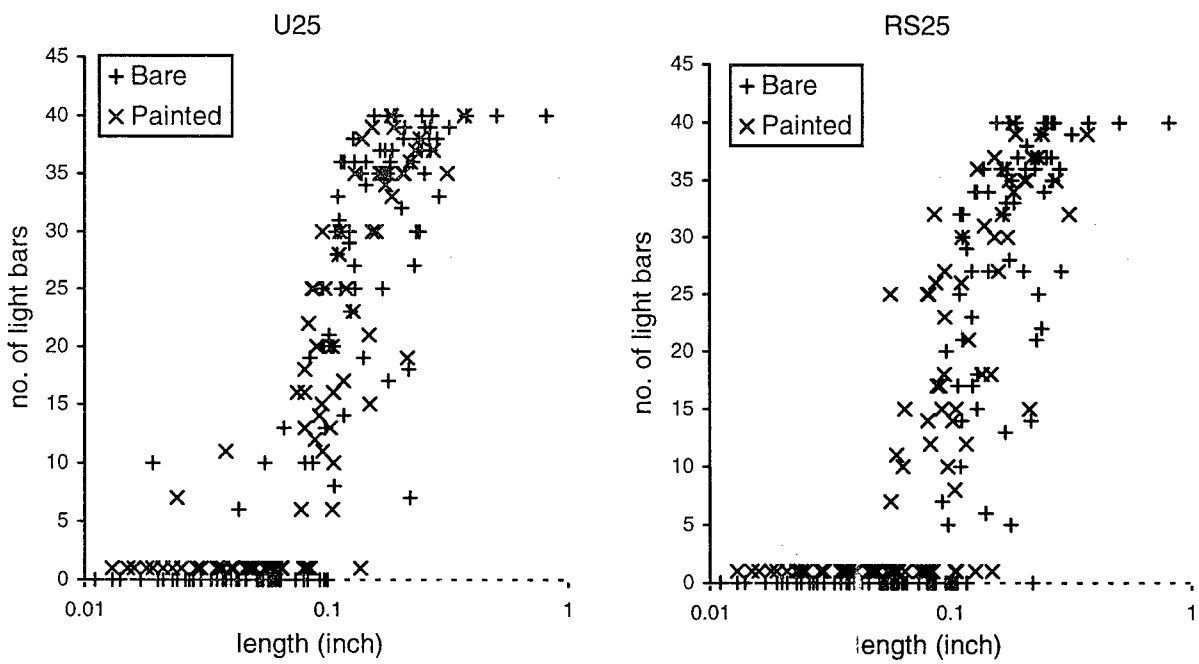

FIGURE 3. CRACKFINDER DATA FOR 0.25-INCH PROBE-UNMODIFIED AND REDUCED SENSITIVITY UNITS

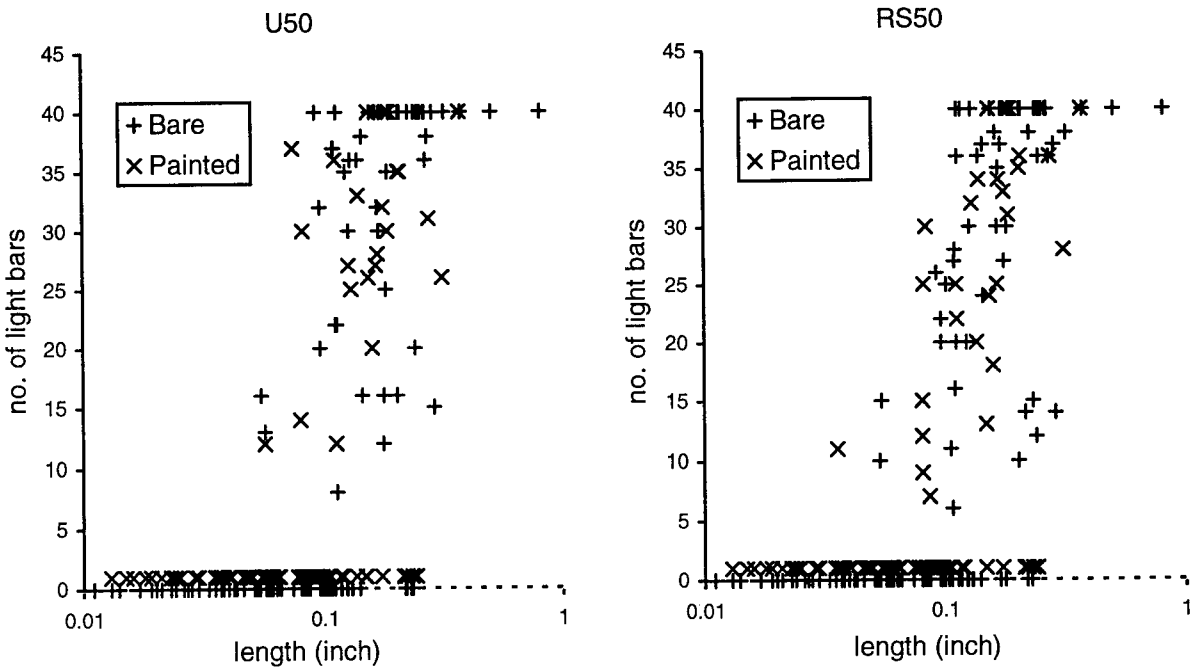

FIGURE 4. CRACKFINDER DATA FOR 0.50-INCH PROBE-UNMODIFIED AND REDUCED SENSITIVITY UNITS

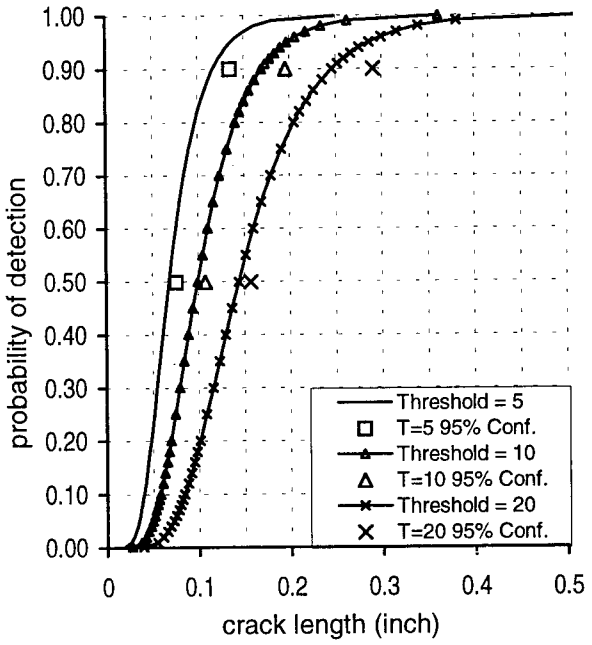

FIGURE 5. CRACKFINDER PoD CURVES FOR UNMODIFIED UNIT WITH 0.25-INCH PROBE-COMBINED INSPECTION SURFACE

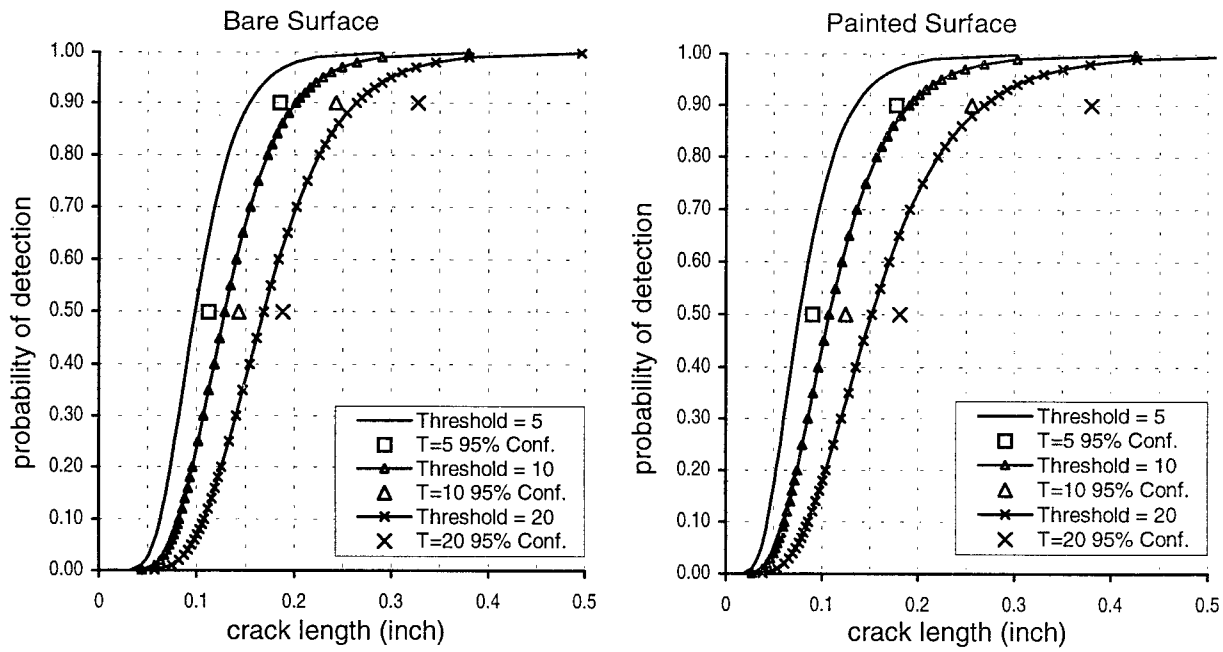

FIGURE 6. CRACKFINDER PoD CURVES FOR REDUCED SENSITIVITY UNIT WITH 0.25-INCH PROBE

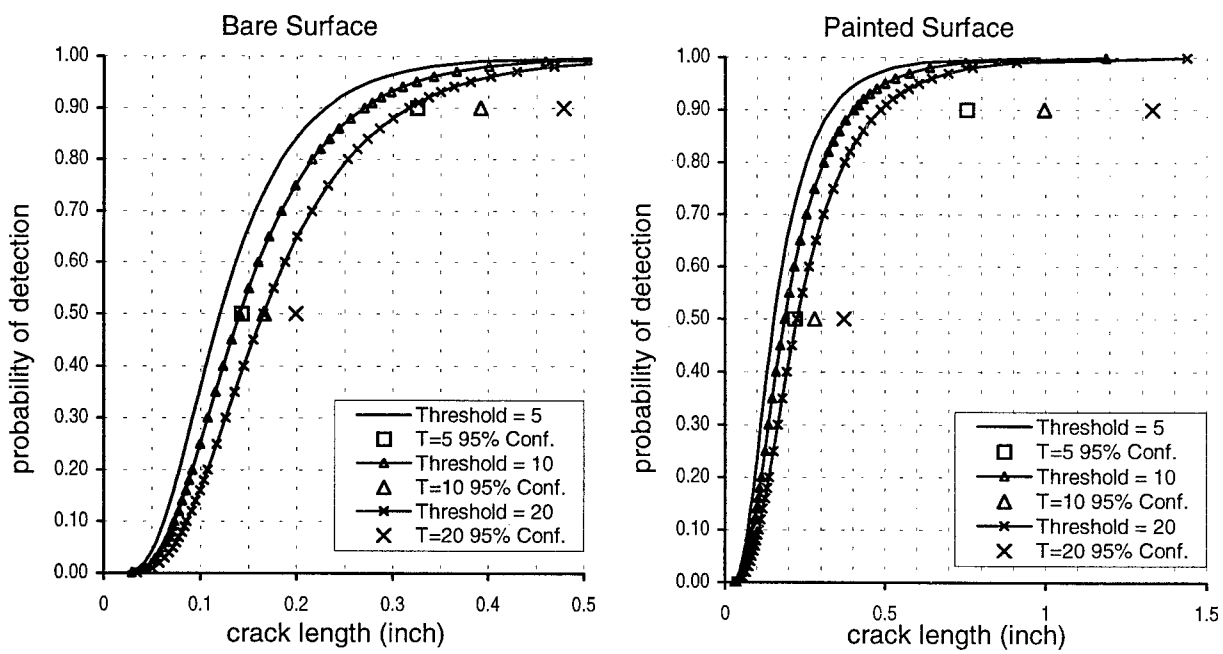

FIGURE 7. CRACKFINDER PoD CURVES FOR UNMODIFIED UNIT WITH 0.50-INCH PROBE

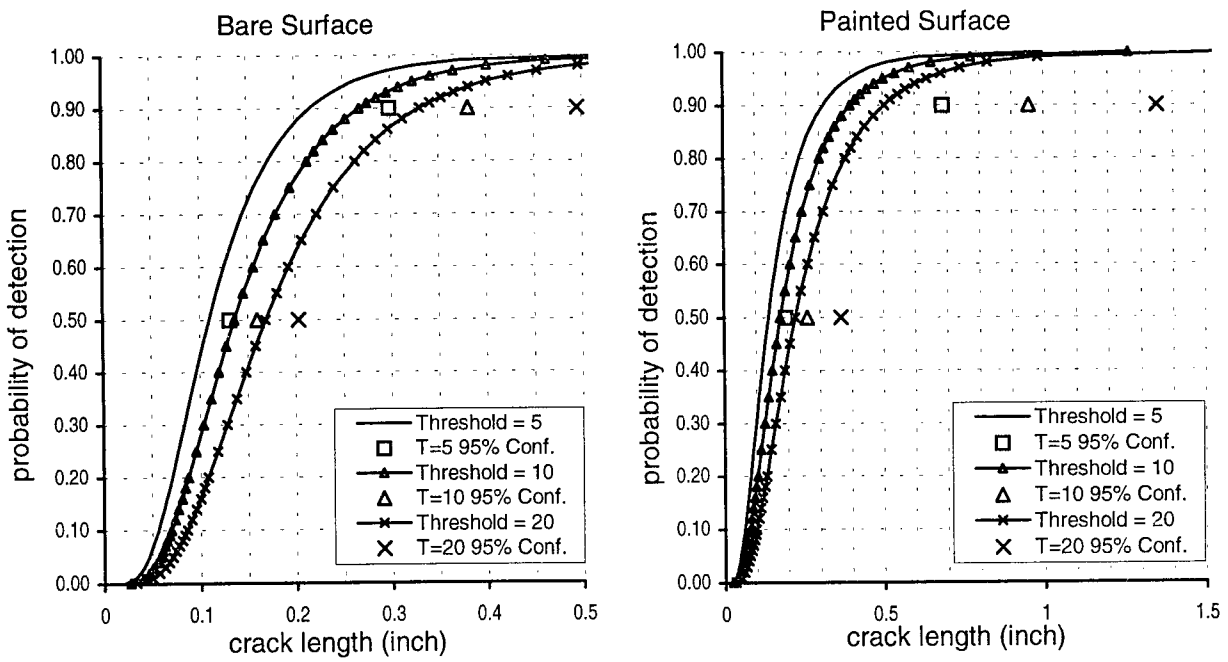

FIGURE 8. CRACKFINDER PoD CURVES FOR REDUCED SENSITIVITY UNIT WITH 0.50-INCH PROBE

TABLE 4. CRACKFINDER FALSE CALL RATES WITH 95% UPPER CONFIDENCE BOUNDS ( )

Unit/Probe	Surface	Threshold 5	Threshold 10	Threshold 20
U25	Both	0.149 (0.172)	0.073 (0.091)	0.010 (0.019)
RS25	Bare	0.011 (0.028)	0.007 (0.023)	0 (0.011)
	Painted	0.189 (0.223)	0.137 (0.168)	0.028 (0.045)
U50	Bare	0.004 (0.017)	0 (0.011)	0 (0.011)
	Painted	0.019 (0.034)	0.017 (0.031)	0.007 (0.018)
RS50	Bare	0.011 (0.028)	0.004 (0.017)	0 (0.011)
	Painted	0.040 (0.060)	0.021 (0.037)	0.007 (0.018)

The assumption about the lower threshold for storing the data will impact the estimated PoD curves. Thus, if a threshold of 5 rather than 4 is assumed, the fitted PoD curves would shift to the left (that is, higher estimated probability of detection). In the cases looked at, the shift was between 0.010 and 0.015 inch. Because of the integral nature of the response, which was the number of lights in the signal, various modeling scenarios could be argued. The modeling error associated with these assumptions is of the order of 10 to 15 mils in the placement and shape of the various PoD curves.

3.2.3 Crackfinder Experiment 2 Background.

A second set of experiments was performed to obtain probability of detection information using different procedures. Two procedures were employed in the second visit. The first procedure will be referred to as the template procedure. For that procedure the inspector used a 13/16-inch aperture on a draftsman template, with the 0.25-inch probe operating at 500 kHz. The template procedure differed from that of the first visit in that it widened the aperture, keeping the probe completely off the rivet head. This was not the case in the first visit. This was done to eliminate the numerous false calls.

The second procedure used the same probe but employed a straight edge to pass the probe along a vertical tangent to the rivet head edge. This procedure will be referred to as the transverse procedure, as the intent was to pass the probe perpendicular to the orientation of the cracks. The known location of cracks on the test panels (left and right sides of the rivet) required the latter procedure to be done in two passes, one for each side of the rivet. A straight edge was placed along the rivet edge and used as a probe guide to assure that the probe itself was not passing over the rivet edge. For efficiency of measurement, one side of the 20 rivets on each test panel was inspected, the panel turned and the other side of the rivets was then inspected. For both procedures, the reduced sensitivity unit from experiment 1 was used.

The Crackfinder was set to light 20 bars (1/2 scale) for a 0.150-inch standard. For the painted panels the instrument was renulled and was checked on a known defect-free site where the signal was a single bar. To simulate the affect of the painted surfaces the 0.15-inch standard was checked through two sheets of paper. The signal was not noticeably degraded.

3.2.4 Crackfinder Experiment 2 Analysis.

The procedures of the second visit were successful in eliminating the false calls. There was one false call made with each of the two procedures. This was out of 700 defect-free rivets. For the transverse procedure each of two sides was inspected. Thus, there were 1496 distinct chances to produce a false call of which only one false call resulted. After the initial template inspection, the inspector was asked to reinspect the rivet on which the one false call was made. The reinspection did not produce a call signal. This implies that the earlier false call was technique related.

The signals as measured by the number of bars versus crack length are presented in figure 9. Of immediate note is that cracks of lengths exceeding 0.2 inch were missed with the template method. Both procedures exhibited a general rise of the signal strength with the crack size. However, the template procedure produced very few nonsaturated signals above the threshold of four lights. Those that did occur were all on bare panels. The transverse readings seemed to plateau at 35 bars rather than 40 bars. This latter phenomenon was observed during the inspection but went unexplained. No intentional resetting of the equipment had taken place. (The transverse procedure followed the template procedure in gathering the data.)

The estimated PoD curves for each of the procedures are shown in figures 10 and 11. The PoD curve for the higher threshold is not shown, as it is unnecessary to go to 20 lights to remove the false calls. In the template procedure there were eight signals that were neither saturated nor

below the assumed detection threshold of four lights. The signals from the eight were all above 10 lights, and therefore the thresholds of 5 and 10 light bars would result in the same detects. The curves for the two thresholds are similar and reflect the largest crack missed of 0.203 inch and the smallest crack detected of 0.128 inch.

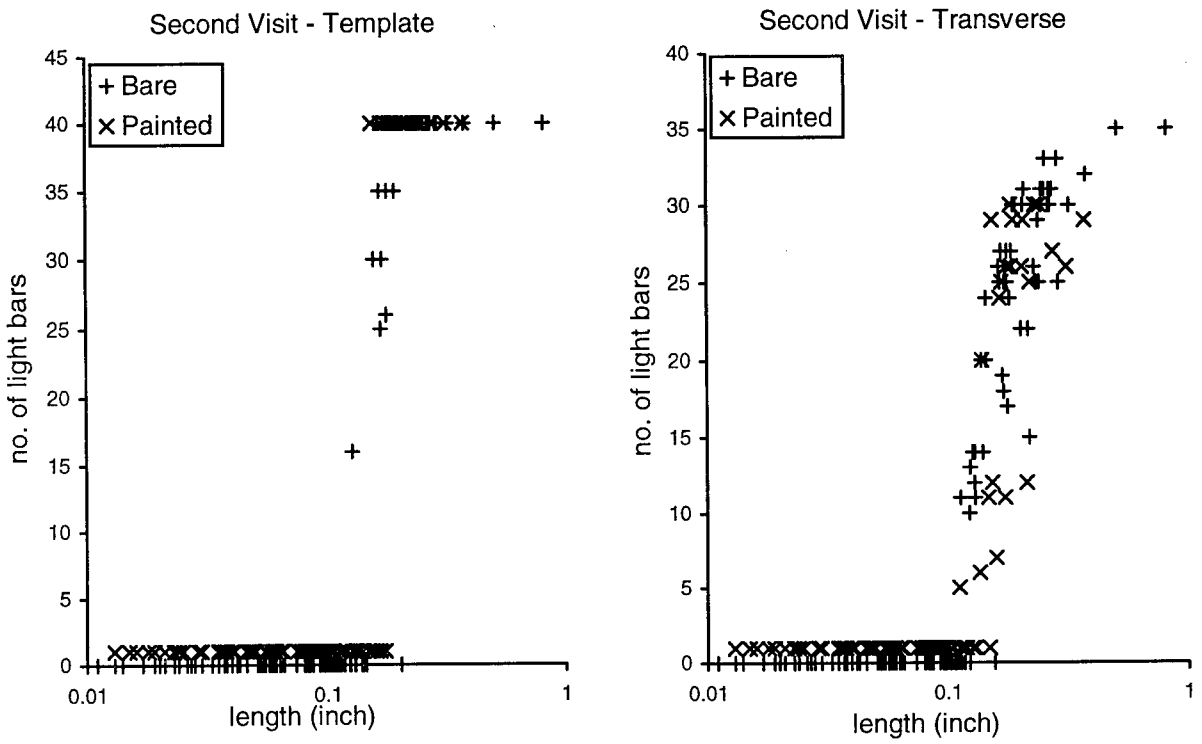

FIGURE 9. CRACKFINDER SIGNALS-TEMPLATE AND TRANSVERSE PROCEDURES

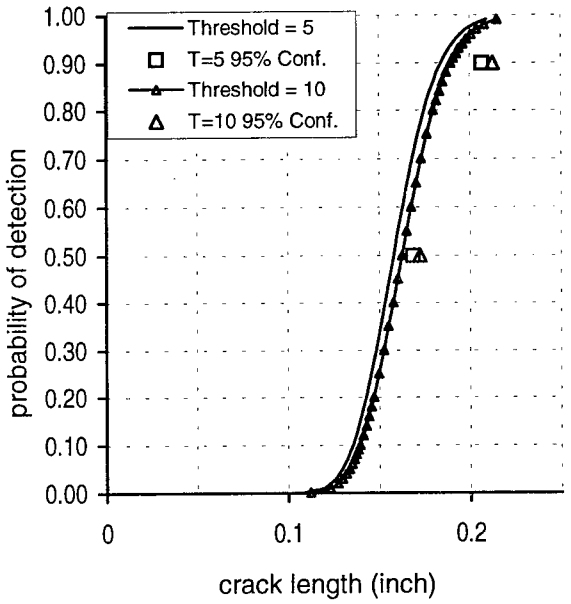

FIGURE 10. SECOND VISIT CRACKFINDER PoD-TEMPLATE PROCEDURE

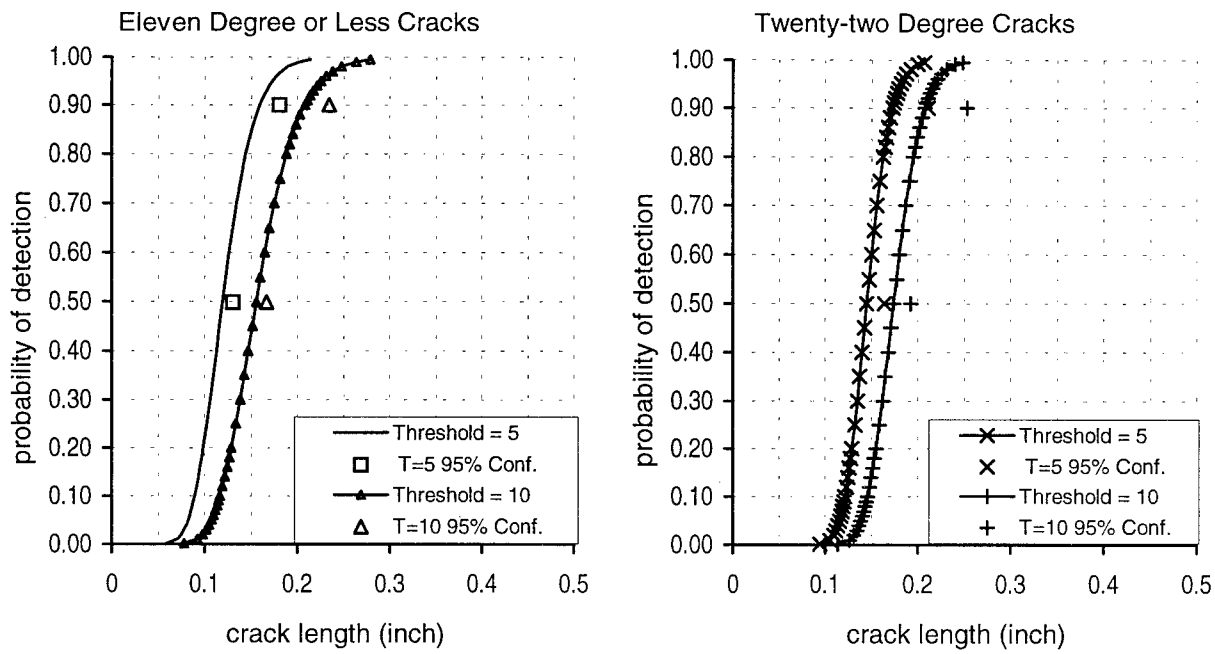

FIGURE 11. SECOND VISIT CRACKFINDER PoDs-TRANSVERSE PROCEDURE

There was no statistically significant difference in the effect of the surface condition on inspection results for either procedure used during the second visit. However, because the transverse procedure results in the probe moving along a tangent to the rivet head, geometrical considerations suggests that the angle of the crack could have an influence on the detection rate. The panels have flaws nominally at 0, 11, and 22 degrees from horizontal. Therefore the effect of off-angle cracks could be studied. The crack angle effect was non existent within the template procedure, as would be suggested by its nature. For the transverse procedure the effect of going from horizontal to 11 degrees off-angle was not significant, but the effect of the 22 degree off-angle cracks from the others was significant ($p < 0.05$). The estimated PoD curves for these two categories are shown in figure 11. The PoD curves for the 22 degree cracks are steeper, but have approximately the same 90 percent detection crack lengths as was estimated from the horizontal and 11 degree cracks. Thus, the smaller cracks are less likely to be detected.

3.3 HOCKING FASTSCAN.

3.3.1 Hocking FastScan Background.

The FastScan probe consists of the probe assembly and an aperture guide. To inspect for cracks from beneath fasteners, the aperture guide is centered on the rivet and then the probe is placed into the aperture. Bearings on the perimeter of the guide allow the probe to be freely rotated. The probe is then rotated by wrist action for 90 to 180 degrees. Through balanced coils and symmetrical rotation a complete signal from around the rivet is obtained. The probe is designed to minimize the signal from the rivet itself. Dual frequency techniques are also employed to reduce the common mode signal profile. The FastScan probe can be used with various Hocking instruments. In the experiments reported here, it was used with the Phasec 3.4. The Phasec 3.4 provides three frequencies and a four-channel operation.

A representative from Hocking performed the inspections using the Hocking FastScan over a two-day period. The aperture guide and probes come in several sizes to match the size of the rivets being inspected. In the inspections reported here, Hocking probe 851P103 and aperture guide 851A003 were used.

Results from three different setups are reported. In all cases, the probe liftoff signal was in a horizontal direction and crack signals were in a vertical direction. The first setup was based on a standard containing a 0.040-inch notch. The signal for the 0.040-inch standard covered approximately 8 divisions on the instrument screen. Twelve panels were inspected using this setup on the first day. Parameters for this and subsequent setups are given in table 5. The Hocking naming conventions for the four channels are used in table 5.

TABLE 5. HOCKING FASTSCAN INSPECTION PARAMETERS

	Channel	Freq. (kHz)	X-gain (dB)	Y-gain (dB)	Phase (angle)	X Sft	Y Sft
Setup 1 (0.040-inch standard)	1	6.00	51.0	51.0	2.5	0	0
	2	30.0	60.0	60.0	197.5	0	0
	2'		11.6	6.2	337.0	0	0
	ΣA		5.0		301.5	0	-40
Setup 2 (0.080-inch standard)	1	10.0	46.0	46.0	333.0	0	0
	2	30.0	52.0	52.0	197.5	0	0
	2'		9.0	0.8	11.0	0	0
	ΣA		5.0		345.5	0	0
Setup 3 (0.040-inch standard)	1	10.0	52.0	52.0	333.0	0	0
	2	30.0	58.0	58.0	197.5	0	0
	2'		9.0	0.8	11.0	0	0
	ΣA		8.5		345.5	0	0

In the original setup, signals from rivets at the edge of the panels were very similar to the signals from the 0.040-inch standard. The inspector was forced to judge the signal as being due to edge effects or due to the presence of a crack. Ten percent of the rivet sites were on the edges. After the first days' inspection the inspector experimented with the setup and eliminated the edge effect with setup 2. The parameters of setup 2 were chosen to make the 0.080-inch standard produce a signal that was close to full screen. The setup 2 standard signals were approximately 11 divisions for the 0.080-inch notch, 6 divisions for the 0.060-inch notch, and 3 divisions for the 0.040-inch notch.

Setup 3 differed from setup 2 in that the gains were increased so that the 0.04-inch standard produced a signal of approximately 10 divisions. Only the six bare panels were inspected with this setup.

3.3.2 Hocking FastScan Analysis.

The ΣA channel signal was used to make calls. The inspector would call out the number of screen divisions for the signal at each rivet site. An experiment monitor recorded the values. The inspector did not attempt to identify individual cracks at each rivet site, as this information was not readily available from the displayed signal. In figure 12 the signal magnitudes are plotted versus the maximum crack length at each rivet for each of the setups. Any signals above 16 divisions were reported as saturated and are plotted at 17.

In setup 1 the smallest crack (0.026-inch) with a saturated signal occurred at one of the edge sites. The inspector attributed the signal to an edge effect and did not make the call. Of the 24 rivet sites that were inspected on the edge of panels, 10 of them had cracks. All but two of these cracks were properly identified in setup 1 as being cracked. The two smallest cracks (0.026 and 0.042 inch in length) were missed. Setup 2 signals on these cracks were 0.5 and 2.5, respectively.

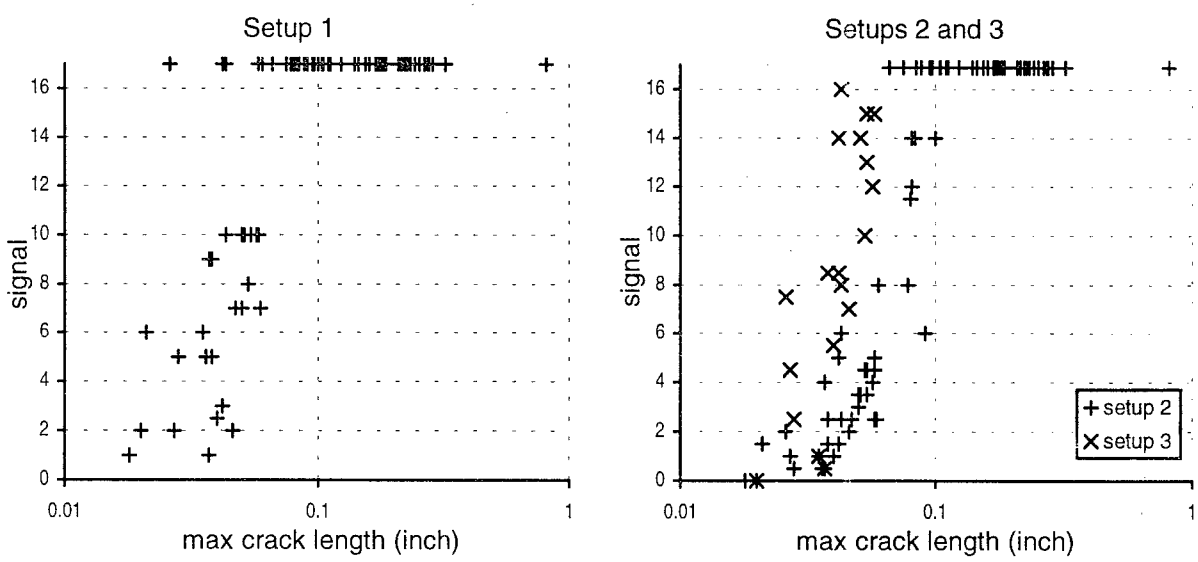

FIGURE 12. HOCKING FASTSCAN SIGNALS ON SMALL PANELS

Table 6 shows the number and rate of false calls at different signal thresholds. These are important when considering the PoD values estimated for the various signal thresholds. The false call rates for bare versus painted surfaces were significantly different for setup 1. Both are shown. For setup 2 they were not significantly different, and setup 3 was a limited experiment performed only on bare panels.

TABLE 6. HOCKING FASTSCAN FALSE CALL RATES VERSUS SIGNAL THRESHOLDS

Threshold	Setup 1 - Painted 82 Unflawed Sites Inspected	Setup 1- Bare 79 Unflawed Sites Inspected	Setup 2 161 Unflawed Sites Inspected	Setup 3 79 Unflawed Sites Inspected
	No.- (rate %)	No.- (rate %)	No. (rate %)	No. (rate %)
10	1 - (1)	0 - (0)	0 - (0)	0 - (0)
5	8 - (10)	1 - (1)	0 - (0)	0 - (0)
3	20 - (24)	8 - (10)	2 - (1.2)	3 - (3.8)
2	40 - (49)	33 - (42)	12 - (7.5)	26 - (33)

Since the signal data are by rivet site, there is a question as to the appropriate value to use for the explanatory variable *crack length* at those sites having two cracks. We considered two different models. In one model we use the sum of the crack lengths as the independent variable with the number of cracks present being an additional explanatory factor. In the second model we consider the maximum crack length present as the independent variable with the number of cracks, again, being an additional explanatory factor. In both cases, if a single crack was present the independent variable was its length.

When using the *sum of crack lengths* as the explanatory variable, the factor *number of cracks* is statistically significant (p=0.01). Using a call threshold of 10, the mean (50% detection rate) for a single crack is 0.052 inch and the mean for 2 cracks is 0.070 inch.

When using the *maximum crack length* as the chief explanatory variable and using a call threshold of 10, the mean detection length for a single crack is estimated to be 0.051 inch. For a double crack, the mean detection length is estimated to be 0.044 inch. However, this difference was not statistically significant (p=0.08). Therefore, the variable *maximum crack length* was chosen for additional analysis.

The surface condition (paint versus bare) was not a significant factor. Neither was the interaction between surface condition and the number of cracks present. The estimated PoD curves for the three setups are shown in figures 13 and 14. The observed false call rates (FCR) and upper 95 percent confidence values (FCR_{95}) are given in table 7.

A threshold of 2 leads to excessive false call rates in all cases. A threshold of 3 leads to excessive false calls in the two setups that used the 0.040-inch standard, but the 1.2 percent false call rate in setup 2 might be considered tolerable. Using a threshold of 5 divisions for the more conservative setup 2 led to an estimated 90 percent detectable crack length as being 0.074 inch.

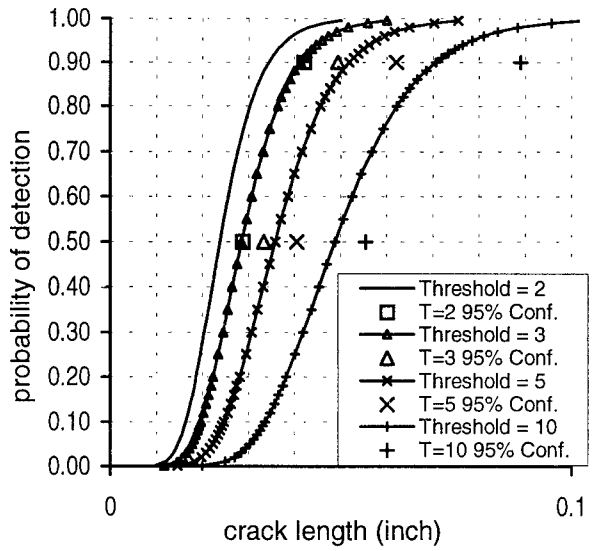

FIGURE 13. HOCKING FASTSCAN PoD CURVES-SETUP 1

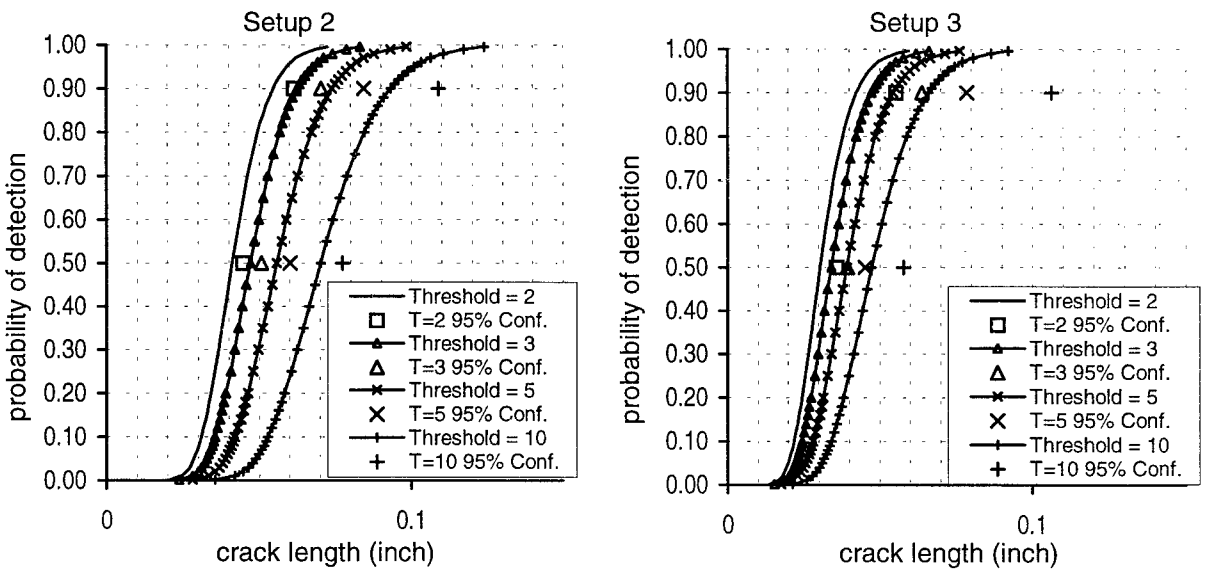

FIGURE 14. HOCKING FASTSCAN PoD CURVES-SETUPS 2 and 3

TABLE 7. HOCKING FASTSCAN FALSE CALL RATES

Setup		Threshold (divisions) 2	3	5	10
1	FCR (FCR_{95})	0.45 (0.52)	0.17 (0.23)	0.056 (0.096)	0.006 (0.029)
2	FCR (FCR_{95})	0.075 (0.12)	0.012 (0.039)	0 (0.018)	0 (0.018)
3	FCR (FCR_{95})	0.33 (0.43)	0.038 (0.096)	0 (0.037)	0 (0.037)

Not reflected in the PoD data is the time taken to do inspections. The wrist movement needed to turn the probe in the aperture guide was not taxing and the probe was easily moved. However, a good inspection is based on having the probe centered over the rivet. Centering requires proper placement of the aperture guide. This was done visually through a plastic film at the bottom of the guide. This was not an easy process due to shadowing within the aperture guide. Low light levels exacerbated the problem. The result was a slow overall inspection time because the inspector took many readings at rivet sites to ensure that a signal was not the result of poor centering.

3.4 NORTEC 19e WITH SLIDING PROBE.

3.4.1 Nortec 19e Background.

The use of a sliding probe to quickly inspect large numbers of rivets is common in aircraft inspection and the variation of detection capabilities observed in the field for this procedure is also substantial [1]. The purpose of this experiment was to obtain signal data that could be analyzed after the inspection in an attempt to correlate signal characteristics with crack characteristics. Thus, multiple "$\hat{a}$ versus a" types of analysis could be performed without the burden of multiple inspections.

An NDI technician performed inspections using the Nortec 19e. The signals from the Nortec 19e inspections were captured by an analog to digital converter and stored as (x,y) points. Eighteen of the small panels, evenly divided between bare and painted inspection surfaces, were used in the experiment. The panels were inspected at frequencies of 20 and 30 kHz. Since the intent was to analyze the signal after the inspection, the inspector made no real-time calls. He only gathered the signal data. This was done by attaching a straight edge with double sided tape to each panel in turn and then scanning all 20 rivet sites on the panel, while resting the sliding probe against the straight edge and scanning from left to right. The instrument settings used are given in table 8. Signals obtained from the setup standards are shown in figure 15.

TABLE 8. NORTEC 19e INSTRUMENT SETTINGS

	Run 1	Run 2
Frequency	20 kHz	30 kHz
Horizontal gain	66 dB	66 dB
Vertical gain	77 dB	77 dB
Rotation	317	328
Low pass filter	100	100
high pass filter	0	0
Probe - SPO-3806 (1 kHz-100 kHz)		

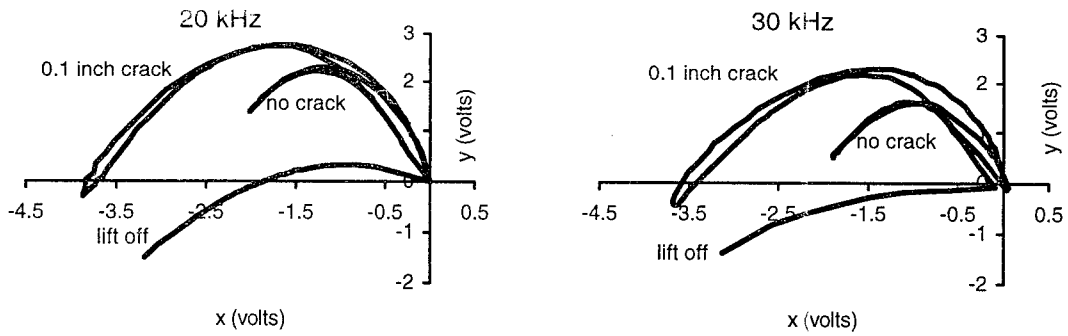

FIGURE 15. NORTEC 19e SLIDING PROBE SETUP SIGNALS

From figure 15 it is apparent that with the given setup, most of the signal movement is in the x-axis. However, the upward swoop of the curve is necessary to distinguish a rivet signal from lift off. For analysis purposes three points (x,y values) were recorded from these curves. The points are the maximum height point in the upward rise, the point farthest to the left, and the maximum height point on the return. The logarithm of the minimum negative x, (that is, the maximum x movement) versus maximum crack length are shown for each of the two operating frequencies in figure 16. In figure 16 data for all signals are shown, including the no flaw signals at x=0.

The signal strengths of approximately 1.6 in figure 16 resulted from saturated signals. Not shown in each of the plots are the 16 cracks with lengths greater than 0.2 inch. All resulted in saturated signals. It is also clear that the variation in the nonflawed signals is substantial enough to hide the smaller flaws. That is, if thresholds were set high enough to keep the false call rate down, then many of the smaller flaws would go undetected.

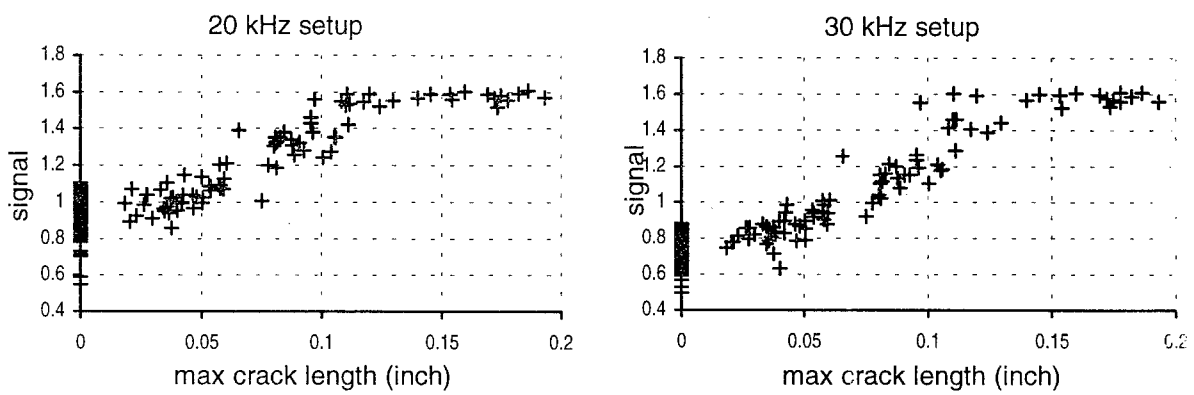

FIGURE 16. NORTEC 19e SLIDING PROBE SIGNAL

3.4.2 Nortec 19e Analysis.

The signal response variable graphed in figure 16 is the natural logarithm of $-x_{min}$. This one-dimensional summary of the whole signal (see figure 15) is used in the initial analysis. Additional aspects of the signal are used in latter analysis. The analysis is initially based on the

signals from non flawed rivets. That analysis is followed by a regression analysis on the signals from the flawed sites. In the regression, the signal responses are the dependent variables and the crack lengths are the independent variables.

3.4.2.1 Analysis of Signals From Noncracked Rivets.

Before analyzing the crack data, we analyzed the inspection signals from the rivet sites with no flaws. The effect of frequency can be seen in figure 16, with the lower inspection frequency producing larger signals. In figure 17 the signal movement in the x direction is shown by rivet for two of the test panels. It is apparent from the patterns that there is an individual rivet effect. In the second instance (panel 16), the trend for the 20-kHz inspection data is consistent with signal losses due to noncentered probe on the rivets.

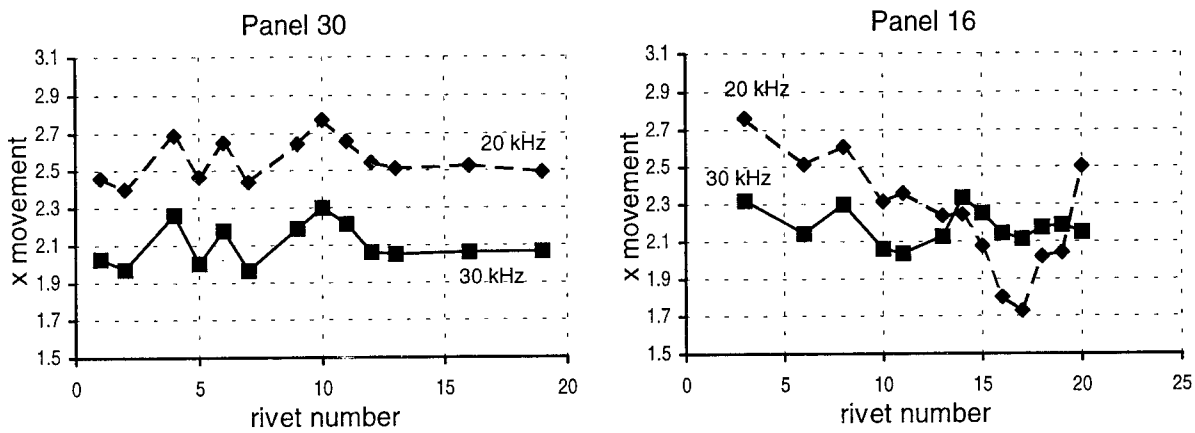

FIGURE 17. NORTEC 19e—NO FLAW SIGNALS BY RIVET WITHIN PANEL

In addition to the rivet effect, there was also a panel effect. That is, for a given inspection the signals exhibited panel-to-panel variations beyond that accounted for by rivet-to-rivet inspection. This is seen in figure 18, where the maximum movement in the x direction of the signal (see figure 15) for each nonflawed rivet site is graphed.

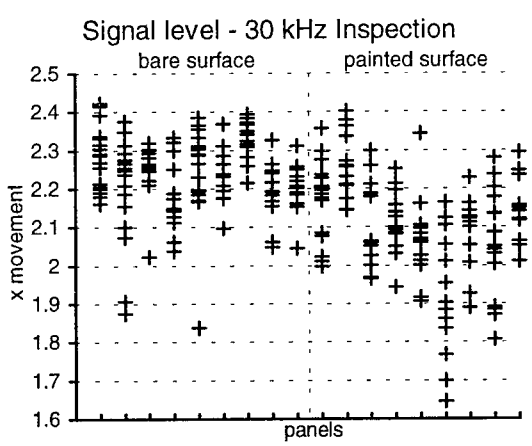

FIGURE 18. NORTEC 19e—NO FLAW SIGNALS BY PANEL

Table 9 contains the estimates of variance from a one-way analysis of variance of the maximum x movement in the signals from the nonflawed sites. Also given is the total variation estimated from each full set of data. This total variation is then used to estimate the response levels that would result in false call rates of 0.05, 0.02, 0.01, and 0.005. For example, for a response that has a Gaussian distribution, approximately 5 percent of the values are more than 1.645 standard deviations from the mean response. This multiple of the total observed standard deviation is added to the mean response level to estimate the response that would lead to a 0.05 false call rate. A similar process, with appropriate multipliers, is followed to determine the other false call levels. The data from 20-kHz inspection of panel 16 (see figure 17) was not included in the estimates of table 9. It was judged to be indicative of an atypical situation, thought to be a straight edge misalignment that resulted in the probe not being centered over all of the rivets.

TABLE 9. NORTEC 19e—NO FLAW VARIATIONS AND FALSE CALL LEVELS

Inspection frequency	20 kHz		30 kHz	
Surface	Bare	Painted	Bare	Painted
Standard Deviations-				
within panel	0.091	0.121	0.078	0.096
between panel	0.100	0.127	0.042	0.073
total observed	0.131	0.171	0.088	0.118
Mean Response Level	2.668	2.616	2.248	2.112
False Call Levels				
0.05	2.883	2.897	2.393	2.307
0.02	2.937	2.966	2.428	2.355
0.01	2.972	3.013	2.452	2.388
0.005	3.005	3.056	2.474	2.417

Fully specified procedures for an inspection would include setting a response level for making a call. The call level is often specified in terms of the signals obtained from the no crack signal and the setup crack signal. A commonly applied threshold for making calls is to use the level half-way between a nonflawed signal and the signal from a flaw of known size. From the signals of figure 15, for the 20-kHz inspection the response level halfway between the nonflawed signal and the signal from the 0.100-inch setup would be 2.943. For the 30-kHz inspection the halfway response would be 2.787. From table 9 we see that applying this halfway criteria for making a call would result in a false call rate of approximately 0.02 for the 20-kHz inspection, but less than 0.005 for the 30-kHz inspection. Of course, decision criteria other than setting a level halfway between the no crack and the crack signals could be applied.

3.4.2.2 Analysis ($\hat{a}$ Versus a) of Signals From Cracked Rivets.

In the following analysis we use the natural logarithm of the x excursion as the response variable. The maximum crack length (when two cracks are present) is used as the independent variable rather than its logarithm because the relationship with the signal is more linear (see figure 16). The one-crack versus two-crack effect was significant when the total crack length was used in the analysis, whereas the effect was not significant at level p=0.05 when the

maximum crack length was used. For this reason the maximum crack length is a better explanatory factor than is the total length.

The data for both inspections were analyzed in a general linear model. The effects modeled were the angle of the crack from horizontal (3 levels) and the surface condition (2 levels) with the maximum crack length included as a covariate. In the analysis presented here only the nonsaturated signals are included.

Crack angle was initially included in the model as a three-level factor (horizontal, 11 degrees, and 22 degrees). The 11-degree and 22-degree off-angle cracks had similar effects that were significantly different from the horizontal cracks. Therefore these two were combined and the model was fit to a two-level angle factor (horizontal and off-angle). The interaction of angle and surface condition was included in the model and was significant. However, the test specimens with bare surfaces did not contain any of the 11-degree cracks and the off angle cracks within the painted surfaces were mostly 11-degree cracks with a few 22-degree cracks. It is possible that the effect due to the 11-degree off-angle cracks is closer to the horizontal effect than to the 22-degree off-angle effect and that the original indication to the contrary was due to partial confounding of the surface condition with the angle affect.

Figures 19 through 22 show the estimated PoD curves when the thresholds are set at false call rate levels determined from the nonflawed signals. As discussed earlier, the off-angle category for the bare surface reflects an influence of 22-degree cracks, whereas it is 11-degree off-angle cracks that are reflected in the painted category.

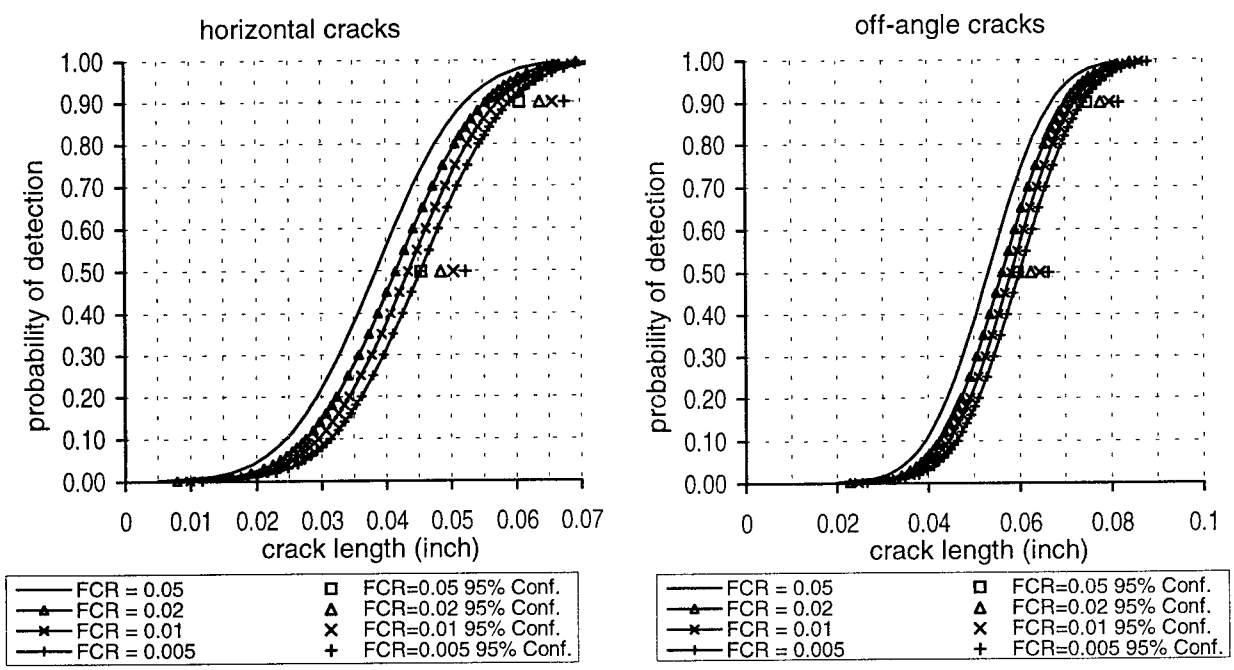

FIGURE 19. PoD ESTIMATES FOR NORTEC 19e SLIDING PROBE 20-kHz INSPECTIONS ON BARE SURFACE

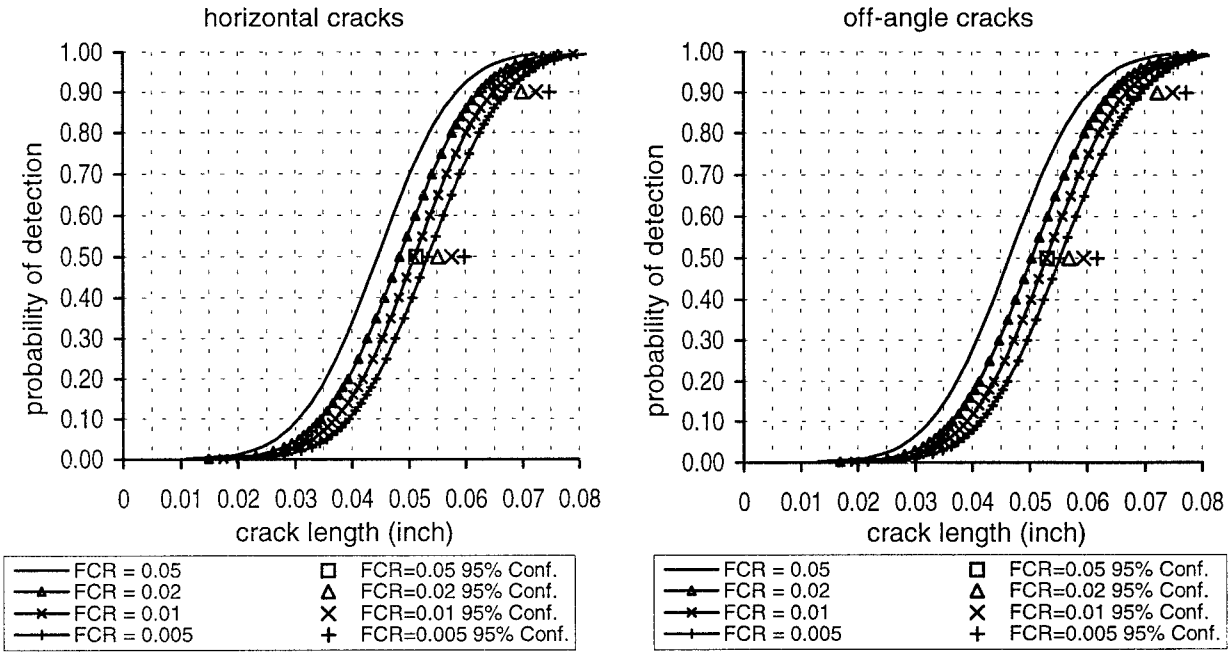

FIGURE 20. PoD ESTIMATES FOR NORTEC 19e SLIDING PROBE 20-kHz INSPECTIONS ON PAINTED SURFACE

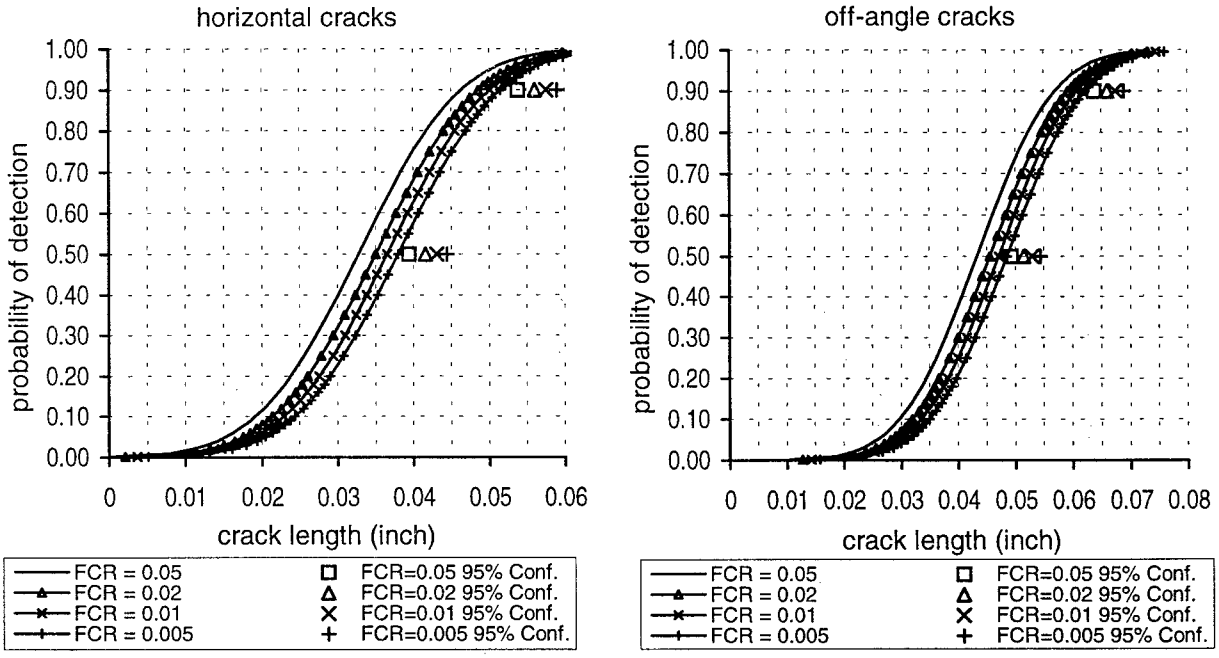

FIGURE 21. PoD ESTIMATES FOR NORTEC 19e SLIDING PROBE 30-kHz INSPECTIONS ON BARE SURFACE

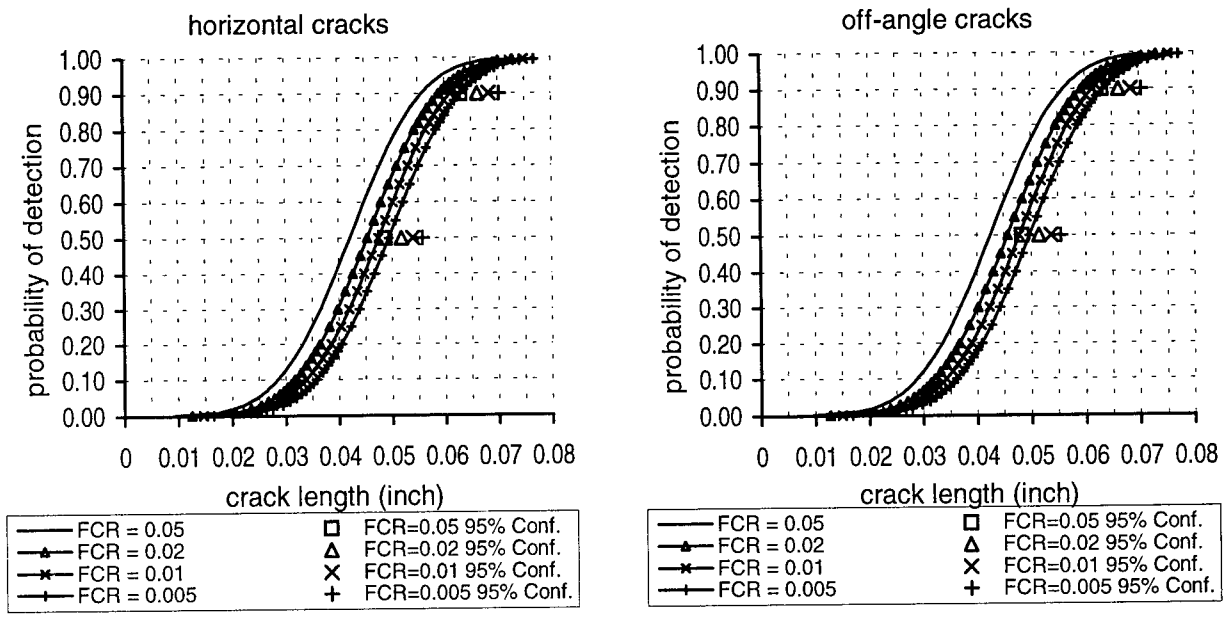

FIGURE 22. PoD ESTIMATES FOR NORTEC 19e SLIDING PROBE 30-kHz INSPECTIONS ON PAINTED SURFACE

A clearer indication of the relationship between false call rates and detection rates can be determined by constructing a relative operating characteristic (ROC) curve showing the change of probability of detection versus the false call rate [8,9]. These curves are constructed by setting a signal level (determined by crack length) and then varying the threshold for making a call. The false call rate for a given threshold is modeled by the normal distribution as discussed earlier. To model the probability of detection it is assumed that the signal, *ln*(maximum x movement) is distributed as a normal random variable with mean equal to $\beta_0 + \beta_1 \cdot cracklength$ and with variance, δ^2. The ROC curves for two of the bare surface inspections are shown in figure 23 for five signal levels determined by cracks in 0.01 inch increments.

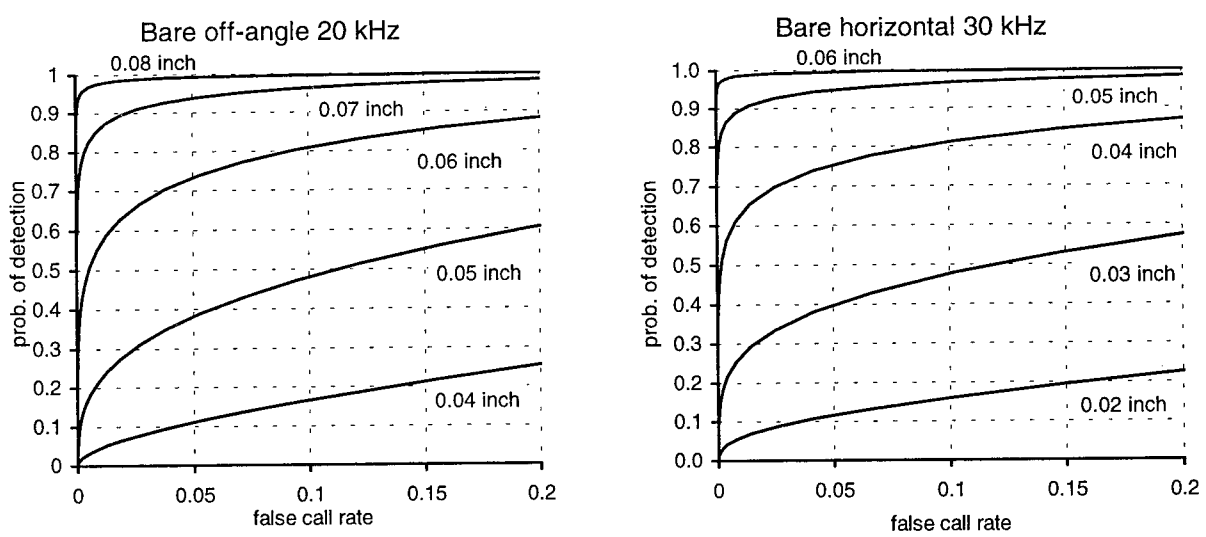

FIGURE 23. NORTEC 19e ROC CURVES

Up to this point only a single characteristic, the horizontal movement of the signal, has been used as a response to establish reliability estimates. Could additional signal characteristics, such as height or vertical movement, be included to increase the reliability of the inspection? To answer this question we look at using more of the signal to predict the crack length. This is done through an $\hat{a}$ versus a analysis that includes an inverse regression step. This technique is new and is explained in more detail in the following section.

3.4.2.3 Use of Inverse Regression in Analysis of Signals.

In the $\hat{a}$ versus a analysis the crack length, a, is assumed to be known without error and the signal response, $\hat{a}$, is measured with error. Here we consider the straight-line regression expressed as $y = \beta_0 + \beta_1 \cdot x + \varepsilon$, where ε has normal distribution with mean 0 and variance δ^2. The same type of model has been considered in the calibration literature where the aim is to provide a prediction of the underlying x value after calibrating the curve by estimating β_0 and β_1. The suggested prediction is $(y - \beta_0)/\beta_1$. In the statistical literature, this is referred to as inverse regression. (See reference 10 and citations given there.) The above quantity (with y = T) is the mean of the normal distribution curve that estimates the probability of detection.

Since the signal, denoted y, is meant to predict the unknown quantity x, we can regress the known x's against y and derive a calibration line that transforms y into an estimate of the x's. That is, fit the equation $x = \gamma_0 + \gamma_1 \cdot y$. There are a number of citations in reference 10 for articles addressing the conditions for which the predictive equation with γ's does better than the predictive equation with the β's. We do not pursue this aspect of the subject, but rather we consider how the regression of x onto y can be used to derive the probability of detection curve.

The proposed analysis consists of three major steps as follows.

Step 1. Let crack length (or suitable function of flaw characteristic) be the dependent variable and be denoted by x. Let the signal characteristic be denoted by y. Let $\hat{\gamma}_0$ and $\hat{\gamma}_1$ be the maximum likelihood (least squares) estimates of the coefficients γ_0 and γ_1 from the linear equation $x = \gamma_0 + \gamma_1 \cdot y$.

Step 2. Using the fitted equation of step 1, derive the fits given by $y' = \hat{\gamma}_0 + \hat{\gamma}_1 \cdot y$. The y' values can be thought of as the predicted crack length (or any other flaw characteristic, such as depth, that is being used as the explanatory variable) corresponding to the signal y.

Step 3. Use the y' values of step 2 in an $\hat{a}$ versus a analysis. That is, the linear equation $y' = \beta_0 + \beta_1 \cdot x$ is fit and the usual PoD curve estimated from the parameters of the fit. Let $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\delta}$ be the maximum likelihood estimates of β_0, β_1, and δ, the standard deviation of the residual error. For a threshold of T on y use the corresponding threshold of $\hat{\gamma}_0 + \hat{\gamma}_1 \cdot T$ for y'. The estimate of the probability of detection curve is then given by the cumulative normal distribution function with mean $\hat{\mu} = (\hat{\gamma}_0 + \hat{\gamma}_1 \cdot T - \hat{\beta}) / \hat{\beta}$ and standard deviation $\hat{\sigma} = \hat{\delta} / \hat{\beta}_1$.

The PoD function that is derived from these three steps is the same as that derived from fitting the equation $y = \beta_0 + \beta_1 \cdot x$ and following the usual $\hat{a}$ versus a analysis. (See Appendix A for proof.) The value of the above formulation is that it can readily be extended to multidimensional signal data by replacing the linear equation of step 1 with a more general linear regression equation $x = \gamma_0 + \Sigma\gamma_i \cdot y_i$, where the y_i's can be different aspects of a signal.

We will illustrate the above analysis with the Nortec 19e data. Additional aspects of the signal shape other than just the excursion in a single dimension are included in the regression to determine if probability of detection can be increased.

The typical signal shape for inspecting a rivet site can be seen in figure 15. The impedance plane trace rises to the left, then falls while continuing to the left, and then nearly retraces the same path as it returns to the null position. We characterize this signal with 6 variables. Considering the null position as (0,0) we denote the point of maximum x excursion as (x_0, y_0), the initial high point in the curve as (x_1, y_1), and the return high point as (x_2, y_2). We characterize the points in the sense of distances so that the x values are positive. In the previous analysis we used x_0 as $\hat{a}$ and maximum crack length, maxlen, as a. We consider two extensions of this analysis. The first extension will be to consider the regression of maxlen on $\{x_0, y_0, x_1, y_1, x_2, y_2\}$, and the second extension will include the above six variables as well as their interactions. In both cases factors that are significant (p=0.05) are retained in the model. In the interaction case the first order terms that are part of the significant interactions are also retained in the model.

Table 10 shows the fits that were obtained. The last two columns give the mean and standard deviation of the cumulative normal distribution that estimates the probability of detection curve when a predictive threshold of 0.050 inch is used. As more of the signal is used to predict crack length, the PoD curves move to the right (increasing estimate of μ) and become less variable (decreasing estimate of σ) but not by large amounts.

TABLE 10. NORTEC 19e PoDs FROM PREDICTIVE FITS

Case	Freq	Predictive Fits (n=60 for 20 kHz, n=68 for 30 kHz)	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\delta}$	$\hat{\mu}$ (T'=0.05)	$\hat{\sigma}$
1.	20	maxlen = -0.0743 + 0.0422*x_0	0.0108	0.8292	0.0106	0.0473	0.0128
2.	20	maxlen = -0.1231 + 0.0490*x_0 + 0.0211*y_0	0.0089	0.8591	0.0098	0.0478	0.0114
3.	20	maxlen = -0.3852 + 0.1089*x_0 + 0.0212*y_0 + 0.1796*y_1 + 0.1592*x_2 + 0.0140*y_2 + 0.0312*x_0*x_2 + -0.0517*y_1*y_2 + -0.0781*x_2*y_2	0.0075	0.8825	0.0091	0.0482	0.0103
4.	30	maxlen = -0.0691 + 0.0096*I{p} + -0.0460*x_0	0.0092	0.8573	0.0103	0.0476	0.0120
5.	30	maxlen = -0.0831 + 0.0099*I{p} + -0.0470*x_0 + 0.0143*y_0	0.0070	0.8908	0.0092	0.0483	0.0103
6.	30	maxlen = -0.2685 + 0.0072*I{p} + 0.0533*x_0 + -0.0352*y_0 + 0.0671*x_1 + 0.1993*y_1 + 0.0043*x_2 + 0.1481*y_2 + -0.0606*x_0*x_1 + -0.0694*x_0*y_1 + 0.0260*x_0*x_2 + 0.0287*y_0*x_1 + -0.0427*y_0*x_2 + 0.0163*y_0*y_2 + -0.1111*x_1*y_2 + 0.0466*y_1*x_2 + -0.1427* y_1*y_2	0.0034	0.9476	0.0066	0.0492	0.0070

I{p} in the fits for cases 4-6 is an indicator variable for the inspection over paint

The lack of major shifts in the estimated PoD curves in using more of the inspection signal indicates that one dimension of the signal is adequate. However, detection is not the only factor in an inspection. The inspection should also not be falsely identifying crack-free sites as having cracks. This was the reason for considering the trade-offs between false call rates and detection rates reflected in the modeling of figure 23. We therefore also look at the predictions from the nonflawed sites that result from using the predictive equations given in table 10. In figure 24 the predictions at the nonflawed sites are presented in probability plots. Graphed on the y-axis is the standard normal z score associated with the empirical cumulative distribution of the predictions, which are graphed on the x-axis. In this format a population that follows a Gaussian distribution would plot as a straight line.

In both the 20-kHz inspections and the 30-kHz inspections the multidimensional use of the signal to make predictions resulted in lower average crack length predictions at the noncracked rivets. However, on the 20-kHz signals there was very little change for the upper portion of the population. Thus, there would be no net change of the prediction level when setting the false call rate at levels no greater than 0.05 (that is, z values greater than 1.64). For the 30-kHz inspections, the predictive equations that were fit to the flawed data also performed better on the unflawed data (that is, decrease the estimated crack length). However, in this case there is more separation in the curves at the higher z values. The practical consequence of this observation is that by using multidimensional signal data in the manner prescribed, PoD can be increased, without increasing false alarms, by slightly lowering the predictive level (~0.003 inch) that is used to make calls.

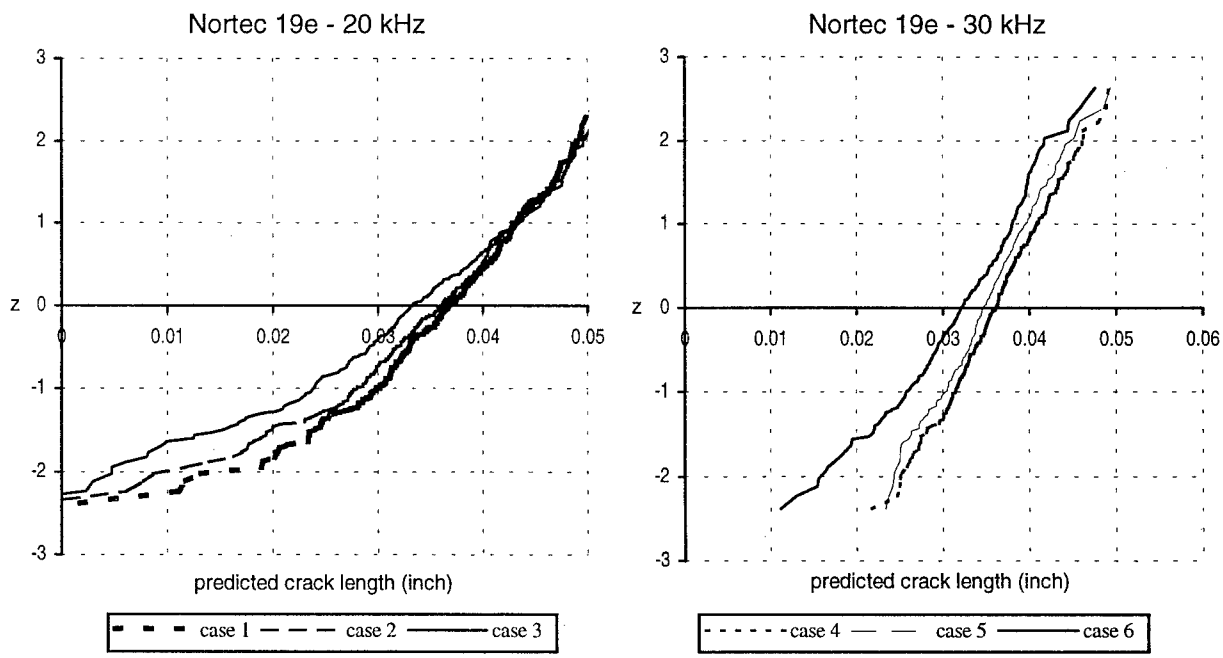

FIGURE 24. NORTEC 19e PREDICTED CRACK LENGTH DISTRIBUTION AT NONFLAWED RIVETS

The analysis presented here for incorporating more of the signal than a single dimension was meant to be illustrative of a new proposed method of analysis. Some gains in probability of

detection were achieved but they were modest gains. However, specific functions of the signal other than those considered here might produce better PoD results. This may be the case if appropriate functions of the signal were identified through physical modeling of the signal.

3.5 PENCIL PROBE WITH ZETEC MIZ-22.

3.5.1 Zetec MIZ-22 Background.

An alternative to the sliding probe procedure given in section 3.4 is a template procedure. In the template procedure a pencil probe is guided around the edge of each rivet. This procedure was followed using a Zetec MIZ-22 instrument. An NDI technician performed the inspections over a period of approximately one month. The same test specimens used in the sliding probe study were used. The signals were captured by an analog to digital converter and stored as (x,y) points.

The Zetec MIZ-22 was set at a frequency of 500 kHz and used with a differential 50-500 hertz angled pencil probe. The phase was set at 160 degrees and the gain at 24 decibels with the vertical to horizontal ratio set at 10/10. The inspection was set up so that a 0.100-inch notch produced a vertical signal of approximately 50 percent of full scale. The template aperture diameter was $^{7}/_{16}$ inch. The signal taken from the calibration standard is shown in figure 25. The first graph is of the signal that would be displayed on the instrument panel. The second shows the x and y channels individually as a function of time. From the second graph we know that the notch was on the right side of the rivet, as the inspector started all inspections from the top of the rivet and then circled the rivet clockwise with the probe.

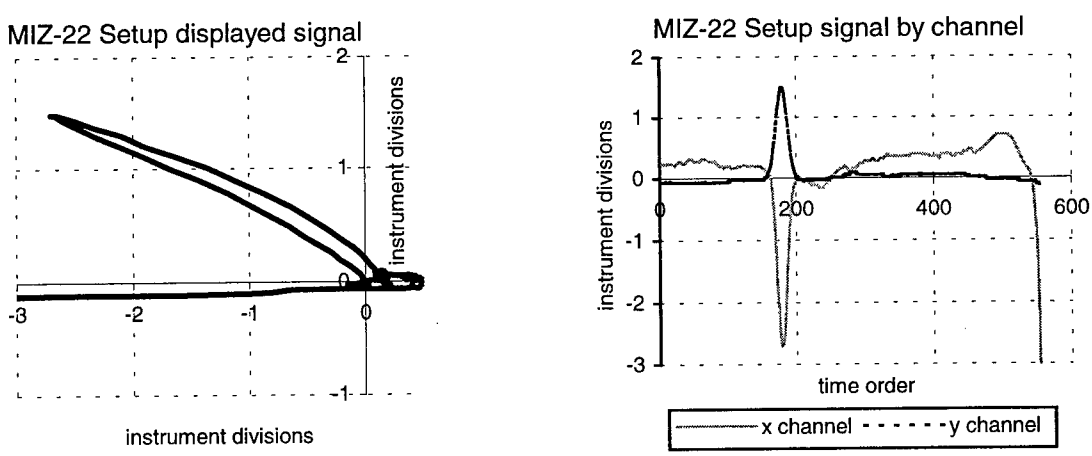

FIGURE 25. ZETEC MIZ-22 SIGNALS

3.5.2 Zetec MIZ-22 Analysis.

The cracks in the test panels emanate from either the right side or the left side of each rivet. The pencil probe inspection produces a signal that can be correlated with the points around the rivet. We therefore consider each side of the rivet as an inspection. The 360 rivet sites inspected thus yielded 720 inspection points of which 145 contained cracks.

The signals from unflawed rivets do not have the structure noted in figure 25. They are more like random walks around the null point. The x and y values from signals produced at cracks are correlated. We looked at predictions using the y portion of the signal, the x portion of the signal, and the total distance that the tip of the signal was from the null point. All produced about the same results in the analysis on the signals produced by cracks, but there were substantial differences in the number of false calls associated with the same detection rates. There were substantially fewer false calls in the use of the y-value than with the x-value or the total distance. This is expected, as the x direction is the direction of lift off. The y variable is graphed versus the logarithm of the crack length in figure 26. The signals from the sides of the rivets at which there were no flaws are graphed at -4.6.

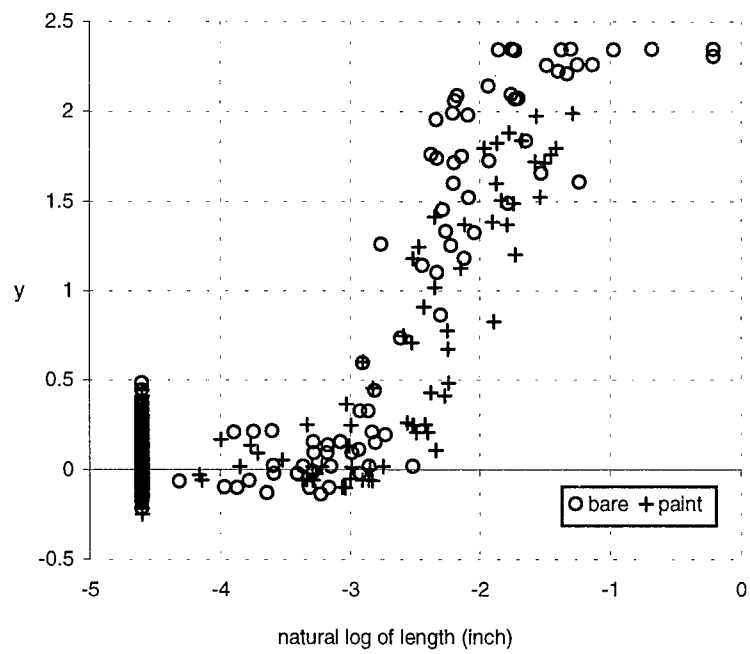

FIGURE 26. SIGNAL DATA FROM ZETEC MIZ-22 PENCIL PROBE INSPECTION

The nine signals graphed across the top of figure 26 are saturated signals. That is, the recorded signal strength was limited by the instrumentation. These values are properly analyzed by considering them as censored values (that is, known only to be at least as large as the recorded value). However, there was very little difference in the fits from deleting these data versus treating them as censored data. The results presented here are with these values removed. The data look consistent with a linear model given by,

$$y = c + \varepsilon, \qquad a < a_{\text{det}}$$
$$y = \beta_0 + \beta_1 \cdot \ln(a) + \varepsilon, \qquad a > a_{\text{det}}$$

where the error, ε, is a random variable with zero mean and standard deviation δ. The a_{det} is that crack length for which $c = \beta_0 + \beta_1 \cdot \ln(a)$, or where the signal, y, starts to depend on the crack length. The parameter estimates are given in table 11, where the model was fit to the painted and bare panels separately and the corresponding PoD curves are shown in figure 27.

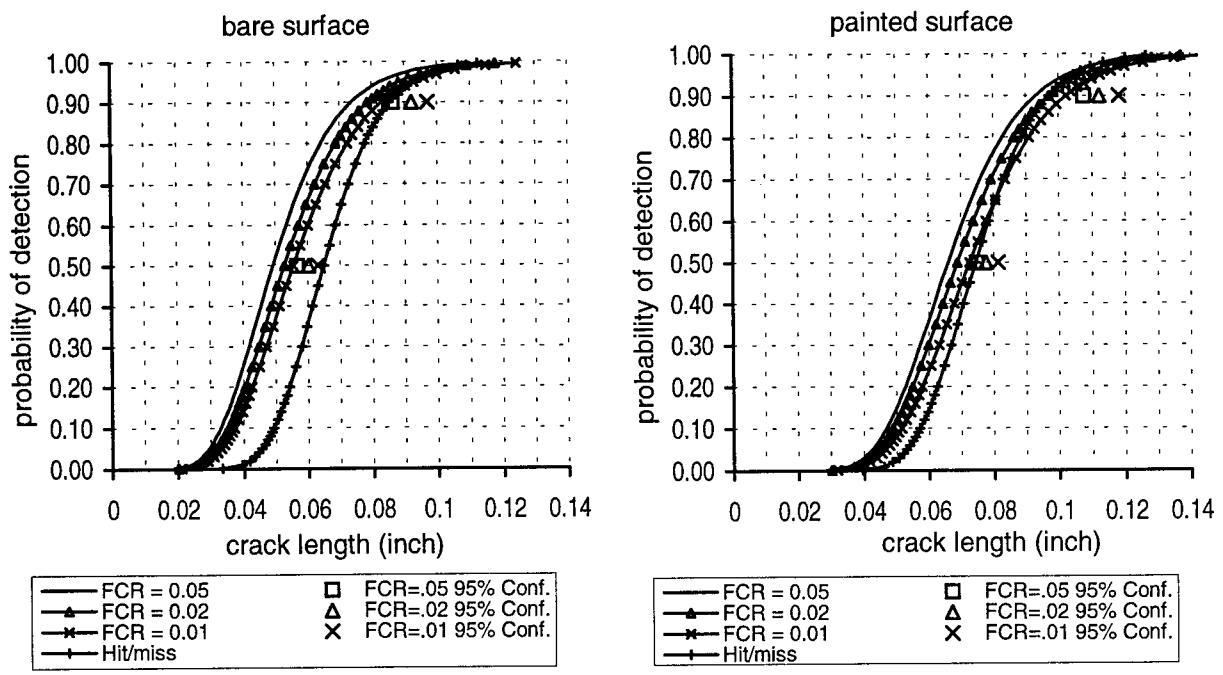

FIGURE 27. MIZ-22 PENCIL PROBE ESTIMATED PoD CURVES

The false call rates given in table 11 are determined empirically from the signals resulting from the inspections of the unflawed sides. For example, in approximately 5 percent of the signals on the bare sites without cracks, y was greater than or equal to 0.294.

TABLE 11. ZETEC MIZ-22 PENCIL PROBE PoD PARAMETERS

	Bare	Painted
c (mean when $a < a_{det}$)	0.0189	0.04614
intercept, β_0	4.0884	3.7170
ln(crack length), β_1	1.2604	1.2631
scale, δ	0.3500	0.2883
a_{det}	0.0396	0.0547
False call rate=0.05 threshold	0.294	0.284
a_{50} (inch)	0.049	0.066
a_{90} (inch)	0.074	0.093
False call rate=0.02 threshold	0.379	0.338
a_{50} (inch)	0.053	0.069
a_{90} (inch)	0.079	0.097
False call rate=0.01 threshold	0.445	0.404
a_{50} (inch)	0.056	0.073
a_{90} (inch)	0.083	0.102
Hit/miss analysis (no false calls)		
a_{50} (inch)	0.065	0.074
a_{90} (inch)	0.084	0.096

The inspector made calls as he performed the inspection. The fits, using binary logistic regression, to the hit/miss data are also given in table 11. There were no false calls among the detections. The PoD curves as determined from the hit/miss analysis are consistent with those derived from the signal regression analysis when a high signal threshold (low false call rate) is used.

3.6 LOW-FREQUENCY EDDY-CURRENT ARRAY (LFECA).

3.6.1 LFECA Inspection Background.

The Low-Frequency Eddy-Current Array (LFECA) was developed at Northrop for the improved detection of cracks under fasteners on the vertical stabilizers of the F/A-18 [11]. A segmented 16 coil outer receiver array is used in conjunction with a center driver coil. A signal obtained for an unflawed rivet site is used as a reference. Inspection signals are then compared to this reference signal and the differences are displayed in a graphical waveform.

Two different inspectors used the LFECA system to inspect the lap splice test specimens at AANC. The inspections were done at different times and separated by several months. The second inspection resulted because a review of the setup and conditions employed in the first inspection were believed to be less than adequate to demonstrate the instrument capabilities.

3.6.2 LFECA Analysis.

The displayed signals are waveforms constructed from the 16 outer array receivers. From the waveforms it is possible to determine from which area of a rivet that a crack is emanating. For this reason the detections are analyzed by each individual crack rather than by rivet site.

Although the LFECA system has the capability to store signals, the analysis presented here is based solely on the inspectors' calls at the time of inspection. In both inspections the inspectors were asked to categorize each inspection into four categories. Those categories are 0, no flaw present, 1, a slight indication of an anomaly, 2, moderately high indication, and 3, sure call for a flaw. Figure 28 shows PoD curves from the logistic regression fits using the probit link function. The fits are shown for each of the inspectors and by considering detection with respect to relaxing the call criteria to include various levels of calls discussed above.

The estimated PoD curves in figure 28 are based on the detections of the individual cracks in the test specimens. Therefore, a miss of a crack on one side of a rivet affected the fit even though a second crack at the same rivet site was detected. For inspector 2, the three largest cracks missed were of lengths 0.057, 0.055, and 0.49 inch. All occurred at rivet sites where a second crack was present (0.210, 0.193, 0.154 inch, respectively). Upon review of the signals obtained at each of the inspection sites, there is a strong indication that second, smaller cracks are masked when a much larger crack is also present at the rivet.

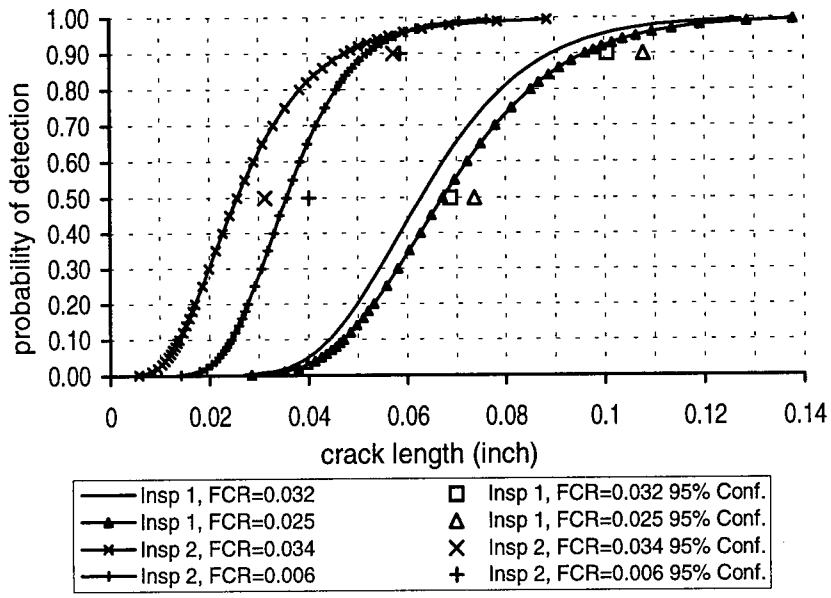

FIGURE 28. LFECA ESTIMATED PoD CURVES—INSPECTIONS BY CRACK

To ascertain the impact of the individual crack misses in the presence of the larger cracks, we also fit the inspection results of inspector 2 to rivet site data. That is, each rivet site was considered as a hit or miss and was characterized by the length of the largest crack present. Those PoD curves are shown in figure 29. Of particular interest is the sharpness of the PoD curve for the criteria defined by the sure call level 3, which resulted in a false call rate of 0.006. In fact, the inspector had stated that through his experience he felt confident that he could find cracks as small as 0.040 inch. This is consistent with the test results.

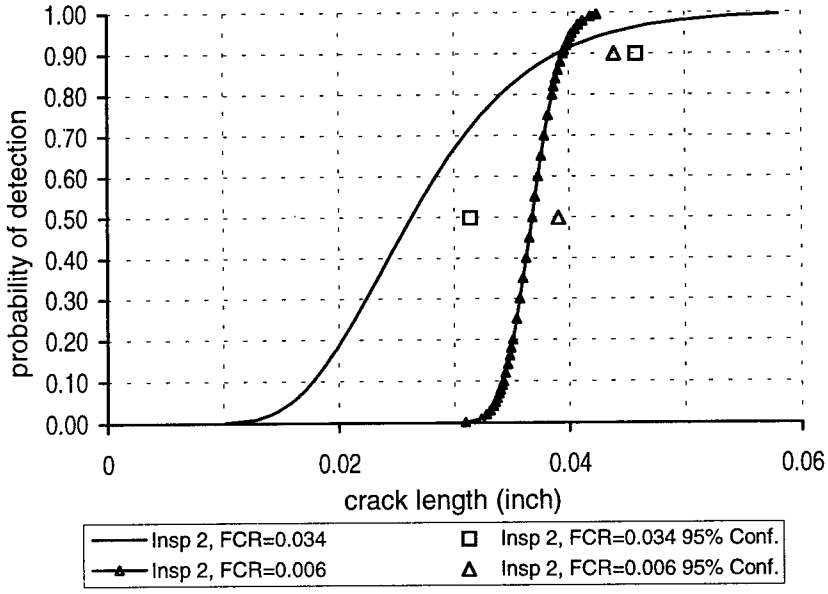

FIGURE 29. LFECA ESTIMATED PoD CURVES—INSPECTIONS BY RIVET FOR INSPECTOR 2

From the comparisons of the curves in figure 28 we see that the suspicion of a less than optimal setup in the original inspection was justified. The second inspection by a different inspector was substantially better. This illustrates the point that reliability is very much a function of the complete inspection process, of which the equipment is only the starting point.

3.7 NASA SELF-NULLING ROTATING PROBE.

A NASA Langley Research Center group of developers visited AANC twice during development work of a new eddy-current probe. The probe has been reported in the literature as the self-nulling electromagnetic flaw detector [12]. It is sometimes referred to as the Simpson probe in deference to its inventor. Taken from reference 12, a short description of the working principle follows.

> "The Self-Nulling Electromagnetic Flaw Detector induces a high density eddy-current ring in the sample under test. A ferromagnetic flux focusing lens is incorporated such that in the absence of any inhomogeneities in the material under test only a minimal magnetic field will reach the interior of the probe. A magnetometer (pickup coil) located in the center of the probe therefore registers a null voltage in the absence of material defects. When a fatigue crack or other discontinuity is present in the test article the path of the eddy-currents in the material is changed. The magnetic field associated with these eddy currents then enter into the interior of the probe, producing a large output voltage across the pickup coil leads."

Details of the probe design can be found in published literature [12,13]. Here we review the major aspects of the design. The instrument has a 1/8-inch-diameter pickup coil, and the outside diameter of the entire probe is approximately 0.25 inch. The operating frequency of the probe was 125 kHz. The motor driving the probe was a 180-rpm AC synchronous motor.

There were noticeable improvements in the instrument between the first and second inspections. The major hardware changes were:

a. A switch to a synchronous motor providing a more stable rotation frequency.
b. A change in the rotation frequency from 80 to 180 revolutions per minute.
c. A reduction in the weight of the probe head from 3 to just under 2 pounds.

The software was also modified; the main change was the addition of a display of a centering vector and the modification of the filter used to isolate the fatigue crack signal from this centering vector. During the inspections of the first visit, the inspectors made subjective decisions looking at the displayed signal as to whether the probe was properly centered. On the second visit, data were not acquired until the probe was adequately centered.

3.7.1 NASA Self-Nulling Probe Inspection Background.

A total of 18 of the AANC panels with surface cracks were used. The panels were mounted on frames to simulate a lap splice. All were mounted on the upper row and therefore all inspections were done with the inspector operating the probe from a standing position with the probe at or

slightly lower than eye level. The panels were in a random pattern. Nine of the panels were painted and nine had the paint stripped from the lap area.

While one inspector operated the probe, a second operator sat at a computer console and captured the signal data and entered the level of the call in a data book. (The nature of the signals will be discussed later.) Both had access to the computer presentation, as the probe operator used a heads-up display as he manipulated the probe. The experimenters switched tasks between themselves during the inspection.

Figures 30–32 show the basic signals available to the operators from a nominally small crack. Figure 30 shows the raw signal in polar coordinates. In the absence of a crack it is expected that this signal would be close to circular. (Note: Figure 30 was reproduced from recorded data but the original aspect ratio is not necessarily maintained.) Figure 31 shows a processed signal where deviations from circular are shown as a function of angular position. (The zero angle is straight up with respect to the rotating probe head and positive angles are in the clockwise direction.) Figure 32 shows the same data as that of the processed signal in figure 31 but in polar coordinates. The direction of the crack (left side) is readily apparent from figure 32. The smaller lobes in quadrants 1 and 4 are the negative portions of the signal from the crack on the left.

In addition to the signals shown, the inspector would also have the maximum height from the signal of figure 31 displayed. For example, a 9.7 would be displayed for the signal of figure 31. The ultimate decision concerning the presence of a flaw would be based on this maximum voltage display.

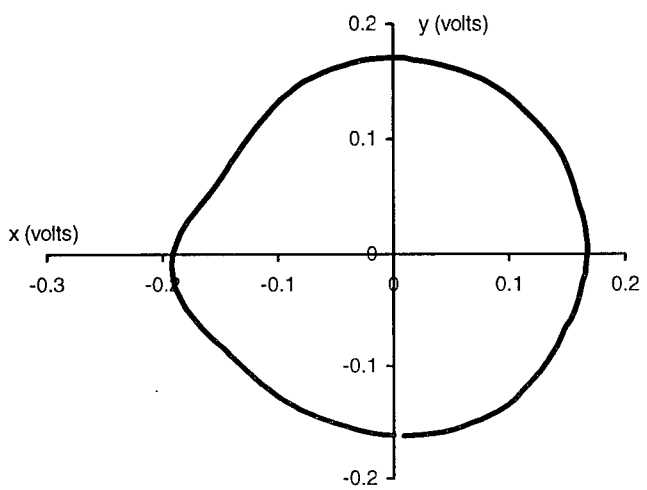

FIGURE 30. NASA SELF-NULLING ROTATING PROBE RAW SIGNAL

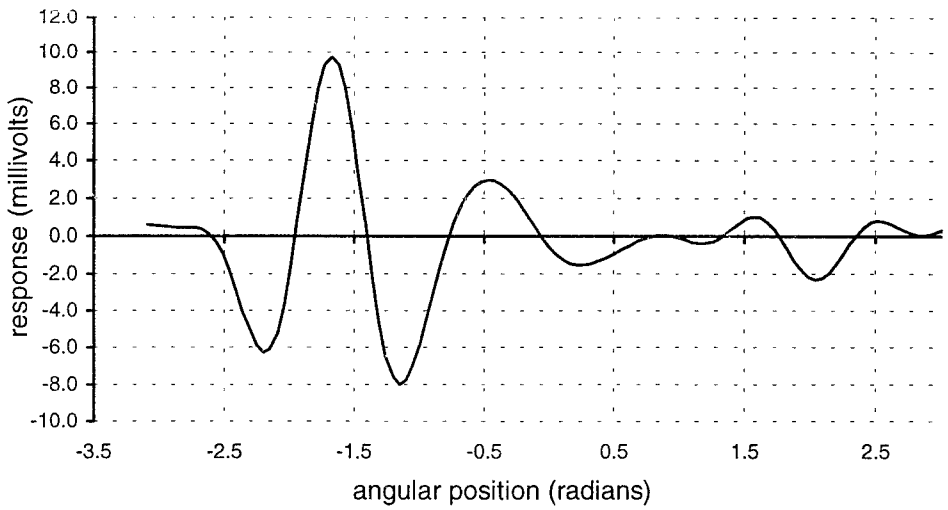

FIGURE 31. NASA SELF-NULLING ROTATING PROBE PROCESSED SIGNAL

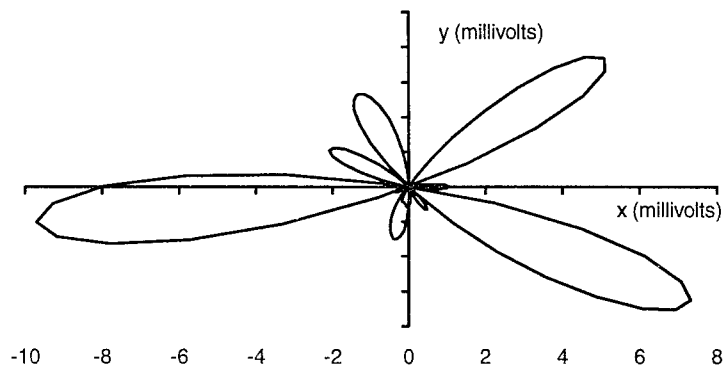

FIGURE 32. NASA SELF-NULLING ROTATING PROBE PROCESSED SIGNAL IN POLAR COORDINATES

The signals are updated continuously until one of the operators signals the program to capture the data, at which time a snapshot of one revolution is recorded. Thus, while the operator is getting the probe aligned over the rivet there is continuous movement with these curves. The alignment of the probe over the rivet and the decision as to when proper alignment had been achieved has a direct effect on the outcome. The method of making that decision was a major factor in the changes that the NASA researchers instituted between the two visits. In the first visit the operator was essentially making the decision concerning proper alignment from the signal of figure 30. That is, he would manipulate the probe until he had as close to a circular signal as he could obtain. The determination of circularity was subjective. By the second visit, the NASA investigators had added an analysis of the raw signal for circularity. The result of the analysis was a vector from the center of the figure with a crosshair at the end of the vector. The magnitude of the vector reflected how far off center and the direction that the probe was off. Therefore the inspector now had a better defined criteria (i.e., center or nearly center the crosshair) for assessing the adequacy of the probe alignment over the rivet head.

3.7.2 NASA Self-Nulling Probe Data Analysis.

The analysis and results presented here are for data gathered during the second visit. The data are analyzed by considering the relationship of the maximum deviation amplitude (for example, the 9.7 from figure 31) with the known crack lengths in an $\hat{a}$ versus a analysis. A summary table of results is given in table 12 and shown graphically in figure 33. The response range in table 12 is given in units of millivolts, whereas the graph of figure 33 is the logarithm of the response versus the logarithm of the crack size. Included in table 12 is the range of the response for each group of rivets examined. The crack lengths are in terms of the largest crack at each rivet site. (Note that the signals from the self-nulling probe would indicate the presence of two cracks but only an overall maximum from the processed signal would be given.)

TABLE 12. NASA SELF-NULLING ROTATING PROBE—SUMMARY OF RESPONSES

Flaw Length Range (inch)	Number in Data Set	Range of Response (millivolts)
> 0.100	44	17 to 97
0.090 to 0.100	6	30 to 65
0.080 to 0.090	11	21 to 60
0.070 to 0.080	2	22 to 50
0.060 to 0.070	3	13 to 50
0.050 to 0.060	13	11 to 46
0.040 to 0.050	10	9.4 to 27
0.030 to 0.040	12	1.3 to 20
< 0.03	8	1.0 to 9.2
no flaws	271	0.4 to 3.7

From table 12 it is seen that the largest signal from a nonflawed rivet was 3.7 (or $1.31 = ln(3.7)$, in the log response scale of figure 33). This implies that setting thresholds for making a call in the 3 to 4 millivolt range should control false call rates. It is apparent from the general pattern of figure 33 that a linear relationship does not hold up over the complete range of crack sizes. However, the larger crack data of figure 33 show responses well above 3 to 4 millivolts. Therefore, although the mean response-crack length relationship could be modeled by some nonlinear function, the upper portion of that mean curve would have very little effect on the estimation of the increasing portion of the PoD curve. Because of this, only the first part of the data is used (cracks less than 0.09 inch) and we restrict attention to linear models. The linear fit is shown in figure 33, where the line is given only for the range used in its estimation. The equation of the best fit line is $ln(r) = 9.000 + 2.109 * ln(a)$, where r is the response in millivolts and a is crack length in inches. The residual standard error estimate is $\hat{\delta} = 0.522$. The probability of detection curves resulting from these fits are shown in figure 34. There was not a significant surface effect (paint or bare).

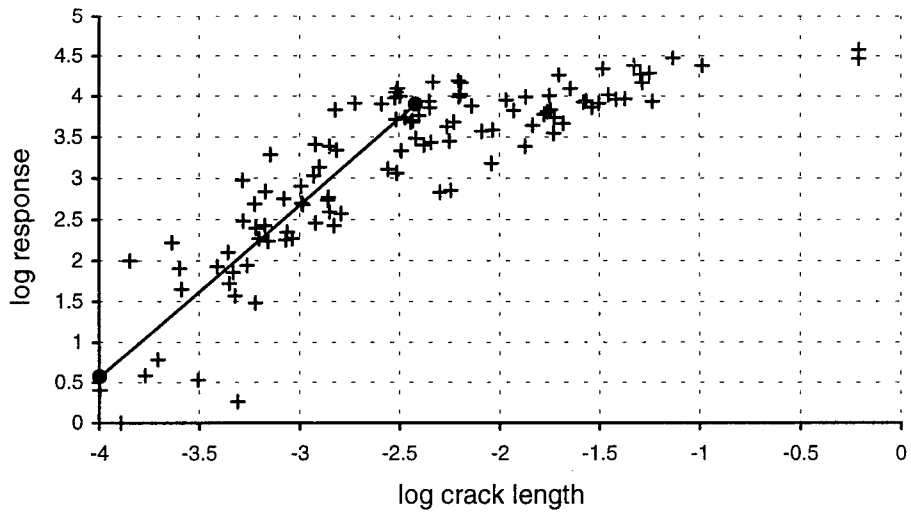

FIGURE 33. NASA SELF-NULLING PROBE RESPONSE VERSUS CRACK LENGTH—SMALL PANELS

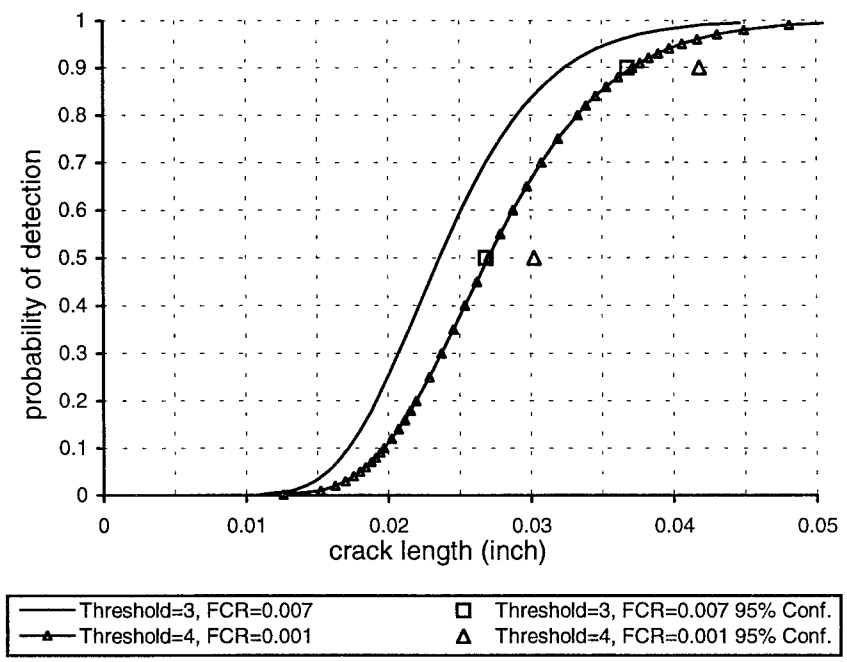

FIGURE 34. ESTIMATED PROBABILITY OF DETECTION CURVES—NASA ROTATING PROBE

During the inspection the operators noted that if they were making calls independent of perceived crack length, that they would use a threshold of 3 to 4 millivolts (that is, 1.099 to 1.386 in the log scale). Using a threshold of 3 with the regression estimates gives the probability of detection of a crack of length a , PoD(a)= Probability{ln(r) >1.099)}. Since ln(r) is modeled as a normal random variable with mean 9.000 + 2.109*$ln(a)$ and standard deviation, δ = 0.522, the

PoD is expressed as $1-\Phi\left(\frac{1.099-(9.000\ +\ 2.109*\ln(a))}{0.522}\right)$, where Φ is the standard normal cumulative distribution function. This curve is graphed in figure 34, as is the similarly derived curve using 4 millivolts as the decision point.

From table 12 it is apparent that no signals from the rivet sites without flaws exceeded 3.7 millivolts. Only one response among the unflawed rivets exceeded 3. Thus, empirically, with a threshold of 3 the false call rate would be about 1 in 271. Applying similar analysis as above, we note that the unflawed signals had a mean response of 0.213 and standard error of .358 (these values are in the log scale). Therefore, the probability of a false alarm with the threshold of 3 would be estimated by the probability that a realization from a normal distribution with mean 0.213 and standard deviation of .358 would exceed 1.099. This value is 0.007. Similarly, the false call rate for using the threshold of 4 is estimated to be lower than 0.001.

The NASA investigators took as much time as they wanted to do the above inspections. In general, they inspected at a rate of between 1 to 3 rivets per minute, with most of the inspection being at the slower rate. They switched probe operators several times during the inspections. It was not uncommon for the computer operator (who was operating the data capture) to initiate the data capture as the probe operator moved the probe, making it necessary to reinspect. Had the probe operator made calls without capturing the data, the inspection time is likely to have been decreased substantially.

After the experience gained, the NASA team has reported making changes in the data capture to speed the inspection as well as to enhance repeatability. Specifically, they have added a centering circle to the raw data display giving the operator an absolute criterion of when the probe is aligned. That is, the operator has to get the centering cross hair inside the displayed circle to ensure alignment. They also changed the flow of the program so that the data processing is only performed when the probe is aligned. This has increased the update rate of the system to near real-time. An option to halt data acquisition and hold the data display once the probe is aligned was also added. With this latter change, the operator doesn't have to hold the probe centered over the rivet for an extended period. He only has to have it centered for a single revolution and the data will be held on the display. A click of the button on the probe head will restart the data acquisition while holding the button in for ~0.5 seconds writes the data to disk.

3.8 MCDONNELL DOUGLAS AEROSPACE/GK ENGINEERING SURFACE SCANNING PROBE.

3.8.1 GK Engineering Probe Background.

An experimenter from McDonnell Douglas Aerospace, visited AANC to test the GK Engineering surface scanning probe with the Elotest B1 minirotor [14,15]. The tests consisted of inspections of all 43 small panels, 2 large panels, and an area of the B737. Eighteen of the small panels had a bare inspection surface. The remaining panels were painted.

Table 13 gives the settings used in the inspection. The inspection required the centering of a probe guide over each rivet. The probe was then inserted in the guide and a waveform obtained.

A flat waveform would result from an unflawed rivet when the probe was properly centered. A flaw would show up as a peak in the waveform. Several fatigue cracks of known length were used to calibrate the instrument. Cracks in the range of 0.040 to 0.050 inch produced signals from 1 to 2 divisions during the calibration. In recording the results of the inspection, the inspector not only made a call, but also called out the height of the peak in terms of the nearest half division on the instrument face. The inspector used a signal threshold for making calls of approximately 3/4 of a division.

TABLE 13. GK PROBE/ELOTEST B1 INSTRUMENT SETTINGS

Frequency	150 kHz
Bandwdth	HF
Pre-amp	24 dB
Horizontal gain	47 dB
Vertical gain	67 dB
Rotation	348
Low pass filter	800
high pass filter	400

The first sample set was the bare panels mounted on the frames in the hangar. This required the inspector to inspect in a standing position with the probe at about eye level and the instrument screen placed on a table at waist level and off to his side.

During the course of the bare-panel inspections, the AANC monitor noticed that the inspector would sometimes have a flat response when he initially inserted the probe in the guide. But, as he held the probe in place and turned to view the monitor, a signal would result that was interpreted as indicating flaws on both sides of the rivet. After finishing the panels on the frames, the painted panels were given to the inspector one at a time. He inspected these on the bench top. The inspection time was much quicker with these panels than with the previous panels and the earlier observed double crack false indication did not occur near as often.

The full set of small panels was completed on the first day. The results were reviewed overnight and it was verified that more false calls had been made on the bare panels in the frames than had been made on the bench top inspections. The next day, the bare panels were reinspected, as well as a few of the painted panels. All the second inspections were performed on the bench top. During these inspections, the inspector took additional time to characterize the signals when the probe was aligned off-center, up and down, and side-to-side. The signals obtained from the probe being off-center in a downward direction were most like the signals obtained when cracks existed from both sides of the rivet. This helped explain the previous days false calls of double cracks when the inspections were done on the frames.

The signal data for the GK Engineering probe are graphed in figure 35.

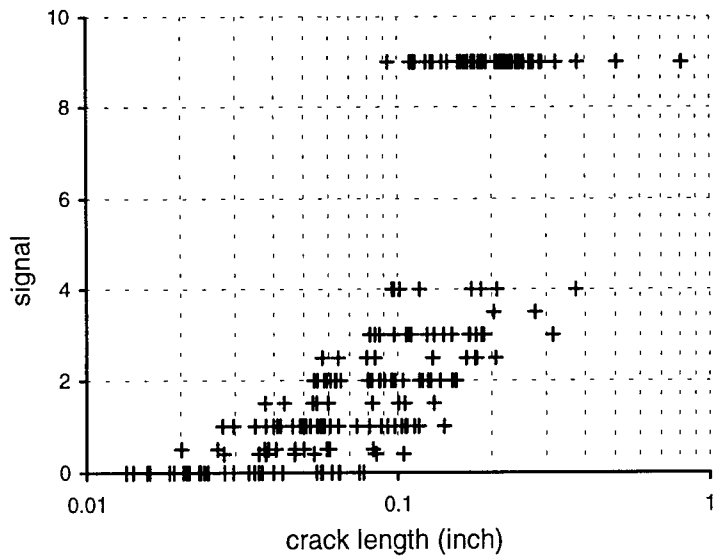

FIGURE 35. GK ENGINEERING PROBE—SIGNAL DATA

3.8.2 GK Probe Analysis.

Figure 36 shows probability of detection fits to the hit/miss data. The first curve is for the data obtained on the first day. There were 25 rivets at which false calls were made and several larger flaws missed. (The largest crack missed was at a rivet where another crack was present. Both cracks were ~0.25 inches.) The second curve substitutes the data for the set of panels done a second time for that of the first. Not only did the detections increase, but the number of false calls decreased to 8 rivets.

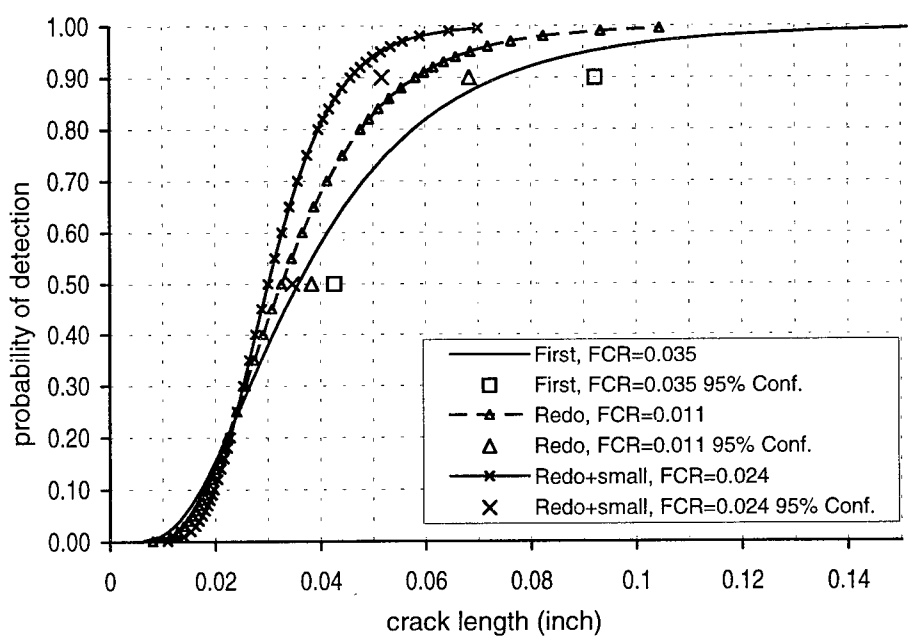

FIGURE 36. GK ENGINEERING PROBE—ESTIMATED PoD CURVES

During the inspection, there were numerous small signals that the inspector said he would not call, but he thought there might be a small crack present. The third curve is for when calls are made under the relaxed criteria. With the increased detection there was an increase in the false calls from 0.011 to 0.024.

There was not a significant effect due to inspecting on a painted surface versus a bare surface.

With the setup used, the middle curve of figure 36 gives the estimated PoD when the false call rate is approximately 1 percent and when there were no problems with maintaining the centering of the probe. The upper or leftmost curve shows the estimated PoD for a false call rate of approximately 2 percent. The lower or right curve gives an indication of the effect that field conditions might have if an inspector had problems in maintaining the centering of the probe. The false call rate would also increase in this latter case.

4. SUMMARY AND DISCUSSION.

We summarize the reliability results of the eight instruments in table 14. The 50 and 90 percent probability of detection crack lengths are for the cases where the false call rate is close to 1 percent.

TABLE 14. SUMMARY PoD VALUES

Technique / Instrument	a_{50}	a_{90}	Comments On Equipment and Procedures Used
Nortec Eddyscan	0.034	0.047	These detection levels require smaller standards (~0.060)
KB Crackfinder	0.119	0.160	PoD estimated using straight edge to guide probe close to rivet on bare surface and requiring probe to travel perpendicular to crack. Detection rates not as good when using template procedure.
Hocking Fastscan	0.047	0.062	Small gains in PoD indicated for setting scale for display to smaller standard, but not enough data to indicate impact on false call rates. Time consuming inspections due to inability to quickly center probe.
Nortec 19e - sliding probe 20 kHz	0.043	0.057	Off-angle cracks (22 degrees) shift PoD curve about 0.015 inch. Inspection over paint also shifts the PoD by about 0.008 inch.
Nortec 19e - sliding probe 30 kHz	0.037	0.050	Off-angle cracks (22 degrees) shift PoD curve about 0.011 inch. Inspection over paint also shifts the PoD by about 0.008 inch.
Zetec MIZ-22 - pencil probe	0.065	0.083	Painted surfaces shift the PoD curves by about 0.018 inch
Northrop LFECA	0.028	0.040	Indications that detection rates declined for smaller cracks (0.04 to 0.06) when in the presence of larger cracks at the same rivet.
NASA Self-nulling rotating probe	0.024	0.032	Latest implementation enables operator to accurately center probe
GK Engineering probe	0.033	0.058	Care has to be taken in centering the probe. Typical off-center indication is like that of having two cracks.

In any summary of this nature it is tempting to make judgments ranking the various instruments or procedures. We reiterate that the PoD analysis is not only for the equipment but also the

procedures and the way the inspection was implemented. Thus, it is possible that the same equipment that was used in the experiments reported here could be used in a different manner with different results. That this is the case is clearly indicated with the multiple inspection results of Northrop's LFECA and the KB CrackFinder and the procedural changes with the Hocking FastScan. In all these cases procedures or setups were altered from initial inspections with changes in the probability of detection resulting.

It is clear that probability of detection depends on the signal level used to make a call as well as the inspector's use of equipment. The false call rate will also depend on the signal level used to make a call. Therefore any comparisons of inspection methods should incorporate both probability of detection as well as false call rate. We do this in table 14 by picking 0.01 as a nominal false call rate for which the probability of detection values are given. This value is subject to uncertainty and this adds to the uncertainty of the values given in table 14.

The effect of inspection through paint was indicated as significant in several of the experiments. In general, the lessening of a signal due to lift off induced by the paint is not the issue. The presence of paint makes an impact on PoD through the added variation on probe alignment with respect to the rivets. Rivet boundaries are more obscure and for procedures that rely on visual alignment the result is more variation. The three systems that performed best (NASA self-nulling, LFECA, and Nortec Eddyscan) had signal displays that aided the inspector in centering the probe.

Hagemaier and Kark [14, 15] have studied the detection capabilities of four of the systems included in this study. They are the Hocking FastScan, Nortec-30 Eddyscan, Northrop LFECA, and GK Engineering/Elotest. Based on analyzing signals obtained using each of the instruments they concluded that the instruments are capable of detecting surface cracks 1.0 mm (0.040 inch) in length under flush-head aluminum rivets. Our results indicate that although capable of detecting this size crack, the probability associated with routinely detecting them (at false call rates < 0.01) are approximately 0.23 (FastScan), 0.74 (Eddyscan), 0.88 (LFECA), and 0.67 (GK).

In the cases we looked at using the maximum crack length as an explanatory variable is preferable to using the *sum of the crack lengths* when two cracks are present. In the *sum of cracks* model the number of cracks present was a significant factor and the PoD curves shifted in the direction making two cracks less detectable than one crack when at the same level for the crack length sum. On the other hand, when using the maximum crack length as an explanatory variable the number of cracks was often not significant. If estimated however, the PoD for two cracks versus one crack would increase (for constant maximum length) which is more intuitive.

5. REFERENCES.

1. Spencer, F.W. and Schurman, D. L., "Reliability Assessment at Airline Inspection Facilities, Vol III: Results of an Eddy-Current Inspection Reliability Experiment," DOT/FAA/CT-92/12, III, May 1995.

2. Proposed MIL-STD (1823) Nondestructive Evaluation System Reliability Assessment, August 1989.

3. Berens, Alan P., "NDE Reliability Data Analysis," Metals Handbook, v. 17, 9th ed., ASM International, 1988.

4. Hovey, P.W. and Berens, A.P., "Statistical Evaluation of NDE Reliability in the Aerospace Industry," *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 7B, Thompson, D.O. and Chimenti, D.E., eds., pp 1761-1768, Plenum Press, 1988.

5. Berens, A.P. and Hovey, P.W., "Statistical Methods for Estimating Crack Detection Probabilities," *Probabilistic Fracture Mechanics and Fatigue Methods: Application for Structural Design and Maintenance*, ASTM STP 798, J. M. Bloom and J. C. Ekvall, eds., American Society for Testing and Materials, 1983, pp 79-84.

6. Spencer, F.W., "Inspection Reliability of Nortec-30 Eddyscan System," DOT/FAA/AR-TN95/1, August 1995.

7. McCool, John I., "Censored Data," Encyclopedia of Statistical Sciences, Vol I, Kotz, S. and Johnson, N.L., eds, pp 389-396, Wiley & Sons, 1982.

8. Swets, J.A., "Assessment of NDT Systems—Part I: The Relationship of True and False Detections," *Materials Evaluation*, 41:1294-1298, 1983.

9. Swets, J.A., "Assessment of NDT Systems—Part II: Indices of Performance," *Materials Evaluation*, 41:1300-1303, 1983.

10. Draper, N.R. and Smith, H., *Applied Regression Analysis*, 1981, Wiley & Sons.

11. Sheppard, W., *WL-TR-94-4006: Eddy Current for Detecting Second Layer Cracks Under Installed Fasteners*, Wright Patterson Air Force Base, OH, 1994.

12. Wincheski, B., et al., "Detection of Fatigue Cracks at Rivets with Self-Nulling Probe," ASNT 1994 Spring Conference, p 125, 1994.

13. Wincheski, B., et al., "Self-Nulling Eddy-Current Probe for Surface and Subsurface Flaw Detection," *Materials Evaluation*, Vol. 52/no. 1, January 1994.

14. Hagemaier, D. and Kark, G.. "Eddy-Current Detection of Short Cracks Under Installed Fasteners," *Material Evaluation*, January 1997, pp 25-30.

15. Hagemaier, D., "Eddy-Current Detection of Short Cracks Under Installed Fasteners," 1995 ATA NDT Forum, September 26-28, 1995, Hartford, CT.

APPENDIX A—USING INVERSE REGRESSION FITS TO ESTABLISH PoD FUNCTIONS

The purpose of this appendix is to provide proof that the three step analysis introduced in section 3.4.2.3 produces the same estimate of probability of detection as does the $\hat{a}$ versus a analysis.

To facilitate the discussion let y be an NDI signal variable and let x be a flaw characteristic. The model considered is $y = \beta_0 + \beta_1 \cdot x + \varepsilon$, where ε has normal distribution with mean 0 and variance δ^2. A detection is said to occur if the signal, y, exceeds some threshold, T. The estimate of the probability of detection curve, using a threshold of T, is given by the cumulative normal distribution function with mean $\mu = \frac{T - \beta_0}{\beta_1}$ and standard deviation $\sigma = \delta / \beta_1$. The probability of detection function is estimated by substituting each of the β_0, β_1, and δ with their estimates from the regression.

It is convenient to express the estimates in terms of basic statistics of the x and y data. To this end we write $S_{xy} = (1/n) \cdot \sum (x_i - \overline{X}) \cdot (y_i - \overline{Y})$, $S_x^2 = (1/n) \cdot \sum (x_i - \overline{X})^2$, and $S_y^2 = (1/n) \cdot \sum (y_i - \overline{Y})^2$ where $\overline{X}$ and $\overline{Y}$ are the means of the x and y data. We also write $r = \frac{S_{xy}}{S_x \cdot S_y}$. It is well known that the maximum likelihood estimates are given by $\hat{\beta}_1 = \frac{S_y}{S_x} \cdot r$, $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \cdot \overline{X}$, and $\hat{\delta}^2 = (1 - r^2) \cdot S_y{}^2$. By substitution, it follows that the parameters of the PoD equation are estimated by

$$\hat{\mu} = \overline{X} + (T - \overline{Y}) \cdot \frac{S_x}{r \cdot S_y} \text{ and } \hat{\sigma} = \frac{(1 - r^2)^{1/2}}{r} \cdot S_x . \qquad (1)$$

(Note that the estimate for δ^2 is the maximum likelihood estimate and not the unbiased estimate that is derived from the expected mean squares in an analysis of variance table.)

Now consider fitting the equation $x = \gamma_0 + \gamma_1 \cdot y$. By reversing the roles of x and y we get $\hat{\gamma}_1 = \frac{S_x}{S_y} \cdot r$ and $\hat{\gamma}_0 = \overline{X} - \hat{\gamma}_1 \cdot \overline{Y}$. The predicted y' $(= \hat{\gamma}_0 + \hat{\gamma}_1 \cdot y)$ value is given by $y' = \overline{X} + \frac{S_x}{S_y} \cdot r \cdot (y - \overline{Y})$. It follows that $\overline{Y'} = \overline{X}$, $S_{y'}^2 = \frac{S_x^2}{S_y^2} \cdot r^2 \cdot S_y^2 = \frac{S_{xy}^2}{S_y^2}$, and $S_{xy'} = \frac{S_x}{S_y} \cdot r \cdot S_{xy} = \frac{S_{xy}^2}{S_y^2}$. We also consider, $r' = \frac{S_{xy'}}{S_x \cdot S_{y'}} = r$.

Let y' be the predicted x value after fitting the parameters in the above inverse regression. Now consider the equation $y' = \beta_0' + \beta_1' \cdot x$, with residual error δ'. The estimates for this equation are given by $\hat{\beta}_1' = \frac{S_{y'}}{S_x} \cdot r' = r^2$, $\hat{\beta}_0' = \overline{X} - r^2 \cdot \overline{X} = (1 - r^2) \cdot \overline{X}$, and $\hat{\delta}'^2 = (1 - r'^2) \cdot S_{y'}{}^2 = (1 - r^2) \cdot r^2 \cdot S_x^2$. Combining these quantities with a threshold on y' of

$\hat{\gamma}_0 + \hat{\gamma}_1 \cdot T = \overline{X} + \frac{S_x}{S_y} \cdot r \cdot (T - \overline{Y})$ and estimating μ and σ for the PoD curve using β_0', β_1', and δ', we obtain the same estimates as given in (1). Therefore, for the purposes of deriving a PoD curve, regressing the predicted flaw characteristic from an inverse regression fit onto the flaw characteristic in an $\hat{a}$ versus a analysis is equivalent to a single step $\hat{a}$ versus a analysis.

NOTE: The above argument addresses the PoD function that results from regression in a $\hat{a}$ versus a analysis. The PoD function is determined by various functions of the underlying parameters in the regression. It should be clear that these individual parameter estimates are not equivalent in the two analyses.

Before leaving this presentation of an alternate analysis using inverse regression, we briefly comment on the estimation of the scale parameter, δ. The $\hat{a}$ versus a analysis as previously presented [2,3] develop the probability of detection curve estimation and the estimation of confidence bounds using maximum likelihood estimates. The maximum likelihood estimate of δ^2 is the sum of squares due to error divided by the number of data points used in the analysis. That is, the estimate does not account for the number of parameters fit in the regression equation. To obtain an unbiased estimate of δ^2, the sum of squares due to error would be divided by the associated degrees of freedom. The difference in the estimate of δ would be a factor of $\sqrt{\frac{n}{n-p}}$, where n is the number of data points and p is the number of estimated parameters. The usual $\hat{a}$ versus a analysis includes two parameters, so that p = 2, and the difference in an unbiased estimate and the maximum likelihood estimate is not much. In case 6 of table 14, where 68 data points were used and 17 parameters were fit, the estimate of δ (and therefore σ) would be 15 percent higher if unbiased estimates rather than maximum likelihood estimates were used. Since the proposed inverse regression model does not limit the number of parameters used in fitting data and the estimated PoD function is dependent scale parameter, δ, we believe it is prudent to use an unbiased estimate of δ.