NUREG/CR-4908
PNL-6196

Ultrasonic Inspection Reliability for Intergranular Stress Corrosion Cracks

A Round Robin Study of the Effects of Personnel, Procedures, Equipment and Crack Characteristics

Prepared by P. G. Heasler, T. T. Taylor, J. C. Spanner, S. R. Doctor, J. D. Deffenbaugh

Pacific Northwest Laboratory
Operated by
Battelle Memorial Institute

Prepared for
U.S. Nuclear Regulatory Commission

AVAILABILITY NOTICE

Availability of Reference Materials Cited in NRC Publications

Most documents cited in NRC publications will be available from one of the following sources:

1. The NRC Public Document Room, 2120 L Street, NW, Lower Level, Washington, DC 20555

2. The Superintendent of Documents, U.S. Government Printing Office, P.O. Box 37082, Washington, DC 20013-7082

3. The National Technical Information Service, Springfield, VA 22161

Although the listing that follows represents the majority of documents cited in NRC publications, it is not intended to be exhaustive.

Referenced documents available for inspection and copying for a fee from the NRC Public Document Room include NRC correspondence and internal NRC memoranda; NRC Office of Inspection and Enforcement bulletins, circulars, information notices, inspection and investigation notices; Licensee Event Reports; vendor reports and correspondence; Commission papers; and applicant and licensee documents and correspondence.

The following documents in the NUREG series are available for purchase from the GPO Sales Program: formal NRC staff and contractor reports, NRC-sponsored conference proceedings, and NRC booklets and brochures. Also available are Regulatory Guides, NRC regulations in the *Code of Federal Regulations*, and *Nuclear Regulatory Commission Issuances*.

Documents available from the National Technical Information Service include NUREG series reports and technical reports prepared by other federal agencies and reports prepared by the Atomic Energy Commission, forerunner agency to the Nuclear Regulatory Commission.

Documents available from public and special technical libraries include all open literature items, such as books, journal and periodical articles, and transactions. *Federal Register* notices, federal and state legislation, and congressional reports can usually be obtained from these libraries.

Documents such as theses, dissertations, foreign reports and translations, and non-NRC conference proceedings are available for purchase from the organization sponsoring the publication cited.

Single copies of NRC draft reports are available free, to the extent of supply, upon written request to the Office of Information Resources Management, Distribution Section, U.S. Nuclear Regulatory Commission, Washington, DC 20555.

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at the NRC Library, 7920 Norfolk Avenue, Bethesda, Maryland, and are available there for reference use by the public. Codes and standards are usually copyrighted and may be purchased from the originating organization or, if they are American National Standards, from the American National Standards Institute, 1430 Broadway, New York, NY 10018.

NUREG/CR-4908
PNL-6196
R5

Ultrasonic Inspection Reliability for Intergranular Stress Corrosion Cracks

A Round Robin Study of the Effects of Personnel, Procedures, Equipment and Crack Characteristics

Manuscript Completed: May 1990
Date Published: July 1990

Prepared by
P. G. Heasler, T. T. Taylor, J. C. Spanner,
S. R. Doctor, J. D. Deffenbaugh

Pacific Northwest Laboratory
Richland, WA 99352

Prepared for
Division of Engineering
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555
NRC FIN B2289

ABSTRACT

A pipe inspection round robin entitled "Mini-Round Robin" was conducted at Pacific Northwest Laboratory from May 1985 through October 1985. The research was sponsored by the U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research under a program entitled "Evaluation and Improvement of NDE Reliability for Inservice Inspection of Light Water Reactors."

The Mini-Round Robin (MRR) measured the intergranular stress corrosion (IGSC) crack detection and sizing capabilities of inservice inspection (ISI) inspectors that had passed the requirements of IEB 83-02 and the Electric Power Research Institute (EPRI) sizing training course. The MRR data base was compared with an earlier Pipe Inspection Round Robin (PIRR) that had measured the performance of inservice inspection prior to 1982. Comparison of the MRR and PIRR data bases indicates no significant change in the inspection capability for detecting IGSCC. Also, when comparing detection of long and short cracks, no difference in detection capability was measured. An improvement in the ability to differentiate between shallow and deeper IGSCC was found when the MRR sizing capability was compared with an earlier sizing round robin conducted by the EPRI.

In addition to the pipe inspection round robin, a human factors study was conducted in conjunction with the Mini-Round Robin. The most important result of the human factors study is that the Relative Operating Characteristics (ROC) curves provide a better methodology for describing inspector performance than only probability of detection (POD) or single-point crack/no crack data.

EXECUTIVE SUMMARY

This report describes the findings from a limited pipe inspection round robin conducted at Pacific Northwest Laboratory (PNL) from May 1985 through October 1985. The research was sponsored by the U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research (RES).

The Mini-Round Robin (MRR) was conducted to provide an engineering data base for UT/ISI that would help:

- quantify the effect of training and performance demonstration testing that resulted from IEB 83-02,
- quantify the differences in capability between detecting long (greater than 3-in.) cracks versus short (less than 2-in.) cracks, and
- quantify the capability of UT/ISI inspectors to determine length and depth of intergranular stress corrosion cracks (IGSCC).

For this study, RES sponsored through PNL 6 two-person manual inspection teams, and industry sponsored the participation of three advanced ultrasonic inspection systems. The Electric Power Research Institute (EPRI) sponsored the participation of Amdata Systems, Inc. personnel with an INTRASPECT-2 (an automated data acquisition system that uses C-scan imaging technology) and J. A. Jones/EPRI NDE Center personnel using an ALN-4060 inspection system (a manual inspection system that uses automated crack/no crack decision analysis). Nuclear Energy Services/Dynacon Systems sponsored the participation of the Ultrasonic Data Recording and Processing System (UDRPS) (an automated data acquisition and signal processing imaging inspection system). The participation of these advanced UT inspection systems provided data that enabled a comparison of the detection capabilities of manual and advanced UT inspection systems.

In addition to the pipe inspection round robin, a human factors study was conducted in conjunction with the MRR. The purpose of the human factors study was to acquire preliminary data on the performance-shaping factors that affect ultrasonic testing (UT) reliability, to test the efficacy of relative operating characteristics analysis for representing UT performance, and to determine the direction of future human factors efforts in the NDT area.

The samples used in the MRR were fabricated from 7- to 9-in. lengths of 10-in. Schedule 80, 12-in. Schedule 80, and 12-in. Schedule 100 Type 304 wrought stainless steel pipe. The lengths of pipe were welded according to nuclear standards. The samples contained a variety of weld preparation conditions and welding techniques that ranged from manual to automated. Intergranular stress corrosion cracks were introduced in the samples under laboratory conditions at PNL and Ishikawajima-Harima Heavy Industries Co., Ltd. (IHI). All samples of the same diameter and wall thickness were fabricated to appear as identical as possible on the outside surface.

The 10-in.-dia. Schedule 80 pipe samples were used in a previous round-robin test, known as the Pipe Inspection Round Robin (PIRR) [1], conducted at PNL prior to the IEB 83-02 requirements. The PIRR samples were used to provide a basis for comparing current manual inspection capabilities with inspection capabilities before the performance demonstration practices were required by IEB 83-02.

Selection of participants for the MRR was designed to provide an unbiased sample of manual UT/ISI inspectors capability after IEB 83-02 requirements were in effect. Six organizations actively engaged in UT/ISI were contracted to send two inspectors. Inspectors participating in the manual inspection portion of the MRR were selected using the following criteria.

- Demonstrated experience in ultrasonic inservice or preservice inspection. All personnel, except one, participating in the MRR had successfully completed the performance demonstrations required by IEB 82-03.

- Demonstrated experience in using ultrasonic techniques in sizing cracks. Each team participating in the MRR had at least one inspector that had successfully completed the EPRI Sizing Training Course.

Operators of the advanced UT equipment were field engineers. Two of the advanced teams had passed the requirements of IEB 83-02. The advanced UT operators generally did not have as much inservice inspection experience as the manual UT participants.

Samples were prepared for presentation in two ways. First, the samples were masked such that only a portion of pipe was accessible, usually an octant (Figure ES.1). Second, three specimens were presented as whole pipe.

The specimen presentations were made to the inspectors in random order with two constraints. First, the same pipe was not to be presented twice in a row because of sample scheduling problems. Second, all pipe samples of the same wall thickness were presented as a block of inspections to minimize calibration time.

Inspectors were encouraged to complete each inspection within a 45-minute time period. However, they were allowed as much time as they desired to make an inspection. It is recognized that field conditions may dictate that an inspection be performed within a given time period due to ALARA or other considerations. However, a strict time limit was not imposed during this study.

Inspections were conducted on a workbench. Each sample was located on a four-roller stand, which allowed the pipe to be rotated to provide optimum location of the sample examination area. The work station was arranged by the inspector as desired.

Each inspector was asked to record his own data. This was done on either a company form or on a form provided by PNL. Data were later transcribed by the inspector to the PNL form for evaluation. In addition to recording information on the crack location and depth, each inspector was required to indicate his degree of confidence of crack <u>detection</u> on the following five-item scale:

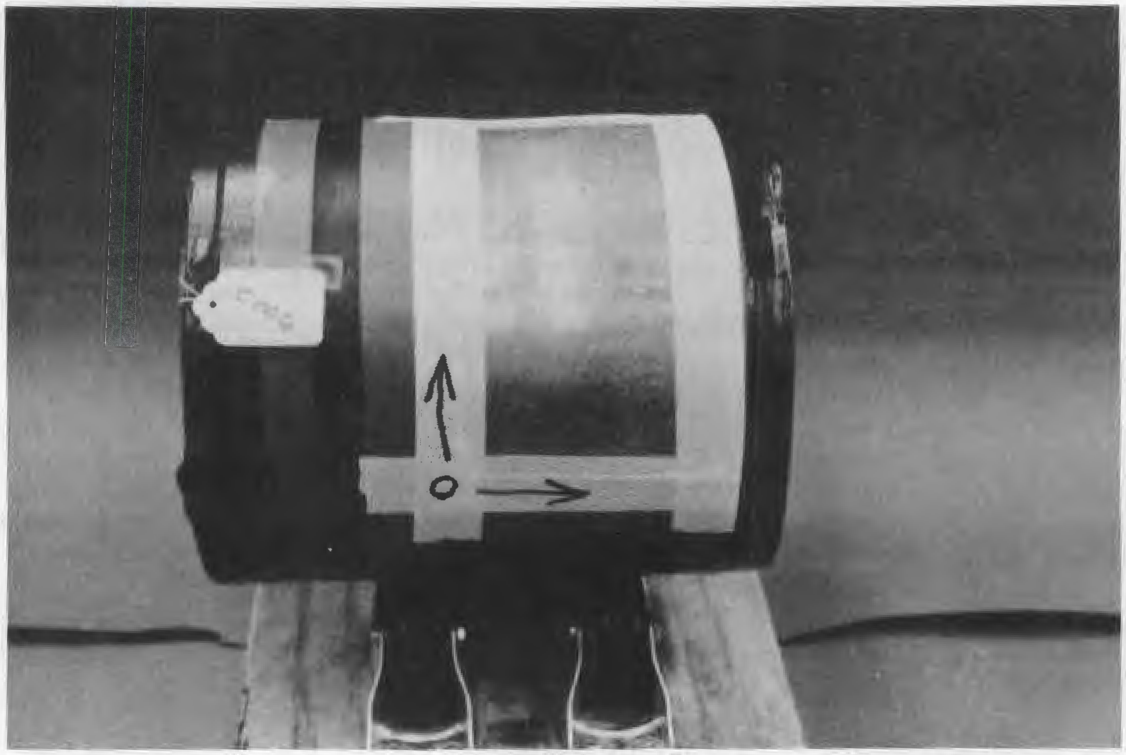


FIGURE ES.1. Typical Sample Presentation for Mini-Round Robin

C-1 = Definitely, or almost definitely, an IGSC crack
C-2 = Probably an IGSC crack
C-3 = Possibly an IGSC crack
N-4 = Probably not an IGSC crack
N-5 = Definitely, or almost definitely, not an IGSC crack

This scale was used in developing Relative Operating Characteristic (ROC) curves which plot Probability of Detection (POD) or the number of cracked grading units called cracked/number of cracked grading units inspected and False Call Probability (FCP) or number of blank grading units called cracked/number of blank grading units inspected and are used in analyzing technician performance.

Care was taken to check the transcription process and to remove any obvious data recording errors (i.e., simple and obvious transpositions and out-of-range entries) from the data. Though this "data cleanup" may not represent the real world problem of recording errors during an ISI, it was desired to eliminate as many extraneous errors as possible from the detection problem. The errors corrected were of a nature that would normally be caught during an administrative or supervisory review of UT/ISI data.

The scope of the human factors evaluation of the MRR was:

- To acquire preliminary data with which to evaluate the performance-shaping factors that affect the reliability of the UT/ISI process.

- To test the efficacy of the relative operating characteristic (ROC) analysis technique for measuring the reliability and accuracy of UT/ISI results.

- To gain insight regarding needed direction and focus of future research and regulatory efforts in the human factors-NDT area.

Data from the human factors study provided preliminary answers to the following questions:

- How do organizational and supervisory characteristics of the job affect UT performance?

- How do IGSCC training, performance demonstrations, and field experience affect UT performance?

- How does equipment design affect UT performance?

- How does the inspection environment affect UT performance?

- How is UT/ISI conducted to detect IGSCC?

- Is it feasible to use Relative Operating Characteristic (ROC) analysis on ultrasonic systems?

During the human factors study, each inspector was observed as he was examining pipe and occasionally queried as to his rationale for decision criteria (i.e., why does that particular signal on the instrument display warrant attention?). Each inspector also participated in individual, structured interviews, which focused on their experiences in UT/ISI (see Appendix A). Each subject also filled out a survey questionnaire dealing with performance-shaping factors in UT/ISI (see Appendix B). In addition, human factors specialists evaluated the UT equipment for compliance with established human factors principles using a specially prepared equipment evaluation checklist (see Appendix C).

INSPECTION RESULTS

Figure ES.2 shows ROC curves for the three "best," three "worst" inspectors, and "average" result from the twelve teams that participated in the MRR. Figure ES.2 illustrates that some manual UT/ISI technicians are capable of adequate detection of IGSC cracks.

Figure ES.3 shows a comparison of MRR inspectors and PIRR inspectors. Figure ES.3 illustrates that there is little difference in detection capability between inspectors that participated in the MRR (post-IEB 83-02) versus inspectors that participated in the PIRR (pre-IEB 83-02).

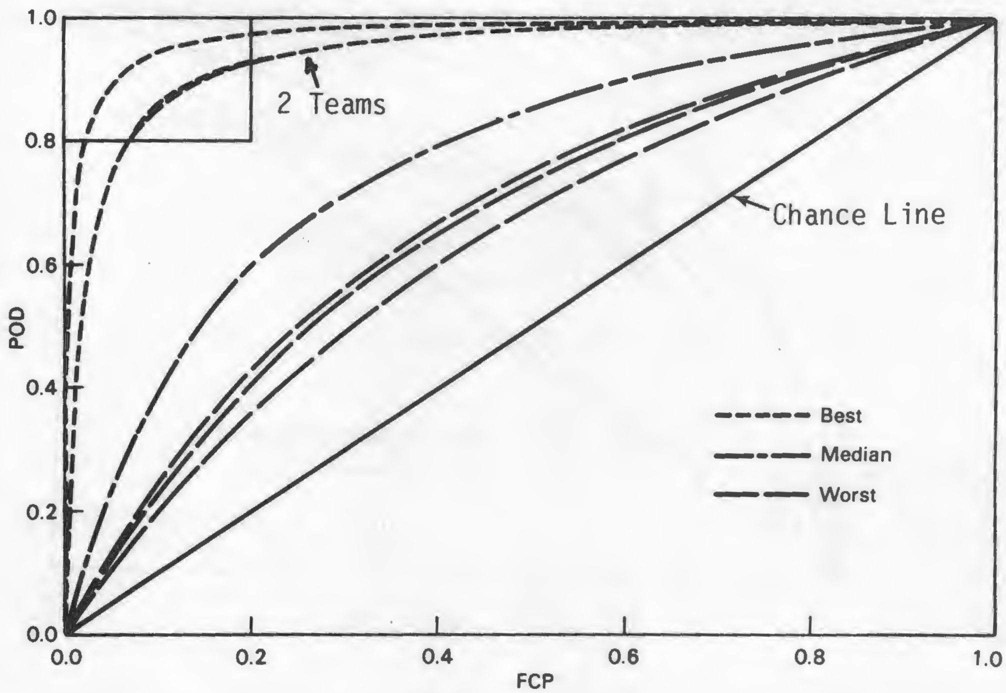


FIGURE ES.2. ROC Curves Showing Three Best, Average, and Three Worst Manual Inspection Performance for Near-Side Examination of Cracks Measured by Mini-Round Robin

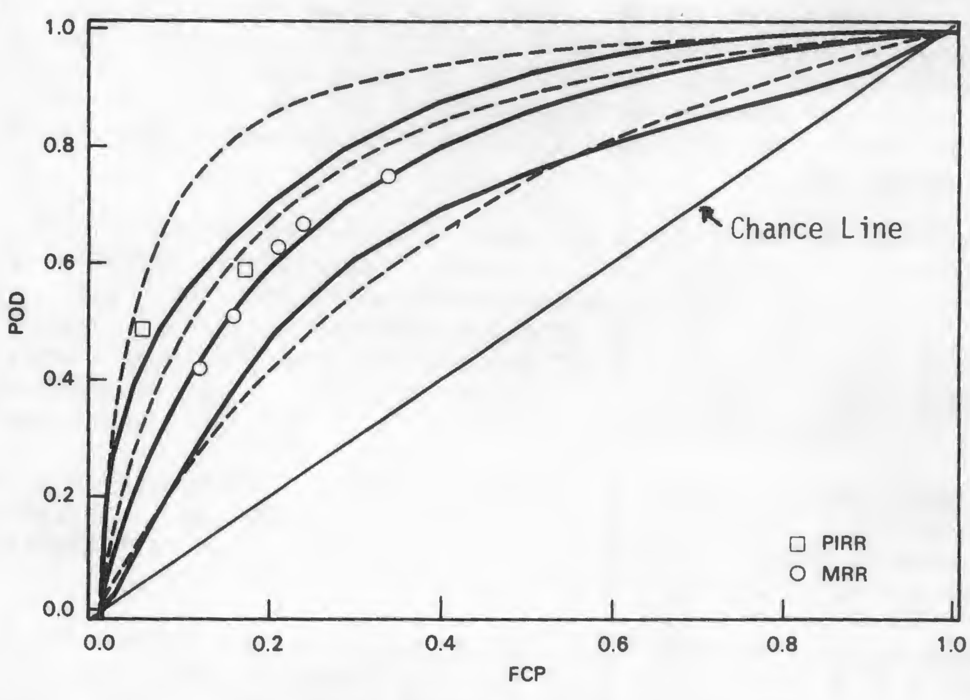


FIGURE ES.3. Comparison of MRR (——) and PIRR (---) Inspectors showing ROC and 95% Confidence Limits

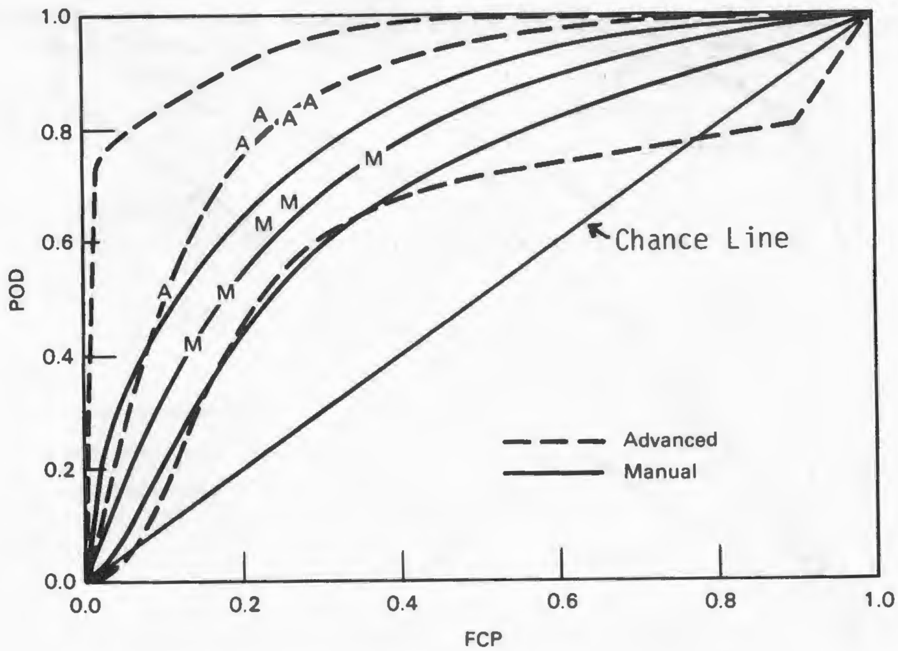


FIGURE ES.4. Advanced Versus Manual ROC for Near-Side Access for Short Cracks

Figure ES.4 shows the "averaged" performance of MRR manual inspectors and advanced inspection systems. Data from Figure ES.4 indicates that the detection performance of advanced UT inspection systems was slightly better than the MRR manual teams. A t-test of the area under the curves indicates that separation of curves was significant at the 18% level, or one was 82% confident that the average advanced UT inspection was better than the average manual inspection.

Data from the MRR project was used to determine whether "long" IGSC cracks were easier to detect than "short" cracks. Our analysis showed no significant difference in detection performance between short and long cracks. This hypothesis is bolstered by the fact that the inspectors tended to report cracks in the long-crack samples to be of roughly the same length as cracks in the short-crack sample. They also exhibited considerably less confidence in their calls in the long-crack samples than they did in the short-crack samples.

In addition to conducting a detection test, one inspector from each team participated in a sizing round robin. Inspectors that participated in the sizing round robin had successfully completed the sizing performance demonstrations conducted at the EPRI NDE Center.

There appears to be definite improvement in sizing capability based on comparing a 1983 EPRI sizing study [5] with the MRR sizing data base. The MRR data, however, does not support the conclusion that sizing is repeatable and reliable. The MRR data base shows that sizing cracks, especially cracks in the 20% to 50% through-wall range, is very difficult. Sometimes crack tip signals can be found; often, however, crack tip signals cannot be distinguished from noise. If crack tip signals are detected, accurate results may be obtained. The MRR data shows that crack tip signals are not always easy to detect; hence, inspectors' sizing results showed significant sample-to-sample variation. The MRR data base also shows that crack tip sizing procedures seem to have an inherent resolution limit estimated at 0.08 inch for usable frequencies in austenitic material. This resolution limit needs to be recognized when interpreting sizing results and designing performance demonstrations.

CONCLUSIONS

The data from the MRR suggests the following:

1. The MRR data base shows that the best manual UT/ISI is capable of acceptable crack detection performance. However, the MRR data base shows a wide variation between best and worst inspectors with the majority of inspectors having unacceptable performance.

2. The MRR data base suggests that there has been little improvement in the POD/FCP capability of inspectors since IEB 83-02 requirements (when passing the test required detecting 4 of 5 cracks) versus before these requirements.

3. The average IGSC crack detection capability of advanced UT technology is slightly better than the "average" performance of manual inspectors that participated in the MRR. This does not mean that manual inspectors are categorically poor; in fact, the best manual inspector's performance exceeded the performance of all advanced inspection systems. However, the "average" performance of advanced technology was better than the average performance of manual inspection. This is important to note because the experience in using advanced technology on actual cracks is limited. The assumption has been that advanced systems place more intelligence into the system and, thus, reduce the experience needs of the inspectors. This data supports this assumption.

4. The MRR demonstrates that ROC analysis can be used to measure the performance of UT inspection systems and illustrates the relationship between POD and FCP.

5. The MRR analysis indicates that no significant difference exists in the POD performance of inspection for long versus short IGSC cracks of similar depths.

6. As expected, the short/shallow cracks were very difficult to detect and did not differ much from the chance performance line in the ROC.

7. The MRR data shows a systematic undersizing (by approximately 1 in.) in the length of long IGSC cracks.

8. The MRR data base shows that far-side inspection is very poor.

9. Crack depth sizing performance has improved in this study compared with round-robin data taken in 1982, before the EPRI crack sizing course. However, the MRR data shows that sizing is not reliable.

10. Crack depth sizing of cracks in the 20-50% through-wall range is very difficult.

11. If crack-tip signals are detected, then accurate sizing may be achieved. However, only a small percentage of the time were crack-tip signals recognized.

12. Large crack-to-crack variability was found as seen by the large standard deviation.

13. All sizing is based on acoustic systems, which have resolution limits (estimated at approximately 0.08 in. for usable frequencies in austenitic piping). These limits must be recognized when interpreting sizing results and designing performance demonstration tests.

14. The MRR data analysis showed that there is a significant need for improving the reliability and reducing the variability of the UT/ISI process and one solution may be by applying the concepts and principles of the human factors engineering discipline. The human factors work conducted in conjunction with the MRR showed that this is a complicated problem with many factors requiring correction; and in order to address these, a comprehensive program should be initiated to reduce their influence on UT/ISI reliability to acceptable levels.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the following people for their support, cooperation, and contributions:

- Dr. Joe Muscara of the U.S. Nuclear Regulatory Commission.

- The participating inspectors and their companies:

  - Amdata Systems, Inc.
  - BESTCO
  - Dynacon Systems
  - EBASCO Corp.
  - J. A. Jones/EPRI NDE Center
  - Lambert, McGill & Thomas, Inc.
  - Nuclear Energy Services
  - Southwest Research Institute
  - Virginia Corp.

- Mr. Larry D. Reid of Pacific Northwest Laboratory for checking the data base true-state values.

- Ms. Kay E. Hass of Pacific Northwest Laboratory for her critical editorial comments and even temperament during the arduous task of editing, typing, and re-typing.

- A special thanks to the Electric Power Research Institute and NES/Dynacon Systems for their sponsorship of the inspections using the advanced UT systems.

- A number of PNL staff were instrumental in the preparation of weld samples, the growing of IGSCC, and the characterization of the cracking. They included: Ivar Husa, Linnea Bickford, Larry Van Fleet, Bob Bowey, and Howard Hartzog.

- Other human factors staff who worked on the MRR were Richard Badalamente, Bill Rankin, and Bill Wheeler. We thank you for your contributions.

CONTENTS

FIGURES

TABLES

PREVIOUS REPORTS IN THIS SERIES

Doctor, S. R., J. D. Deffenbaugh, M. S. Good, E. R. Green, P. G. Heasler, F. A. Simonen, J. C. Spanner, and T. T. Taylor. 1989. Nondestructive Examination (NDE) Reliability for Inservice Inspection of Light Water Reactors, Semi-Annual Report October 1987-March 1988. NUREG/CR-4469, PNL-5711, Vol. 8. Pacific Northwest Laboratory, Richland, Washington.

Doctor, S. R., D. J. Bates, J. D. Deffenbaugh, M. S. Good, P. G. Heasler, F. A. Simonen, J. C. Spanner, T. T. Taylor, and L. G. Van Fleet. 1988. Nondestructive Examination (NDE) Reliability for Inservice Inspection of Light Water Reactors, Semi-Annual Report April 1987-September 1987. NUREG/CR-4469, PNL-5711, Vol. 7. Pacific Northwest Laboratory, Richland, Washington.

Doctor, S. R., J. D. Deffenbaugh, M. S. Good, E. R. Green, P. G. Heasler, G. A. Mart, F. A. Simonen, J. C. Spanner, T. T. Taylor, and L. G. Van Fleet. 1987. Nondestructive Examination (NDE) Reliability for Inservice Inspection of Light Water Reactors, Semi-Annual Report October 1986-March 1987. NUREG/CR-4469, PNL-5711, Vol. 6. Pacific Northwest Laboratory, Richland, Washington.

Doctor, S. R., D. J. Bates, J. D. Deffenbaugh, M. S. Good, P. G. Heasler, G. A. Mart, F. A. Simonen, J. C. Spanner, T. T. Taylor, and L. G. Van Fleet. 1987. Nondestructive Examination (NDE) Reliability for Inservice Inspection of Light Water Reactors, Semi-Annual Report April 1986-September 1986. NUREG/CR-4469, PNL-5711, Vol. 5. Pacific Northwest Laboratory, Richland, Washington.

Doctor, S. R., D. J. Bates, J. D. Deffenbaugh, M. S. Good, P. G. Heasler, G. A. Mart, F. A. Simonen, J. C. Spanner, A. S. Tabatabai, T. T. Taylor, and L. G. Van Fleet. 1987. Nondestructive Examination (NDE) Reliability for Inservice Inspection of Light Water Reactors, Semi-Annual Report October 1985-March 1986. NUREG/CR-4469, PNL-5711, Vol. 4. Pacific Northwest Laboratory, Richland, Washington.

Collins, H. D. and R. P. Gribble. 1986. Siamese Imaging Technique for Quasi-Vertical Type (QVT) Defects in Nuclear Reactor Piping. NUREG/CR-4472, PNL-5717. Pacific Northwest Laboratory, Richland, Washington.

Doctor, S. R., D. J. Bates, R. L. Bickford, L. A. Charlot, J. D. Deffenbaugh, M. S. Good, P. G. Heasler, G. A. Mart, F. A. Simonen, J. C. Spanner, A. S. Tabatabai, T. T. Taylor, and L. G. Van Fleet. 1986. Nondestructive Examination (NDE) Reliability for Inservice Inspection of Light Water Reactors, Semi-Annual Report April 1985-September 1985. NUREG/CR-4469, PNL-5711, Vol. 3. Pacific Northwest Laboratory, Richland, Washington.

Doctor, S. R., D. J. Bates, L. A. Charlot, M. S. Good, H. R. Hartzog, P. G. Heasler, G. A. Mart, F. A. Simonen, J. C. Spanner, A. S. Tabatabai, and T. T. Taylor. 1986. Evaluation and Improvement of NDE Reliability for Inservice Inspection of Light Water Reactors, Semi-Annual Report, October 1984-March 1985. NUREG/CR-4469, PNL-5711, Vol. 2. Pacific Northwest Laboratory, Richland, Washington.

Doctor, S. R., D. J. Bates, L. A. Charlot, H. D. Collins, M. S. Good, H. R. Hartzog, P. G. Heasler, G. A. Mart, F. A. Simonen, J. C. Spanner, and T. T. Taylor. 1986. Integration of Nondestructive Examination (NDE) Reliability and Fracture Mechanics, Semi-Annual Report, April 1984 - September 1984. NUREG/CR-4469, PNL-5711, Vol. 1. Pacific Northwest Laboratory, Richland, Washington.

Good, M. S. and L. G. Van Fleet. 1986. Status of Activities for Inspecting Weld Overlaid Pipe Joints. NUREG/CR-4484, PNL-5729. Pacific Northwest Laboratory, Richland, Washington.

Heasler, P. G., D. J. Bates, T. T. Taylor, and S. R. Doctor. 1986. Performance Demonstration Tests for Detection of Intergranular Stress Corrosion Cracking. NUREG/CR-4464, PNL-5705, Pacific Northwest Laboratory, Richland, Washington.

Simonen, F. A. 1984. The Impact of Nondestructive Examination Unreliability on Pressure Vessel Fracture Predictions. NUREG/CR-3743, PNL-5062. Pacific Northwest Laboratory, Richland, Washington.

Simonen, F. A. and H. H. Woo. 1984. Analyses of the Impact of Inservice Inspection Using Piping Reliability Model. NUREG/CR-3753, PNL-5070. Pacific Northwest Laboratory, Richland, Washington.

Taylor, T. T. 1984. An Evaluation of Manual Ultrasonic Inspection of Cast Stainless Steel Piping. NUREG/CR-3753, PNL-5070. Pacific Northwest Laboratory, Richland, Washington.

Bush, S. H. 1983. Reliability of Nondestructive Examination, Volumes I, II, and III. NUREG/CR-3110-1, -2, and -3; PNL-4584. Pacific Northwest Laboratory, Richland, Washington.

Simonen, F. A. and C. W. Goodrich. 1983. Parametric Calculations of Fatigue Crack Growth in Piping. NUREG/CR-3059, PNL-4537. Pacific Northwest Laboratory, Richland, Washington.

Simonen, F. A., M. E. Mayfield, T. P. Forte, and D. Jones. 1983. Crack Growth Evaluation for Small Cracks in Reactor-Coolant Piping. NUREG/CR-3176, PNL-4642. Pacific Northwest Laboratory, Richland, Washington.

Taylor, T. T., S. L. Crawford, S. R. Doctor, and G. J. Posakony. 1983. Detection of Small-Sized Near-Surface Under-Clad Cracks for Reactor Pressure Vessels. NUREG/CR-2878, PNL-4373. Pacific Northwest Laboratory, Richland, Washington.

Busse, L. J., F. L. Becker, R. E. Bowey, S. R. Doctor, R. P. Gribble, and G. J. Posakony. 1982. Characterization Methods for Ultrasonic Test Systems. NUREG/CR-2264, PNL-4215. Pacific Northwest Laboratory, Richland, Washington.

Morris, C. J. and F. L. Becker. 1982. State-of-Practice Review of Ultrasonic In-service Inspection of Class I System Piping in Commercial Nuclear Power Plants. NUREG/CR-2468, PNL-4026. Pacific Northwest Laboratory, Richland, Washington.

Becker, F. L., S. R. Doctor, P. G. Heasler, C. J. Morris, S. G. Pitman, G. P. Selby, and F. A. Simonen. 1981. Integration of NDE Reliability and Fracture Mechanics, Phase I Report. NUREG/CR-1696-1, PNL-3469. Pacific Northwest Laboratory, Richland, Washington.

Taylor, T. T. and G. P. Selby. 1981. Evaluation of ASME Section XI Reference Level Sensitivity for Initiation of Ultrasonic Inspection Examination. NUREG/CR-1957, PNL-3692. Pacific Northwest Laboratory, Richland, Washington.

1.0 INTRODUCTION

After ultrasonic inservice inspection (UT/ISI) failed to detect intergranular stress corrosion (IGSC) cracking in the 28-in.-dia. recirculation piping of Nine Mile Point, the NRC issued requirements for crack detection performance demonstrations for inspection of BWR primary piping systems. These requirements were first issued in 1982 in Inspection and Enforcement Bulletin (IEB) 82-03 and then updated in 1983 by IEB 83-02. The nuclear power industry responded to the requirements of these bulletins by developing a training and performance demonstration program. The Mini-Round Robin (MRR) was conducted from May 1985 to October 1985 to provide an engineering data base for UT/ISI that would help:

- quantify the effect of training and performance demonstration testing that resulted from IEB 83-02,
- quantify the differences in capability between detecting long (greater than 3-in.) cracks versus short (less than 2-in.) cracks, and
- quantify the capability of UT/ISI inspectors to determine the length and depth of intergranular stress corrosion cracks (IGSCC).

For this study, the NRC Office of Nuclear Regulatory Research through PNL sponsored 6 two-person manual inspection teams, and industry sponsored the participation of three advanced ultrasonic inspection systems. The Electric Power Research Institute (EPRI) sponsored the participation of Amdata Systems, Inc. personnel with an INTRASPECT-2 (an automated data acquisition system that uses C-scan imaging technology) and J. A. Jones/EPRI NDE Center personnel using an ALN-4060 inspection system (a manual inspection system that uses automated crack/no crack decision analysis). Nuclear Energy Services/Dynacon Systems sponsored the participation of the Ultrasonic Data Recording and Processing System (UDRPS) (an automated data acquisition and signal processing imaging inspection system). The participation of these advanced UT inspection systems provided data that enabled a comparison of the detection capabilities of manual and advanced UT inspection systems.

In addition to the pipe inspection round robin, a human factors study was conducted during the MRR. The purpose of the human factors study was to acquire preliminary data on performance-shaping factors that affect ultrasonic testing (UT) reliability, to test the efficacy of relative operating characteristics analysis for representing UT performance, and to determine the direction of future human factors efforts in the NDT area.

This report has been organized in the following manner: Chapter 2 of this report describes the design of the MRR and the associated human factors study. Chapter 3 describes the parameters and analysis scheme developed based on ROC curves to evaluate the inspection capability. Chapter 4 describes the results for the manual and advanced teams. Chapter 5 is the analysis of the human factors study. Finally, Chapter 6 provides a list of conclusions and recommendations arising from the study.

blank material (see Section 3.3.2, False Call Signals) have been destructively analyzed, so the results reported in this report are based on destructive data of the samples.

The inspection area consisted of a smooth surface weld (i.e., no crown or outer-diameter undercut) and a single scribe line, used as a reference line for flaw location measurement. It is recognized that field conditions often include difficult weld conditions, such as weld reinforcement and ripple. However, in this study all weld crowns were prepared to appear identical so that replicate studies could be conducted; and thus, it was not possible to include variable difficult crown conditions because the inspectors would easily recognize them.

2.1.2 MRR Participants

Selection of participants for the MRR was designed to provide an unbiased sample of manual UT/ISI inspector capability after IEB 83-02 requirements were in effect. Six organizations actively engaged in UT/ISI were contracted to send two inspectors. Inspectors participating in the manual inspection portion of the MRR were selected using the following criteria.

- Demonstrated experience in ultrasonic inservice or preservice inspection. All personnel, except one, participating in the MRR had successfully completed the performance demonstrations required by IEB 83-02.

- Demonstrated experience in using ultrasonic techniques in sizing cracks. Each team participating in the MRR had at least one inspector that had successfully completed the EPRI Sizing Training Course.

Operators of the advanced UT equipment were field engineers. Two of the advanced teams had passed the requirements of IEB 83-02. The advanced UT operators generally did not have as much inservice inspection experience as the manual UT participants.

2.1.3 Test Protocol

Samples were prepared for presentation in two ways. First, the samples were masked such that only a portion of pipe was accessible, usually an octant (Figure 2.1). Second, three specimens were presented as whole pipe.

The specimen presentations were made to the inspectors in random order with two constraints. First, the same pipe was not to be presented twice in a row because of sample scheduling problems. Second, all pipe samples of the same wall thickness were presented as a block of inspections to minimize time spent in calibrating. Table 2.2 provides an example of the matrix (not the order of presentation) of the short-crack (<2 inches) samples presented to MRR participants.

The detection test presentation for short cracks resulted in a test set that contained:

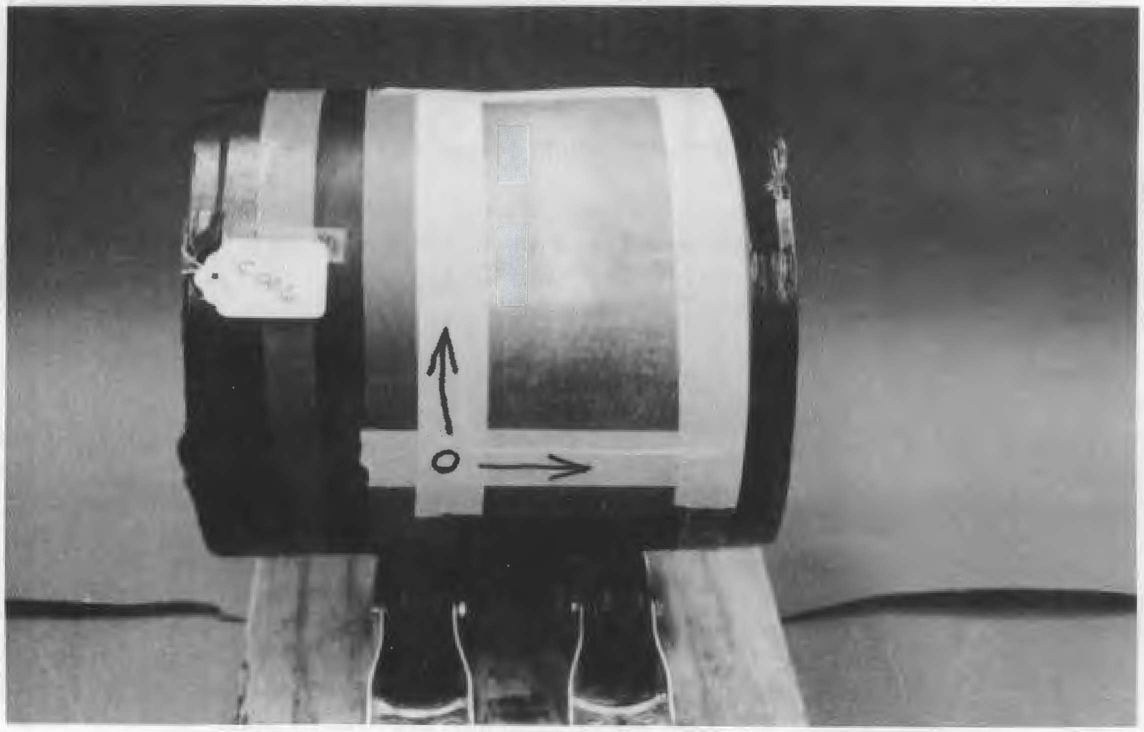


FIGURE 2.1. Typical Sample Presentation for Mini-Round Robin

TABLE 2.2. Example of the Matrix for Short "Crack" (<2 inches) MRR Specimens

No.	Specimen	Side Presented	Cracks Near-Side	Cracks Far-Side	
1	D101-1	A	0	0--	-->repeat test
2	D101-1	A	0	0--	
3	D101-2	B	0	0	
4	D101-4	B	1	0--	-->repeat test
5	D101-4	B	1	0--	
6	D102-1	A	0	0	
7	D102-2	A	1	1--	-->specimens reversed
8	D102-2	B	1	1--	
9	D102-3	B	1	0	
10	D102-4	B	1	0	
11	D105-4	A	1	0	
12	D107-2	A	1	1--	-->specimens reversed
13	D107-2	B	1	1--	
14	B217-2	A	3	1	

4 blank quadrants
12 near-side cracks
5 far-side cracks

Each participant worked independently on the specimens that he was provided.

To avoid specimen conflict within each team, one individual was presented specimens of one pipe wall thickness, while the other individual was presented specimens of another pipe wall thickness.

Inspectors were encouraged to complete each inspection within a 45-minute time period. However, they were allowed as much time as they desired to make an inspection, and a summary of their average inspection times is presented in Table 2.3. It is recognized that field conditions may dictate that an inspection be performed within a given time period due to ALARA or other considerations. From Table 2.3, it is apparent that the average inspection time achieved during the test is fairly close to the 45-minute (0.75-hour) guidance. We had expected the inspection time to be much less (say a factor of 2) for blank specimens as opposed to cracked specimens. Surprisingly, Table 2.3 shows that blank specimens take approximately the same time as cracked specimens, although seven of the inspectors took more time on the cracked specimens and five took more time on the blank specimens.

TABLE 2.3. Average Inspection Times for Blank and Cracked Specimens

	Specimen State	
Inspector	Blank, hours	Cracked, hours
1	0.83	1.27
2	0.64	0.85
3	0.44	0.59
4	0.77	0.62
5	0.36	0.41
6	0.23	0.35
7	1.67	0.86
8	0.69	0.89
9	0.94	1.32
10	2.11	1.54
11	0.97	0.59
12	1.34	0.90
Average	0.92	0.85

Inspections were conducted on a workbench. Each sample was located on a four-roller stand, which allowed the pipe to be rotated to provide optimum location of the sample examination area. Each work station was arranged by the inspector as desired. A typical work station is shown in Figure 2.2.

FIGURE 2.2. Typical Work Station showing the PNL observer in the right-hand lower corner and two of the MRR participating inspectors performing the examinations

Each inspector was asked to record his own data. This was done on either a company form or on a form provided by PNL. If a company form was used, then the data were later transcribed by the inspector to the PNL form for evaluation. Figure 2.3 shows the PNL form used for summarizing inspection data. In addition to recording information on the crack location and depth, each inspector was required to indicate his degree of confidence of crack detection or non-detection on the following five-item scale:

C-1 = Definitely, or almost definitely, an IGSC crack
C-2 = Probably an IGSC crack
C-3 = Possibly an IGSC crack
N-4 = Probably not an IGSC crack
N-5 = Definitely, or almost definitely, not an IGSC crack

This scale was used in developing Relative Operating Characteristic (ROC) curves for analyzing performance.

Areas were preselected to be used for grading performance for crack detection and false call rate. The data analysis procedure deals with the classification of the calls or dispositions made by the inspectors in these areas called "grading units." To account for all calls and in some cases where no calls were made, the data were placed into a new classification scale. The inspectors' 5-item scale was transformed into a 6-item grading unit scale as listed in Table 2.4. The data analysis then used the systematic

MINI ROUND ROBIN
INSPECTION REPORT FORM

TO BE FILLED OUT BY PNL PERSONNEL ONLY!

TEAM	6	PIPE CODE	D101
MEMBER	2	QUADRANT	1
INSPECTION	6225	SIDE PRESENTED	A
PREPARED BY		OBSERVER	

TO BE FILLED OUT BY PNL PERSONNEL ONLY!

	Hour*		Date		
	Begin	End	Year	Month	Day
INSPEC.		2005	85	7	19

CALIBRATION

*A time of 3:45 p.m. would be written as 15:45

INDICATIONS

MAX FLAW RESP (dB)	POSITION OF MAX RESP L0 (in)	L1 (in)	L2 (in)	WO (in) -FAR +NEAR	DEPTH T-WALL	DISPOSITION (See key below)
80%	3/8	0	7/8	+.25		C-1
100%	1-7/8	1-1/4	2-3/8	+.25		C-1
50%	3-5/8	Spot		+.25		C-1
50%	5-1/4	5-1/8	5-3/8	+.25		C-1
112%	7-1/2	6-7/8	7-7/8	+.25		C-1
36%	6	Axial		+.125		C-1
36%	6-1/8	Axial		+.125		C-1
36%	6-7/8	Axial		+.125		C-1

Level 2   Level 3

Key: C-1 = Definitely, or almost definitely, a crack
C-2 = Probably a crack
C-3 = Possibly a crack
N-4 = Probably not a crack
N-5 = Definitely, or almost definitely, not a crack

FIGURE 2.3. MRR Data Report Form

classification procedure as outlined in Table 2.4. It needs to be recognized that IGSCC are spotty in their UT response and, thus, a grading unit may contain areas where several different inspector dispositions are recorded. The analysis procedure classified a grading unit with the highest classification that the inpector recorded for an ordering where C-1 is the highest and N-5 is the lowest. This is the reason for the seemingly complex relationship criteria that are listed in Table 2.4.

TABLE 2.4. Classification Scale for Grading Units

Classification Assigned to Grading Unit	Relationship to Inspection Results
N6	No indications placed within grading unit
N5	At least one N-5 indication placed in G.U. but no N-4 through C-1 indications
N4	At least one N-4 indication placed in G.U. but no C-3 through C-1 indications
C3	At least one C-3 indication placed in G.U. but no C-2 through C-1 indications
C2	At least one C-2 indication placed in G.U. but no C-1 indication
C1	At least one C-1 indication placed in G.U.

Care was taken to check the transcription process and to remove any obvious data recording errors (i.e., simple and obvious transpositions and out-of-range entries) from the data. Though this "data cleanup" may not represent the real world problem of recording errors during an ISI, it was desired to eliminate as many extraneous errors as possible from the detection problem. The errors corrected were of a nature that would normally be caught during an administrative or supervisory review of UT/ISI data.

2.2 DESIGN OF HUMAN FACTORS STUDY

Human factors is a multidisciplinary field that utilizes the specialized knowledge available from engineering, experimental psychology, biology, sociology, statistics, and operations research [2]. As shown in Figure 2.4, human factors specialists consider the basic human variables as well as the interfaces between the human, the machine, and the environment under which the human-machine system must operate. Human capabilities and limitations, the manner by which humans process and respond to information, and the effect of the environment on overall system performance are important.

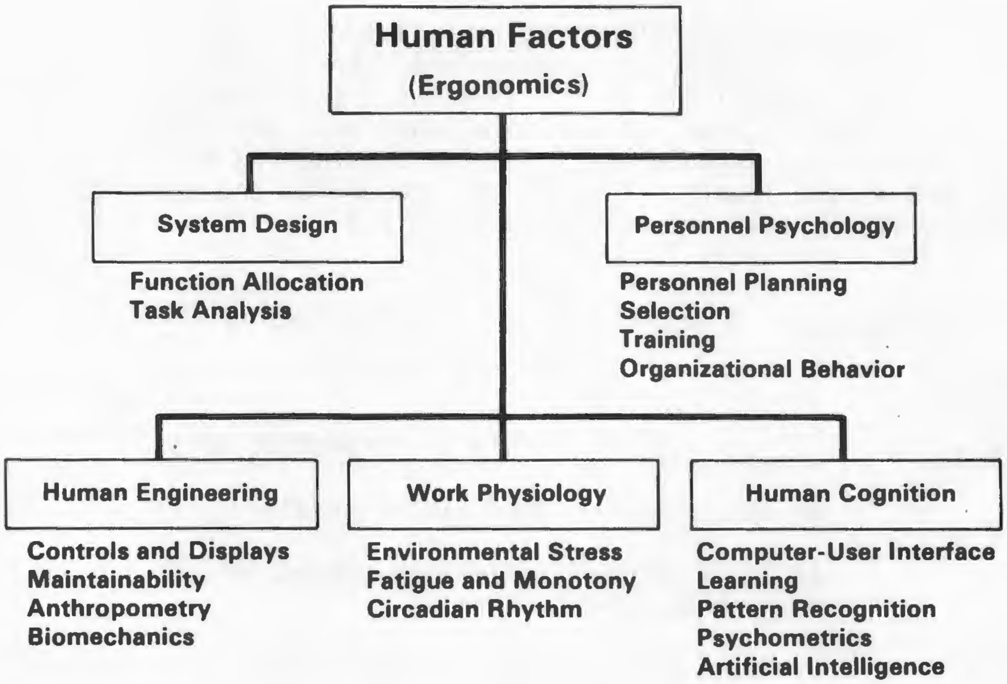


FIGURE 2.4. Five Major Areas of Specialization within the Human Factors Discipline

The scope of the human factors evaluation of the MRR was [3]:

- To acquire preliminary data with which to evaluate the performance-shaping factors that affect the reliability of the UT/ISI process.
- To test the efficacy of the relative operating characteristic (ROC) analysis technique for measuring the reliability and accuracy of UT/ISI results.
- To gain insight regarding needed direction and focus of future research and regulatory efforts in the human factors-NDT area.

Data from the human factors study provided preliminary answers to the following questions:

- How do organizational and supervisory characteristics of the job affect UT performance?
- How do IGSCC training, performance demonstrations, and field experience affect UT performance?
- How does equipment design affect UT performance?

- How does the inspection environment affect UT performance?
- How is UT/ISI conducted to detect IGSCC?
- Is it feasible to use Relative Operating Characteristic (ROC) analysis on ultrasonic systems?

During the human factors study, each inspector was observed as he was examining pipe and occasionally queried as to his rationale for decision criteria (i.e., why does that particular signal on the instrument display warrant attention?). Each inspector also participated in individual, structured interviews, which focused on their experiences in UT/ISI (see Appendix A). Each subject also filled out a survey questionnaire dealing with performance-shaping factors in UT/ISI (see Appendix B). In addition, human factors specialists evaluated the UT equipment for compliance with established human factors principles using a specially prepared equipment evaluation checklist (see Appendix C). The detailed analysis for the human factors evaluation checklist data is found in Chapter 5.

3.0 MEASUREMENT OF INSPECTION RELIABILITY

This chapter provides the analysis methodology used in this study. The first section provides a definition of the terms and measurement parameters used to quantify performance and to graphically display the data in Relative Operating Characteristic (ROC) curves. The second section provides details on the influence of several critical variables that were found to affect the study and analysis of MRR results. These factors include the effect of grading unit size, location of blank grading units with respect to cracked areas, and features of the specimen geometry that caused false calls.

3.1 DESCRIPTION OF FALSE CALL PROBABILITY (FCP), PROBABILITY OF DETECTION (POD), AND RELATIVE OPERATING CHARACTERISTIC (ROC) CURVES

3.1.1 Definition of FCP and POD

Each inspector's detection capabilities are estimated with the use of two statistics, which are defined by:

$$POD = \frac{\text{Number of Correct Detections}}{\text{Number of Cracked Grading Units Inspected}}$$

$$FCP = \frac{\text{Number of Blank Grading Units Called Cracked}}{\text{Number of Blank Grading Units Inspected}}$$

The round robin was constructed to allow these two statistics to be estimated efficiently for different inspection conditions. They are employed in Section 4.0 to examine inspection performance.

These two statistics always refer to a unit of material of particular size, which we call a "grading unit." Therefore, individual detection and false call statistics are always associated with a unique "grading unit." For a particular grading unit, its detection statistic is 1 if a crack was called within the unit and 0 if no crack was called. Individual grading units were separated by a minimum of one inch of blank weld material in the circumferential direction with no separation axially for grading units on opposite sides of the weld, and all the grading units have the same circumferential length. Grading units are randomly located within the test specimens, but are not actually marked on the outside of the specimen so that the inspectors do not know their location. See Appendix E which shows all the grading units on each MRR specimen.

Each grading unit was classified as "defective" if it contained a crack or as "blank" if it was not cracked. A true-state data base describing defective and blank grading units was developed for each test specimen used in the MRR.

3.1.2 Classification of Results

MRR inspections were scored by comparing each inspector's reported results with the true-state grading unit data base. To accomplish this, each inspector was required to classify any indications he found according to the 5-item scale listed in Section 2.1.3, Test Protocol. This 5-item indication scale allows the grading units to be categorized into a 6-item grading unit scale shown in Table 2.4.

Figure 3.1 illustrates an example of an MRR test sample that has been divided into four grading units. Grading unit 202 contains a crack (identified as a triangle and with the letter "T" for true state), while grading units 201, 203, and 204 are blank. Notice that a separate grading unit has been defined on each side of the weld. We have found the weld itself to be a substantial barrier to inspection in wrought stainless steel. It is therefore necessary to define grading units in this manner to allow near- and far-side effects to be measured. The grading units are 3-in. wide, while the test specimen is approximately 8-in. wide.

Inspection results are also displayed in Figure 3.1. The two indications marked "C-1" result from the inspection. The inspector therefore recorded two indications during the inspection and classified both of them as "C-1" indications, signifying that he was very certain that these indications were associated with cracks. This inspection was carried out from the "B" side of the specimen (i.e., negative axial direction from the weld centerline); hence, the inspection was near side for grading units 202 and 204 but far side for grading units 201 and 203.

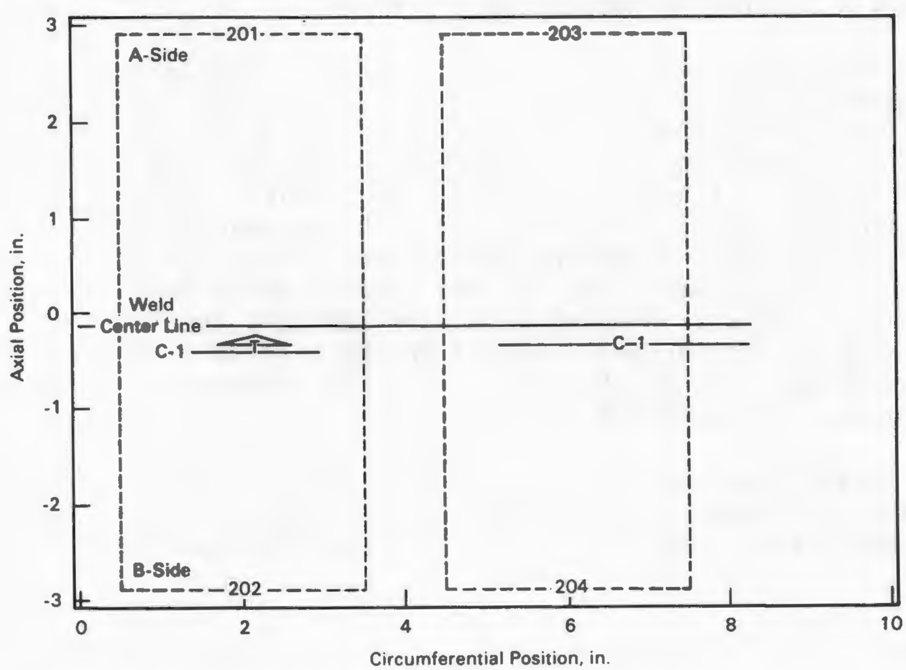


FIGURE 3.1. Scoring Example showing the Location of Cracks (Δ), Crack Calls ——C-1——, and Grading Units (201, 202, 203, and 204)

Since the inspector placed a C-1 indication within grading unit 202, it is classified as C-1 for this inspection. The grading unit 204 is also classified as C-1 for the inspection. Grading units 201 and 203 are classified as N-6 because no indications were recorded within them. Hence, the inspector would be credited with 1 correct detection (G.U. 202) and 1 false call (G.U. 204).

In this study, the majority of the analysis will be performed using 3-in. (circumferential length) grading units as illustrated in Figure 3.1. The axial length of the grading unit is not very important since the defects occur near the weld and typically a value of 3 in. was employed. However, certain analyses that employ 1-, 2-, and 8-in. grading units will also be used (see Section 3.2) to investigate the effect of grading unit size and to handle long cracks.

3.2 ROC CURVES

3.2.1 Explanation of ROC Categories and Model

The ROC model provides a description of how the information assembled during the inspection is transformed into a crack/no-crack decision. In this model, it is assumed that all results from a grading unit can be summarized into a single "signal" z, and this single value is used to obtain the decision by comparing z to a threshold. The distribution of z is assumed to depend on the size of the flaw present in the grading unit; when no flaw is present, z has a distribution given by $f_o(z)$ and when a flaw of size s is present, the distribution of z is defined to be $F_s(z)$.

Figure 3.2 illustrates the model graphically. Given $f_o(z)$ and $f_s(z)$, it is possible to calculate a false call probability and probability of detection associated with the threshold T. The threshold T can be varied to produce the curve traced out by:

$$(FCP(T),POD(T)) = (F_o(T),1-F_s(T))$$

and this happens to be the definition of an ROC curve. Since FCP and POD are the primary measures of detection performance, this curve summarizes the potential capabilities of the system; hence, the name "Relative Operating Characteristics" curve.

Although the concept of an ROC curve requires the notion of a signal z and a threshold T, it is not necessary to explicitly define or measure these values to determine an ROC curve. All that is required are a few (FCP,POD) points on the curve. This is generally how ROC curve fitting proceeds; there is no attempt to directly measure the signal z. An experiment is performed on the system in which the decision criteria is varied (from more to less stringent).

In fact, the ROC model is commonly applied to diagnostic procedures that involve a good deal of expert judgement to come up with the final decision (see Swets). The justification for using the signal detection model is simply that it provides a good analog to what is going on in the expert's head.

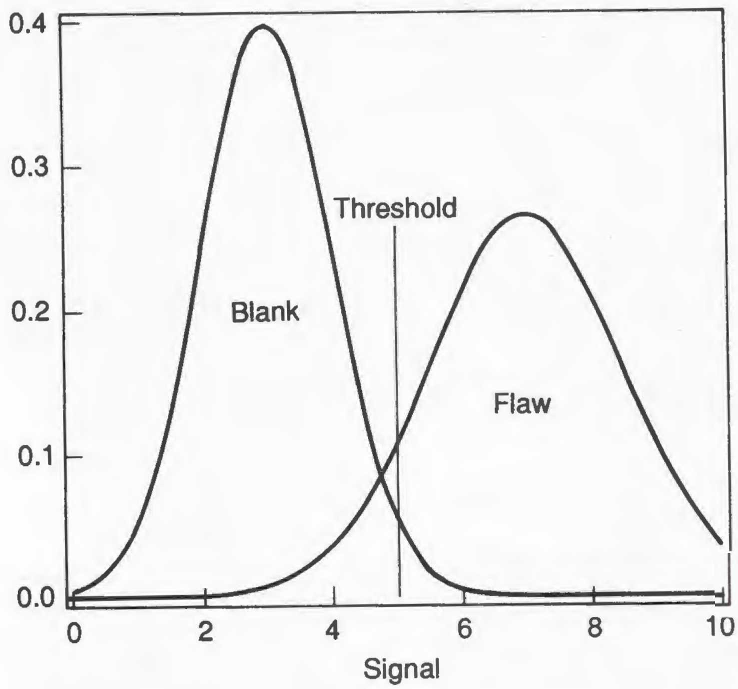


FIGURE 3.2. ROC Signal Detection Model

In the case of flaw detection with ultrasonic methods, there is a more specific reason to utilize a signal detection model. Amplitude drop procedures do in fact use a signal to make the crack/no-crack decision. The simplest amplitude drop procedure will call a crack when the maximum amplitude (as adjusted by the DAC curve) is above a certain threshold.

In the present round robin, we have utilized six thresholds for ROC analysis which, in increasing level of severity, are designated N6, N5, N4, C3, C2, and C1. All but the N6 category summarize the inspector's opinion of the state of the grading unit. (The N6 category signifies that no reportable indications were received for the inspector to classify.) The N5 category identifies an indication that the inspector is most certain is not a crack, while the C1 category is reserved for an indication that the inspector is most certain is a crack. Since multiple indications might occur in a single grading unit, the whole grading unit is categorized according to the most severe category present in it.

The letter portion of the threshold code (i.e., "N" or "C") is used to mark a very important point in the threshold scale. All indications marked "C3" and above represent indications that the inspectors would have called as cracks in the field, while those labeled as N5 and N4 are indications that they would not have called cracks. Thus, the C3 threshold is the threshold used to compute "in-field" false call probability and probability of detection. The other points on the threshold scale can be used to determine what would happen if the inspector's decision criteria were loosened or tightened up.

Thus the N4 and N5 points on the threshold scale might represent his performance if he were exhorted to "do a better job" and find more cracks, while the C2 and C1 point might represent his performance if he were exhorted to "do a better job" and make less false calls.

3.2.2 Estimating the Shape of ROC Curves

The data used to develop an ROC curve can be summarized as illustrated in Table 3.1. The counts in this table describe the number of grading-unit inspections that fall into each decision category. A grading unit not called anything by an inspector is automatically placed in the category (N6).

TABLE 3.1. Example of Inspection Data Required to Develop ROC Curves

Classification Scale	Blank Grading Units			Cracked Grading Units		
	Number Called	Cumulative Sum	FCP	Number Called	Cumulative Sum	POD
C1	22	22	0.11	74	74	0.49
C2	11	33	0.17	18	92	0.61
C3	9	42	0.21	17	109	0.73
N4	7	49	0.25	5	114	0.76
N5	17	66	0.33	12	126	0.84
N6	132	198	1.00	24	150	1.00

These counts can then be used to obtain (FCP,POD) estimates for each individual classification (the calculation is a cumulative distribution). For example, the FCP estimate associated with the C3 criterion is computed by:

$$FCP(C3) = \frac{22 + 11 + 9}{198} = 0.21$$

The (FCP,POD) estimates are assumed to lie upon a curve whose form is given by:

$$POD = F[B_0 + B_1 F^{-1} (FCP)]$$

where: $F(x) = \frac{1}{1 + \exp(-x)}$

$x = C1, C2, etc.$

The function F(x) is sometimes called a logistic function. The unknown parameters (B_0,B_1) determine the shape of the ROC curve and are determined by the regression fit to the data. The motivation for using such a model to represent ROC originates from well-established signal detection theory. If the ultrasonic signals from blank material are distributed with a mean of 0 while the signals from cracked material have a mean of $-B_0/B_1$; and furthermore, if the ratio of the standard deviation of blank to cracked signals is B_1, then the above relationship holds.

The model parameters are estimated from the (FCP,POD) data presented in Table 3.1 using an iteratively re-weighted regression procedure that produces the maximum likelihood estimates for the above model. The fit also produces the asymptotic covariance of the parameter estimates, which allows curve confidence bounds and hypothesis tests to be constructed.

For the example presented in Table 3.1, the fitting procedure produced the results:

$$B_0 = 1.94$$
$$B_1 = 1.04$$
$$\text{Cov}(B_0,B_1) = \begin{bmatrix} 0.09 & 0.04 \\ 0.04 & 0.03 \end{bmatrix}$$

The associated confidence bounds are displayed in Figure 3.3. The calculated curve is the same ROC curve presented in Figure 3.1 for near-side inspections on short cracks. For a more complete description of the maximum likelihood fitting procedure, see Swets [4].

3.2.3 ROC Curve

The scale introduced in Table 2.4 allows a different set of (FCP,POD) statistics to be calculated for each state listed in the scale. These different sets of (FCP,POD) statistics allow us to examine the behavior of (FCP,POD) as the inspector changes his decision criteria based on his confidence in each decision. These (FCP,POD) pairs can be fit to a curve, which is called a relative operating characteristic (ROC) curve and quantifies the relationship between detection performance and decision criteria [2].

Figure 3.4 shows two typical ROC curves; probability of detection (POD) is plotted against false call probability (FCP). The actual (FCP,POD) data points calculated from the inspections are denoted in the plot by their associated decision criteria (i.e., "C1," "C2," etc.). ROC curves have been fitted to these (FCP,POD) data points using iteratively re-weighted regression procedures. A detailed explanation for computing the ROC points and in fitting the ROC curve is contained in this section.

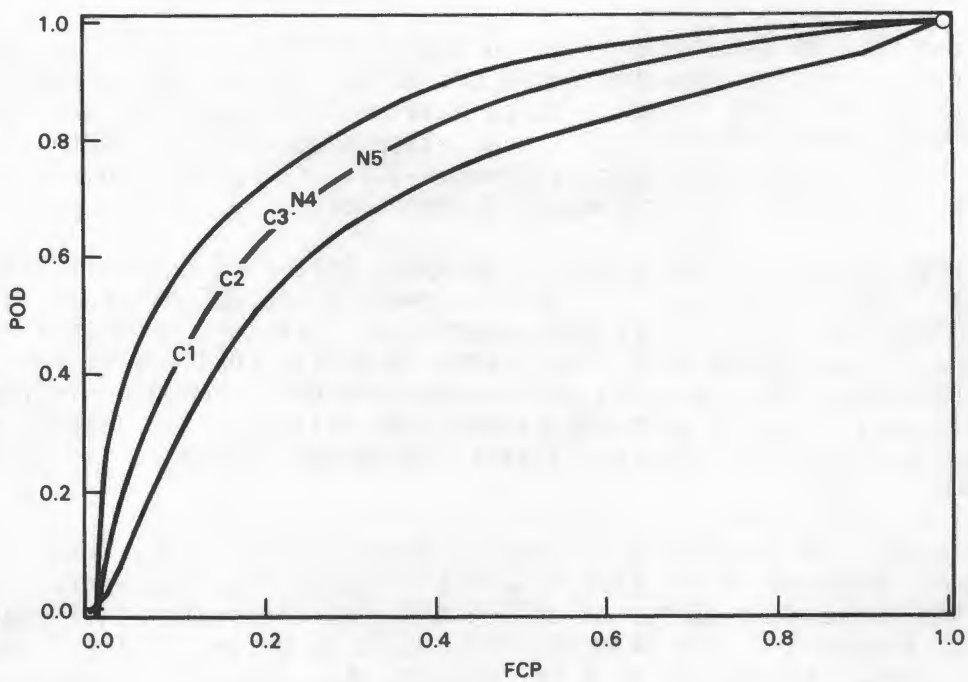


FIGURE 3.3. Example Fit of ROC Curve with 95% Confidence Bounds

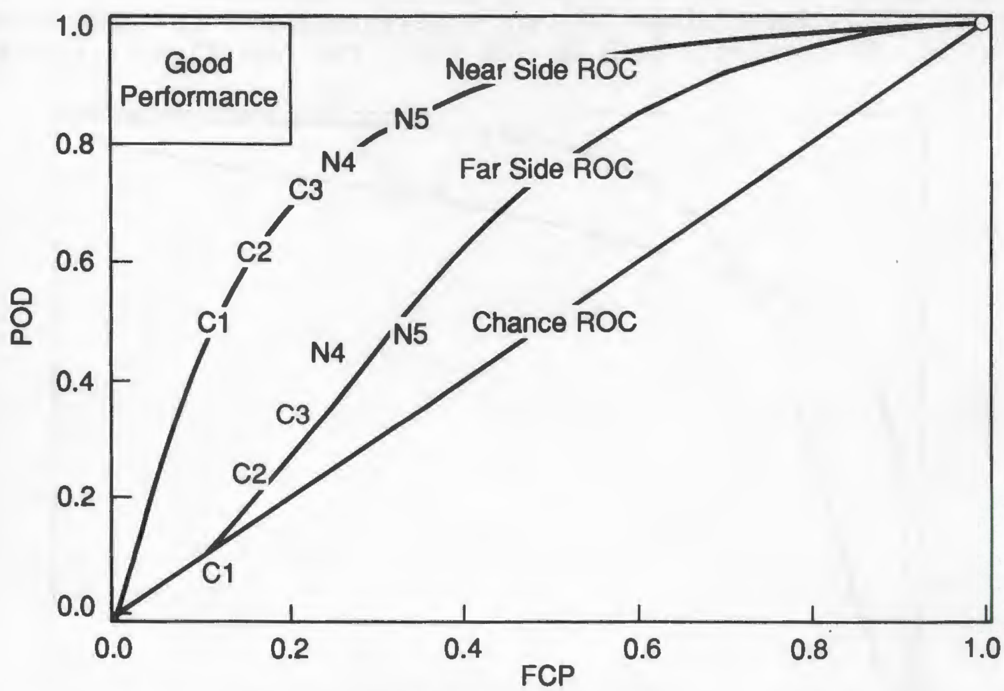


FIGURE 3.4. Example of Typical ROC Curves showing Better Inspection Performance for Near-Side Inspection versus Far-Side Inspection

The two sets of data points are an average (of all teams) for all near- and far-side inspection results on short cracks. Notice that the inspection performance for far side is not particularly good, because the ROC curve is close to the diagonal "Chance ROC" line. (The diagonal line represents performance that can be achieved by chance--i.e., flipping a coin.) On the other hand, the near-side performance is much better.

The best performers would be in the upper left-hand corner as indicated by the box drawn in the upper left-hand corner of the ROC diagram. Good inspection results will produce ROC curves that intersect with this box. Furthermore, a good inspection should intersect with this box at the C3 decision criterion, because this is the decision criterion used in the field (crack/no crack). Notice that the C3 decision criterion for near-side inspections displayed in Figure 3.4 does indeed come close to the "good performance" box.

ROC curves can frequently be used to unambiguously order team or procedure performance. Whenever a set of ROC curves is monotonically ordered, then the "larger" ROC curve must always be associated with the better performance. In the case of Figure 3.4, for example, the near-side curve is always above the far-side. Thus, it would be safe to conclude that near-side inspection performance is better than far-side performance.

Figure 3.5 illustrates a set of ROC curves that cannot be ordered in such an obvious manner. In this figure, ROC A is superior to B when the false call probability is larger than 20%, while performance is switched when FCP is less than 20%. To compare A to B in this case, the investigator is required

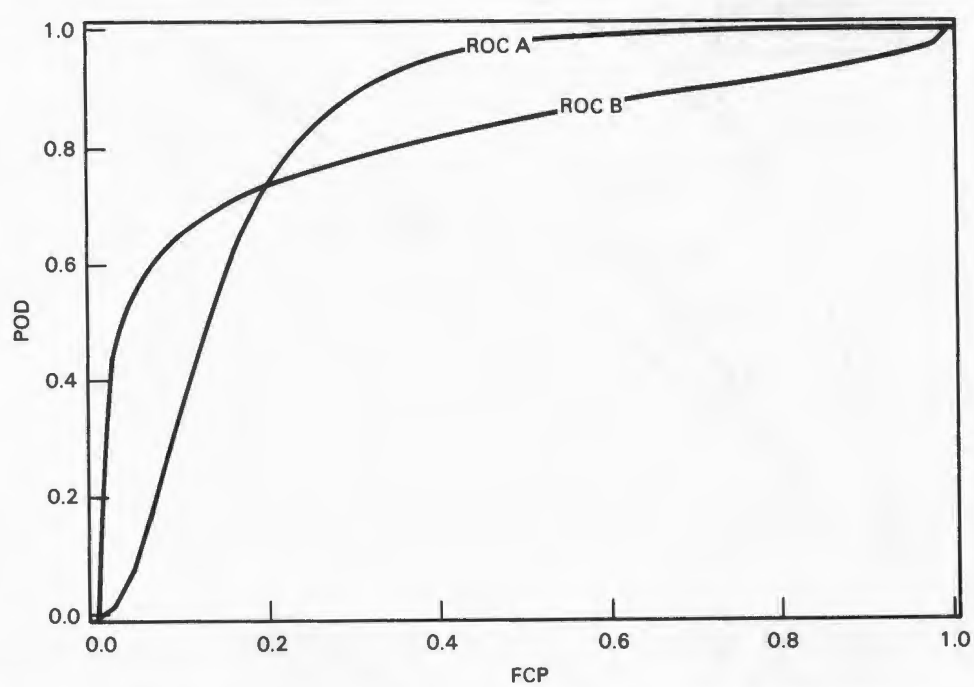


FIGURE 3.5. Example of Ambiguous Curves (i.e., comparing performance of ROC-A and ROC-B is difficult)

to be specific about the relative costs/benefits of false call versus detection performance. Fortunately, the situation illustrated in Figure 3.5 does not frequently occur so that ROC analysis can frequently be used to unambiguously compare results.

One final comment concerning the role ROC curves play in evaluating detection performance. ROC curves will be employed in this analysis when we want to examine a team's (or procedure's) INHERENT detection capability. Inherent detection capability is the performance (FCP,POD) that would be achieved if the "optimal" decision criteria were being employed. Of course, in actual field inspections, a procedure with excellent inherent capabilities may still yield poor results because a poor decision criterion is being employed. A procedure with good inherent capabilities may also yield poor field results because the decision criterion cannot be controlled. (It may vary from inspection to inspection and/or team to team.)

Consequently, ROC curves are not the only statistic of interest; it is also important to determine (FCP,POD) performance associated with the actual crack/no crack decision criterion (C3) for extrapolation to field inspections. In fact the most important single summary of inspection performance is the (FCP,POD) point associated with this decision criterion. An important part of the analysis will therefore be concerned with what variables influence this pair of statistics.

3.2.4 Weaknesses with the ROC Model

The biggest weakness associated with the ROC analysis is that ONE signal z is used to make the decision. Generally, an inspection can produce literally millions of values, and a good decision procedure may require more than one variable to make a reasonable decision. In other words, the "signal" that the decision procedure uses may be multi-dimensional.

A second weakness associated with ROC analysis concerns the exact form of the distributions $f_0(z)$ and $f_S(z)$. In order to produce an ROC curve, the shape of these distributions have to be known. For the present work, we have assumed that both distributions are normal, the most popular distribution used in ROC analysis. However, if this assumption were changed, the shape of the fitted ROC curves would change.

For example, the fact that ROC curves always go through (0,0) and (1,1) on the ROC diagram is a consequence of the normal distribution assumption. If distributions of finite support were used (i.e., distributions that concentrated all their mass in a finite interval), then the ROC curves would not necessarily go through these points.

3.3 PARAMETERS AFFECTING MEASUREMENT OF DETECTION PERFORMANCE

3.3.1 Effect of Grading Unit on MRR Detection Results

Since the concept of a grading unit is fundamental to analyzing results obtained from the MRR experiment, it is important to know how sensitive the results are to the way grading units have been defined (particularly their

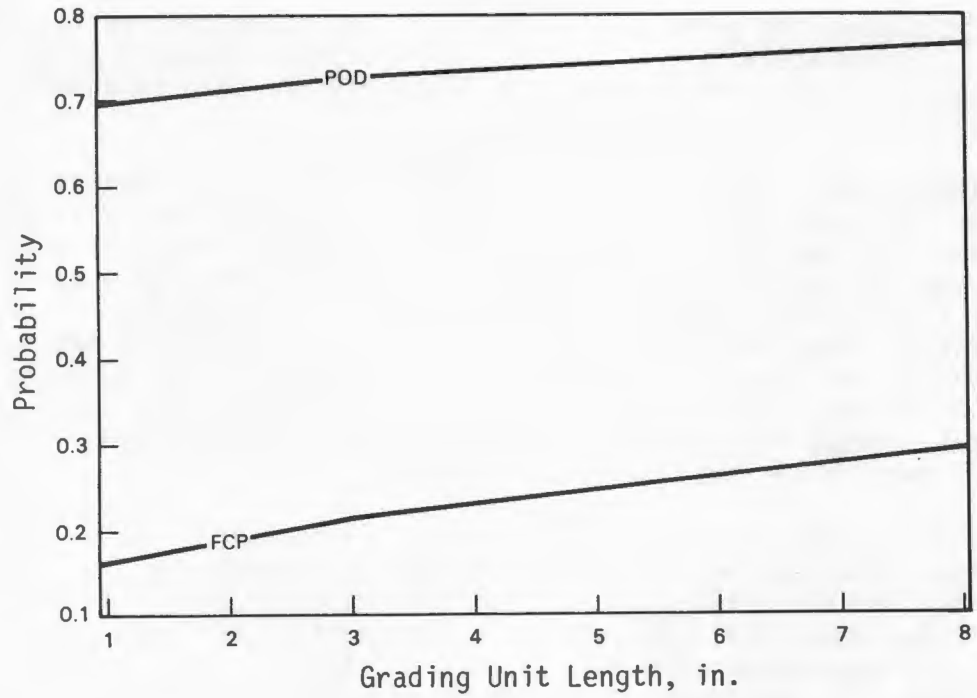


FIGURE 3.6. Effect of Grading Unit Size on FCP and POD

size). In this section, the effect of grading unit size on (FCP,POD) statistics is examined. In general, we would expect both FCP and POD to increase as the size of the grading unit is increased. Figure 3.6 illustrates this effect using all near-side inspections on short cracks with grading unit sizes of 1, 2, 3, and 8 inches (the 8-in. grading units are essentially whole specimens).

The grading unit size does not have a dramatic effect on the POD plotted in Figure 3.6. However, the situation is different for blank grading units (or grading units with small cracks in them). For example, the FCP calculated from 8-in. grading units is approximately double that of a 1-in. unit, although these numbers are small and not unrealistic. Of course, if even larger grading units could have been included in the experiment, we could expect to see an even larger size effect.

Since grading unit size can affect POD and FCP differently, it is important to be aware of the grading unit size employed to define these statistics. For reasons of economy, we have selected a 3-in. grading unit for all short-crack analysis. Thus, the obtained FCP and POD statistics are only relevant to these particular units of material. To determine the performance on different lengths of material, some extrapolating curves, such as those illustrated in Figure 3.6, must be employed.

The effect of grading unit size is less pronounced on ROC curves. Figure 3.7 displays ROC curves calculated from grading units of 1-, 2-, 3-, and 8-in. circumferential lengths (the data consists of near-side inspections of short cracks). Notice the shape of all ROC curves is fairly close, even though individual (FCP,POD) points for the different grading units may not be. Indicated on the plot are the (FCP,POD) points associated with the C3 decision criterion. These points roughly trace out the ROC curves beginning with (.20, .60) for a 1-in. grading unit to (.32, .70) for an 8-in. unit.

A few specimens in the MRR contained long IGSC cracks. These cracks averaged 6 in. in length and were, therefore, too long to fit into a 3-in. grading unit. Because of this, all long crack inspection data was analyzed using whole-specimen grading units. Figure 3.8 shows that there is also little effect of grading unit size on the shape of the ROC curve for long cracks.

It is hard to compare MRR results directly with other studies, since FCP and POD statistics are sensitive to grading unit size. For example, as noted above, we selected a 3-in. grading unit while testing performed under IEB 83-02 was based on a 1-in. grading unit. Furthermore, the functional dependency for POD and FCP on grading unit size is different, resulting in further complications. These relationships will need to be accounted for when results from different studies are compared.

3.3.2 Use of Blank Grading Units

Blank grading units are employed to estimate an FCP that would be applicable to in-field inspections of defect-free welds. Although the blank grading units defined for this analysis never contain defects, they may be located on specimens that do contain cracks. Since blank grading units may be "near to" cracks, the recorded false calls may be influenced by these nearby cracks and the estimated false call probability may be biased.

To investigate this possibility, blank grading units were classified into two categories: isolated and contaminated blanks. An isolated blank grading unit was at least 2-in. away from any crack (on both sides of the weld centerline). Contaminated blanks consisted of all grading units that did not meet the isolated blank requirements. Contaminated blank grading units consisted of blank grading units where a crack existed on the far (opposite) side of the weld.

Table 3.2 presents FCP values calculated using both isolated and contaminated blank grading units for 3-in. grading units.

The table displays some significant differences between isolated and contaminated FCP, particularly those associated with the near-side measurements. Since these values are used in computing the ROC curves, they will greatly change the shape of the lower region of the curve. It was, therefore, decided to employ only "isolated blanks" in the calculations of false call probability.

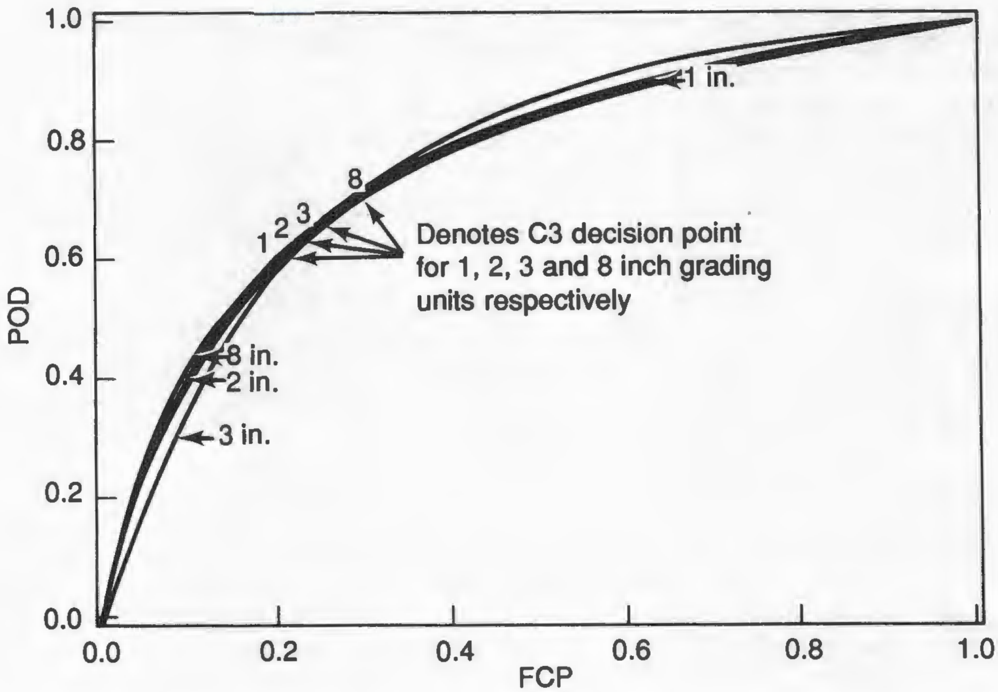


FIGURE 3.7. Effect of Grading Unit Size on ROC Curves for Near-Side Inspection

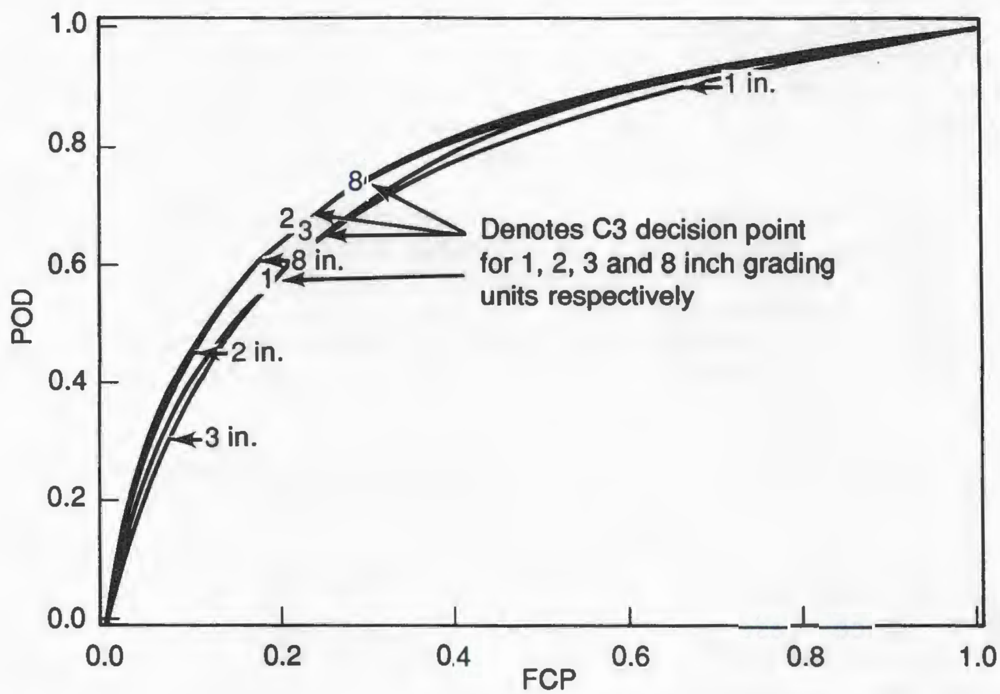


FIGURE 3.8. Effect of Grading Unit Size on ROC Results for Long Cracks

TABLE 3.2. FCP Values Calculated Using Isolated Versus Contaminated Blanks

Classification Scale	Isolated Blanks Near	Isolated Blanks Far	Contaminated Blanks Near	Contaminated Blanks Far
C1	0.11	0.11	0.22	0.10
C2	0.17	0.17	0.28	0.18
C3	0.21	0.25	0.36	0.26
N4	0.25	0.32	0.41	0.31
N5	0.33	0.38	0.47	0.38
Total Units Used	198	198	88	120

4.0 MRR INSPECTION RESULTS

This chapter provides the results for the detection and sizing quantification that was performed. The first section provides an analysis of the manual and advanced technique results for POD and FCP. The second section provides a comparison of the inspection performance for the four variables; manual or advanced technique, near- or far-side inspection access, short or long cracks, and pre- versus post-IEB 83-02 capability. The final section deals with the depth sizing capability of the inspectors who had passed the performance demonstration requirements of IEB 83-02 in place at the EPRI NDE Center for depth sizing of planar flaws.

4.1 MANUAL EXAMINATION RESULTS

Fifteen inspectors participated in the Mini-Round Robin. The inspectors were asked to follow the same procedures as those that they would use during ISI. Some of the procedures were extremely thorough in defining all steps for the inspectors to follow for the UT equipment being used. Others used the generic procedure developed by EPRI. The other three inspectors employed advanced inspection techniques. The MRR study was not constructed to provide very detailed descriptions of each individual inspector's performance. It must be noted that each inspector inspected a total of 23 quadrants and three whole 12-in.-dia. pipe specimens (approximately 21 feet of single-side access of weld length), so false call and probability of detection statistics will have large confidence bounds.

Nevertheless, it was possible to estimate ROC curves for each inspector by employing a restrictive version of the ROC model presented in Section 3.2. To fit ROC curves to individual inspector data, we assumed that $B_1 = 1$; this forces the ROC curve to be symmetric about the line POD = 1-FCP. Swets suggests using a model of this form when data is limited.

In Figures 4.1 through 4.15, the estimated ROC curves are displayed along with the (FCP,POD) data points used to produce the curve. The figures present ROC performance for near-side access on short cracks, the set of conditions that the experimental design measured most extensively in the round robin. Since different sized grading units were used for the short cracks versus the long cracks in the MRR, it is hard to collapse the data for all inspections in a meaningful way. All ROC curves are surrounded by 95% confidence bounds. The reader should be cautioned that these error bounds are most likely to be correct estimates of the true uncertainty for that portion of the ROC curve that is surrounded by actual data. For the extrapolative portions of the ROC curve, the correctness of the error bounds is heavily dependent on the appropriateness of the above choice of the ROC model.

These 15 ROC curves display variation in performance. The three best manual teams (6, 9, and 15) have ROC curves that intersect the FCP < 20% and POD > 80% corner of the ROC diagram (these teams are noted with an asterisk in Table 4.1). The three worst teams (2, 11, and 14) have ROC curves that are not significantly different (at the 95% level) than the POD=FCP chance line.

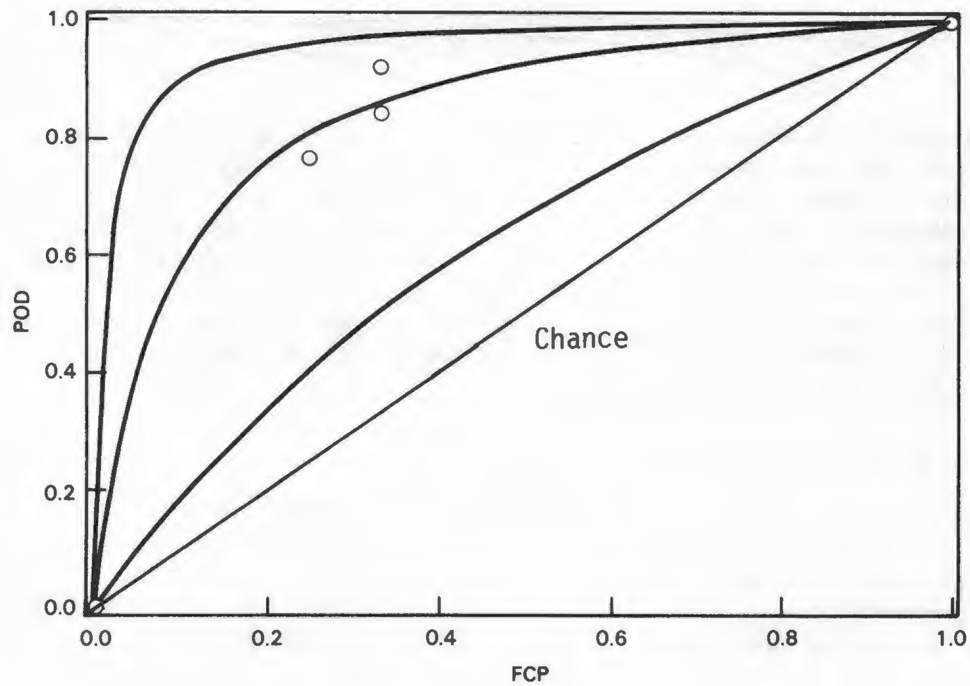


FIGURE 4.1. ROC Performance Curve for Inspector 1 for Near-Side Inspection of Short Cracks

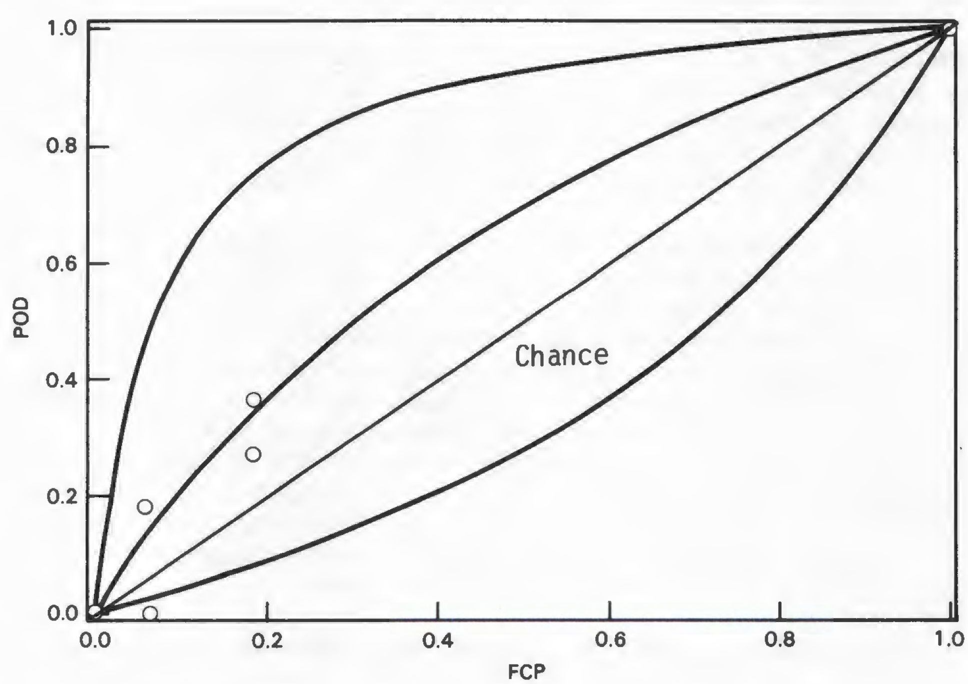


FIGURE 4.2. ROC Performance Curve for Inspector 2 for Near-Side Inspection of Short Cracks

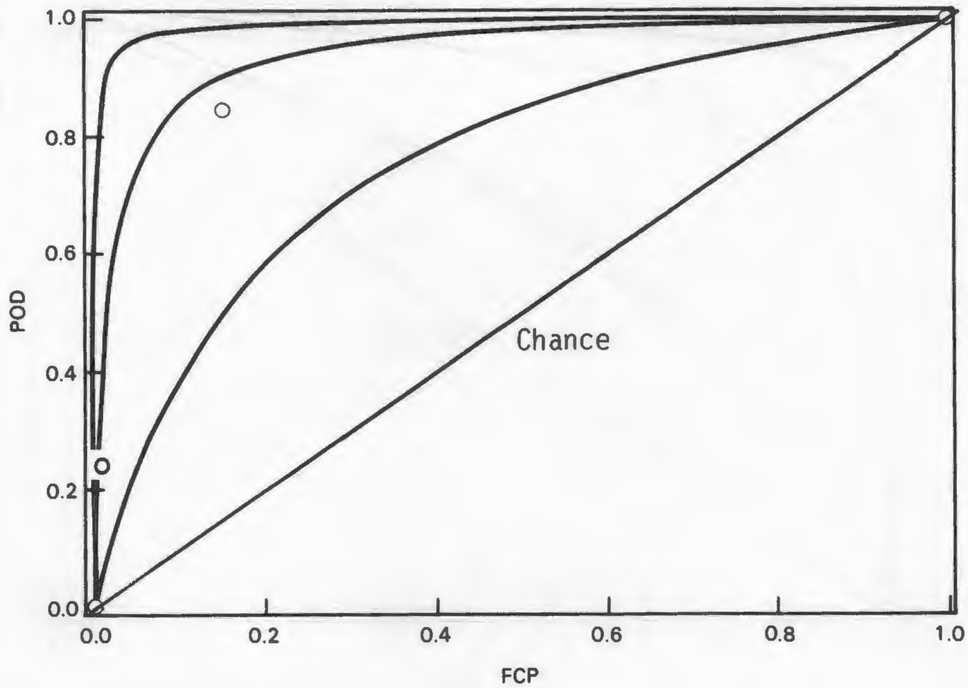


FIGURE 4.3. ROC Performance Curve for Inspector 3 for Near-Side Inspection of Short Cracks

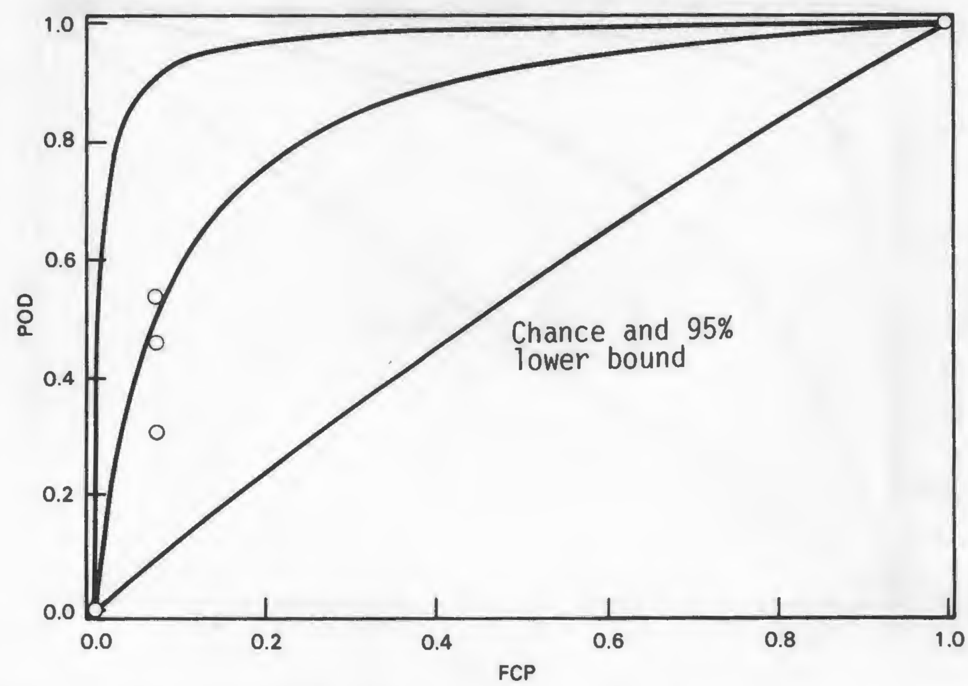


FIGURE 4.4. ROC Performance Curve for Inspector 4 for Near-Side Inspection of Short Cracks

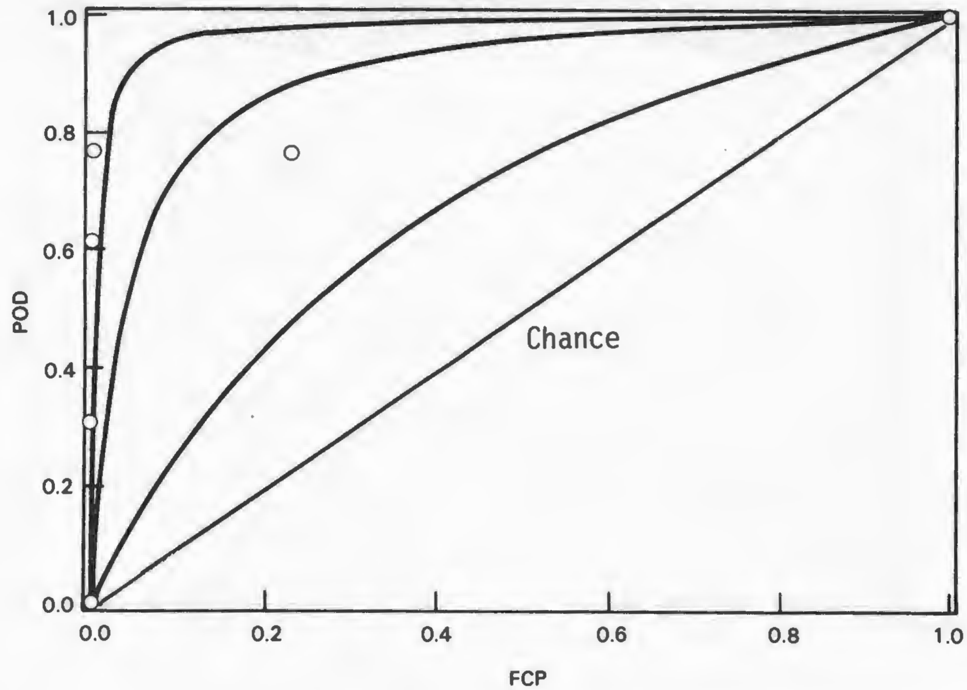


FIGURE 4.5. ROC Performance Curve for Inspector 5 for Near-Side Inspection of Short Cracks

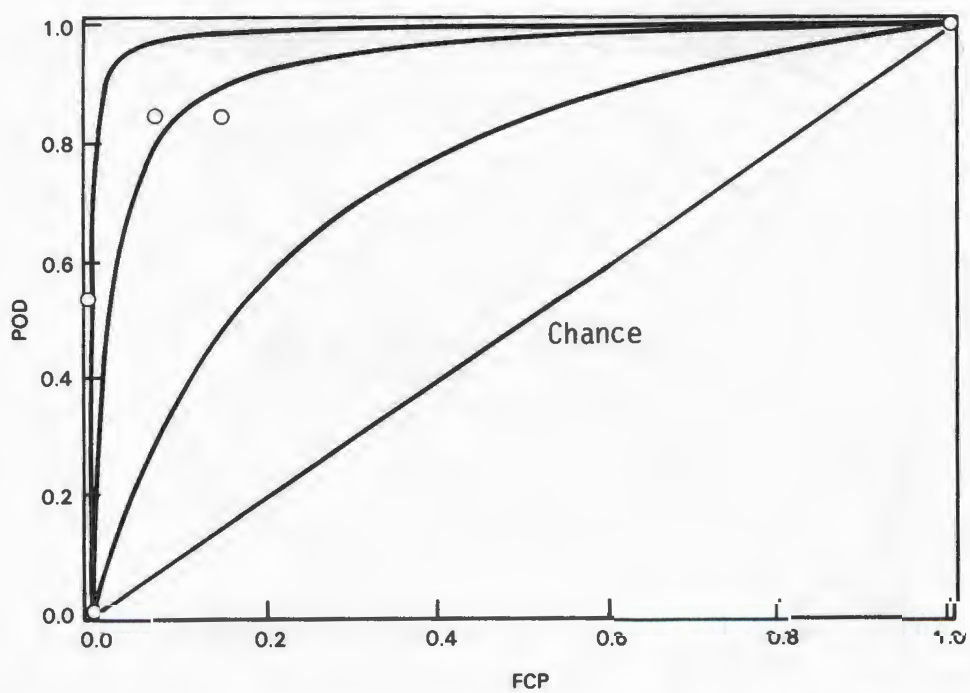


FIGURE 4.6. ROC Performance Curve for Inspector 6 for Near-Side Inspection of Short Cracks

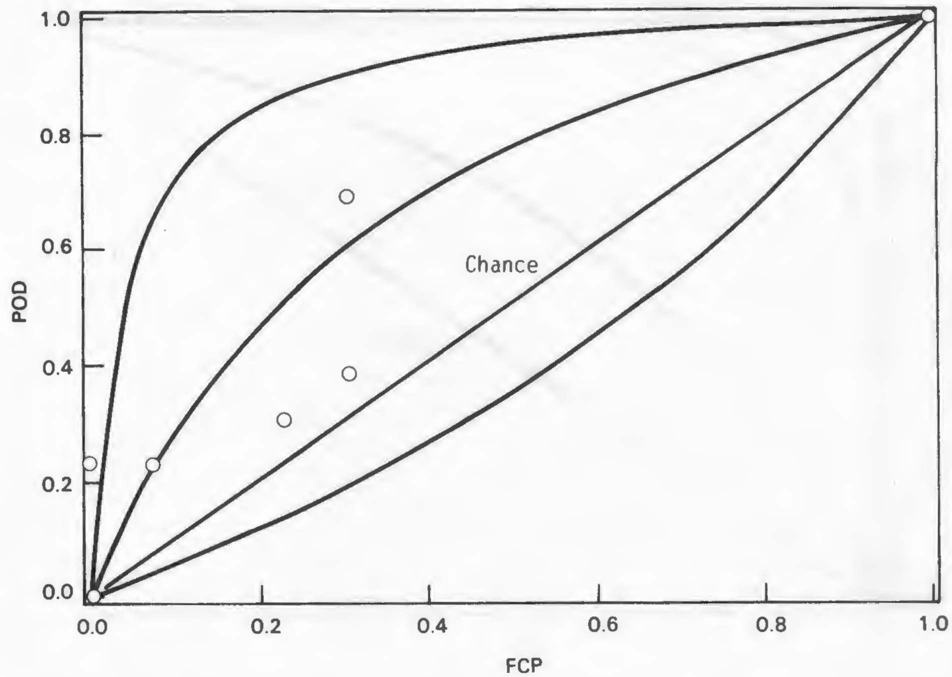


FIGURE 4.7. ROC Performance Curve for Inspector 7 for Near-Side Inspection of Short Cracks

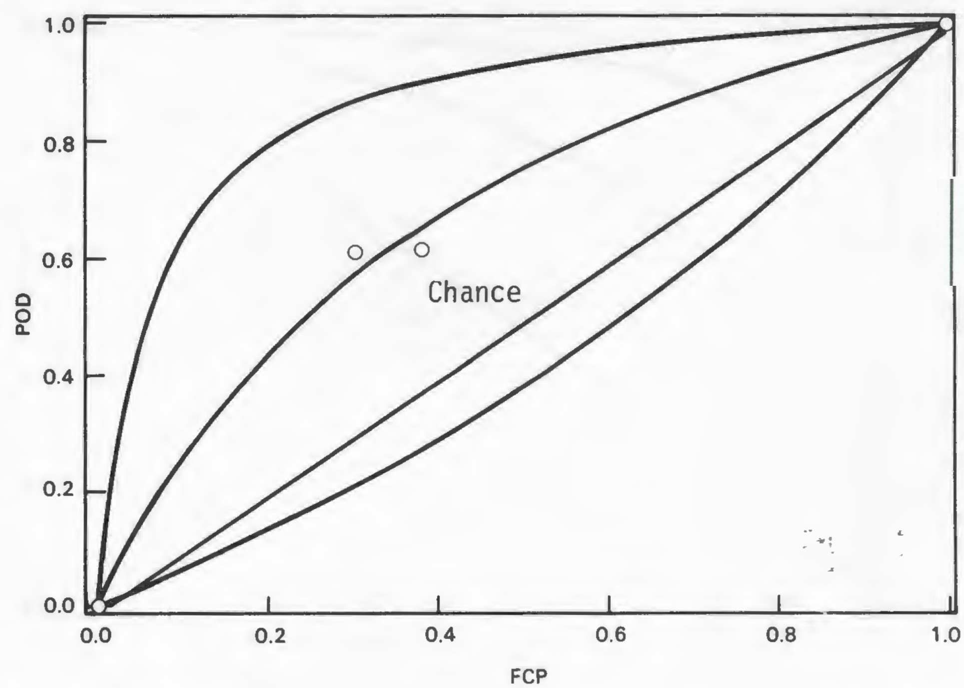


FIGURE 4.8. ROC Performance Curve for Inspector 8 for Near-Side Inspection of Short Cracks

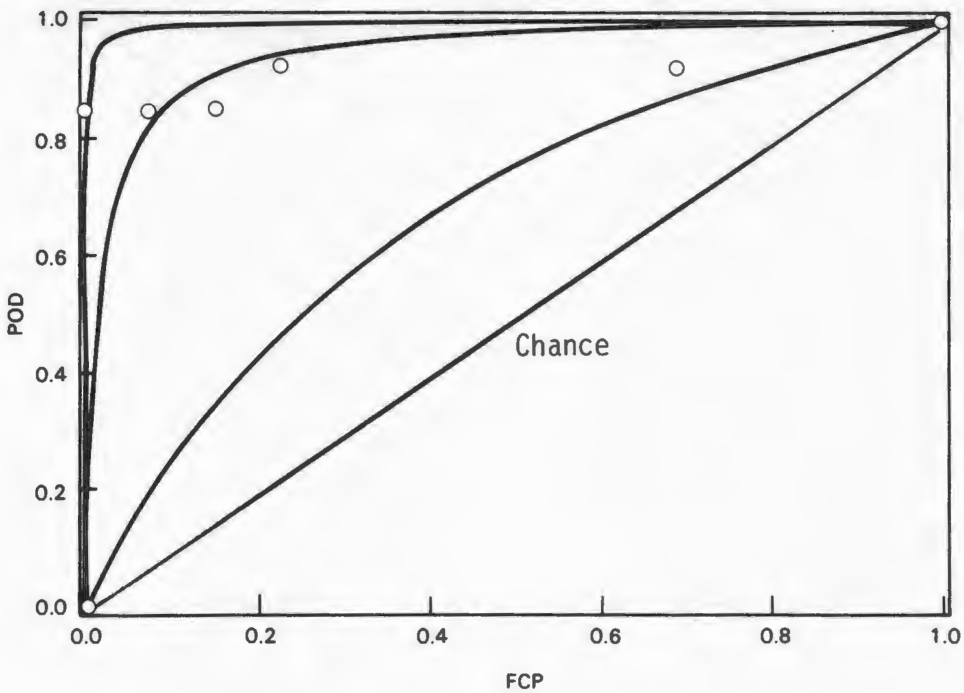


FIGURE 4.9. ROC Performance Curve for Inspector 9 for Near-Side Inspection of Short Cracks

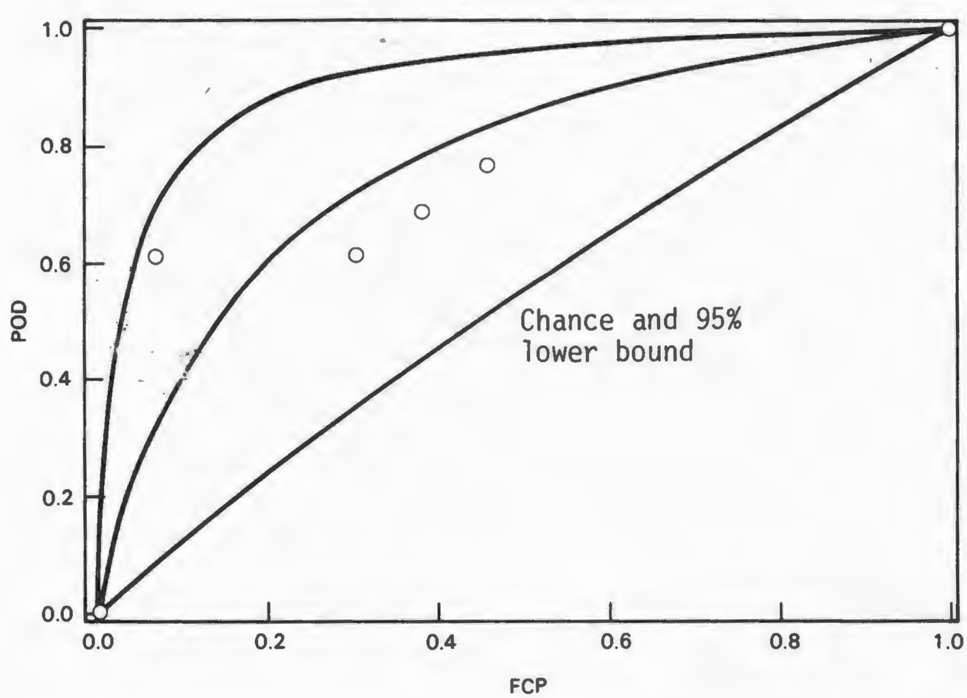


FIGURE 4.10. ROC Performance Curve for Inspector 10 for Near-Side Inspection of Short Cracks

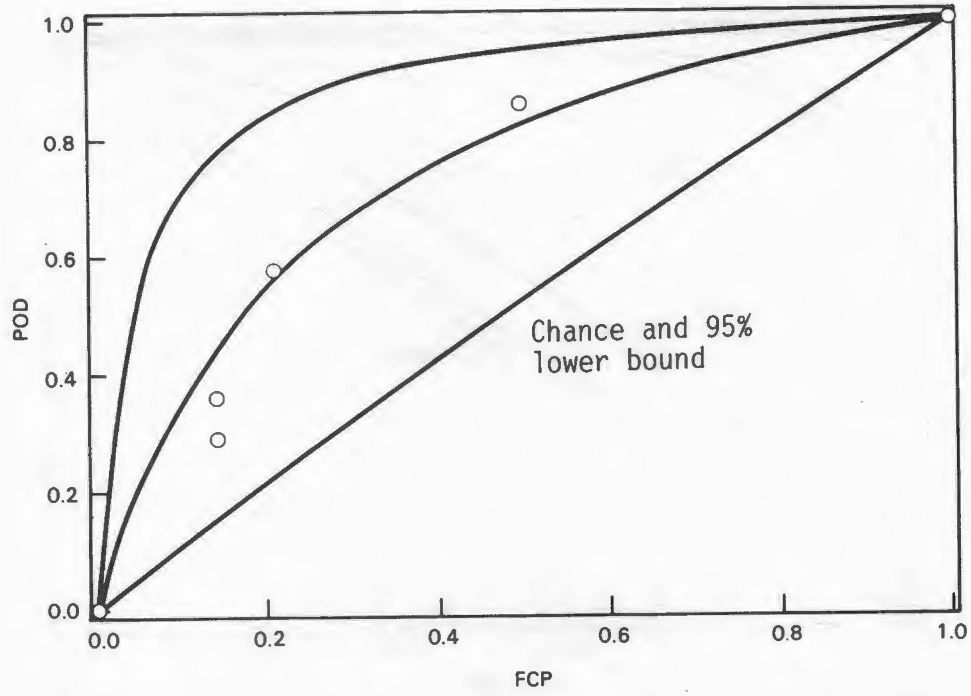


FIGURE 4.11. ROC Performance Curve for Inspector 11 for Near-Side Inspection of Short Cracks

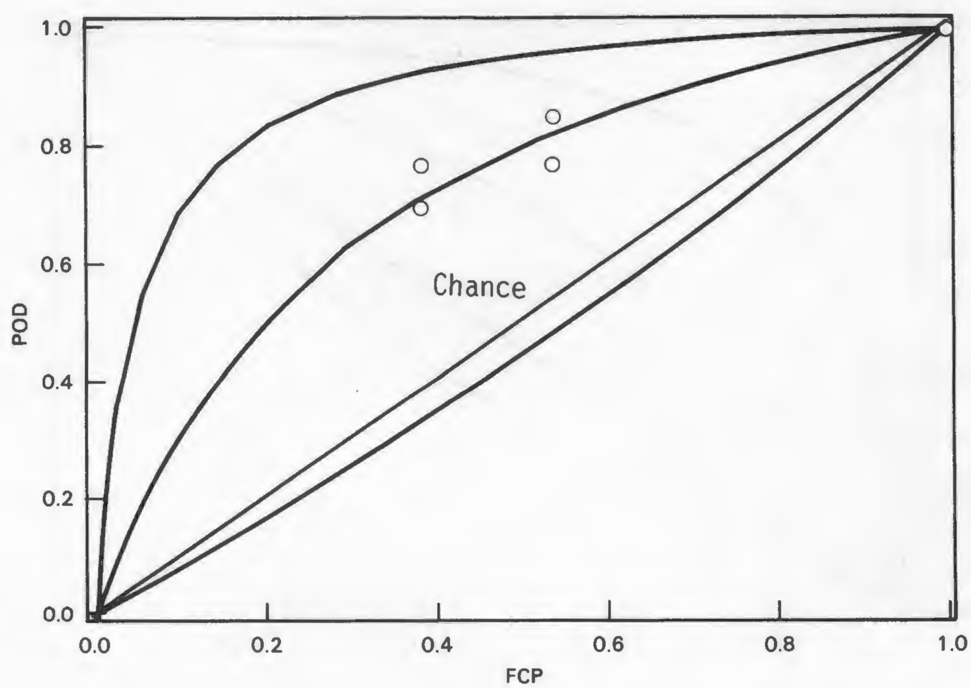


FIGURE 4.12. ROC Performance Curve for Inspector 12 for Near-Side Inspection of Short Cracks

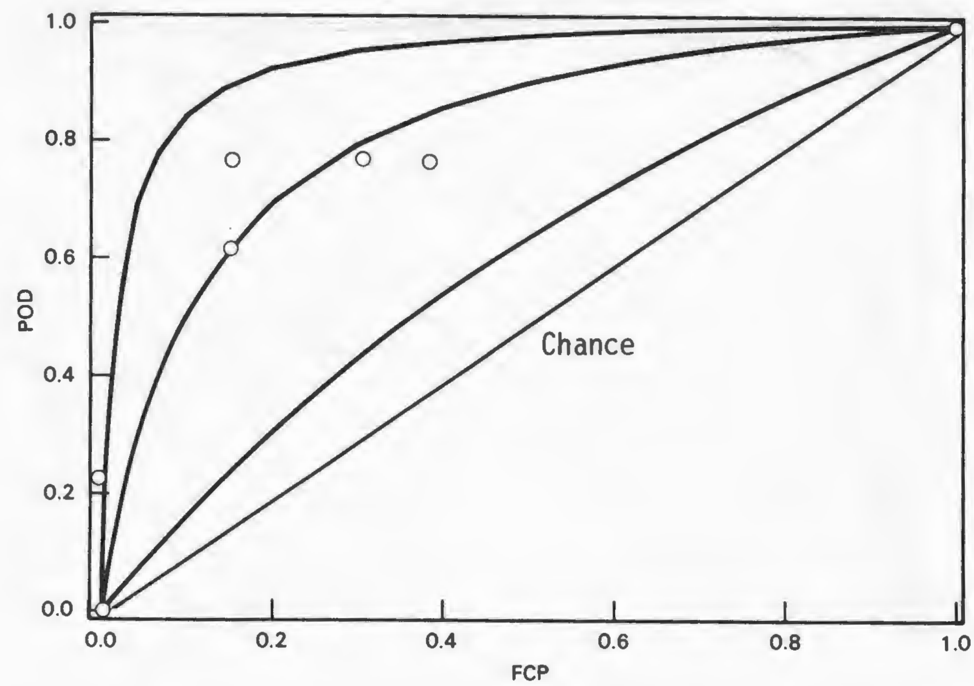


FIGURE 4.13. ROC Performance Curve for Inspector 13 for Near-Side Inspection of Short Cracks

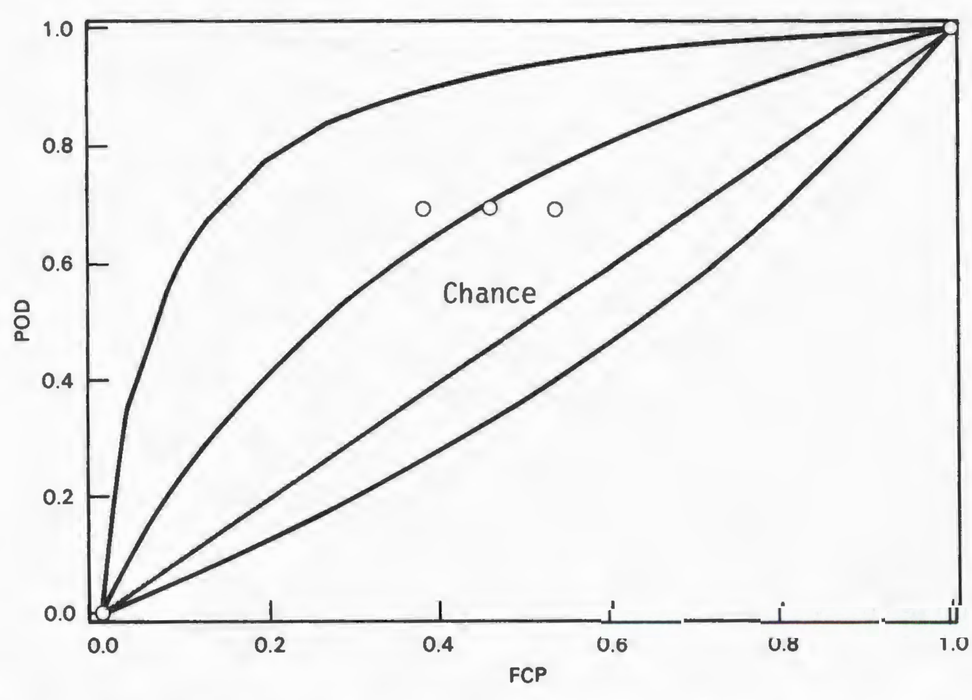


FIGURE 4.14. ROC Performance Curve for Inspector 14 for Near-Side Inspection of Short Cracks

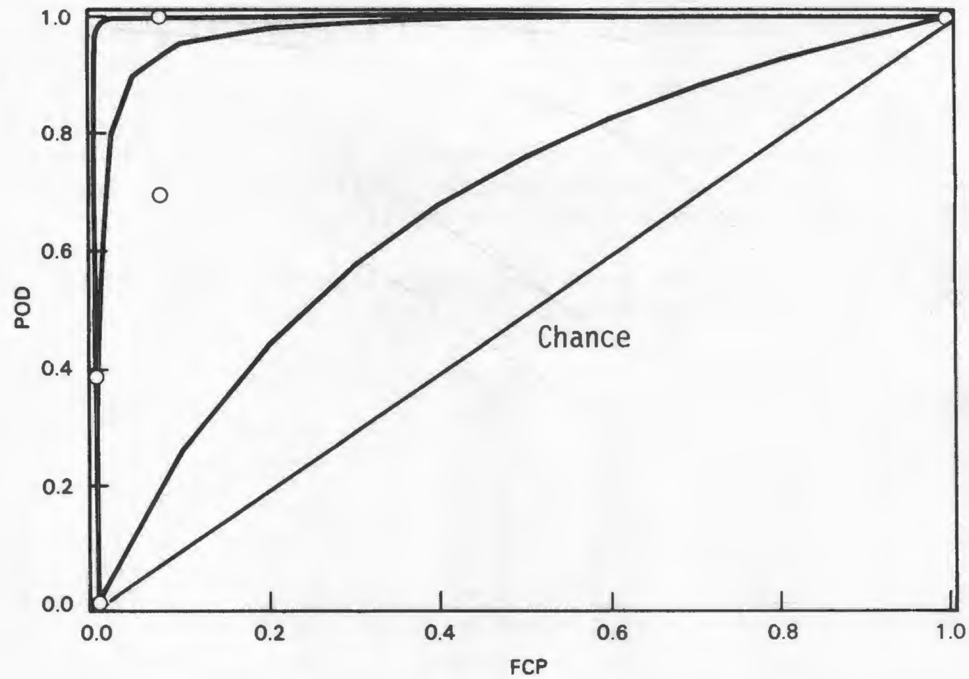


FIGURE 4.15. ROC Performance Curve for Inspector 15 for Near-Side Inspection of Short Cracks

The good performance curves estimated from the MRR data base show that good manual ultrasonic examination is possible. The wide variability of UT/ISI isillustrated by the wide variation between best and worst performance curves. This variability is shown later in Figure 4.16.

There are two potential causes of this team-to-team variability. First, there may be actual differences in team capability that are reflected in these ROC curve estimates. Secondly, a certain amount of the variability is due to experimental error. This experimental error is reflected in the size of the confidence bounds and from these plots, it is apparent that it is substantial. A critical question is therefore, are the "good" teams significantly better than the bad teams, or can all the variation displayed in these 15 plots be ascribed to experimental error?

To answer this question, the ROC curves were tested for equality. The test turned out to be highly significant; and therefore the displayed differences in team ROC curves are not entirely due to experimental error--some inspectors are "better" than others. If UT/ISI is to become a reliable tool for detecting and evaluating generic degradation of nuclear power plants, the reasons "why" some inspectors are better than others must be determined. This study did not address this question.

The ROC curves displayed in Figures 4.1 through 4.15 give a good overview of inherent capability. However, in order for an inspector to do a good job in the field, he must use correct decision criteria in conjunction with a good ROC curve. Tables 4.1 and 4.2 provide a summary of the estimated false call probability and probability of detection associated with crack/no crack classification (i.e., C3). Ninety percent confidence limits are given in both tables. Notice that the bounds are relatively large.

TABLE 4.1. Summary of Individual Near-Side Inspector Performance for Both Manual and Advanced Techniques

Inspector	Lower 90%	FCP	Upper 90%	Lower 90%	POD	Upper 90%
1	0.12	0.33	0.61	0.59	0.85	0.97
2	0.05	0.19	0.42	0.08	0.27	0.56
3	0.00	0.00	0.21	0.51	0.77	0.93
4	0.00	0.08	0.32	0.22	0.46	0.71
5	0.00	0.00	0.21	0.51	0.77	0.93
6*	0.03	0.15	0.41	0.59	0.85	0.97
7	0.07	0.23	0.49	0.11	0.31	0.57
8	0.17	0.38	0.65	0.35	0.62	0.83
9*	0.03	0.15	0.41	0.59	0.85	0.97
10	0.17	0.38	0.65	0.43	0.69	0.89
11	0.03	0.14	0.39	0.15	0.36	0.61
12	0.17	0.38	0.65	0.51	0.77	0.93
13	0.03	0.15	0.41	0.51	0.77	0.93
14	0.29	0.54	0.78	0.43	0.69	0.89
15*	0.00	0.08	0.32	0.79	1.00	1.00

TABLE 4.2. Summary of Individual Far-Side Inspector Performance for Both Manual and Advanced Techniques

Inspector	Lower 90%	FCP	Upper 90%	Lower 90%	POD	Upper 90%
1	0.00	0.00	0.22	0.00	0.00	0.31
2	0.09	0.25	0.48	0.23	0.57	0.87
3	0.11	0.31	0.57	0.00	0.00	0.31
4	0.00	0.00	0.21	0.01	0.13	0.47
5	0.00	0.00	0.21	0.01	0.13	0.47
6	0.11	0.31	0.57	0.05	0.25	0.60
7	0.29	0.54	0.78	0.40	0.75	0.95
8	0.00	0.08	0.32	0.01	0.13	0.47
9	0.43	0.69	0.89	0.19	0.50	0.81
10	0.17	0.38	0.65	0.19	0.50	0.81
11	0.33	0.57	0.79	0.01	0.17	0.58
12	0.11	0.31	0.57	0.05	0.25	0.60
13	0.00	0.08	0.32	0.01	0.13	0.47
14	0.00	0.08	0.32	0.00	0.00	0.31
15	0.00	0.08	0.32	0.00	0.00	0.31

Generally speaking, "good" performance is indicated when FCP < 0.20 and POD > 0.80. From Table 4.1, one can see that three inspectors satisfy this requirement for near-side inspections (those marked with an asterisk). It is interesting to note that high POD values seem to be associated with low FCP values in the near-side inspections. In other words, good inspectors are not increasing their POD by simply calling more weld cracked. The good inspectors simply possess a much better capability to distinguish cracks from nondefective welds. This phenomenon is illustrated in Figure 4.16. Note that the three best teams display performance well within the acceptable criteria (i.e., 80% POD and 20% FCP -- the upper left-hand box of this figure). Also note that the performance of the three worst inspectors is not much better than chance.

The situation is much different for the far-side performance illustrated in Table 4.2. In this case, the only high POD values are achieved at the cost of a correspondingly large FCP. This indicates that the difference in inspector performance illustrated in this table is probably due to differences in the decision criteria employed and not inherent differences in inspector ability. None of the inspectors listed in Table 4.2 demonstrated "good" capability to detect cracks from the far side.

4.2 THE IMPACT OF SELECTED VARIABLES ON INSPECTION PERFORMANCE

4.2.1 Impact of Material and Defect Variables

In this section, the effect of four variables on aggregate inspector performance is examined. These four variables are:

Ultrasonic Technology:	Manual or Advanced
Inspection Access:	Near or Far Side
Inspector Quality:	PIRR inspectors (before IEB 83-02) or MRR inspectors (after IEB 83-02)

Crack Size	No. of Cracks	Average Length, in.	Average Depth, in./% of Wall	Range Length, in.	Range Depth, in.
Short (S)	5	0.29	0.042/4	0.1-0.67	0.011-0.071
Deep (D)	9	0.92	0.177/18	0.16-3.35	0.109-0.247
Long (L)	14	7.39	not measured(a)	3.6-10	0.170-0.440

Figure 4.17 compares the performance of advanced ultrasonic technology (i.e., two imaging systems and an adaptive learning system) versus manual inspectors when inspecting all short cracks from the near side. Notice that

(a) The long cracks were only destructively analyzed over a portion of their length, and thus the location and value of greatest depth is unknown; but the data reported is for the greatest depth in the destructively analyzed zones.

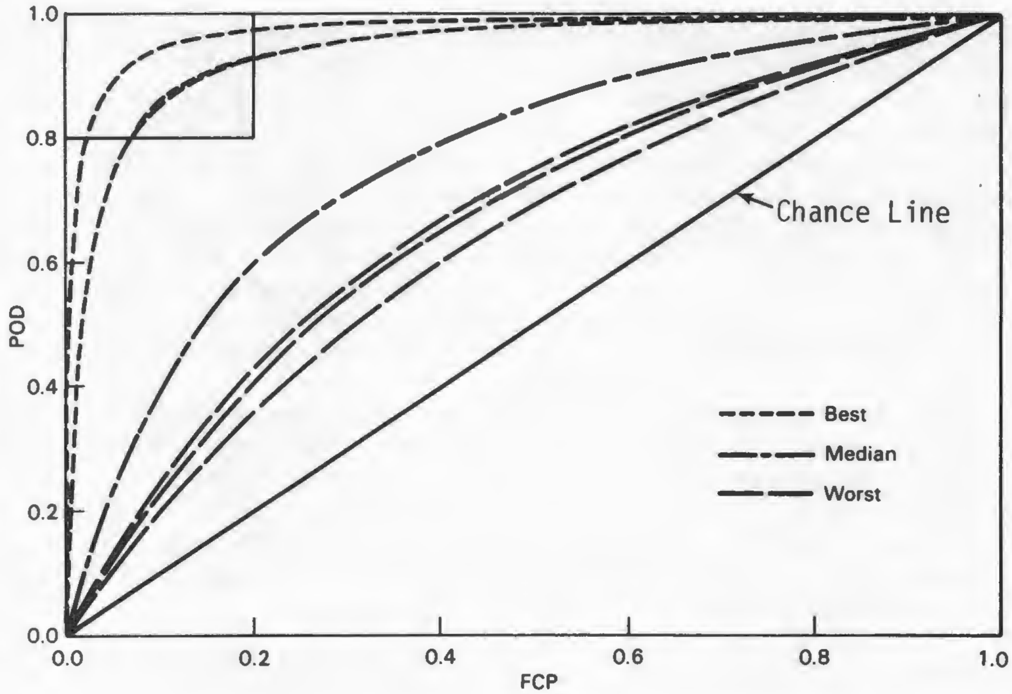


FIGURE 4.16. Variability in Inspection Performance ROC Curves Showing Three Best, Median (based on all 12 manual inspectors), and Three Worst Manual Inspection Performances for Short Cracks

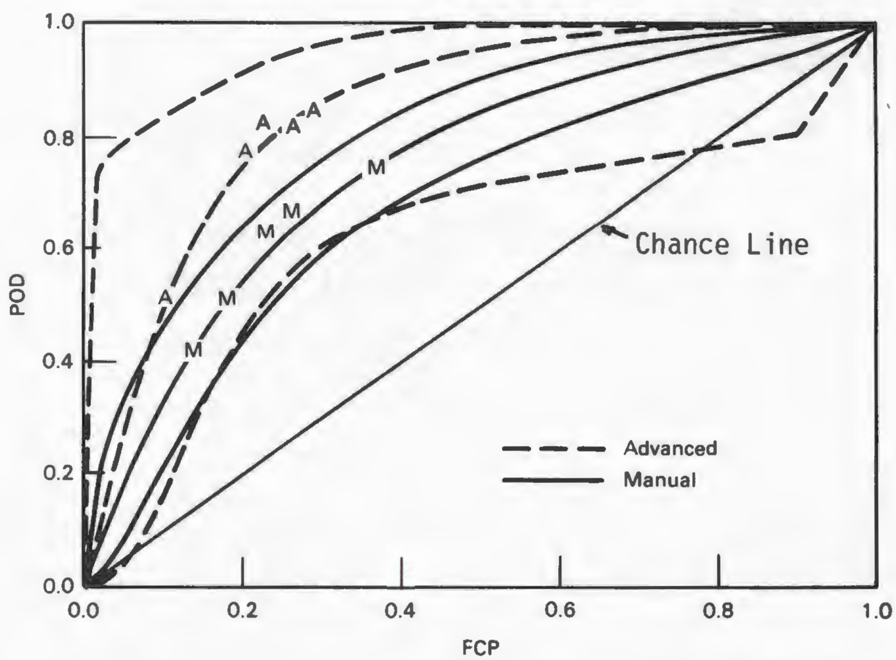


FIGURE 4.17. Advanced Versus Manual ROC for Near-Side Access for (S) and (D) Cracks

95% confidence bounds on the manual curve are much smaller than those for the advanced; there are only three advanced inspectors in the data set versus 12 manual inspectors. The ROC curve for advanced procedures is above the manual, providing some evidence that advanced procedures may be better. However, the experimental error within the advanced procedure's ROC is substantial, as indicated by the large confidence bounds. A formal hypothesis test to determine whether the advanced procedures were better than manual yielded a level of significance of approximately 10%. Therefore, moderate evidence exists to support the conclusion that advanced is better than manual.

Figure 4.18 demonstrates the difference between near- and far-side access. The ROC curves in Figure 4.18 describe the performance of manual inspectors on (S) and (D) cracks. As one can see from this plot, access condition has the most profound influence on inspection performance of all the variables considered in the round robin. The confidence bounds around the ROC curves plotted in Figure 4.18 show that the difference between near- and far-side access conditions is highly significant.

It is also important to note that the far-side performance seems to be indistinguishable from the performance that could be achieved by chance. This conclusion is corroborated by the results in Table 4.3.

One of the objectives of the MRR was to determine what difference, if any, exists in detection capability between different sizes of cracks. Figure 4.19 presents the inspection performance of MRR inspection for three conditions of crack size.

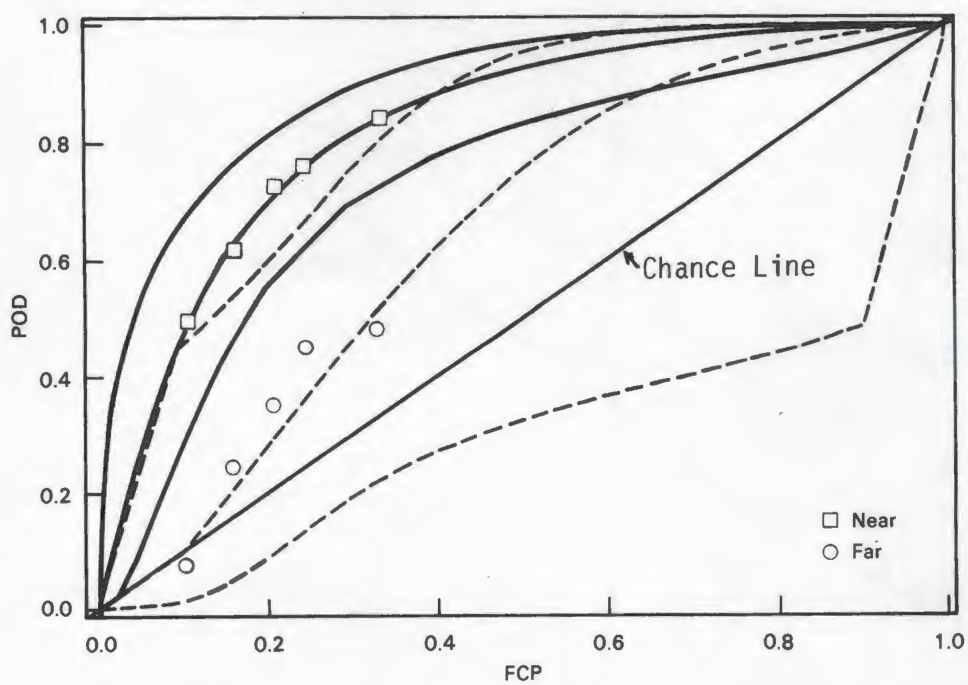


FIGURE 4.18. Effect of Near- (——) versus Far (---) Side Access for Manual Inspectors on (S) and (D) Cracks

Inspection of Figure 4.19 shows little difference in detection capability between (D) cracks and (L) cracks -- the crack/no crack (C3 classification scale) performance for (S) cracks being 0.73 POD and 0.21 FCP, and for (L) and (D) cracks is 0.73 POD and 0.29 FCP. Figure 4.19 also shows that the performance of inspectors is worse for (S) flaws than for either (D) or (L) cracks. It should be noted that the (S) flaws were "spot-like" defects because their size is generally smaller than the resolution of the systems being used.

As a final note, Figure 4.19 illustrates that inspection performance is a function of crack size (length and depth). Since there is little difference in D and L-type cracks, it appears that the depth of a crack seems to have slightly more influence on crack detection than length.

Initially, the MRR was set up to determine the effect that training and performance demonstration testing conducted to meet requirements of IEB 83-02 has had on the population of inspectors. In the Piping Inspection Round Robin (conducted in 1982), six inspectors who had not received the EPRI IGSCC training course and had not gone through a performance demonstration test but were experienced nuclear UT inspectors examined the same set of (S) and (D) cracks as did the MRR inspectors. Since all the MRR inspectors have had recent training and passed a performance demonstration test, these results can be used to evaluate the effectiveness of the training and performance demonstration test. Figures 4.20 and 4.21 compare the ROC curves associated with the PIRR versus MRR inspectors. These ROC curves were calculated for specimens that were common to both the PIRR and MRR.

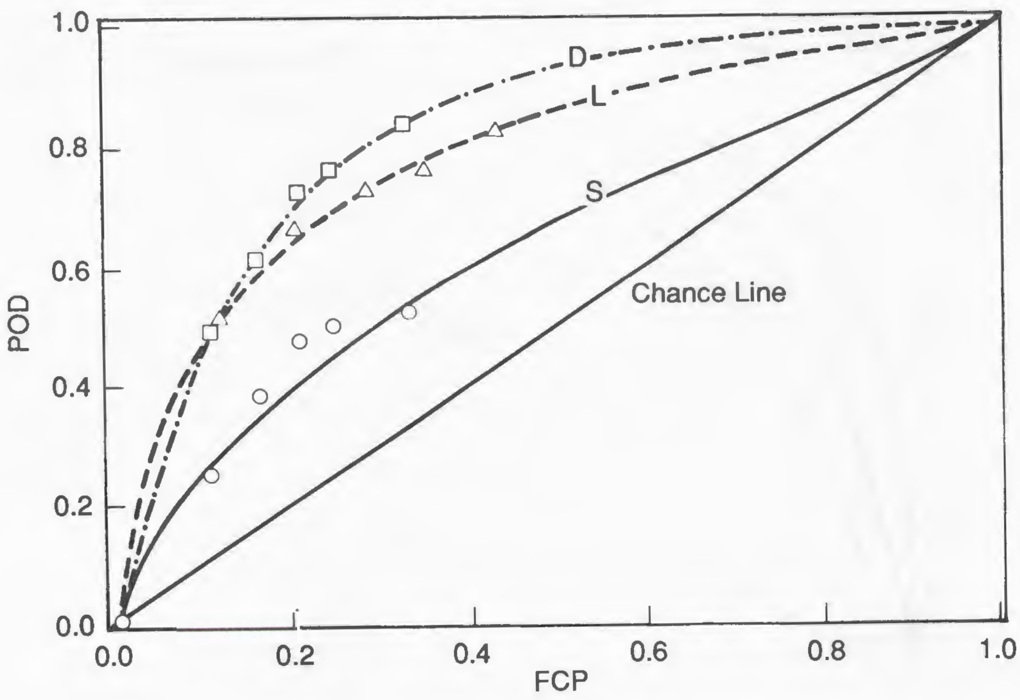


FIGURE 4.19. ROC Curves for the Effect of Crack Size for (S) (——), (D) (-•-•-), and (L) (---) for Manual Inspectors

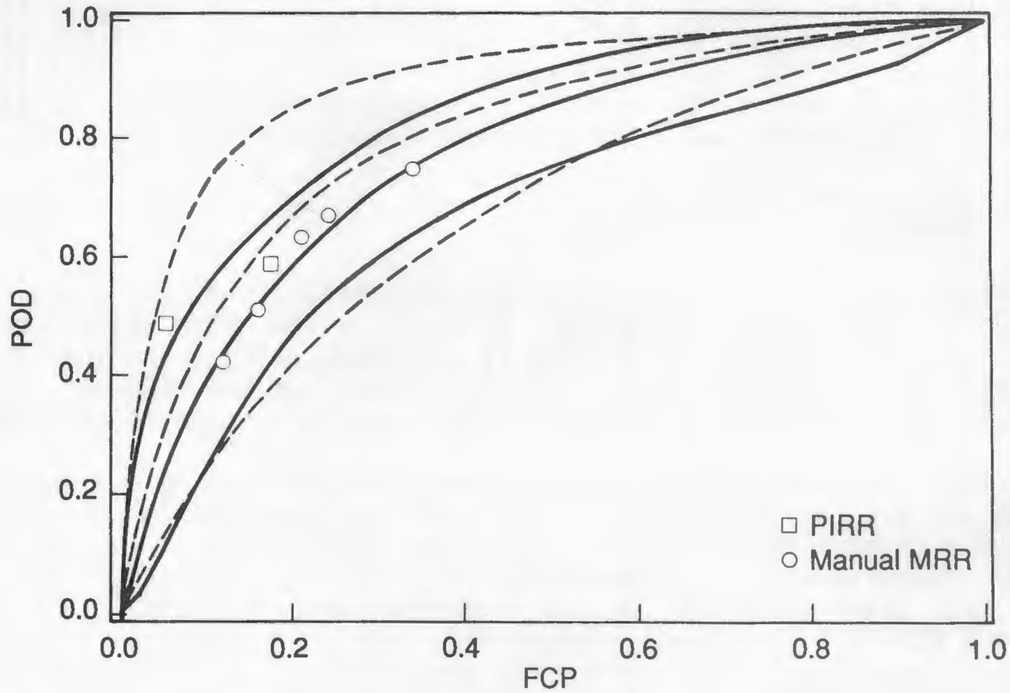


FIGURE 4.20. Comparison of Manual Inspectors in the PIRR (---) and MRR (——) on All Short Cracks

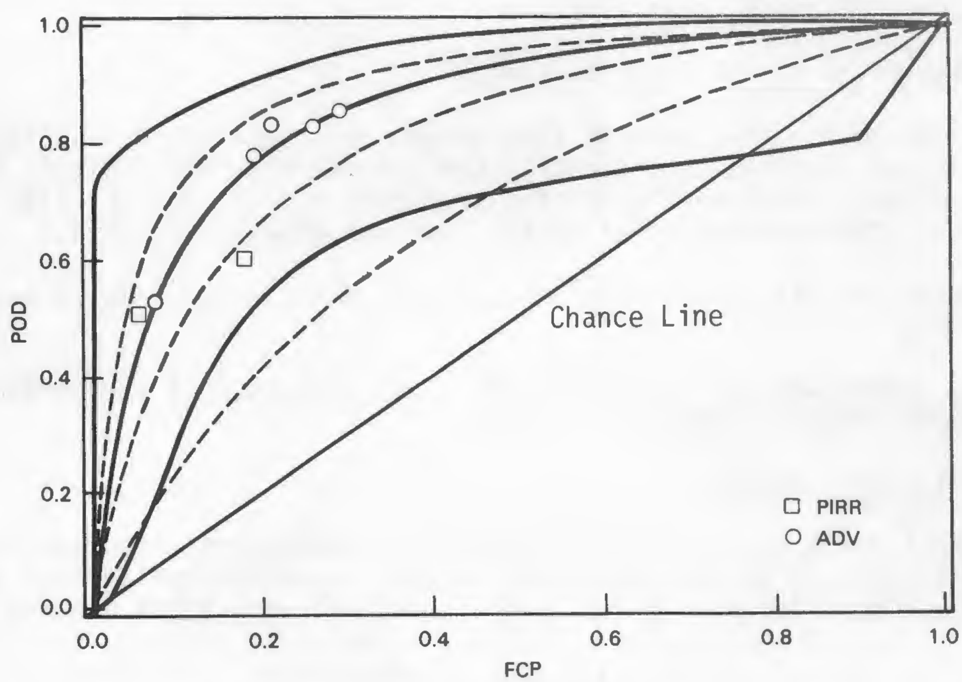


FIGURE 4.21. Comparison of MRR Advanced (——) Inspectors with Manual PIRR (---) Inspectors on All Short Cracks

The PIRR teams did not employ the same classification scale as the MRR inspectors, so the number of (FCP,POD) data points for the PIRR ROC is not the same as the MRR ROC. The PIRR teams employed a two-point classification scale (as opposed to the five-point scale employed in the MRR). In the PIRR, teams were simply requested to mark every indication as either:

C: indication is a crack
N: indication is not a crack.

Because the PIRR ROC is determined from a smaller classification scale, we can be less certain about its shape. However, if the curve shapes are correct, Figures 4.20 and 4.21 show that there is no strong evidence to indicate that the ROC curves for the MRR inspectors are better than the PIRR ROC. In fact, the ROC curves for manual and advanced MRR inspectors straddle the PIRR ROC.

Although training and performance demonstration testing may not have improved the ROC curve, it certainly did have an effect on inspection performance, and a closer look at Figure 4.20 reveals the nature of the effect. In the PIRR inspections, the C classification produced a POD of 0.49 and a FCP of 0.06; but after IEB 83-02, manual inspectors were achieving a POD value of 0.65 and a FCP of 0.22. These changes are not due to experimental error. Apparently, the inspector's performance seems to have moved from one point on the ROC curve to another by altering the decision threshold.

Table 4.3 provides a summary of the inspection variables' effect on FCP and POD. The FCP and POD values listed in this table are calculated using the C3 (crack/no crack) classification.

4.2.2 Comparison of POD for Flaw Length

To compare the influence of flaw length, we constructed a subset of short and long flaws that had approximately the same depth. Flaws within 20% to 30% through-wall depth were extracted from both the "D" and "L" flaw categories. Characteristics of these flaws are given in Table 4.4.

We then calculated POD using this set of flaws. The results are shown in Table 4.5.

Thus, there was very similar performance for IGSCC of the same depth for crack lengths ranging from 0.6 in. to 8 in.

4.2.3 False Call Signals

Analysis of signal classification by the inspectors revealed that the teams made calls in several locations in very consistent patterns. Some of these areas coincided with cracks, some coincided with blank grading units, and others were located in non-scored but blank areas of the pipe. All of these blank areas were examined to try to understand the specimen properties that produced "crack-like" responses, which inspectors used to classify the signals improperly. Destructive plugs were removed and used to verify the presence or lack of any defects. The procedures used resulted in a significant number of zones requiring further analysis. As an example, the side A of

TABLE 4.3. Summary of (FCP,POD) Performance

Access	Procedure		(S) and (D) Cracks FCP	POD, (S) 4% Ave. Depth	POD, (D) 18% Ave. Depth
PIRR (pre-IEB 83-02)					
Near	Manual	lower 90%	0.03	0.05	0.58
		estimate	0.06	0.12	0.67
		upper 90%	0.10	0.22	0.74
Far	Manual	lower 90%	0.10	0.04	0.09
		estimate	0.14	0.09	0.17
		upper 90%	0.19	0.16	0.28

MRR (post-IEB 83-02)			(S) and (D) Cracks			(L) Cracks	
			FCP	POD (S)	POD (D)	FCP	POD
Near	Manual	lower 90%	0.16	0.29	0.62	0.21	0.65
		estimate	0.21	0.43	0.69	0.33	0.75
		upper 90%	0.27	0.58	0.76	0.48	0.83
Far	Manual	lower 90%	0.22	0.14	0.22	0.21	0.25
		estimate	0.28	0.21	0.39	0.33	0.39
		upper 90%	0.34	0.31	0.58	0.48	0.54
Near	Advanced	lower 90%	0.11	0.34	0.72	0.01	0.62
		estimate	0.21	0.67	0.87	0.11	0.83
		upper 90%	0.35	0.90	0.95	0.43	0.95
Far	Advanced	lower 90%	0.05	0.02	0.01	0.04	0.00
		estimate	0.13	0.11	0.17	0.22	0.00
		upper 90%	0.26	0.31	0.58	0.55	0.28

TABLE 4.4. Short Versus Long Flaw Data

Flaw	Depth	Length
"D"		
1	23%	0.6"
2	24%	1.0"
3	25%	3.4"
"L"		
1	24%	5.5"
2	25%	8.0"

NOTE: All flaws in wrought

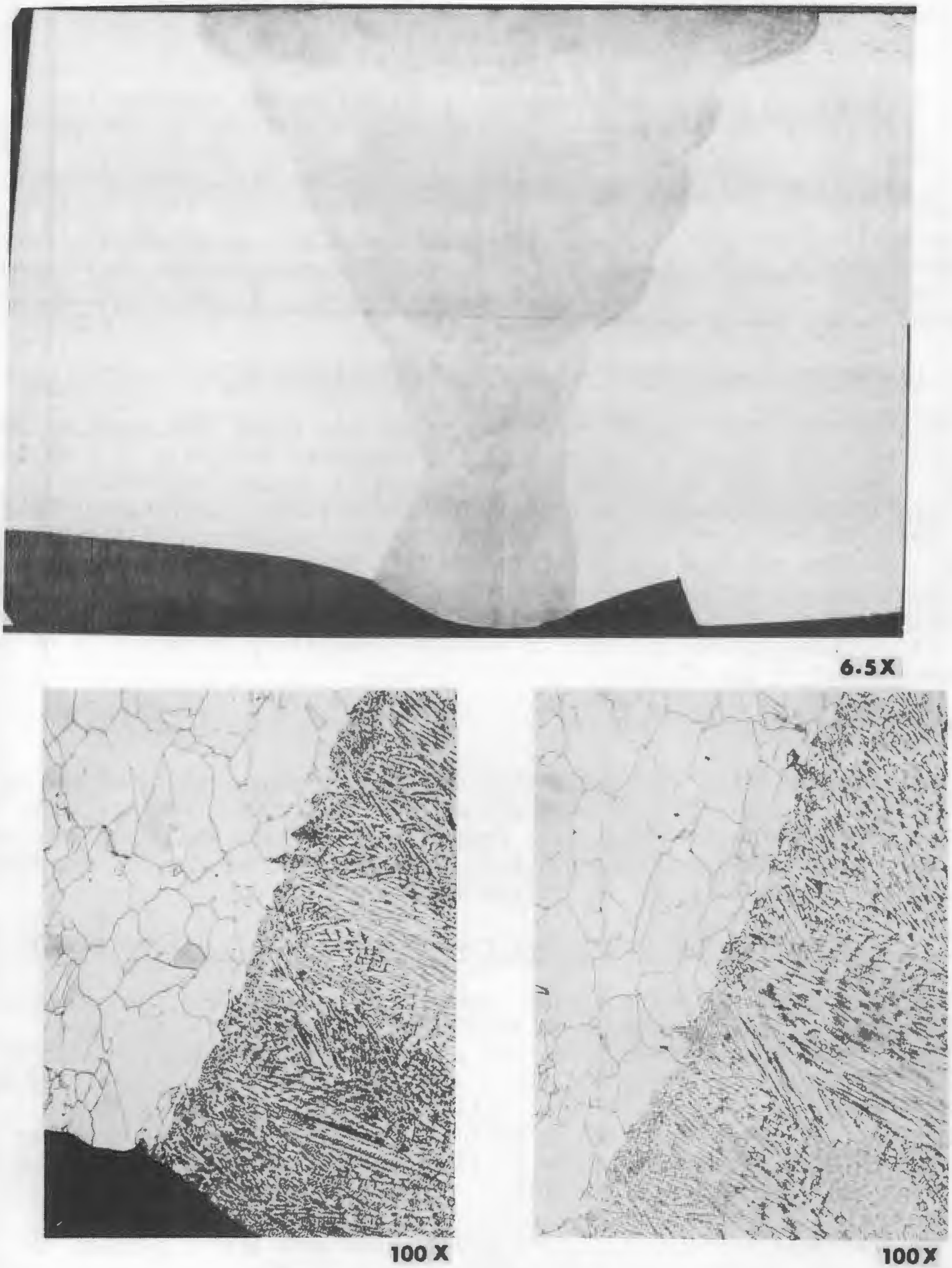
TABLE 4.5. POD for Short Versus Long Flaws

Length	Access Near	Access Far
"S" -- POD	78%	43%
Number of Inspections	46	14
"L" -- POD	84%	43%
Number of Inspections	31	14

specimen D5, quadrant 1 (D5-1A) had in one location 70% of the inspectors identifying this as a potentially defective area. The number of inspectors who classified this as a crack were nearly one-half or 35% of the total inspections. This pattern was repeated on several other specimens with respect to the decision-making process and about one-half the inspectors classified it as a crack of the total number identifying it as a suspect area.

Because of the high incident of calls in D5-1A, this zone was analyzed and a plug was removed from the zone containing the highest frequency of false calls. Then the plug was sectioned and polished to determine the cause of these calls. The micrograph in Figure 4.22 shows a cross section of the weld. The B side of the weld had an abrupt machined counterbore but was not the cause of the false calls on the A side of the weld. The only explanationthat seems plausible is that the ultrasound scatters off the inside of the pipe, to the lower part of the weld, to the upper part of the weld, and then back to the receiver. This double weld metal scattering path exists for transducer sizes commonly used in the MRR. The unique appearance of the signal is that it has the properties of IGSCC. These properties appear to be related to the coarse grain weld-to-weld scattering that the path takes. Ultrasonic tests on austenitic specimens with an inside cladding give a crack-like indication located at the weld fusion line. These conditions give responses that are very large when the transducer is skewed at angles of even 45° to the weld. Furthermore, the signals tend to jump up and down like seeing the facets of IGSCC. This condition is very difficult to analyze and would tend to produce false calls in the field. Possibly an ID creeping wave probe may be able to correctly classify this condition, because it should only provide a single scattering from the weld metal.

In reviewing the ultrasonic responses, there were several material conditions that tended to give results that had "crack-like" properties. Drop-through conditions were present in several of the pipes, but did not appear to give the inspectors any problems. Several of the welds had some suck-back in the weld root area that did create ultrasonic reflections. In only one case, the suck-back condition was located in the same grading unit as a crack, and this crack was consistently oversized. The suck-back provided an ultrasonic signal that plotted near the heat-affected zone weld fusion line, and the signal also had some of the skew properties of IGSCC because of the



D-5 304L S.S.

FIGURE 4.22. Micrograph of MRR Specimen D5-1 with the A side being on the left of the weld centerline

weld ripple pattern in the suck-back. The suck-back present in the MRR samples was a reportable condition based on ASME Code.

The single-most consistent pattern in the ultrasonic response from nondefective areas was the presence of two signals originating from the weld-root area. These signals tended to provide some of the characteristics of the pattern that is normally associated with IGSCC in the heat-affected zone with the root signal falling along later in time on the A-scan. These signals did not possess the skew properties of IGSCC and thus could be properly classified. Signals from weld geometry and material conditions were present in all samples and illustrate the continual decision-making demands that are placed on the inspector during a weld examination.

4.2.4 Estimated Impact of Performance Demonstration Testing

All the teams that participated in the MRR had taken EPRI training and passed a performance demonstration test that required detection of 4 of 5 cracks. The EPRI test has since been made more stringent (performance demonstration now requires detection of 8 of 10 cracks). We have estimated the effects of this more stringent test on the MRR population of inspectors. According to the calculations presented in [6], roughly 25% of the present MRR population would pass the more stringent detection test. If this passing population had taken the MRR test, we would expect it to display an average false call probability of 14% and an average probability of detection of 86%. The average POD of the MRR population is 67%.

4.3 MRR SIZING RESULTS

This section deals with evaluating the depth sizing results of the MRR participants. Two analysis methods are used in this evaluation; relative measurements in percent through-wall described in Section 4.3.1 and absolute measurements described in Section 4.3.2. A summary of the conclusions from the sizing analysis is given in Section 4.3.3.

4.3.1 Sizing Results Analyzed in Relative Measurements (% Through-Wall)

In order to evaluate the sizing capabilities of IGSCC-trained inspectors, each inspector also was required to assess indications at 13 specified locations in the specimens and size these indications if he ascertained it originated from IGSCC. These locations were then destructively analyzed to determine true crack depths. The results of this sizing experiment are presented in Table 4.6 with results measured in relative depths. Figures 4.23 through 4.30 show least square plots of the sizing data in Table 4.6 (excluding the first four data points).

The measurements recorded by the eight inspectors are in percent of through-wall as are the true-state depths. Three of the flaw locations (0%) given to the inspectors to size contained crack-like signals but were destructively analyzed and found to not be cracked.

TABLE 4.6. Summary of Relative Sizing Performance

Destructive Depth	UT Crack Depth Estimate Inspector 1	2	3	4	5	6	7	8
0%	80.0	24.0	32.0	34.0	75.0	10.5	19.0	9.0
0%	15.0	18.0	47.0	21.0	50.0	33.0	58.0	45.0
0%	15.0	18.0	48.0	17.0	25.0	7.0	50.0	29.0
6%	64.0	31.0	20.0	29.0	75.0	18.0	24.0	36.0
14%	16.0	45.0	5.0	11.0	5.0	11.0	38.0	31.0
23%	24.0	18.0	5.0	0.0	10.0	30.0	60.0	35.0
25%	40.0	23.0	32.0	10.0	50.0	5.0	38.0	26.0
26%	46.0	25.0	23.0	10.0	75.0	13.4	38.0	20.0
28%	75.0	40.0	44.0	8.0	75.0	70.0	58.0	23.0
33%	84.0	18.0	36.0	22.0	75.0	5.0	25.0	23.0
37%	15.0	8.0	56.0	11.0	50.0	7.0	17.0	16.0
38%	7.0	29.0	27.0	23.0	80.0	16.0	38.0	21.0
52%	78.0	55.0	56.0	13.0	75.0	50.0	68.0	69.0

Analysis of the data in Table 4.6 provides several useful insights for understanding crack sizing capability. The performance of inspectors on blank sizing specimens shows that sizing flaws is not a simple task and can be very unreliable. The blank samples were selected from areas in a pipe that produced "crack-like" ultrasonic responses. The inspectors were told that the 13 areas selected for sizing had "crack-like" UT signals, but it was not known if they were cracked. Each team was also told that if they thought an area was not cracked to report it as a 0% crack. Only Team 4 did this for one crack (unfortunately, it was 23% deep). Since tip diffracted signals, which are used to size flaws, are very low in amplitude (often little larger in amplitude than noise), it is quite possible that inspectors would, under the pressure of a test situation, find a tip signal associated with the crack-like signals.

Table 4.7 provides a summary of performance on blank material. In the detection phase of the round robin, these blank areas produced a false call probability of 27%, which is typical for the test.

Table 4.8 provides a summary of inspector performance on the sizing data with blanks removed from the analysis.

The performance shown in Table 4.8 is not very good; with slope values in one case close to 1 and others very far away from 1, and high variability with an average standard deviation of 20%. The R^2 statistic in Table 4.8 is a square of the multiple correlation coefficient. R^2, which can vary between 0 and 1, provides a measure of how well the data fits a linear regression model. If R^2 were 1, there would be a perfect fit; if R^2 were zero, no fit to a linear regression is indicated. R^2 is discussed in more detail in section 4.3.2. The performance reflected in Table 4.8 includes a very small flaw (i.e., 6 percent through-wall for a pipe wall thickness of 0.600 in. or 0.030 in.). When considering the physics of sizing flaws using crack tip

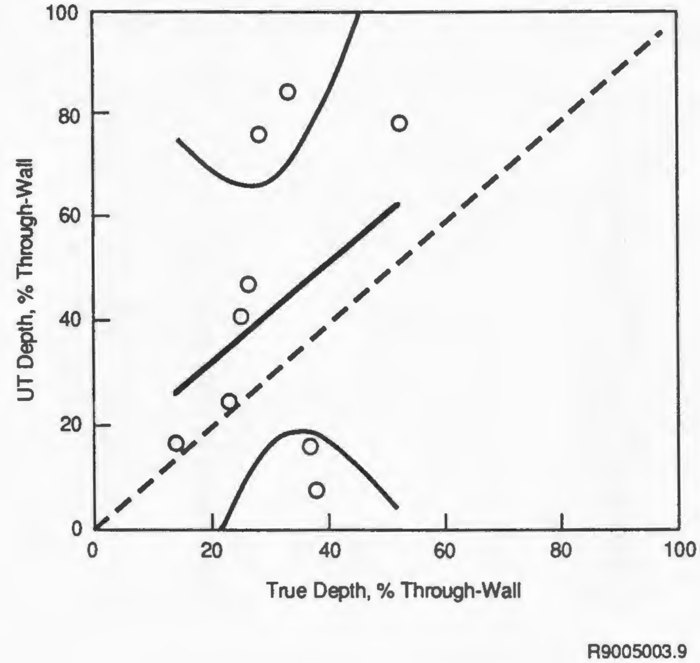


FIGURE 4.23. Relative Sizing Results for Inspector 1, 90% Confidence Bounds

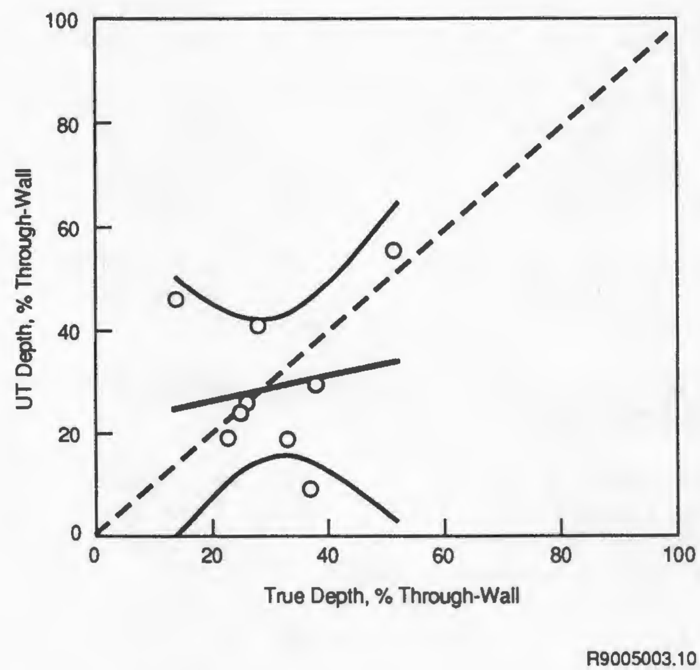
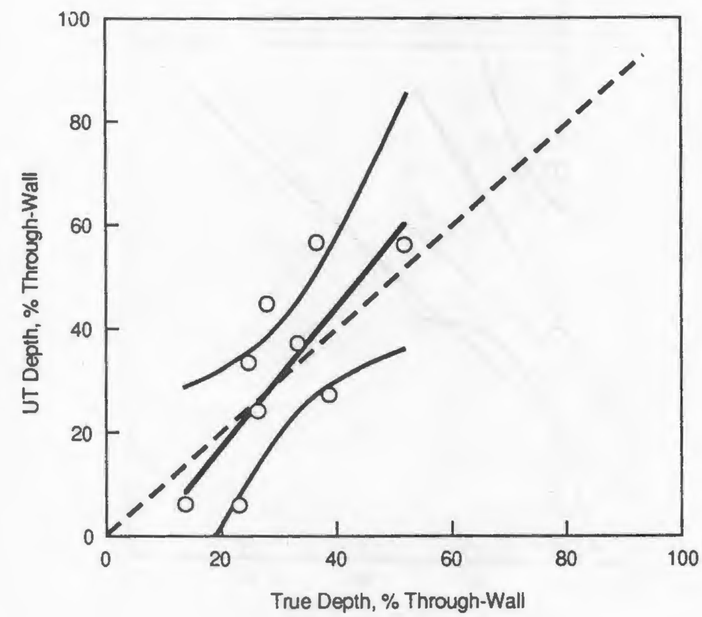
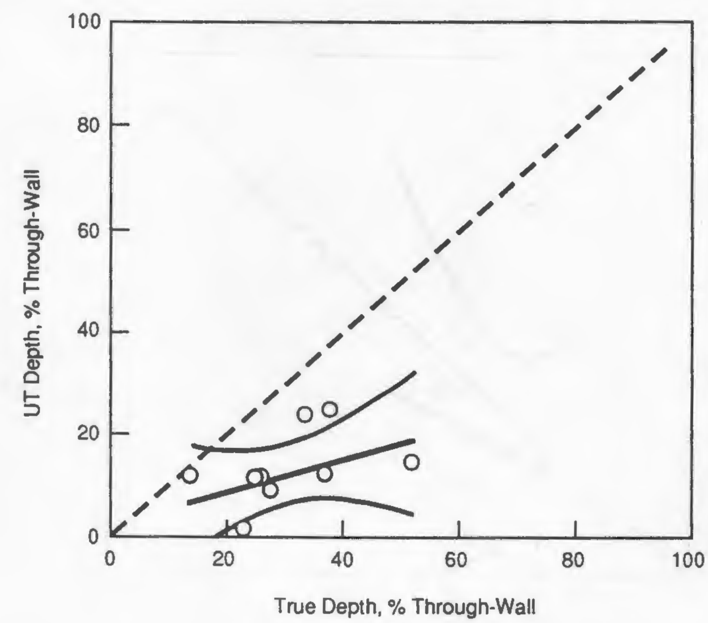


FIGURE 4.24. Relative Sizing Results for Inspector 2, 90% Confidence Bounds



R9005003.11

FIGURE 4.25. Relative Sizing Results for Inspector 3, 90% Confidence Bounds



R9005003.12

FIGURE 4.26. Relative Sizing Results for Inspector 4, 90% Confidence Bounds

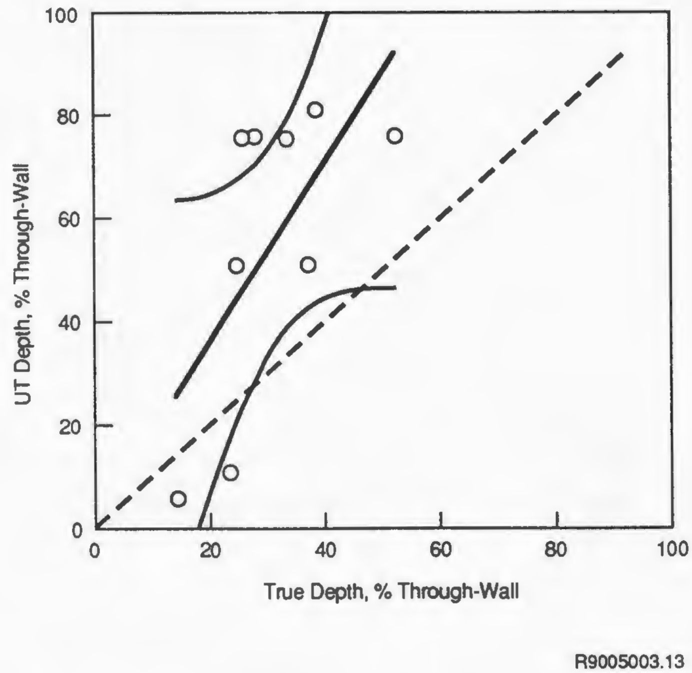


FIGURE 4.27. Relative Sizing Results for Inspector 5, 90% Confidence Bounds

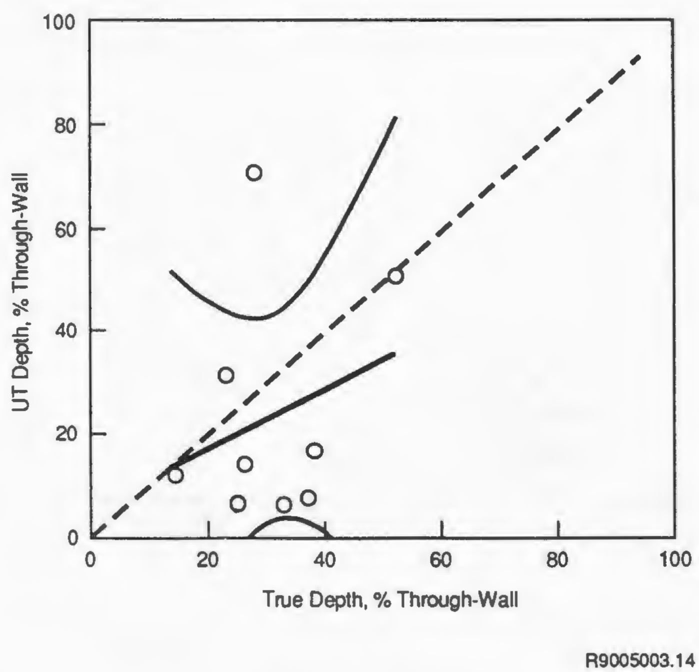


FIGURE 4.28. Relative Sizing Results for Inspector 6, 90% Confidence Bounds

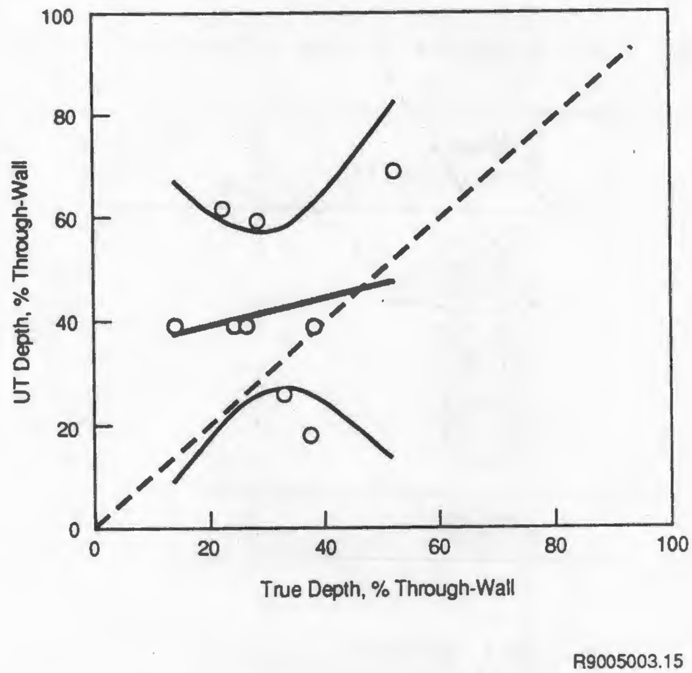


FIGURE 4.29. Relative Sizing Results for Inspector 7, 90% Confidence Bounds

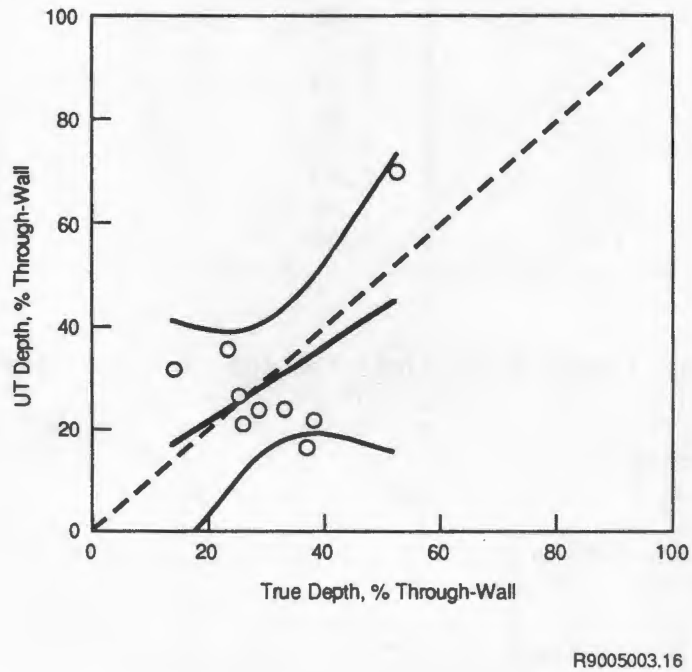


FIGURE 4.30. Relative Sizing Results for Inspector 8, 90% Confidence Bounds

TABLE 4.7. Summary for Relative Sizing Estimates in Blank Material

Inspector	Mean (% through-wall)	Standard Deviation
1	36.67	37.53
2	20.00	3.46
3	42.33	8.96
4	24.00	8.88
5	50.00	25.00
6	16.83	14.10
7	42.33	20.59
8	27.67	18.04
Mean	32.48	19.85

TABLE 4.8. Summary of Inspector Regression Fits for the Relative Sizing Data with the Blanks Removed

Inspector	Intercept (% through-wall)	Slope	Standard Deviation	R^2
1	36.57	0.30	30.40	0.02
2	25.58	0.13	14.86	0.01
3	1.06	1.04	13.09	0.54
4	15.93	-0.08	8.96	0.02
5	33.66	0.83	27.69	0.14
6	10.11	0.44	22.10	0.07
7	28.00	0.44	16.74	0.12
8	19.73	0.36	15.31	0.10

diffraction, there is a resolution limit for the smallest flaw that can be sized. The resolution limit is dependent on the wavelength of sound in the test material. The test material used in the MRR was wrought austenitic steel. This material effectively limits testing frequencies to 5 MHz or lower because of severe attenuation of sound at higher frequencies.

The practical resolution limit for sizing cracks using 5 MHz is approximately 0.040 in. Therefore, one would not expect to be able to resolve tip signals from the 6% through-wall crack. In other words, if one used a test frequency of 5 MHz for cracks 0.040 in. or less, the corner trap would be seen but it would be difficult to detect the crack-tip signal.

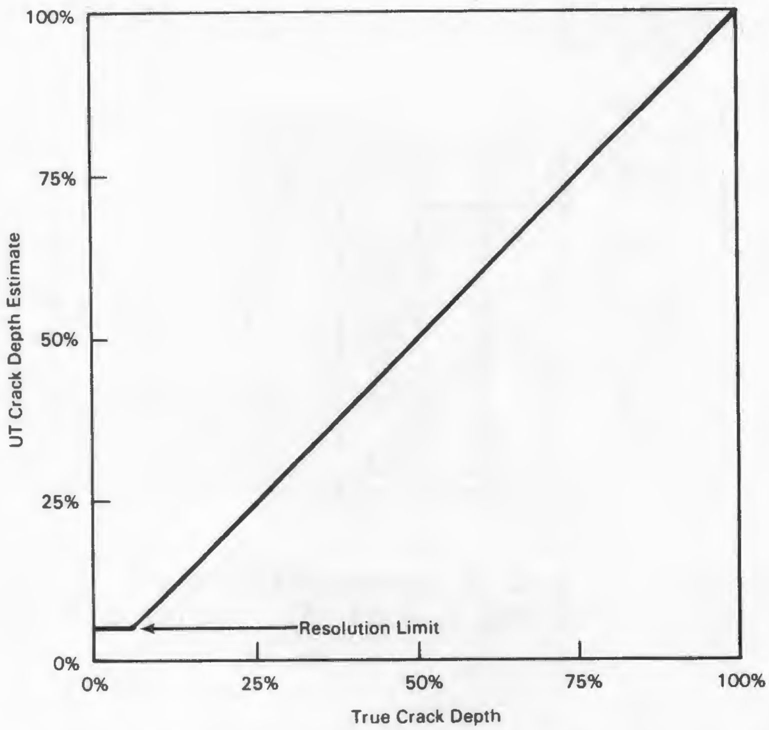


FIGURE 4.31. Model for Fitting Sizing Data that includes Resolution Limit

The average depth attributed to the 6% crack was 37%, remarkably close to the 32% depth attributed to blank material. To further illustrate the effect of resolution limits, if one were to use 2.25 MHz as a test frequency, the resolution limit for cracks would be approximately 0.085 in. An 0.085-in. deep crack in a 10-in.-diameter, Schedule 80 pipe represents approximately a 17% through-wall crack.

When one recognizes that there is an inherent resolution limit in sizing cracks, then one must also realize that a linear model (i.e., straight line) really does not fit the sizing data in Table 4.3. A more realistic model is shown in Figure 4.31, which has a slope of 1 (i.e., is linear) from the resolution limit to 100% through-wall and has a slope of zero from the resolution limit to 0% through-wall.

To account for the nonlinear (zero slope) portion of the model, the data in Table 4.6 was analyzed excluding the blank material and the crack below the resolution limit of 5 MHz (which was used by some inspectors); sizing results look a bit more promising as Table 4.9 shows.

The results are very dispersed in that one inspector had a slope of 1.77 and another was 0.25, the intercepts range from -10.44 to 34.02, and the standard deviation goes from 6.78 to 29.86. Inspector 3 did about the best overall job of sizing more accurately than the other inspectors. The significance of this data is that the reliability and accuracy for sizing the cracks in this study was quite poor.

TABLE 4.9. Summary of Regression Fits for Sizing Data with Blanks and Small Cracks (<.030") Removed

Inspector	Intercept (% through-wall)	Slope	Standard Deviation	R^2
1	13.32	0.96	29.86	0.12
2	21.37	0.25	15.72	0.03
3	-10.44	1.37	12.47	0.62
4	3.66	0.27	6.78	0.18
5	0.72	1.77	23.35	0.44
6	5.36	0.58	23.48	0.08
7	34.02	0.27	17.59	0.03
8	6.98	0.73	14.77	0.25

The results of Table 4.9 show an improvement over that obtained by a sizing round robin conducted by EPRI in 1983 [5]. In the EPRI exercise, the results were:

Standard Deviation = 19.25%
Average Intercept = 19%
Average Slope = 0.36

In reviewing the EPRI crack samples, they did not have any cracks smaller than the resolution limit suggested in the MRR.

4.3.2 Sizing Results Analyzed in Absolute Measurements

The previous section examined flaw sizing capability using relative (% of through-wall) measurements. One may also evaluate sizing capability in terms of absolute (i.e., inch) measurements. Table 4.10 presents a summary of the same measurements discussed in Section 4.3.1 but expressed in absolute units.

If all inspections were performed on material of the same thickness, both relative and absolute sizing measurements would give the same results. However, in the current set of specimens, two are constructed of thicker material (0.9 in. versus 0.6 in.); and furthermore, one of these thicker specimens also contains the largest crack in the data set, a circumstance that could cause differences. It is also important to note that the regressions performed with the two types of measurements rest on different assumptions about measurement error variability. The absolute model assumes that variability remains the same as thickness changes; the relative model assumes that variability increases as wall thickness increases.

Consequently, the absolute model "weights" observations from the thick sections more heavily than the relative. In the current data set, absolute regression will weight the 0.45-in. and 0.242-in. cracks more heavily than the relative.

TABLE 4.10. Summary of Absolute Sizing Performance

Destructive Depth, in.	UT Crack Depth Estimate (inches)							
	Team 1	Team 2	Team 3	Team 4	Team 5	Team 6	Team 7	Team 8
0.000	0.550	0.165	0.220	0.234	0.516	0.072	0.131	0.062
0.000	0.103	0.124	0.330	0.117	0.172	0.048	0.344	0.200
0.000	0.103	0.124	0.323	0.144	0.344	0.227	0.399	0.310
0.033	0.352	0.171	0.110	0.160	0.413	0.099	0.132	0.198
0.077	0.088	0.248	0.028	0.061	0.028	0.061	0.209	0.171
0.127	0.132	0.099	0.028	0.000	0.055	0.165	0.330	0.193
0.143	0.253	0.138	0.127	0.055	0.413	0.074	0.209	0.110
0.172	0.275	0.158	0.220	0.069	0.344	0.034	0.261	0.179
0.209	0.039	0.160	0.149	0.127	0.440	0.088	0.209	0.116
0.227	0.578	0.124	0.248	0.151	0.516	0.034	0.172	0.158
0.242	0.649	0.346	0.381	0.069	0.649	0.606	0.502	0.199
0.255	0.103	0.055	0.385	0.076	0.344	0.048	0.117	0.110
0.450	0.675	0.476	0.485	0.113	0.649	0.433	0.589	0.598

Another important difference between the two models occurs when results of a regression fit are used to predict sizing performance in other thicknesses of material, particularly thicker material. The absolute sizing model, when extrapolated to thicker material, produces a smaller variance for the measurements than the relative. Thus, extrapolations using the absolute model are less conservative than the relative model regression.

In Table 4.11, regression parameters obtained from the absolute measurements are presented. These regressions utilized inspections on cracks larger than 0.04 in. -- the same observations employed by the regressions listed in Table 4.9. For the purposes of comparison, important results listed in Table 4.9 are also repeated in Table 4.11. From these results, we see that the slope, and more importantly, the R^2 statistic is consistently larger for the absolute regressions, indicating a better fit to the data.

The physics of UT inspections for sizing would lead one to expect to find results that reflected a tendency as shown in the above analysis to oversize the small cracks due to the system resolution limits and a tendency for an undersizing of the larger cracks. The undersizing is based on the fact that the crack tip is tight and branched and, thus, may not give a large signal response so that the response from crack regions that are more open may result in more usable signals giving the inspector more confidence in selecting them for sizing estimates. Experience would suggest that the deeper cracks in this study should fall into the transition from sizing accurately to systematic undersizing. In reviewing the data, it is interesting to note that the majority of the data for the three largest cracks is significantly larger or significantly smaller than the true size. The best performance is for the largest crack, and it has three of the values within 10% of true size but the other five values are at a minimum greater than 30% in error and some are much larger than this.

TABLE 4.11. Summary of Team Regression Fits for Absolute Sizing Data with Small Cracks (< 0.040-in.) Removed

	Absolute Results				Relative Results	
	Intercept, in.	Slope	Standard Deviation, in.	R^2*	Slope	R^2*
Team 1	-0.02	1.57	0.21	0.43	0.96	0.12
Team 2	0.04	0.78	0.11	0.38	0.25	0.03
Team 3	-0.06	1.36	0.08	0.80	1.37	0.62
Team 4	0.03	0.22	0.04	0.27	0.27	0.18
Team 5	0.04	1.61	0.15	0.59	1.77	0.44
Team 6	-0.05	1.05	0.18	0.30	0.58	0.08
Team 7	0.10	0.91	0.13	0.38	0.27	0.03
Team 8	-0.03	1.12	0.10	0.62	0.73	0.25

*R^2 = square of multiple correlation coefficient

It is fairly apparent from the graphs for the sizing performance that the data is widely distributed about the estimated linear sizing line. This is true for both the absolute and the relative sizing plots. It is clear that there is substantial error in these fits and that neither simple model would explain this data. The data was evaluated to see what could be learned from this data with regard to the two models. It was found that neither model provided a good fit to the data and, furthermore, a more detailed and larger study would be needed to develop an adequate data base for establishing the best model.

Figures 4.32 through 4.39 present a visual confirmation of the regression results presented in Table 4.11. Perhaps the most important observations (and these also follow from the relative regressions) are that:

1. There is substantial team-to-team variation: regression slopes, intercepts, and standard deviations vary greatly from team to team.

2. The error in an individual measurement is quite large in relation to the line slope.

4.3.3 Summary of Sizing Results

Results from the MRR sizing round robin suggest the following.

1. There appears to be some improvement in sizing capability based on comparing the EPRI sizing study with the MRR sizing data base. The MRR data, however, does not support the conclusion that sizing is repeatable and reliable. The MRR data base shows that sizing cracks, especially cracks in the 20% to 50% through-wall range, is very difficult. Sometimes crack tip signals can be detected; often, however, crack tip signals

simply cannot be distinguished from noise. If crack tip signals are detected, accurate results may be obtained.

2. The MRR data shows that crack tip signals are not always easy to detect; hence, the large standard deviation in sizing results. The MRR data base also shows that crack tip sizing procedures seem to have an inherent resolution limit. This resolution limit needs to be recognized when interpreting sizing results and designing performance demonstrations.

3. There is substantial team-to-team variation: regression slopes, intercepts, and standard deviations vary greatly from team to team.

4. The error in an individual measurement is quite large in relation to the regression slope.

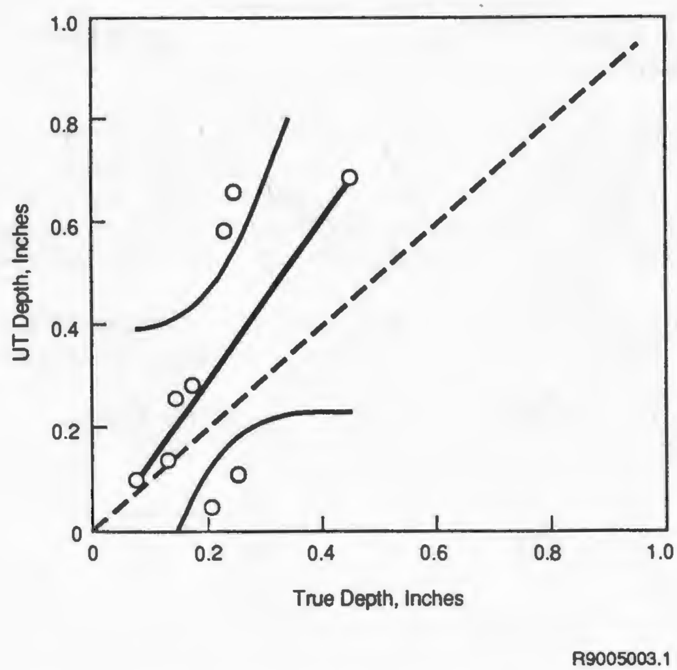


FIGURE 4.32. Absolute Sizing Results for Team 1 with 90% Confidence Bounds

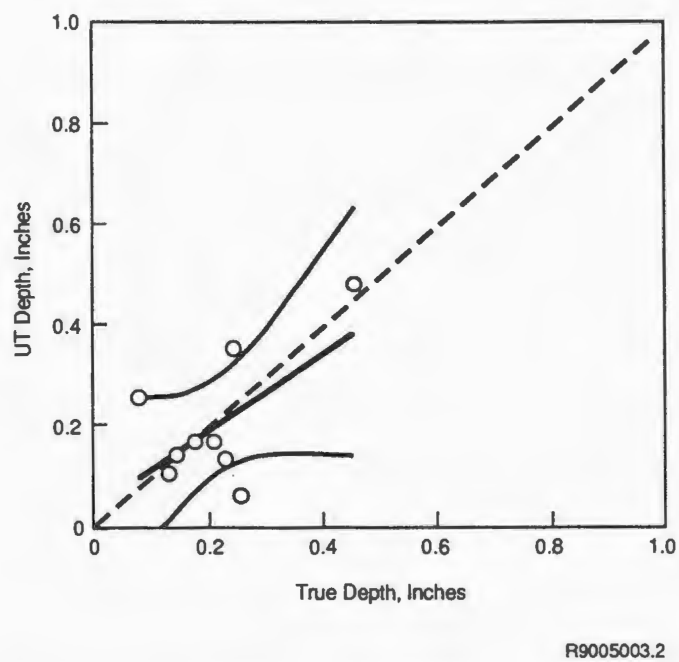


FIGURE 4.33. Absolute Sizing Results for Team 2 with 90% Confidence Bounds

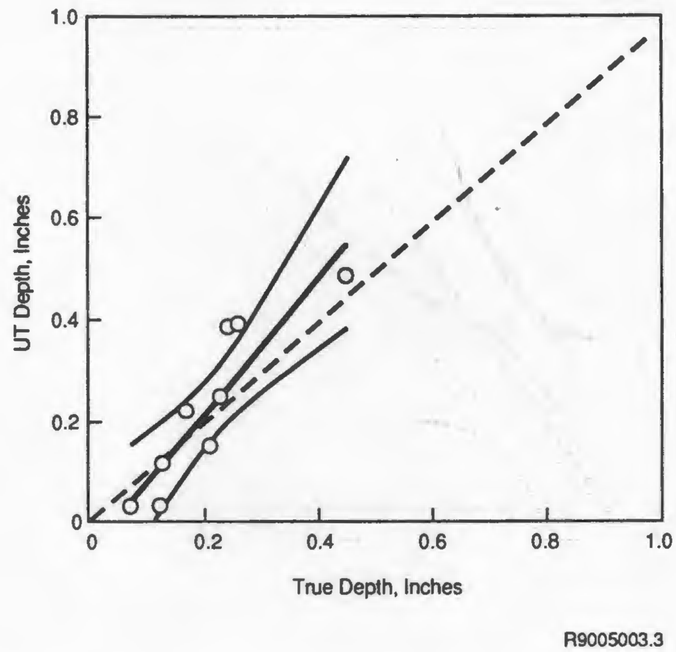


FIGURE 4.34. Absolute Sizing Results for Team 3 with 90% Confidence Bounds

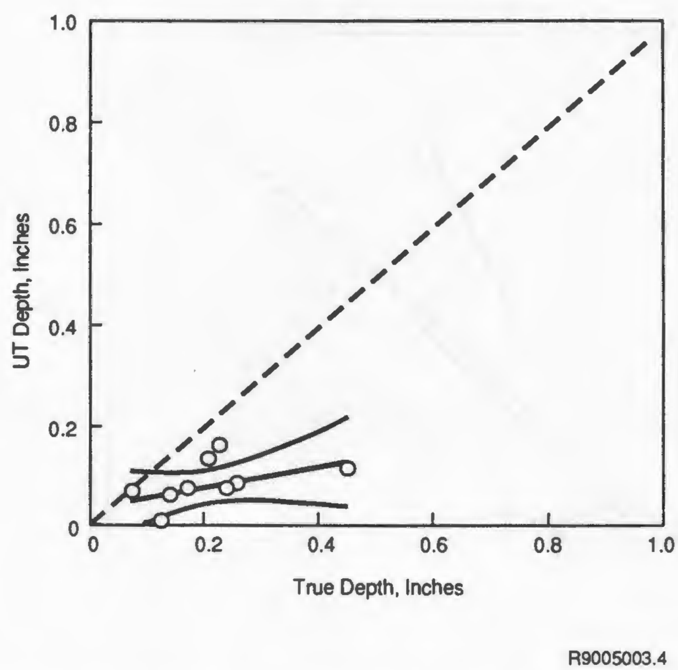


FIGURE 4.35. Absolute Sizing Results for Team 4 with 90% Confidence Bounds

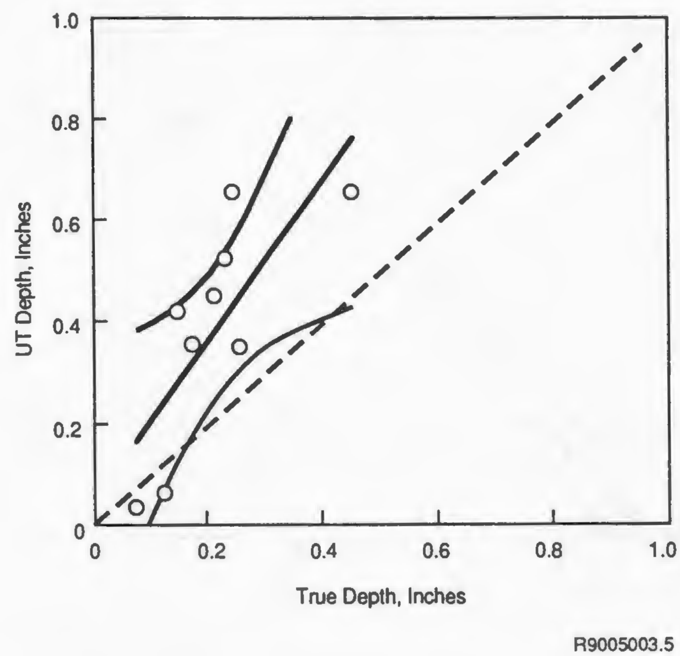


FIGURE 4.36. Absolute Sizing Results for Team 5 with 90% Confidence Bounds

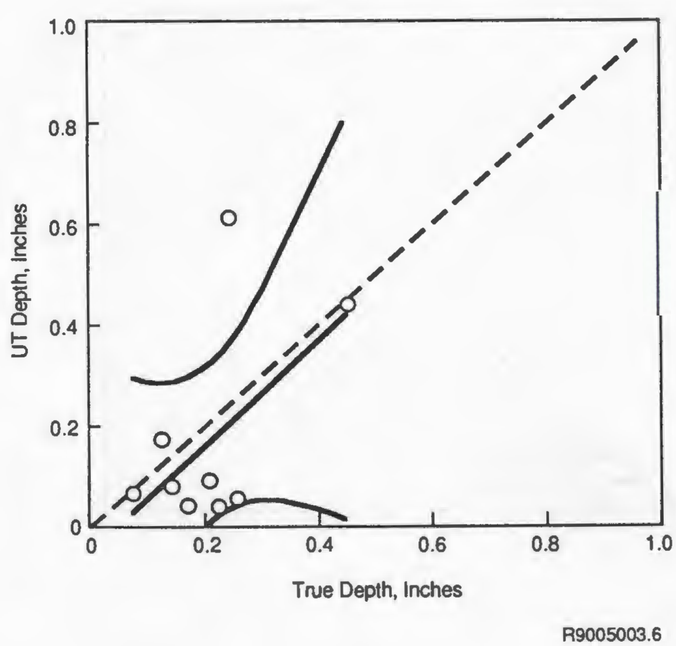


FIGURE 4.37. Absolute Sizing Results for Team 6 with 90% Confidence Bounds

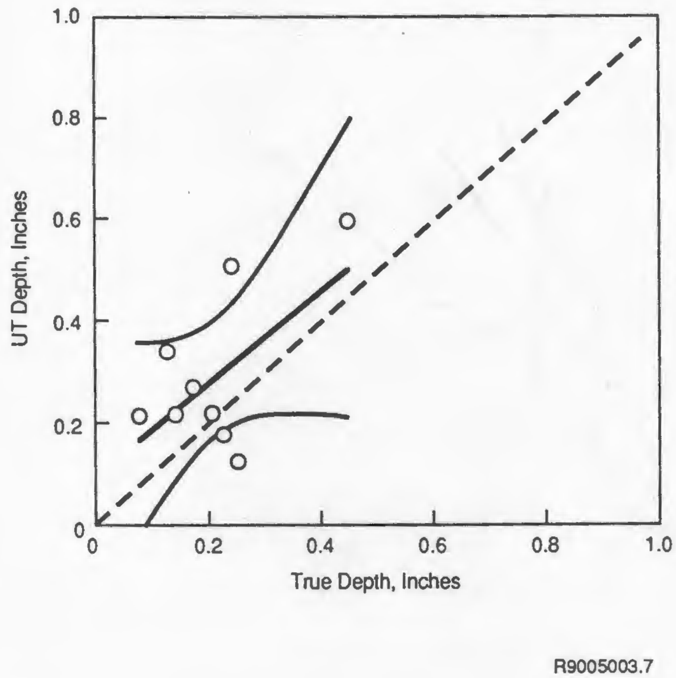


FIGURE 4.38. Absolute Sizing Results for Team 7 with 90% Confidence Bounds

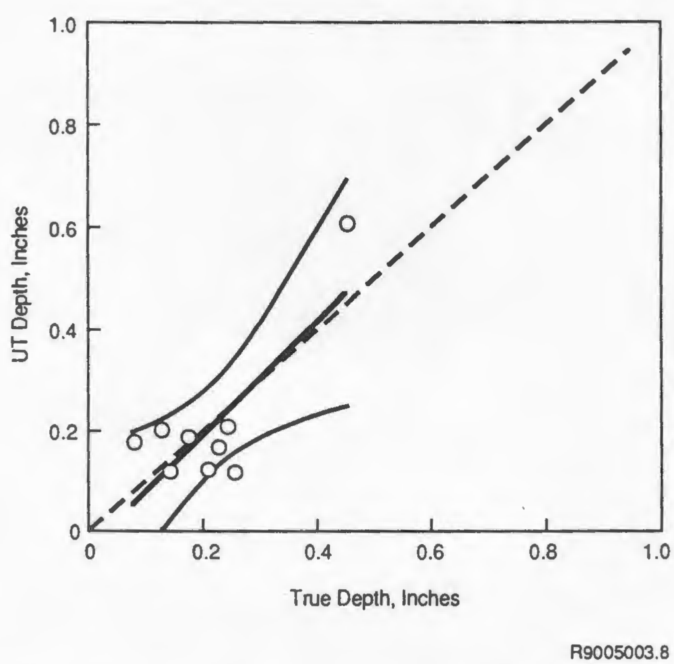


FIGURE 4.39. Absolute Sizing Results for Team 8 with 90% Confidence Bounds

5.0 HUMAN FACTORS STUDY

The human factors study consisted of obtaining and evaluating the responses the technicians provided during one-on-one interviews and a supplementary written questionnaire. A limited evaluation of the equipment features, from the human factors engineering perspective, was conducted.

Inspector Interview - Each MRR technician participated in a one-hour interview with a human factors (HF) specialist. This was a structured type of interview that is usually referred to as a "critical incident interview" by HF specialists. At the beginning of the interview, the technician is asked to describe a significant (good or bad) incident concerning his UT/ISI inspection experience involving IGSCC. The technician is then encouraged to elaborate on the details of the incident, and the interviewer notes those aspects that might be of particular interest to human performance/human reliability. Notes taken during the interview are promptly transcribed using a consistent format to provide a coherent series of interview summaries.

Written Questionnaire - Each MRR technician also completed a written questionnaire to provide information on their background, experience, opinions of the equipment and procedures they had used during their careers, etc. Their opinions were also sought concerning the effect of various elements of the work environment. Other questions were asked to determine their perceptions regarding the task elements involved in performing UT/ISI in nuclear power plants. Each completed questionnaire was pre-reviewed by a human factors specialist, and areas of interpretation and uncertainty were promptly resolved with the technician, as necessary.

5.1 RESPONSE TO SIGNIFICANT INCIDENT INTERVIEW AND WRITTEN QUESTIONNAIRE

5.1.1 Personal, Social, and Organization Factors

Personal, social, and organizational factors can affect technician performance in several ways. Previous training and experience will usually impact performance in a positive manner. The social influence of contemporaries may affect the decision-making process by confirming or disputing signals that have been interpreted. Organizational factors such as the extent and quality of supervision, the support or frustration attributed to the organization, and the faith and trust placed in individuals or teams by the employer or contracting organizations all influence the human reliability aspects of UT/ISI.

The twelve technicians that participated in the MRR ranged in age from 27 to 50 with an average age of 35. They had an average of 7.4 years of experience in performing NDT in the nuclear industry, with an experience range from 1 to 12 years (only two technicians had done extensive NDT work outside of the nuclear industry). Eight of the technicians had UT Level III certification, and four had UT Level II certification. Three were high-school graduates, eight had some college, and one had a college degree. The total number of weeks of formal NDT training (all NDT methods) ranged from 1 to 74 weeks, with the average being about 16 weeks.

As expected, the Level III personnel were somewhat older and had more training and general NDT experience than the Level II technicians. Interestingly, the Level II technicians had slightly more experience in nuclear NDT than did the Level III personnel, and they also had received considerably more training since last certified. It should be noted that, except for the recent IEB 83-02 requirements, there are no general industry requirements for the training of Level III personnel, either prerequisite or periodic.

Table 5.1 shows a demographic comparison of the Level II and Level III MRR participants.

TABLE 5.1. Average Demographic Data for MRR Participants

	Certification Level	
	II	III
Average Age	33.0 years	36.9 years
Average NDT Training (total during career)	7.3 weeks	19.5 weeks
Average General NDT Experience	6.5 years	7.9 years
Average Nuclear NDT Experience	6.5 years	6.0 years
Average NDT Training (since certified to present Level)	5.75 weeks	2.25 weeks

The MRR technicians were asked to rate the extent to which higher level personnel would beneficially assist in evaluating the existence and size of cracks. Nine of 12 felt that consultation with higher level personnel was useful in both these situations. Seven of the 12 expected to have their work checked when suspicious indications were reported. Eight of 12 expected their work to be reinspected only if they reported a rejectable flaw; whereas, only one of 12 thought their work would be reinspected if no indications had been reported. While all technicians agreed that supervisors stressed the importance of finding actual flaws, nine of 12 felt that supervision placed equal, or greater, importance on the correctness of reported flaws. These results imply a general perception that both supervisors and the utilities are more concerned with false calls than with missed flaws.

The master-slave technique involves complex coordination and decision elements not present during normal manual scanning. Several of the technicians indicated that they never trust someone else's scanning techniques and proficiency whenever IGSCC signals are expected. The MRR technicians consistently perceived that a utility's standards of excellence are reflected by overall plant cleanliness and orderliness, and they perceived that their personal performance would respond (positively or negatively) to the expectations of the utility.

The MRR technicians expressed only mild concern regarding unusual work schedules due to plant priorities and work involving other plant personnel. In response to questions concerning fatigue due to shift work and long hours, most technicians felt their performance would remain good for a workday length of up to 12 hours. The combination of too many work hours and too few sleep hours during a given week were expected to substantially reduce performance and overall effectiveness. Five or more continuous weeks of work without a day off was also expected to reduce overall performance.

The MRR technicians considered the following to be important characteristics for "good" NDT technicians:

- ability to concentrate
- understanding of NDT theory
- personal patience
- tolerance of environmental conditions
- manual dexterity
- mathematical ability

5.1.2 Training and Certification Factors

The MRR participants acknowledged the importance of the training they had received at the EPRI NDE Center. The most valuable aspect of this training was the opportunity to work with pipe samples containing known IGSCC, as well as the opportunity to compare IGSCC detection experience with other technicians.

A summary of the responses obtained in this area from the significant incident interview and the written questionnaire is as follows:

- The IEB 83-02 qualification process did not measure ability to detect IGSCC because the technique used for qualification was different than the technique used in the field.

- Utilities will probably verify crack calls but not non-crack calls.

- Better written procedures would improve UT/ISI performance.

- Most valuable aspects of the EPRI NDE Center training:

  - practice on specimens with actual cracks
  - compare techniques with other personnel

- Least valuable aspects of the EPRI NDE Center training:

  - classroom lectures
  - not enough feedback on personal performance

- Average training received -- 2 days/year (classroom: 56%, hands-on: 44%)

- The training at the EPRI NDE Center was very useful, but not specifically related to actual UT/ISI.

- NDT certification levels are not related to UT/ISI ability.

- Different technicians often report different UT/ISI results.

5.1.3 Equipment Factors

Most of the MRR technicians had used a variety of equipment during their careers; usually as many as four or five different systems. Responses to questions regarding equipment preference generally suggested a loyalty to the brand and model with which they were most familiar.

The MRR technicians were generally satisfied with the equipment currently available to them in industry. They preferred equipment with a calibrated time base, along with a display that was clear and sharp. Their experience suggests that keeping the batteries of portable equipment maintained at a proper charge level is an extremely important factor.

5.1.4 Environmental Factors

Environmental conditions in the work place were of concern to the technicians due to the detrimental effects such conditions can have. Protective clothing and gloves make it harder to function in the work area, and a high ambient noise level can interfere with communications between team members. The technicians also mentioned that the face masks of fresh air systems fog due to heat and humidity, which impairs visibility when watching the UT screen.

The MRR technicians mentioned that radiation hazards while conducting IGSCC inspections were of some, but not major, concern. None of these personnel reported reaching their maximum radiation dose during the past few years.

Since UT examination involves tactile feedback of probe location and contact with the surface, bulky gloves plus small search units compound this problem. The equipment presently used for IGSCC detection and sizing is not well suited for use in the hostile environment of nuclear power plant containment.

5.1.5 Procedure Factors

All of the MRR technicians indicated they were using written procedures that exceeded ASME Code requirements. However, they felt that their procedures were not updated often enough to reflect technology and field application changes. When procedures must be approved by the utility but such approval does not include verification of adequacy for specific applications, the technicians felt that procedure deficiencies can become performance-limiting factors.

5.2 HUMAN ENGINEERING EVALUATION OF EQUIPMENT

Most UT equipment that is used for IGSCC detection and sizing involves multi-purpose systems that were designed to perform a broad range of UT inspection tasks. Such equipment designs rarely include basic human engineering considerations. Particular equipment features may be suitable for general usage but not for specific UT/ISI applications; for example, the knobs may be too small to manipulate while wearing gloves.

5.2.1 Purpose

The purpose of the human engineering equipment evaluation was to assess the degree to which conventional UT equipment meets both general human engineering principles and specific requirements for nuclear ISI. Since most of the equipment evaluated was provided by the inspection teams involved in the MRR, this study included a fair representation of equipment currently used for UT/ISI work. It must be recognized that no attempt was made to compare one item of ultrasonic equipment with another, nor to evaluate all of the different makes and models of equipment presently being used for IGSCC inspection.

5.2.2 Equipment Evaluated

Eight commercially available ultrasonic systems, plus a wide variety of search units, were evaluated. Based on responses to the questionnaire, this equipment was representative of the systems most commonly used for UT/ISI work. All equipment items were evaluated using a rating form on which various human engineering criteria were listed. These criteria included:

- Physical characteristics (portability, contamination resistance, attachments)
- Controls (layout, operation, resistance to inadvertent actuation)
- Displays (brightness, clarity, scales, etc.)
- Search unit characteristics (markings, size, and shape).

In addition to the eight units evaluated using the rating form, the control functions of four other UT systems were also examined. The equipment characteristics were evaluated in accordance with the following rating criteria:

1 - <u>Fails to Meet Criteria</u> - Major violations of human engineering design guidelines that are <u>likely</u> to contribute to inappropriate equipment operation or misinterpretation of signals.

2 - <u>Marginal</u> - Minor violations of human engineering design guidelines that <u>may</u> contribute to inappropriate equipment operation or misinterpretation of signals.

3 - Adequate - Violations of human engineering design guidelines are minimized. Equipment design neither increases nor decreases likelihood of operating errors.

4 - Good - Human engineering design guidelines met in most instances. Design features decrease the likelihood of operating errors.

5 - Excellent - Human engineering design guidelines consistently met. Design contributes to proper equipment operation and data interpretation.

5.2.3 Results

Of the eight units that were evaluated, none was so poorly designed as to cause unacceptable performance during UT/ISI work. Conversely, none of the units was so well designed, from a human engineering standpoint, as to recommend its exclusive use for UT/ISI work. It was apparent that most of the equipment was designed on the basis of instrument, circuit design, and manufacturing considerations. The necessity to minimize the size and maximize the portability was also reflected in the small controls and display processes. This problem was complicated by the use of generic ultrasonic equipment with complex controls, which are needed for a variety of applications, rather than the generally simpler but more highly specialized controls needed for a specific application.

Most units were relatively small in size and designed for durability. One problem noted was the difficulty in protecting the equipment from contamination, and for ease of decontamination should it become contaminated. It should be noted that no single piece of equipment was judged totally acceptable on the basis of physical characteristics or equipment control features.

The major area of concern in the operation of complex equipment, such as UT systems, was the way by which the necessary controls were presented and designed. Controls should be arranged to discourage or preclude inadvertent actuation. Where calibration or measurement depends on reading a number associated with a knob or switch position, scales and indicators should facilitate accurate readings. Only two units were judged to be fully adequate with respect to the equipment control features.

Specific problem areas, and their potential impact on UT system performance, were:

1. Size and/or spacing of controls inappropriate for operation while wearing gloves.

2. Positioning of controls does not minimize the opportunity for inadvertent actuation. The use of miniature toggle switches and other fingertip controls compounds this problem when gloves must be worn to perform UT/ISI.

3. Small, difficult to read scales, labels, and displays. Again, this problem is compounded when a protective face mask must be worn for UT/ISI.

Effective ultrasonic inspection is dependent on the ability of the technician to accurately detect minute changes in the display patterns. The three most significant deficient areas found during an evaluation of the UT equipment display characteristics were as follows:

1. Screen design fails to minimize reflected glare. This burdens the inspector with the difficulty of "seeing through" the reflection.

2. Screen brightness, contrast, and focus are fixed via internal controls. Since the inspector cannot adjust these controls for personal comfort or to accommodate changes in ambient conditions, fatigue and distraction are likely to result.

3. Screen size insufficient to allow signal/noise discrimination. The screen size of miniature equipment is such that close screen viewing distances are essential. During UT/ISI in a nuclear power plant, viewing distances of 5 feet or more are not uncommon. Detecting and discriminating the very small signal changes typical of IGSCC under conditions of a small screen size and long viewing distance is very difficult.

This limited human engineering evaluation of eight ultrasonic instruments currently used for UT/ISI applications revealed numerous design deficiencies. The most significant deficiencies included:

- Lack of design consideration for using the equipment in a radiation-contaminated environment.

- Controls that are small and hard to manipulate. These are difficult to use when wearing protective gloves and are also subject to accidental actuation.

- Displays, labels, and scales that are difficult to read at the viewing distances that may be encountered during UT/ISI applications.

- Search units that are difficult to hold and control because of their small size and lack of grip aids.

6.0 CONCLUSIONS AND RECOMMENDATIONS

Following the analysis scheme outlined in this report, a number of conclusions and recommendations can be derived from this study. Other factors surfaced in the course of this study that were identified as important but, unfortunately, the data was not designed to definitely answer these new questions. Further work is needed to provide the data for resolving their impact and influence on inservice inspection.

1. The MRR data base shows that the best manual UT/ISI is capable of acceptable crack detection performance. However, the MRR data base shows a wide variation between best and worst inspectors with the majority of inspectors having unacceptable performance.

2. The MRR data base suggests that there has been little improvement in the POD/FCP capability of inspectors since IEB 83-02 requirements (when passing the test required detecting 4 of 5 cracks) versus before these requirements.

3. The average IGSC crack detection capability of advanced UT technology is slightly better than the "average" performance of manual inspectors that participated in the MRR. This does not mean that manual inspectors are categorically poor; in fact, the best manual inspector's performance exceeded the performance of all advanced inspection systems. However, the "average" performance of advanced technology was better than the average performance of manual inspection. This is important to note because the experience in using advanced technology on actual cracks is limited. The assumption has been that advanced systems place more intelligence into the system and, thus, reduce the experience needs of the inspectors. This data supports this assumption.

4. The MRR demonstrates that ROC analysis can be used to measure the performance of UT inspection systems and illustrates the relationship between POD and FCP.

5. The MRR analysis indicates that no significant difference exists in the POD performance of inspection for long versus short IGSC cracks of similar depths.

6. As expected, the short/shallow cracks were very difficult to detect and did not differ much from the chance performance line in the ROC.

7. The MRR data shows a systematic undersizing (by approximately 1 in.) in the length of long IGSC cracks.

8. The MRR data base shows that far-side inspection is very poor.

9. Crack depth sizing performance has improved in this study compared with round-robin data taken in 1982, before the EPRI crack sizing course. However, the MRR data shows that sizing is not reliable.

10. Crack depth sizing of cracks in the 20-50% through-wall range is very difficult.

11. If crack-tip signals are detected, then accurate sizing may be achieved. However, only a small percentage of the time were crack-tip signals recognized.

12. Large crack-to-crack variability was found as seen by the large standard deviation.

13. All sizing is based on acoustic systems, which have resolution limits (estimated at approximately 0.08 in. for usable frequencies in austenitic piping). These limits must be recognized when interpreting sizing results and designing performance demonstration tests.

14. The MRR data analysis showed that there is a significant need for improving the reliability and reducing the variability of the UT/ISI process and one solution may be by applying the concepts and principles of the human factors engineering discipline. The human factors work conducted in conjunction with the MRR showed that this is a complicated problem with many factors requiring correction; and in order to address these, a comprehensive program should be initiated to reduce their influence on UT/ISI reliability to acceptable levels.

7.0 REFERENCES

1. S. R. Doctor, G. P. Selby, P. G. Heasler, and F. L. Becker. 1982. "Effectiveness and Reliability of Inservice Inspection, A Round Robin Test," Proc. of the Fifth International Conference on Nondestructive Evaluation in the Nuclear Industry. American Society of Metals, Metals Park, Ohio.

2. J. C. Spanner, et al. 1986. Human Reliability Impact on Inservice Inspection, Vol. 1 and 2. NUREG/CR-4436, Pacific Northwest Laboratory, Richland, Washington.

3. W. A. Wheeler, J. C. Spanner, et al. 1986. Human Factors Study Conducted in Conjunction with a Mini-Round Robin Assessment of Ultrasonic Technician Performance. NUREG/CR-4600, Pacific Northwest Laboratory, Richland, Washington.

4. J. A. Swets and R. M. Pickett. 1982. Evaluation of Diagnostic Systems, Methods from Signal Detection Theory, Academic Press, New York.

5. G. J. Dau. 1983. "Ultrasonic Sizing Capability of IGSCC and Its Relation to Flaw Evaluation Procedures," Research Project 1570-2, NDE Center, Electric Power Research Institute, Charlotte, North Carolina, p. 5-2.

6. P. G. Heasler, D. J. Bates, T. T. Taylor, and S. R. Doctor. 1986. Performance Demonstration Tests for Detection of Intergranular Stress Corrosion Cracking., NUREG/CR-4464, Pacific Northwest Laboratory, Richland, Washington.

APPENDIX A

CRITICAL INCIDENT INTERVIEW

CRITICAL INCIDENT INTERVIEW

INSTRUCTIONS TO INTERVIEWEE

Purpose of the Interview

I am involved in a research project designed to identify human factors problems that may affect the reliability of ultrasonic testing for inservice inspection (UT/ISI). We are currently focusing on inspection of BWR piping for intergranular stress corrosion cracking (IGSCC). I'm gathering data from people like yourself who have first-hand, field experience in such inspections.

Non-Attribution of Results

I will not be recording your name or any other identifying information. When I write up the results of these interviews, I will omit any reference to particular plants, contract organizations, or the like.

CRITICAL INCIDENT STATEMENT

I would like you to describe an NDE UT/ISI job-related incident regarding an IGSCC inspection that stands out in your memory as particularly important, either because a job was especially successful or especially unsuccessful. A situation, for example, in which a particularly difficult inspection resulted in the discovery of a serious flaw, or, on the other hand, where an inspection resulted in the removal of a pipe section that turned out to be good. Please describe one incident only and focus on one that you are especially familiar with. Take a few minutes to think about it, then begin by giving me some general background on the incident.

Interviewee was directly involved: yes_____ no_____

Inspection was done by
- contractor ______________________
- utility ______________________

and supervised by
- contractor ______________________
- utility ______________________

Incident took place:
- 0-2 years ago ______________________
- 2-4 years ago ______________________
- >4 years ago ______________________

Incident involved
- pipe exam only ______________________
- exam and interpretation ______________________
- exam through sizing ______________________
- other (specify) ______________________

Qualifications of technicians were:

	Tech 1	Tech 2
Level I		
Level II		
Level III		

Technicians had attended EPRI NDE IGSCC training course yes_____ no_____

Inspection was conducted in a radiation zone yes_____ no_____

Protective gear worn:
- None ______________________
- Full (including breathing apparatus) ______________________
- Partial (specify) ______________________
- Suit ______________________
- Shoe covers ______________________
- Gloves ______________________
- Face Mask ______________________
- Communications gear ______________________

UT equipment used:

UT Instrument (Brand) ____________________

2nd Instrument (Brand) ____________________

Search Units

Straight beam (type) ____________________

Straight beam (type) ____________________

Angle beam (type) ____________________

45 ____________________

60 ____________________

Other ____________________

Was a procedure used? yes____ no____

(If yes, describe) ____________________

Were previous records reviewed yes____ no____

(If yes, what kind?) ____________________

Prior to inspection ____________________

After inspection ____________________

Plant status was:

At power ____________________
Shut down ____________________
In construction ____________________

What were the working conditions like?

High temp/humidity ____________
Poor lighting/visibility ________
High noise ____________________
Cramped quarters/awkward working positions ____________

Technicians working hours were:

____________ hours between breaks
____________ hours on the job

Job took place from:

7:00 a.m. to 3:00 p.m. ____________
3:00 p.m. to 11:00 p.m. ____________
11:00 p.m. to 7:00 a.m. ____________

Management/supervision stressed:

Doing careful exam ________________
Being 100% confident before calling crack ________________

What specific factors does interviewee attribute success or failure to?

____

At task level, what is task frequency, difficulty, and importance?

____

Did inspectors receive feedback on results of this inspection?

____

APPENDIX B

NDT TECHNICIANS SURVEY

NDT TECHNICIANS SURVEY

In recent years there has been increased importance placed on nondestructive testing (NDT) in the nuclear field. We are interested in finding out what kinds of factors influence your performance during ultrasonic testing for inservice inspection (UT/ISI) at nuclear power plants. We are particularly concerned with the detection of intergranular stress corrosion cracking (IGSCC). Please answer all of the following questions to the best of your ability. These questions ask your opinion on various issues. There are no right or wrong answers. We are only interested in your opinion. Your answers are confidential, and the results of this survey will be aggregated and individual and company anonymity will be maintained.

Before you fill out the survey questions regarding NDT, we would like to find out some things about your background. This information will be used to help report the information collected in the survey.

A. We would like to know what kind of and how much training you have had for doing UT/ISI for IGSCC and other flaws. In the spaces below, please indicate where you received your training and how long the training lasted (in weeks).

Received Training at ...	Weeks of Training
a. --	------------------- Weeks
b. --	------------------- Weeks
c. --	------------------- Weeks
d. --	------------------- Weeks
e. --	------------------- Weeks
f. --	------------------- Weeks
g. --	------------------- Weeks
h. --	------------------- Weeks

B. What is your highest level of NDT certification for UT? (circle one)

1. Level 1
2. Level 2
3. Level 3

C. What is your age in years?

_______ years

D. How many years of experience have you had doing UT/ISI work? (please write in)

I have worked _______ years doing UT/ISI.

E. How many years of experience have you had doing UT/ISI at nuclear power plants either as utility staff or contractor staff? (please write in)

I have worked _______ years doing UT/ISI at nuclear power plants.

F. Do you wear ______ while doing UT/ISI? (circle one)

1. Glasses
2. Contact lenses
3. Do not wear glasses or contact lenses

G. Please circle one number below to indicate how much formal education you have had.

1. Less than a high school education
2. High school degree or equivalent
3. Some college
4. Two-year college degree
5. Four-year college degree
6. Some graduate training

H. How many weeks/hours of formal training have you received since you were Certified to your present level?

______ weeks, or
______ hours

I. What percentage of this training was on theory (i.e., classroom training), and what percentage was on the practical aspects (i.e., hands-on laboratory training, or on-the-job training)?

Classroom training _____%
Hands-on-training _____%

Total 100 %

A. Please read the list of ultrasonic testing instruments below. For every type of equipment, circle one number that best states your opinion on how good the equipment is for use in ultrasonic testing during inservice inspections (UT/ISI) for intergranular stress corrosion cracking (IGSCC) and other flaws. If you have never used the equipment, please circle the number 6. For the equipment that you have used, circle one number from 1 through 5 according to the following scale.

1 = Very Good
2 = Good
3 = Average
4 = Poor
5 = Very Poor

6 = Have Never Used This Equipment Before

	How good is equipment for UT/ISI applications?					
EQUIPMENT TYPE	Very Good	Good	Average	Poor	Very Poor	Never Used
a. Nortec NDT-131 Ultrascope.......................	1	2	3	4	5	6
b. Nortec NDT-131D Digital Ultrascope..............	1	2	3	4	5	6
c. Automation Industries Model UJ Ultrasonic Reflectoscope.......................	1	2	3	4	5	6
d. Automation Industries M90 Reflectoscope.........	1	2	3	4	5	6
e. Automation Industries M91 Reflectoscope.........	1	2	3	4	5	6
f. Automation Industries M91C Reflectoscope........	1	2	3	4	5	6
g. Automation Industries SM90D Reflectoscope.......	1	2	3	4	5	6
h. Sonic Mark I Series B Portable Ultrasonic Tester..	1	2	3	4	5	6
i. Sonic Mark IV/Model 120 Portable Ultrasonic Flaw Detector.................................	1	2	3	4	5	6
j. Krautkramer-Branson USK7 Ultrasonic Flaw Detector......................................	1	2	3	4	5	6
k. Krautkramer-Branson USL-38 Ultrasonic Flaw Detector.................................	1	2	3	4	5	6
l. Krautkramer-Branson USL42 Ultrasonic Flaw Detector and Thickness Tester..................	1	2	3	4	5	6
m. Krautkramer-Branson USL48 Ultrasonic Flaw Detector and Thickness Tester..................	1	2	3	4	5	6
Other UT flaw detection systems (please specify)						
n.	1	2	3	4	5	6
o.	1	2	3	4	5	6

B. Considering, in general, the equipment that you have used for UT/ISI, what types of problems, if any, have you had with different aspects of the equipment? Use the following code and circle one number from 1 through 5 for each potential problem area. While a problem area is probably dependent on the type of equipment, please rate the problem areas below by averaging over all the equipment that you have used.

1 = No Problem
2 = Minor Problem
3 = Somewhat of a Problem
4 = Important Problem
5 = Major Problem

POTENTIAL PROBLEM AREA	No Problem	Minor Problem	Somewhat Problem	Important Problem	Major Problem
a. size of equipment................	1	2	3	4	5
b. weight of equipment..............	1	2	3	4	5
c. cables and wires.................	1	2	3	4	5
d. instability......................	1	2	3	4	5
e. complexity.......................	1	2	3	4	5
f. display..........................	1	2	3	4	5
g. controls.........................	1	2	3	4	5
h. reliability......................	1	2	3	4	5
i. ruggedness.......................	1	2	3	4	5
j. lack of calibrated time base on the RF display...........	1	2	3	4	5

C. When do you refer to written procedures during the actual UT/ISI process for IGSCC? (circle all that apply)

1. To verify scanning requirements
2. To verify amplitude of recordable signals
3. To assist in signal interpretation
4. To assist in sizing defects
5. During preparation of the data report

D. If well-written procedures with more explicit instructions and illustrations were available for all inspections and you carefully followed these during the inspection, how do you think that this would affect your ability to detect flaws (IGSCC and other flaws)? (circle one number)

1. Would make it much harder to find flaws.
2. Would make it somewhat harder to find flaws.
3. Would have no effect on finding flaws.
4. Would make it somewhat easier to find flaws.
5. Would make it much easier to find flaws.

E. Which type of procedures do you use? (circle one)

1. Minimum ASME code procedures
2. Procedures that exceed the ASME code minimum requirements

F. When doing UT/ISI, you have to record your findings on a data recording form. Because of the way the form is written and the information needs laid out on the recording form, the forms may be easy or hard to fill out. How easy or hard is it to fill out YOUR EMPLOYER'S FORMS? (circle one)

1. Very easy to fill out
2. Easy to fill out
3. Average to fill out
4. Hard to fill out
5. Very hard to fill out

G. When doing UT/ISI, you have to record your findings on a data recording form. Because of the way the form is written and the information needs laid out on the recording form, the forms may be easy or hard to fill out. On the average, how easy or hard has it been to fill out THE UTILITY FORMS? (circle one)

1. Very easy to fill out
2. Easy to fill out
3. Average to fill out
4. Hard to fill out
5. Very hard to fill out
6. Never filled out utility form

H. If data recording forms are poorly designed or laid out, this might cause an error in recording the findings. If the forms are well designed or laid out, then these errors might not happen. To what extent do you believe that the design and layout of the data recording forms cause errors in recording your findings? (circle one)

1. Form design and layout lead to no errors in data recording.
2. Form design and layout lead to very few errors in data recording.
3. Form design and layout lead to some errors in data recording.
4. Form design and layout lead to a number of errors in data recording.
5. Form design and layout lead to many errors in data recording.

I. Over the course of many inspections, what percent of the time do you think errors are made in recording your findings? (please fill in your response)

Errors are made in recording findings approximately _____% of the time.

J. While scanning pipe during a UT/ISI, what is the approximate viewing distance, on the average, between you and the UT instrument display screen? (circle one)

1. Less than 12 inches
2. 12 inches to 24 inches
3. 24 inches to 36 inches
4. 36 inches to 48 inches
5. 48 inches to 60 inches
6. More than 60 inches

K. UT/ISI is often done under less-than-ideal working conditions. We would like your opinion about how the following working conditions affect performance. Please circle one number for each work element or condition below using the following scale:

1 = Has No Effect on inspection performance
2 = Makes inspection a Little Harder
3 = Makes inspection Harder
4 = Makes inspection Much Harder
5 = Makes inspection Very Much Harder

	Effect on Performance				
WORK ELEMENT OR CONDITION	No Effect	Little Harder	Harder	Much Harder	Very Much
a. work space..........................	1	2	3	4	5
b. awkward working position...........	1	2	3	4	5
c. temperature........................	1	2	3	4	5
d. humidity...........................	1	2	3	4	5
e. lighting...........................	1	2	3	4	5
f. noise..............................	1	2	3	4	5
g. vibration..........................	1	2	3	4	5
h. radiation..........................	1	2	3	4	5
i. wearing gloves.....................	1	2	3	4	5
j. wearing face mask..................	1	2	3	4	5
k. wearing protective suit and boots...	1	2	3	4	5
l. wearing breathing apparatus.........	1	2	3	4	5
m. working around union personnel work schedules.....................	1	2	3	4	5

L. Listed below are the major tasks that are carried out during UT/ISI for IGSCC. In the following three questions we are interested in your opinion on how important the task is to overall performance, how difficult the task is, and how frequently the task is carried out.

a. Please rate the following inspection subtasks according to the importance of the subtask to overall UT/ISI performance:

1 = Not Important
2 = Of Minor Importance
3 = Somewhat Important
4 = Very Important
5 = Of Critical Importance

	Importance of Subtask to Overall UT/ISI Performance for IGSCC				
INSPECTION SUBTASKS	Not	Minor	Somewhat	Very	Critical
a. checking pipe for leakage or irregularities	1	2	3	4	5
b. review of previous records.....................	1	2	3	4	5
c. equipment calibration..........................	1	2	3	4	5
d. weld examination...............................	1	2	3	4	5
e. interpretation of signals......................	1	2	3	4	5
f. sizing...	1	2	3	4	5
g. data recording.................................	1	2	3	4	5

b. Now, please rate the same six subtasks according to how difficult or easy the subtasks are, on the average, to perform correctly:

1 = Very Easy
2 = Somewhat Easy
3 = Neither too Easy nor too Difficult
4 = Somewhat Difficult
5 = Very Difficult

INSPECTION SUBTASKS	Difficulty of Carrying out Subtask Correctly during UT/ISI for IGSCC				
	Very Easy	Somewhat Easy	Neither	Somewhat Diff.	Very Diff.
a. checking pipe for leakage or irregularities	1	2	3	4	5
b. review of previous records....................	1	2	3	4	5
c. equipment calibration.........................	1	2	3	4	5
d. weld examination..............................	1	2	3	4	5
e. interpretation of signals.....................	1	2	3	4	5
f. sizing..	1	2	3	4	5
g. data recording................................	1	2	3	4	5

c. Finally, please rate the same six subtasks according to how frequently the subtasks are performed over the course of one pipe inspection:

1 = Do not perform
2 = Perform 1 or 2 times per inspection
3 = Perform 3 or 4 times per inspection
4 = Perform 5 or 6 times per inspection
5 = Perform more than 6 times per inspection

INSPECTION SUBTASKS	Frequency That Subtask Is Carried Out for UT/ISI for One Inspection				
	Do not Perform	1-2	3-4	5-6	6
a. checking pipe for leakage or irregularities	1	2	3	4	5
b. review of previous records....................	1	2	3	4	5
c. equipment calibration.........................	1	2	3	4	5
d. weld examination..............................	1	2	3	4	5

Thinking back over all of the ISIs that you performed over the past year, on the average, how many indications did you see per ISI? (please write in)

On the average, I saw approximately ______ indications per ISI.

Thinking back over all of the ISIs that you performed over the past year, on the average, how many flaws did you size per ISI? (please write in)

On the average, I sized approximately ______ flaws per ISI.

Thinking back over all of the ISIs that you performed over the past year, on the average, how many pipe welds did you inspect per ISI?

On the average, I inspected approximately ______ welds per ISI.

M. The calibration procedure used in preparing for UT/ISI contains several areas of potential importance to accuracy in conducting the inspection. Please rate each area of calibration according to the following scale:

1 = Not important
2 = Of Minor Importance
3 = Somewhat Important
4 = Very Important
5 = Of Critical Importance

	Importance of Calibration Subtask to Overall UT/ISI Accuracy				
CALIBRATION AREA	Not	Minor	Some-what	Very	Critical
a. temperature of calibration blocks................	1	2	3	4	5
b. UT properties of calibration block the same as properties of component or weld..............	1	2	3	4	5
c. adjustment of time base and time delay...........	1	2	3	4	5
d. set up of scale and DAC curve....................	1	2	3	4	5
e. verification of beam angle and departure point...	1	2	3	4	5
f. adjustment of gain control.......................	1	2	3	4	5
g. filling out calibration record..................	1	2	3	4	5

N. Listed below are a number of subtasks carried out during the inspection step for UT/ISI. In the following three questions we are interested in your opinion on how important the subtask is to the overall inspection task, how difficult the subtask is, and how frequently you carry out the subtask.

a. Please rate the following inspection subtasks according to how important the subtasks are to carrying out the overall inspection task correctly using the following code:

1 = Not Important
2 = Of Minor Importance
3 = Somewhat Important
4 = Very Important
5 = Of Critical Importance

	Importance of Subtask to Inspection Accuracy				
INSPECTION STEP SUBTASKS	Not	Minor	Some-what	Very	Critical
a. locating the top of the pipe.....................	1	2	3	4	5
b. marking off coordinates on the pipe..............	1	2	3	4	5
c. assessing the thickness of the pipe..............	1	2	3	4	5
d. localizing and marking off counterbore...........	1	2	3	4	5
e. scanning for cracks..............................	1	2	3	4	5
f. estimating signal strength in relation to the distance amplitude correction (DAC)..........	1	2	3	4	5
g. checking calibration.............................	1	2	3	4	5
h. recording location of indications................	1	2	3	4	5

b. Please rate the subtasks according to how difficult the subtasks are to perform correctly, on the average:

1 = Very Easy
2 = Somewhat Easy
3 = Neither too Easy nor too Difficult
4 = Somewhat Difficult
5 = Very Difficult

INSPECTION STEP SUBTASKS	Difficulty in Carrying out Subtask Accurately				
	Very Easy	Somewhat Easy	Neither	Somewhat Diff.	Very Diff.
a. locating the top of the pipe....................	1	2	3	4	5
b. marking off coordinates on the pipe.............	1	2	3	4	5
c. assessing the thickness of the pipe.............	1	2	3	4	5
d. localizing and marking off counterbore..........	1	2	3	4	5
e. scanning for cracks.............................	1	2	3	4	5
f. estimating signal strength in relation to the distance amplitude correction (DAC)..........	1	2	3	4	5
g. checking calibration............................	1	2	3	4	5
h. recording location of indications...............	1	2	3	4	5

c. Finally, please rate the same subtasks according to how frequently the subtasks are carried out over the course of one inspection:

1 = Do not perform
2 = Perform 1 or 2 times per inspection
3 = Perform 3 or 4 times per inspection
4 = Perform 5 or 6 times per inspection
5 = Perform more than 6 times per inspection

INSPECTION STEP SUBTASKS	Frequency That Subtask Is Carried Out for UT/ISI for One Inspection				
	Do not Perform	1-2	3-4	5-6	6
a. locating the top of the pipe....................	1	2	3	4	5
b. marking off coordinates on the pipe.............	1	2	3	4	5
c. assessing the thickness of the pipe.............	1	2	3	4	5
d. localizing and marking off counterbore..........	1	2	3	4	5
e. scanning for cracks.............................	1	2	3	4	5
f. checking calibration............................	1	2	3	4	5

O. When working on UT/ISI, how many hours a day (including normal rest and lunch breaks) do you feel you can work and still meet acceptable quality standards? (circle one)

1. 4 hours or less
2. 5 or 6 hours
3. 7 or 8 hours
4. 9 or 10 hours
5. 11 or 12 hours
6. 13 or 14 hours
7. 15 hours or more

P. People vary in the amount of time off (that is, time for meals, leisure activities, and sleep) they require between daily jobs. How much time off do you feel you need in order to still meet acceptable quality standards for UT/ISI from one day to the next? (circle one)

1. Less than 4 hours
2. 4 or 5 hours
3. 6 or 7 hours
4. 8 or 9 hours
5. 10 or 11 hours
6. 12 hours or more

Q. People vary in the amount of sleep that they require between daily jobs. How much sleep do you feel you need after working the following shifts?

a. After an 8-hour shift, I need about _____ hours of sleep in order to meet acceptable quality standards on my job the next day.

b. After a 12-hour shift, I need about ______ hours of sleep in order to meet acceptable quality standards on my job the next day.

c. After a 16-hour shift, I need about ______ hours of sleep in order to meet acceptable quality standards on my job the next day.

R. On the average, how do you feel overtime (that is, working in excess of 8 hours/day and/or in excess of a 5-day work week) affects the quality of your inspections? (circle one)

1. Decreases the quality of my work greatly
2. Decreases the quality of my work to some extent
3. Does not affect the quality of my work
4. Increases the quality of my work to some extent
5. Increases the quality of my work greatly

S. On the average, how much overtime (that is, time over 40 hours in a 5-day work week) do you feel you can work during a normal outage period and still meet acceptable quality standards? (circle one)

1. Less than 8 hours per week
2. 8 to 12 hours per week
3. 13 to 16 hours per week
4. 17 to 20 hours per week
5. 21 to 24 hours per week
6. More than 24 hours per week

T. People may feel particularly sharp and alert at different times of the day or night; at other times they may feel sleepy or run down. For each of the times of day listed below circle how you usually feel during that time using the following code:

1 = Very Alert
2 = Somewhat Alert
3 = Neither Alert nor Sleepy
4 = Somewhat Sleepy or Run Down
5 = Very Sleepy or Run Down

TIME	At that time of day, I feel ... Very Alert	Somewhat Alert	Neither	Somewhat Sleepy	Very Sleepy
a. 8 a.m. to 12 noon............	1	2	3	4	5
b. 12 noon to 4 p.m.............	1	2	3	4	5
c. 4 p.m. to 8 p.m..............	1	2	3	4	5
d. 8 p.m. to 12 midnight.........	1	2	3	4	5
e. 12 midnight to 4 a.m..........	1	2	3	4	5
f. 4 a.m. to 8 a.m..............	1	2	3	4	5

U. During what hours do you normally work when performing UT/ISI? (please write in)

I typically work from ______ a.m. to ______ p.m.

V. What do you think is the MOST useful phase of the EPRI IGSCC training? (circle one)

1. Theory
2. Classroom lecture on technique
3. Hands-on experience with known standards
4. Hands-on experience with unknown samples
5. Other (please write in) ________________________________

W. What do you think is the LEAST useful phase of the EPRI IGSCC training? (circle one)

1. Theory
2. Classroom lecture on technique
3. Hands-on experience with known standards
4. Hands-on experience with unknown samples
5. Other (please write in) ________________________________

X. What aspect of the EPRI IGSCC training do you feel is most important? (circle one)

1. Availability of good equipment
2. Availability of high-quality instructors
3. Availability of samples and standards to practice on
4. Opportunity to share ideas and techniques with other technicians
5. Other (please write in) ________________________________

Y. If you were to do UT/ISI for IGSCC directly behind another inspector who has the same certification level as you do, how likely do you think it would be that you would find flaws he had not found or that you would not find flaws that he had found? (circle one)

1. Very Unlikely
2. Somewhat Unlikely
3. Hard to Tell
4. Somewhat Likely
5. Very Likely

Z. If you were to do UT/ISI for IGSCC directly behind another inspector who had the next lower certification level as you do, how likely do you think it would be that you would find flaws he had not found or that you would not find flaws that he had found? (circle one)

1. Very Unlikely
2. Somewhat Unlikely
3. Hard to Tell
4. Somewhat Likely
5. Very Likely

AA. When doing UT/ISI for IGSCC, how useful do you find the following aids in making a determination of the existence of a crack? Use the following code:

1 = Not at all Useful
2 = Slightly Useful
3 = Somewhat Useful
4 = Very Useful
5 = Extremely Useful

	How useful is aid in determining the existence of a crack?				
AID	Not Useful	Slightly Useful	Somewhat Useful	Very Useful	Extremely Useful
a. other NDT inspectors at same level.......	1	2	3	4	5
b. other NDT inspectors at higher level.....	1	2	3	4	5
c. records of past ultrasonic inspections...	1	2	3	4	5
d. as-built drawings.......................	1	2	3	4	5
e. plant engineering staff..................	1	2	3	4	5
f. radiographic records of the weld area....	1	2	3	4	5

BB. When doing UT/ISI for IGSCC, how useful do you find the following aids in making a determination of the size of a crack? Use the following code:

1 = Not at all Useful
2 = Slightly Useful
3 = Somewhat Useful
4 = Very Useful
5 = Extremely Useful

AID	Not Useful	Slightly Useful	Somewhat Useful	Very Useful	Extremely Useful
	How useful is aid in determining the size of a crack?				
a. other NDT inspectors at same level.......	1	2	3	4	5
b. other NDT inspectors at higher level.....	1	2	3	4	5
c. records of past ultrasonic inspections...	1	2	3	4	5
d. as-built drawings........................	1	2	3	4	5
e. plant engineering staff..................	1	2	3	4	5
f. radiographic records of the weld area....	1	2	3	4	5

CC. In an average work day how many hours do you spend actually scanning pipe during an UT/ISI? (circle one)

1. Less than 2 hours
2. 3 to 4 hours
3. 5 to 6 hours
4. 7 to 8 hours
5. More than 8 hours

DD. When you took the EPRI IGSCC qualifying examination, did you use the same indications and criteria for calling an indication a "crack" as you use during a UT/ISI at a nuclear power plant? (circle one)

1. No, I called "crack" with much less evidence at EPRI than at a nuclear power plant.
2. No, I called "crack" with less evidence at EPRI than at a nuclear power plant.
3. Yes, I called "crack" using about the same evidence at EPRI as I require at a nuclear power plant.
4. No, I required more evidence for calling "crack" at EPRI than at a nuclear power plant.
5. No, I required much more evidence for calling "crack" at EPRI than at a nuclear power plant.

EE. Considering the specific training at EPRI and other places concerning how to detect IGSCC, how do you feel about such training? (circle one)

1. This type of training is of no real value to us in the field.
2. This type of training is a little help to us in the field.
3. This type of training is of some help to us in the field.
4. This type of training is very helpful to us in the field.
5. This type of training is essential to us in the field.

FF. How important do you think the following characteristics are to being a good NDT technician? Use the following code.

1 = Not at all Important
2 = Of Minor Importance
3 = Somewhat Important
4 = Very Important
5 = Of Critical Importance

CHARACTERISTIC	How important are the characteristics to being a good NDT Technician? Not	Minor	Some-what	Very	Critical
a. muscle strength	1	2	3	4	5
b. manual (finger and hand) dexterity	1	2	3	4	5
c. small body size	1	2	3	4	5
d. long arm reach	1	2	3	4	5
e. ability to concentrate	1	2	3	4	5
f. knowledge of equipment	1	2	3	4	5
g. experience in doing this type of work	1	2	3	4	5
h. understanding of NDT theory	1	2	3	4	5
i. following a sequence exactly (methodical)	1	2	3	4	5
j. patience	1	2	3	4	5
k. physical agility	1	2	3	4	5
l. good vision	1	2	3	4	5
m. ability to get along with others	1	2	3	4	5
n. good hearing	1	2	3	4	5
o. mathematical ability	1	2	3	4	5
p. ability to work in poor working conditions	1	2	3	4	5
q. ability to work long hours	1	2	3	4	5

GG. Some NDT technicians believe that it is possible to reliably detect IGSCC through a weld in austenitic stainless steel pipe (that is, far-side detection) with present equipment. How likely do you think it is that you can detect far-side IGSCC cracks in austenitic stainless steel pipe? (circle one)

1. I can probably do far-side detection less than 10% of the time in austenitic stainless steel with present equipment.
2. I can probably do far-side detection about 10% to 30% of the time in austenitic stainless steel with present equipment.
3. I can probably do far-side detection about 30% to 50% of the time in austenitic stainless steel with present equipment.
4. I can probably do far-side detection about 50% to 70% of the time in austenitic stainless steel with present equipment.
5. I can probably do far-side detection more than 70% of the time in austenitic stainless steel with present equipment.

HH. When you have reached your allowable radiation dosage, what sort of work do you do? (circle one)

1. Work NDT in a non-radiation environment
2. Do work unrelated to NDT
3. Take or instruct NDT courses
4. Do not work
5. Other (please write in) ______________________________

II. For the statements below, please state your agreement or disagreement with each statement by circling one number to the right of the statement using the following code:

1 = Strongly Agree
2 = Agree
3 = Neither Agree nor Disagree
4 = Disagree
5 = Strongly Disagree

Statement	Strongly Agree	Agree	Neither	Dis-agree	Strongly Disagree
a. If I record a suspicious indication during UT/ISI for IGSCC at a nuclear power plant, I know I will be carefully questioned and my work re-inspected by utility or plant staff.	1	2	3	4	5
b. If I record a rejectable flaw during UT/ISI for IGSCC at a nuclear power plant, I know I will be carefully questioned and my work re-inspected by utility or plant staff.	1	2	3	4	5
c. If I record no suspicious indications during UT/ISI for IGSCC at a nuclear power plant, I know I will be carefully questioned and my work re-inspected by utility or plant staff.	1	2	3	4	5
d. NDT technician certification accurately represents a technician's ability to perform reliable UT/ISI for IGSCC.	1	2	3	4	5
e. Utility management constantly stresses the importance of finding any flaws that may exist.	1	2	3	4	5
f. The good NDT technicians got high scores on the EPRI training course exam.	1	2	3	4	5
g. The radiation dose that I get while doing my job at a nuclear power plant does not negatively affect my performance.	1	2	3	4	5

Statement	Strongly Agree	Agree	Neither	Dis-agree	Strongly Disagree
h. The poor NDT technicians got low scores....... on the EPRI training course exam.	1	2	3	4	5
i. My company supervisors constantly stress...... the importance of finding any flaws that may exist.	1	2	3	4	5
j. The overall NDT certification process is...... highly rated to on-the-job performance.	1	2	3	4	5
k. The EPRI training course exam scores are...... unrelated to a technician's UT/ISI performance.	1	2	3	4	5
l. My company supervisors constantly stress...... the importance of being certain about the existence of any flaws that I detect.	1	2	3	4	5
m. Utility management constantly stresses........ the importance of being certain about the existence of any flaws that I detect.	1	2	3	4	5

JJ. When I perform UT/ISI on the piping of a boiling water reactor that has operated for approximately 10 years, I generally expect to: (circle one)

1. Find no signs of IGSCC
2. Find some signs of IGSCC
3. Find numerous signs of IGSCC
4. Find rejectable IGSCC flaws
5. I have no expectations regarding the presence or absence of IGSCC at boiling water reactors.

KK. When I took the EPRI IGSCC training course examination, I expected to:

1. Find some signs of IGSCC
2. Find numerous signs of IGSCC
3. Find some rejectable IGSCC flaws
4. Find numerous rejectable IGSCC flaws
5. I had no expectations regarding the presence or absence of IGSCC on the EPRI samples.

LL. In general, during my UT/ISI work, ... : (circle one).

1. I only call an indication an IGSCC crack if I am definitely sure that I am right.
2. If an indication is probably an IGSCC crack, then I call it a crack.
3. If an indication is possibly an IGSCC crack, then I call it a crack.

MM. How would you rate the examination that you took at the completion of your EPRI training course? (circle one)

1. Highly related to the situations that I see in the field
2. Somewhat related to the situations that I see in the field
3. Somewhat unrelated to the situations that I see in the field
4. Highly unrelated to the situations that I see in the field

NN. EPRI has written a generic procedure for UT/ISI to detect IGSCC. If you have used this procedure in the field, please rate its usefulness using the code below: (circle one)

1. Have not used EPRI IGSCC procedure in the field
2. Not at all useful
3. A little useful
4. Somewhat useful
5. Very useful
6. Essential

OO. How often do you receive feedback on the accuracy of your inspection performance during the course of a year?

1. I never receive feedback on the accuracy of my inspection performance.
2. I receive feedback on the accuracy of my inspection performance on less than 10% of the jobs that I do.
3. I receive feedback on the accuracy of my inspection performance on about 10% to 30% of the jobs that I do.
4. I receive feedback on the accuracy of my inspection performance on about 30% to 50% of the jobs that I do.
5. I receive feedback on the accuracy of my inspection performance on about 50% to 70% of the jobs that I do.
6. I receive feedback on the accuracy of my inspection performance on over 70% of the jobs that I do.

PP. You probably have some thoughts about how you could improve the speed and accuracy of your performance doing UT/ISI. What two or three things can you think of that you think might help improve the speed and accuracy of your UT/ISI performance? (please write in)

a. --

----

b. --

----

c. --

----

d. --

QQ. Two years ago IGSCC cracks were somewhat unknown. Then the existence of IGSCC cracks was recognized and the importance of finding such cracks has been increased. This new knowledge and the increased importance on IGSCC have possibly changed the way that you carry out UT/ISI for IGSCC. In your own words, please explain what it is, if anything, that you do differently now compared to two years ago.

RR. If you have any additional comments that you would like to make--either regarding UT/ISI or this survey--please make them in the space provided below.

APPENDIX C

HUMAN ENGINEERING CRITERIA FOR UT INSTRUMENTS

HUMAN ENGINEERING CRITERIA
FOR ULTRASONIC TESTING INSTRUMENTS

1	2	3	4	5
FAILS TO MEET CRITERIA	MARGINAL	ADEQUATE	GOOD	EXCELLENT

TEAM: 1 2 3 4 5 6
TECH: 1 2 3 4 5 6

BRAND: ______________________________

USED FOR DETECTION____, GEOMETRY____.

CONTROLS	EVALUATION	
	SCORE	REMARKS
1. controls functionally grouped		
2. controls clearly labeled		
3. controls easy to manipulate wearing gloves for radiation protection		
4. controls guarded or oriented in a way that minimizes the chance of inadvertent actuation		
5. control movements conform to technician sterotypes		
6. control scales clearly and appropriately labeled		
7. control scales readable while wearing face masks for radiation protection		
8. controls provided with appropriate resistance or detents		
9. shape coding of controls used where it may be advantageous		
10. control pointer-scale relationships appropriate for control use (e.g., Rotary selector knobs should have a moving pointer and fixed scale for maximum readability)		

DISPLAYS (CRT)

	EVALUATION	
	SCORE	REMARKS
11. screen has easy-to-read scale indications (graticules) scribed both vertically and horizontally		
12. screen design minimizes reflected glare (uses polarized overlay or glare screen)		
13. instrument allows technician adjustment of screen brightness, contrast, and focus		
14. screen size is sufficient to allow signal-noise discriminability (average is 2.2"H x 2.75"W)		(Indicate Screen Size)
15. visibility of trace on scope is adequate		

NOTES:

	EVALUATION	
OTHER	SCORE	REMARKS
16. instrument has good portability features (e.g., handles, lightweight)		
17. instrument features facilitate protection from decontamination		
18. cable connectors/sockets are coded to reduce probabilty of improper connection		
SEARCH UNITS*		
19. design of the SU facilitates ease of handling/manipulation		
20. SU connectors permit easy and fast connection/disconnection from the UT instrument		
21. SU's are clearly and appropriately labeled and marked to show frequency, angle, and beam entry point		

***Specify Type of SU Used:**

Brand________________________________

________________________________.

Angle

0_____

45_____

52_____

60_____

70 _____

Other(Specify)_____

APPENDIX D

DESTRUCTIVE ANALYSIS OF MRR SPECIMENS

APPENDIX D

DESTRUCTIVE ANALYSIS OF MRR SPECIMENS

Destructive metallography was performed on the MRR specimens to verify the length and depth of cracks used in the MRR study and to provide limited data on sample microstructure. This section of the report documents the methodology and results of the destructive test performed on the MRR samples.

Extensive analysis of MRR inspection results and prior documentation was performed before the MRR samples were laid out for the machinist to cut. Each sample was sketched on drawing paper and the existence of known cracks was overlaid on the sketch. Areas surrounding known cracks were circumscribed with a scribe circle for destructive analysis. Next, the results of all MRR technicians were overlaid on the sketch to correlate inspection results with known cracks. Any additional areas on the sample sketch where several inspectors indicated cracks (where cracks were not known to exist) were also circumscribed as coupons for analysis. Figure D.1 shows a sketch of sample B217-2A with known cracks and inspection results, which illustrate the information used to mark the MRR samples for destructive analysis. This approach to destructive analysis was used because a minimum amount of each sample would actually be destroyed and allow the sample to be used for additional analyses and experiments.

Once coupons were cut from each sample, surface profiles were taken of both I.D. and O.D. surfaces. The coupons were then put in a press and bent open. Figure D.2 shows sample B217 after cracks 5 and 6 were bent open. Crack depth was measured using an optical microscope; crack length was measuring using dial calipers. A good correlation was found in optical microscope measurements for a number of the specimens, which were subsequently broken apart. The approach of bending cracks open was taken because the bending process would clearly expose any cracks not previously found by PT and which could affect POD and FCP. A number of cracks has been selected for the sizing study, and these were all destructively analyzed. For MRR samples that contained long cracks, only those areas of each long crack that were used for sizing were destructively analyzed.

Figure D.3 provides a listing of typical cracks used in the MRR relative to depth, length, and average POD; and shows the I.D. and O.D. profiles taken on all cracks.

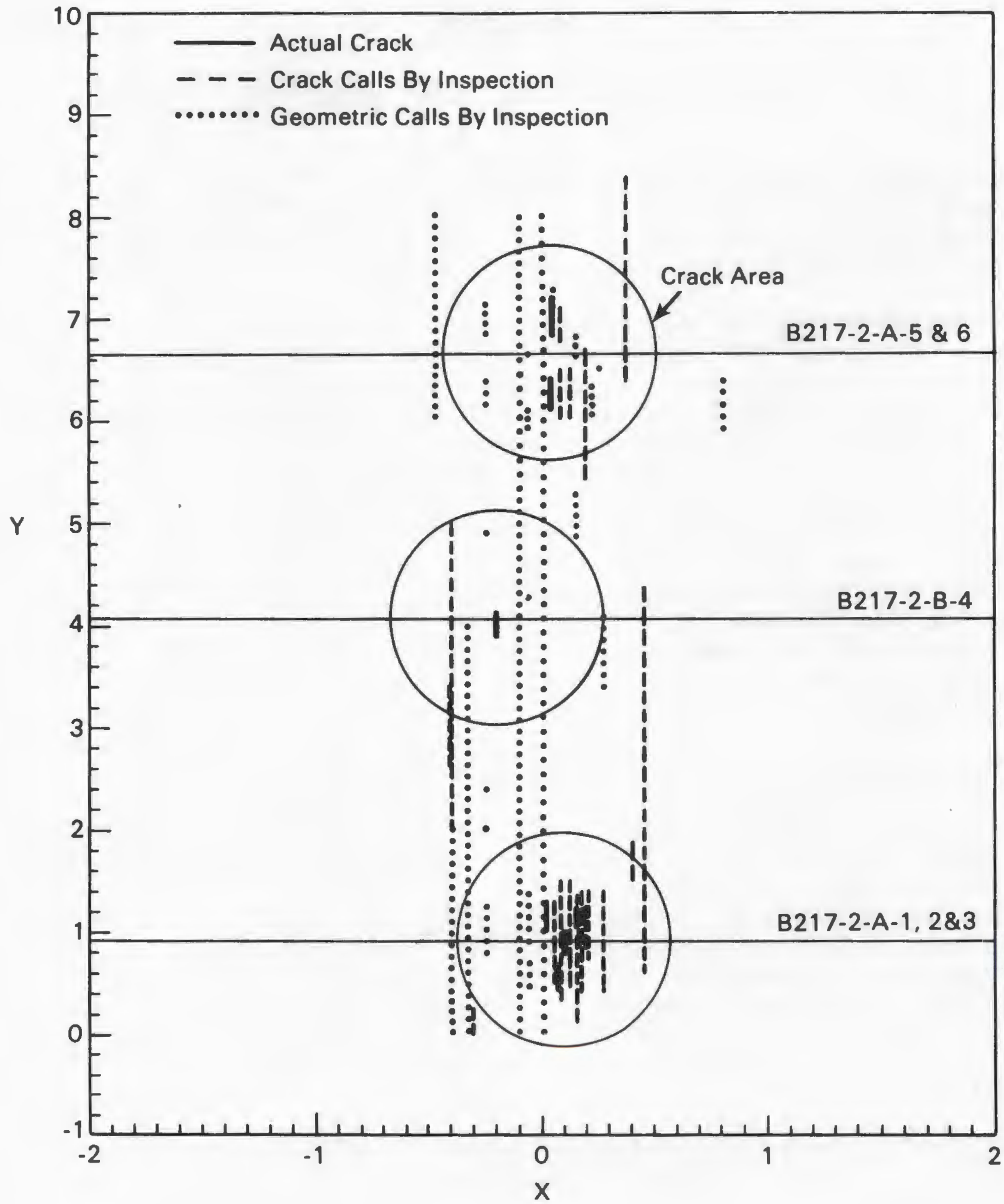


FIGURE D.1. Sketch of Sample as Prepared for Destructive Analysis

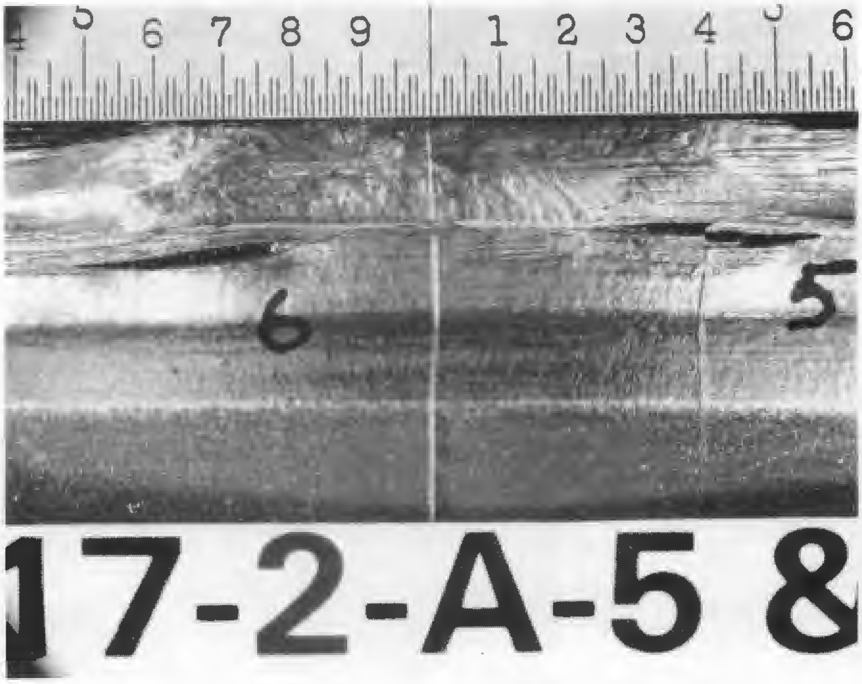


FIGURE D.2. Sample B217 - Cracks 5 and 6 After Being Bent Open in Press

Short Crack Specimens

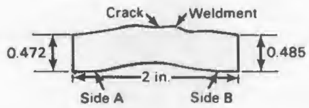
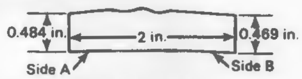
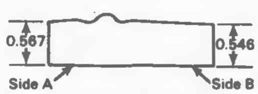
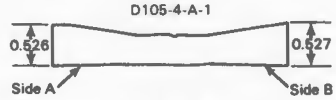
Specimen I.D.#	Specimen Profile of 2 in. Diameter	Length	Maximum Depth	POD	FCP
B217-2	 B217-2-A-1,2, & 3	Crack Cluster 1† 0.26 in. 2# 0.21 in. 3# 0.29 in.	 0.070 in. 0.078 in. 0.085 in.	POD for 80%	
B217-2	 B217-2-B-4	0.25 in.	0.66 in.	Far Side	
B217-2	 B271-2-A-5 & 6	1# 0.32 in. 2# 0.39 in.	0.066 in. 0.072 in.	53%	
D101-1	 D101-1-B-1	1.2 in.	0.006 in.	Far Side	Range of 0%-32% Average 17%
D101-4	 D101-4-B-1	2.61 in.	0.095 in.	80%	
D102-2	 D102-2-A-1	0.24 in.	0.027 in.	50%	Range of 20-40% Average 29%
D102-2	 D102-2-B2	1# 0.13 in. 2# 0.60 in.	0.19 in. 0.124 in.	25% 81%	
D102-3	 D102-3-B-1	1.15 in.	0.130 in.	93%	
D102-4	 D102-4-B-1	0.45 in.	0.074 in.	93%	
D105-4	 D105-4-A-1	0.22 in.	0.60 in.	0.29 in.	7%
D107-2	 D107-2-A-1 & 2	1† 0.68 in. 2# 0.36 in.	0.39 in. 0.88 in.	71% A Side 73% B Side	

FIGURE D.3. Cross-sectional Profiles of the Weld Area in the Vicinity of the IGSCC, IGSCC Size Data (length and depth), and Average Inspector Performance for These Areas (POD,FCP)

APPENDIX E

SUMMARY OF CRACK CALL LOCATIONS

APPENDIX E

SUMMARY OF CRACK CALL LOCATIONS

This appendix provides a summary of the location of each grading unit and crack calls for each of the specimens used in the MRR study. This information is summarized in three figures for each specimen. The middle figure shows a diagram of the individual specimen. The diagram shows the location of all grading units used to score MRR inspection results and the location of cracks within the specimen. As an example, Figure E.1(b) shows the true state of cracks in Specimen B217 and the location of grading units 99, 100, 101, and 102. Figure E.1(a) shows a histogram of all cracked calls made on Specimen B217, quadrant side A versus circumferential location of the calls. Note that B217-2A was inspected 15 times (once by each inspector) during the MRR. Figure E.1(c) shows that no inspections were performed on B217, quadrant 2, B side.

A final table is included that contains the true-state data for all of the grading units that contain IGSCC.

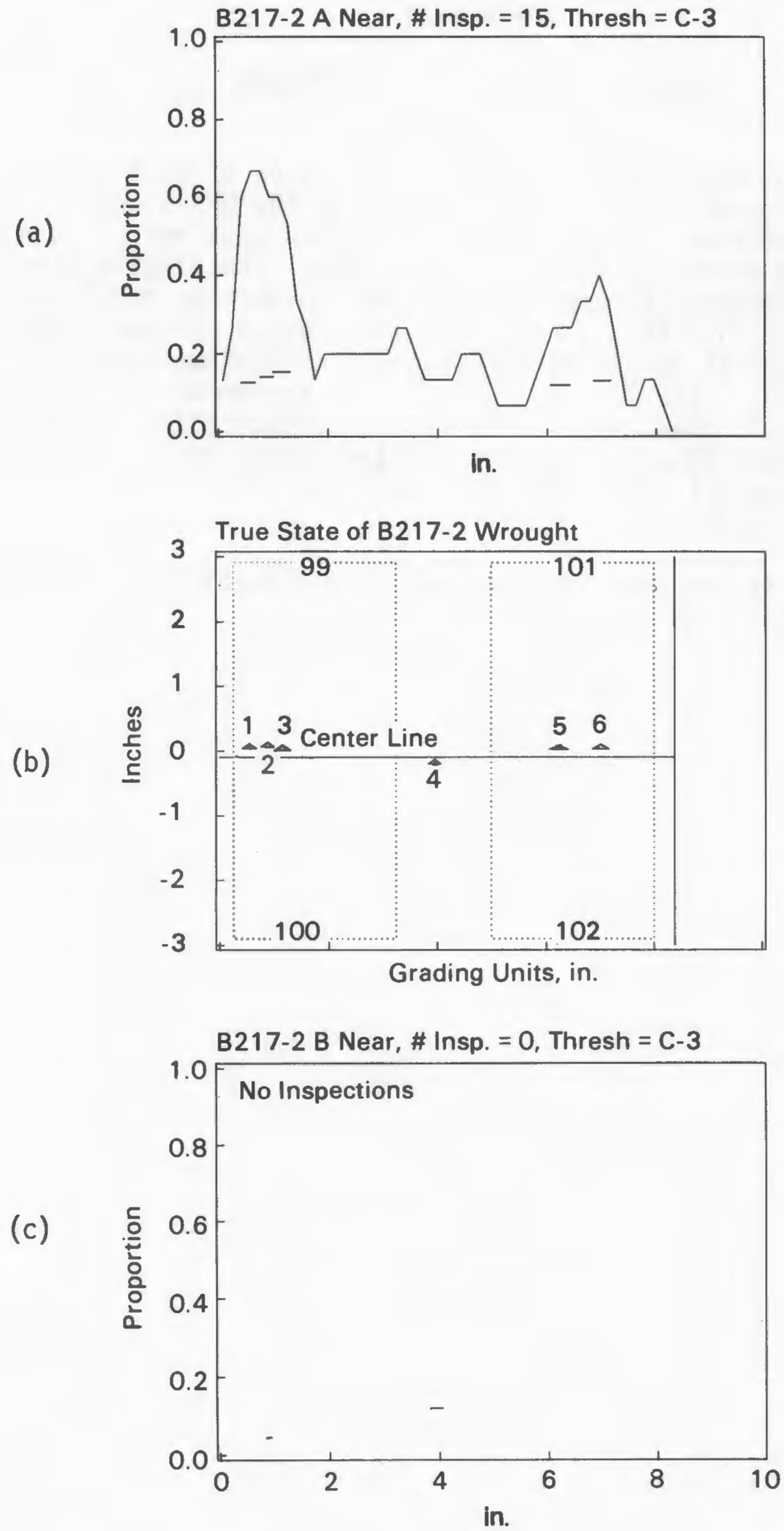


FIGURE E.1. Crack Call Data for Specimen B217

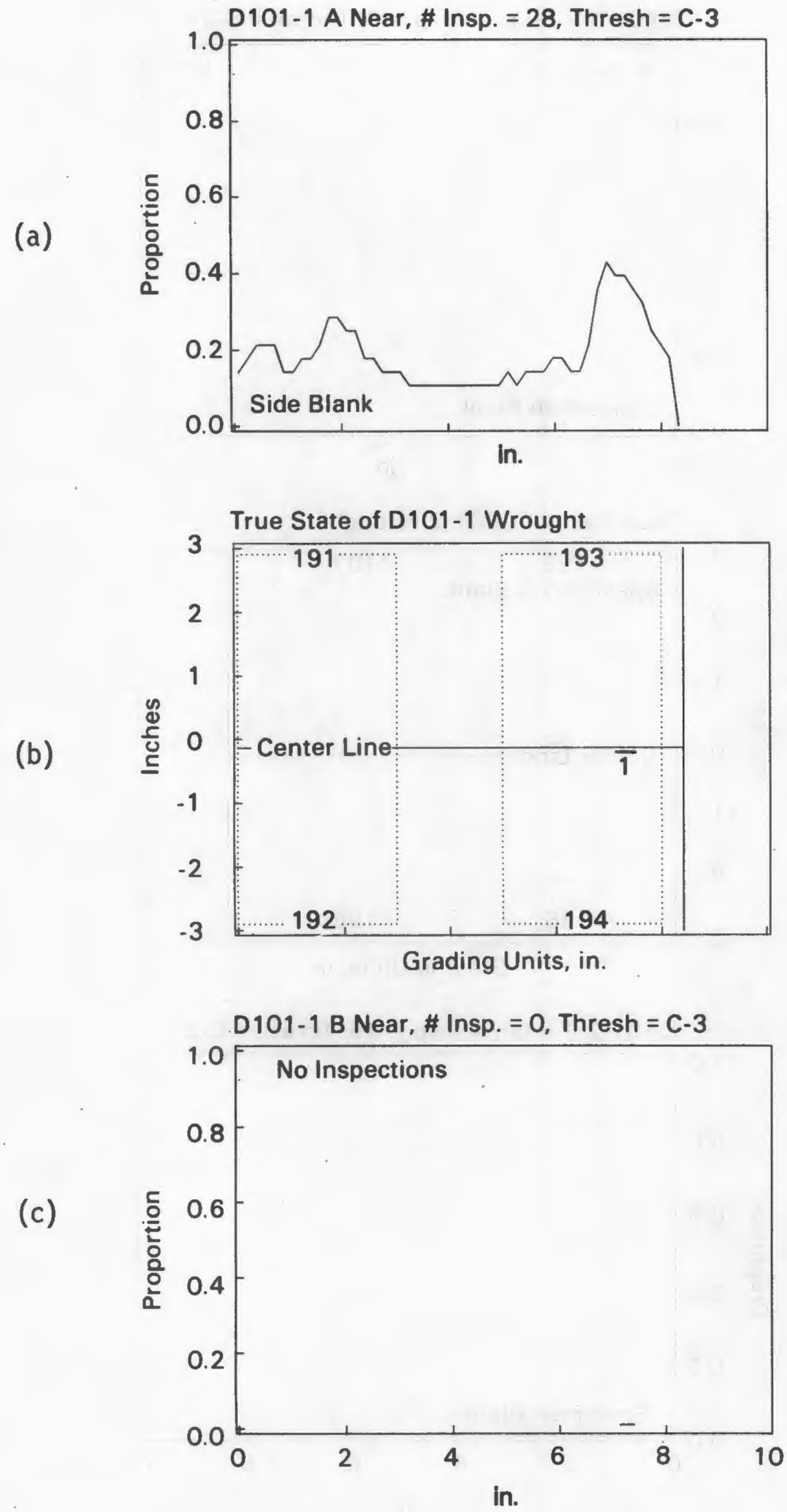


FIGURE E.2. Crack Call Data for Specimen D101-1

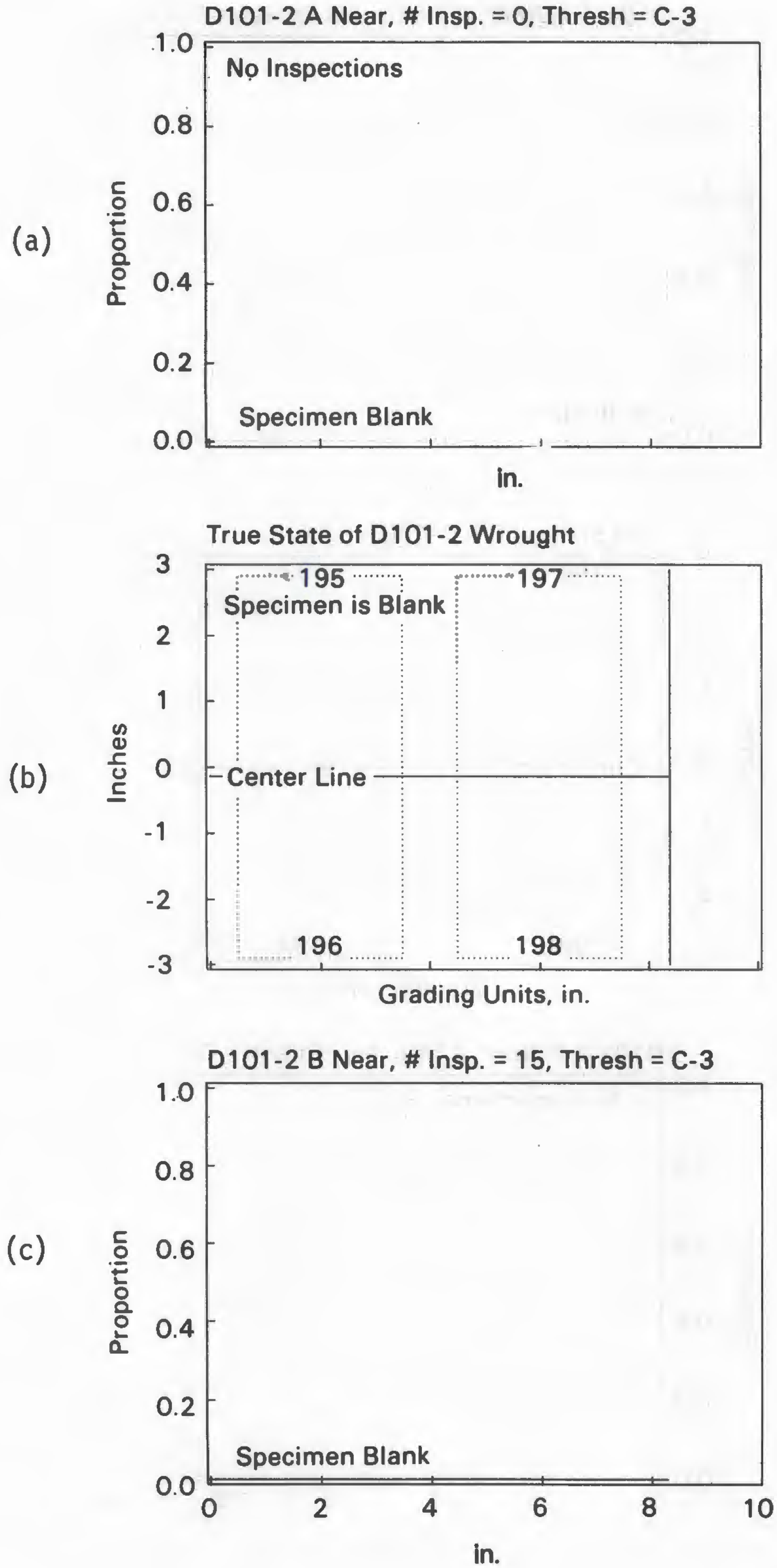


FIGURE E.3. Crack Call Data for Specimen D101-2

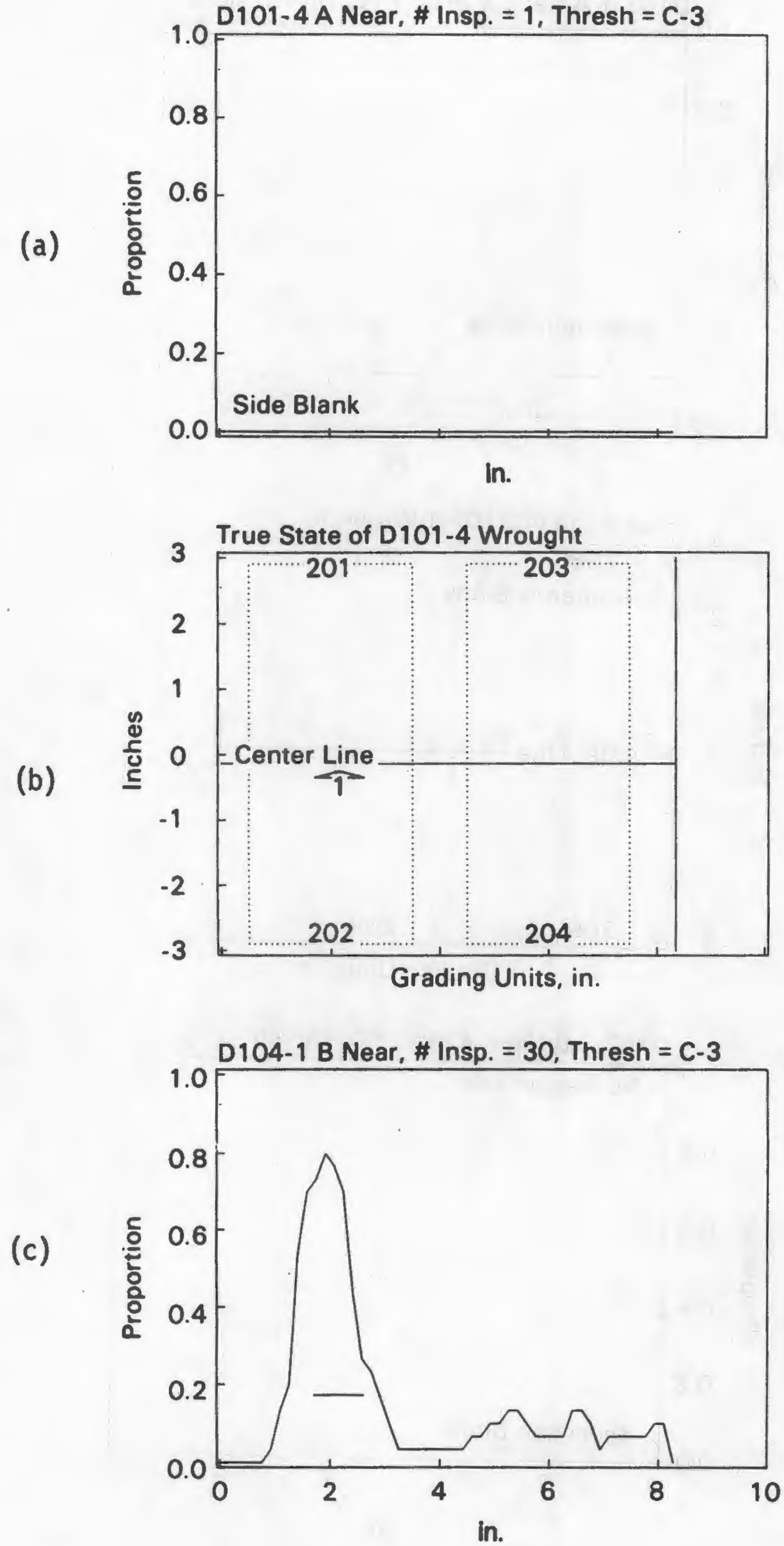


FIGURE E.4. Crack Call Data for Specimen D101-4

(a)

(b)

(c)

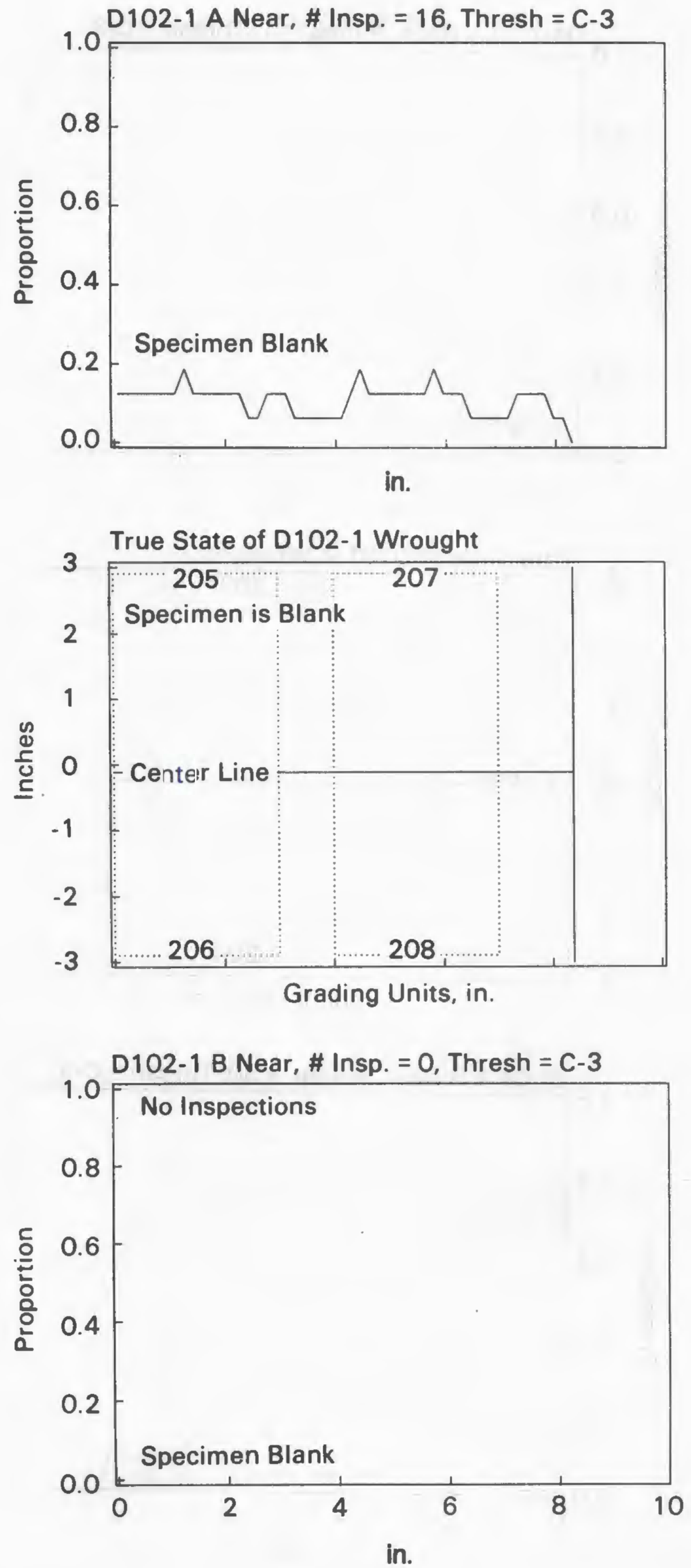


FIGURE E.5. Crack Call Data for Specimen D102-1

(a)

(b)

(c)

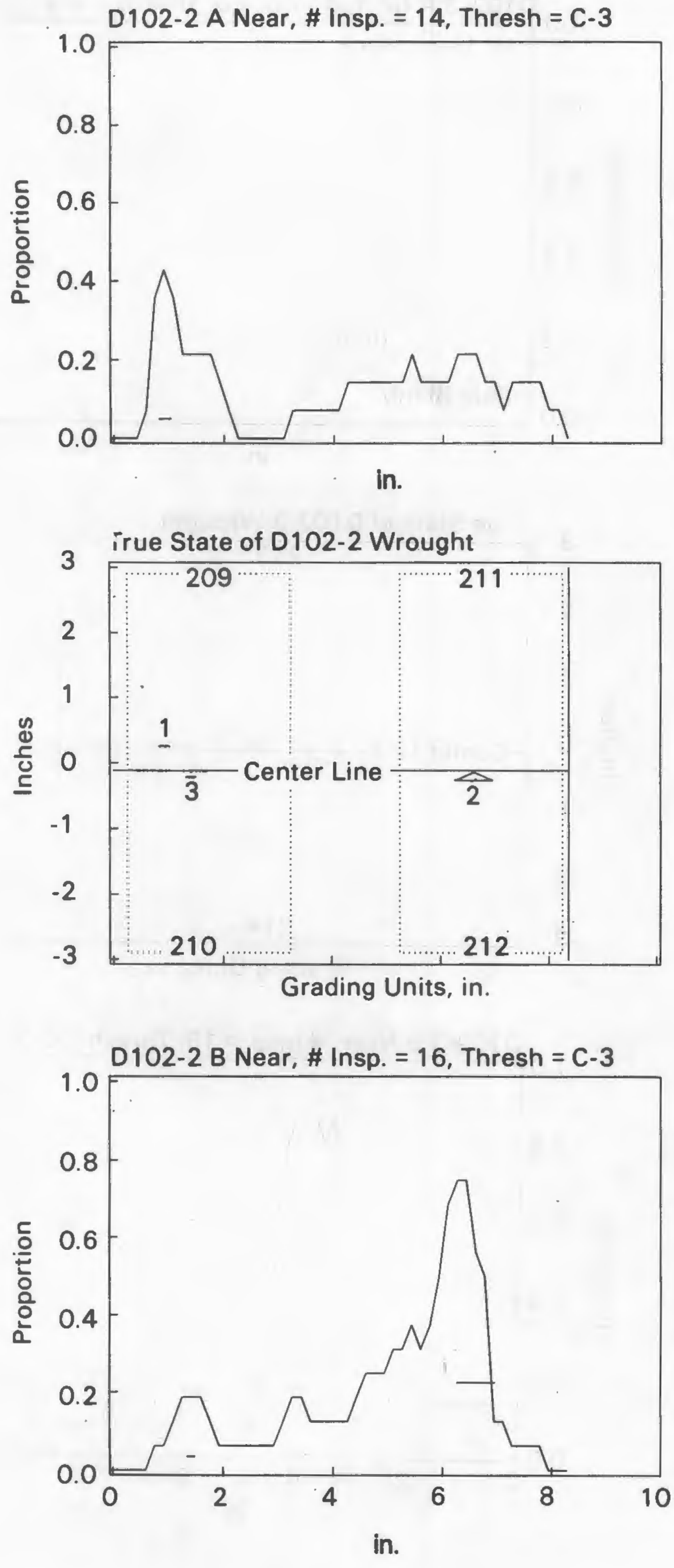
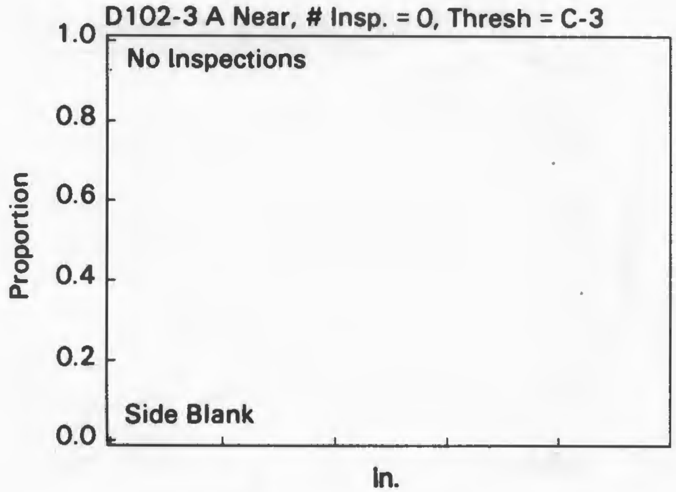
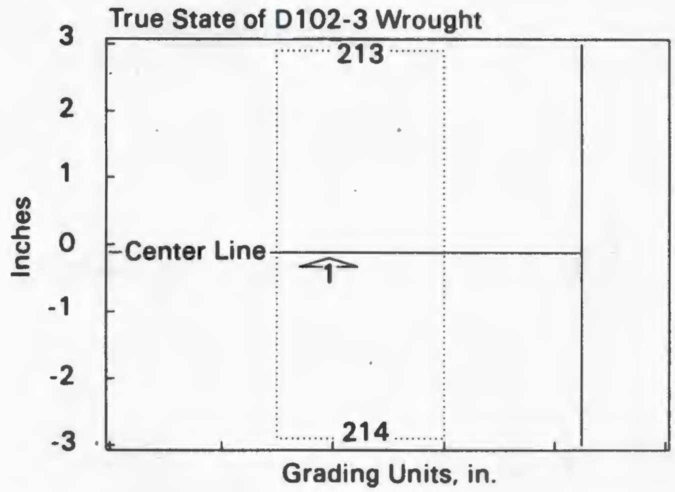


FIGURE E.6. Crack Call Data for Specimen D102-2

(a)



(b)



(c)

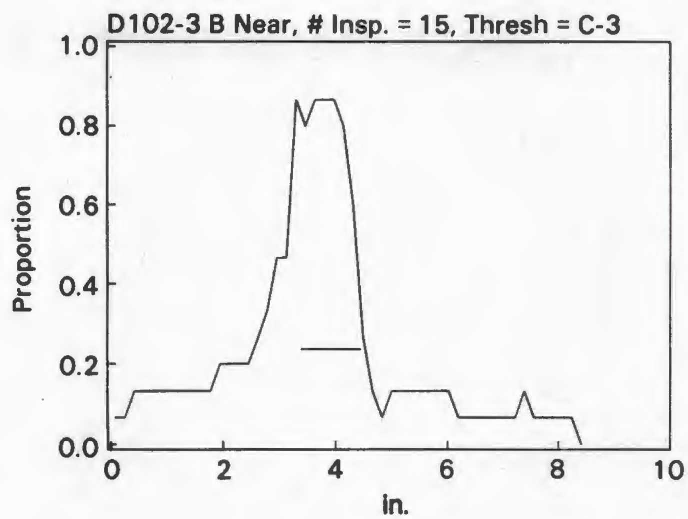


FIGURE E.7. Crack Call Data for Specimen D102-3

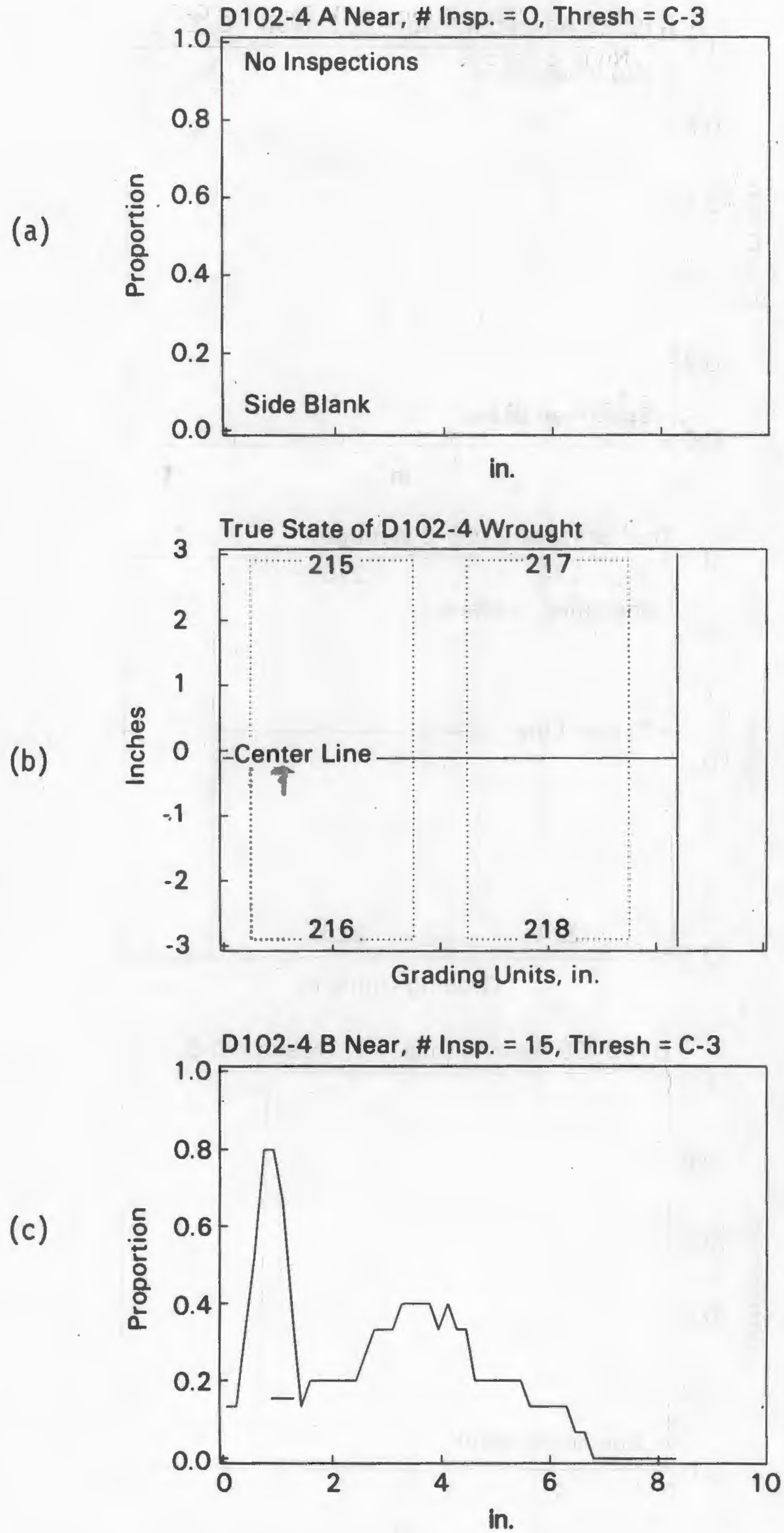


FIGURE E.8. Crack Call Data for Specimen D102-4

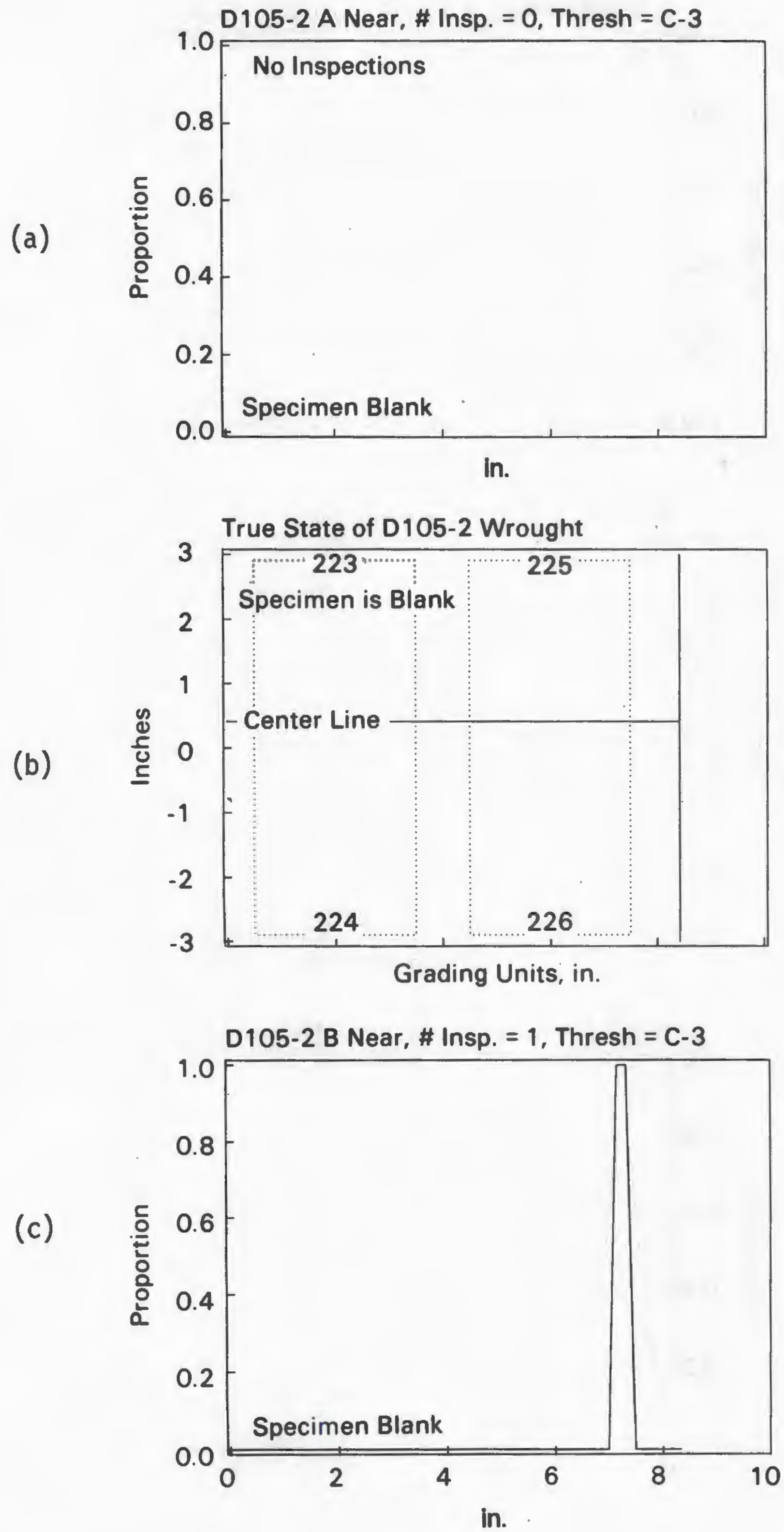


FIGURE E.9. Crack Call Data for Specimen D105-2

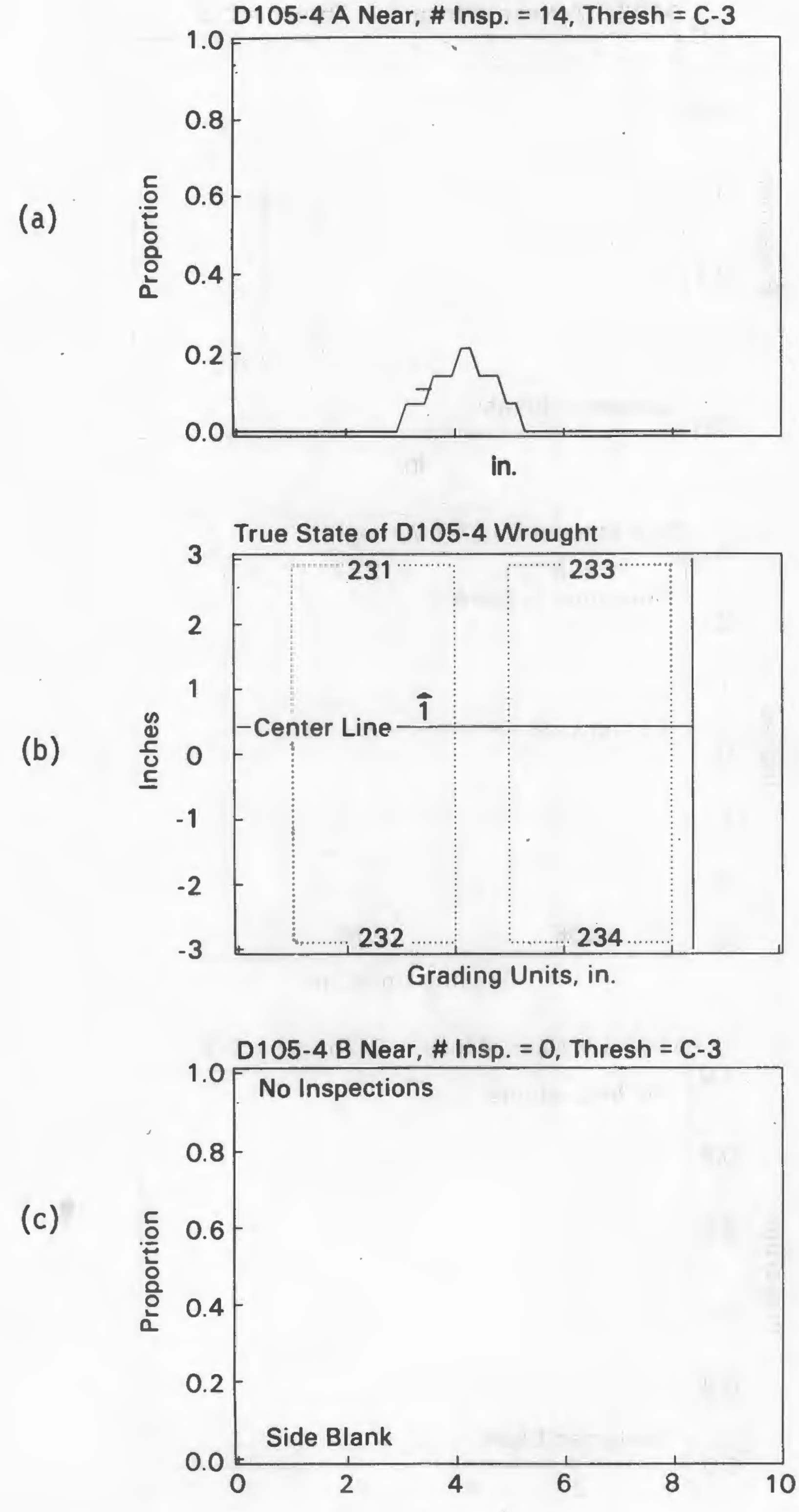


FIGURE E.10. Crack Call Data for Specimen D105-4

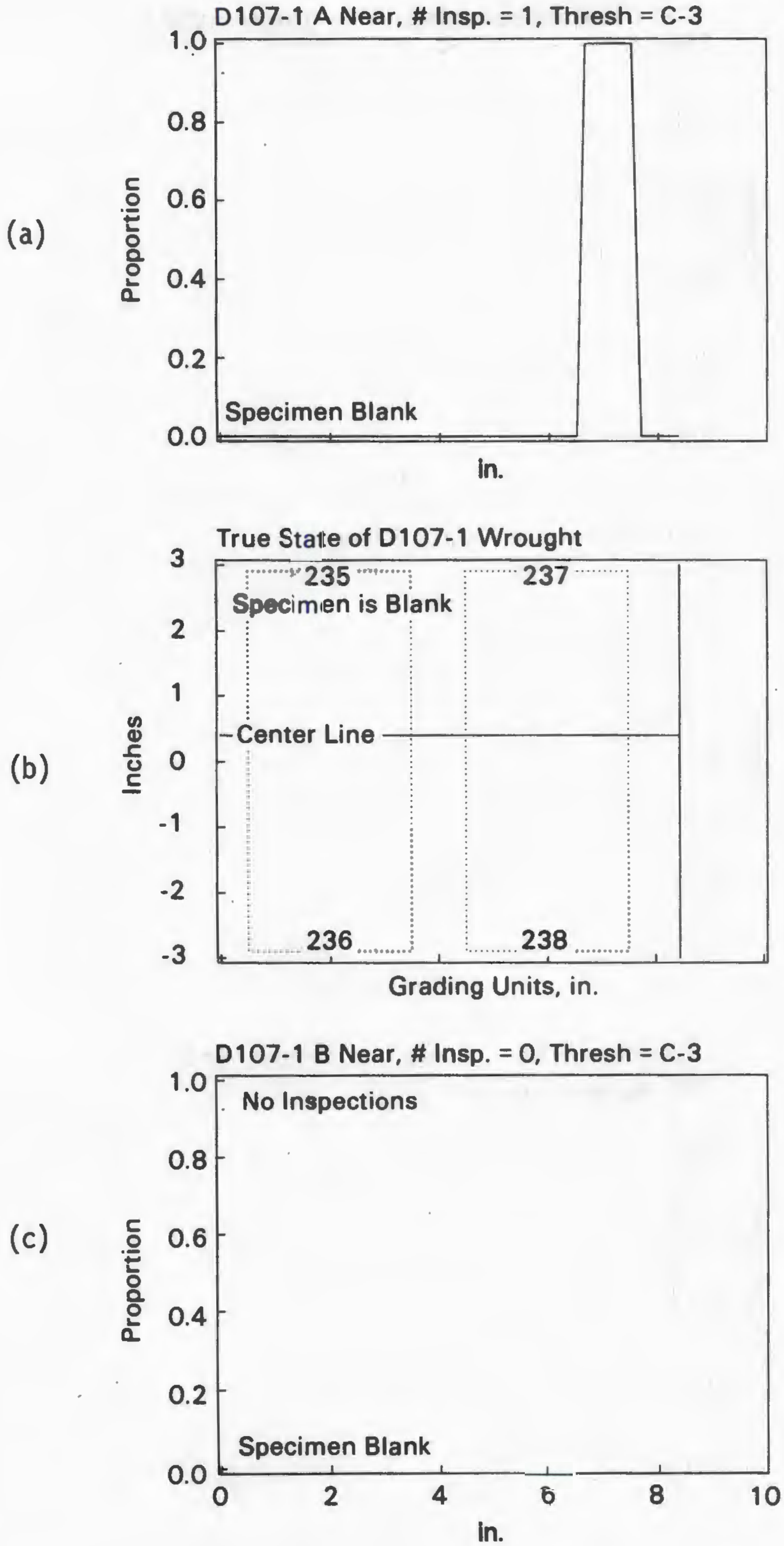


FIGURE E.11. Crack Call Data for Specimen D107-1

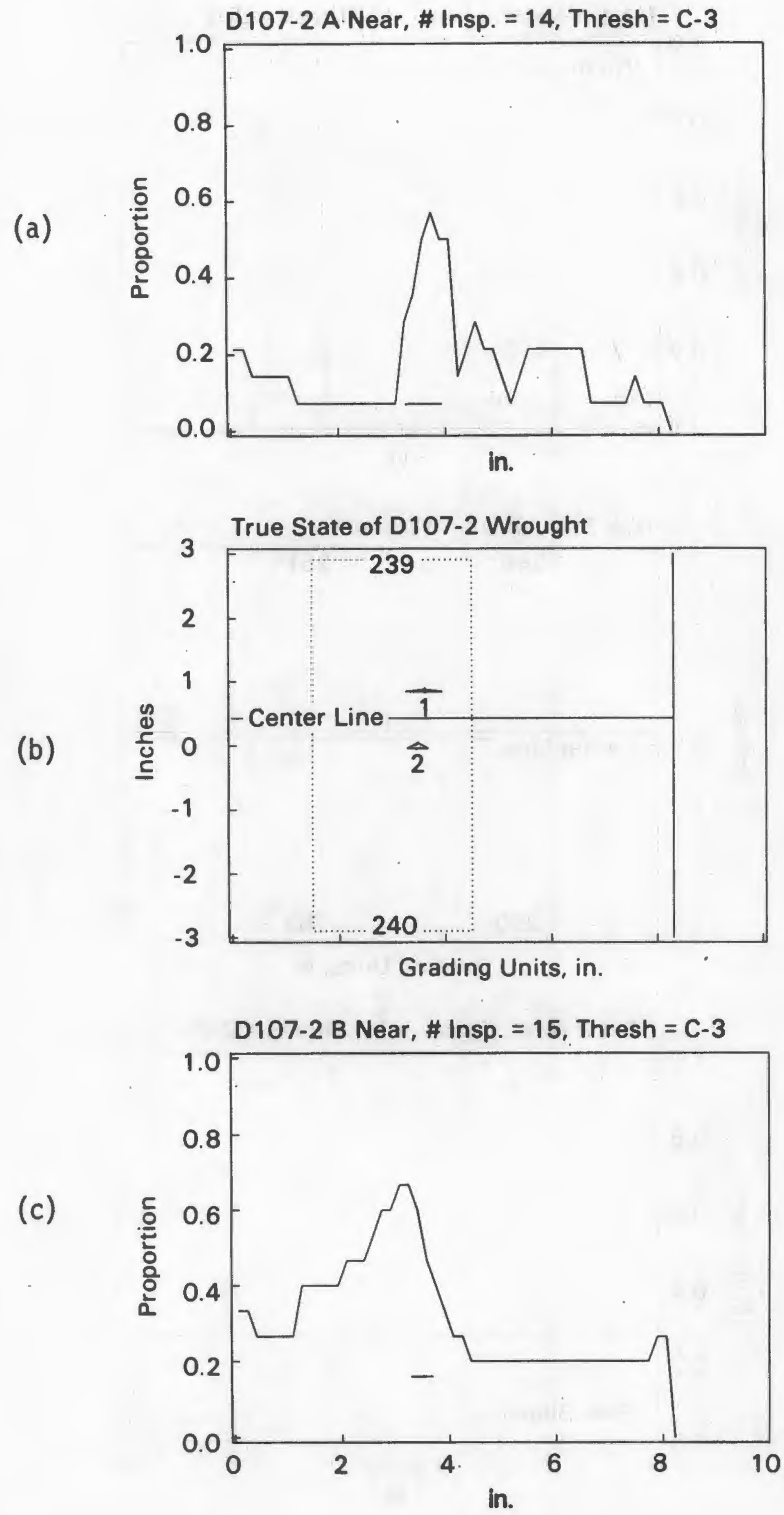


FIGURE E.12. Crack Call Data for Specimen D107-2

(a)

(b)

(c)

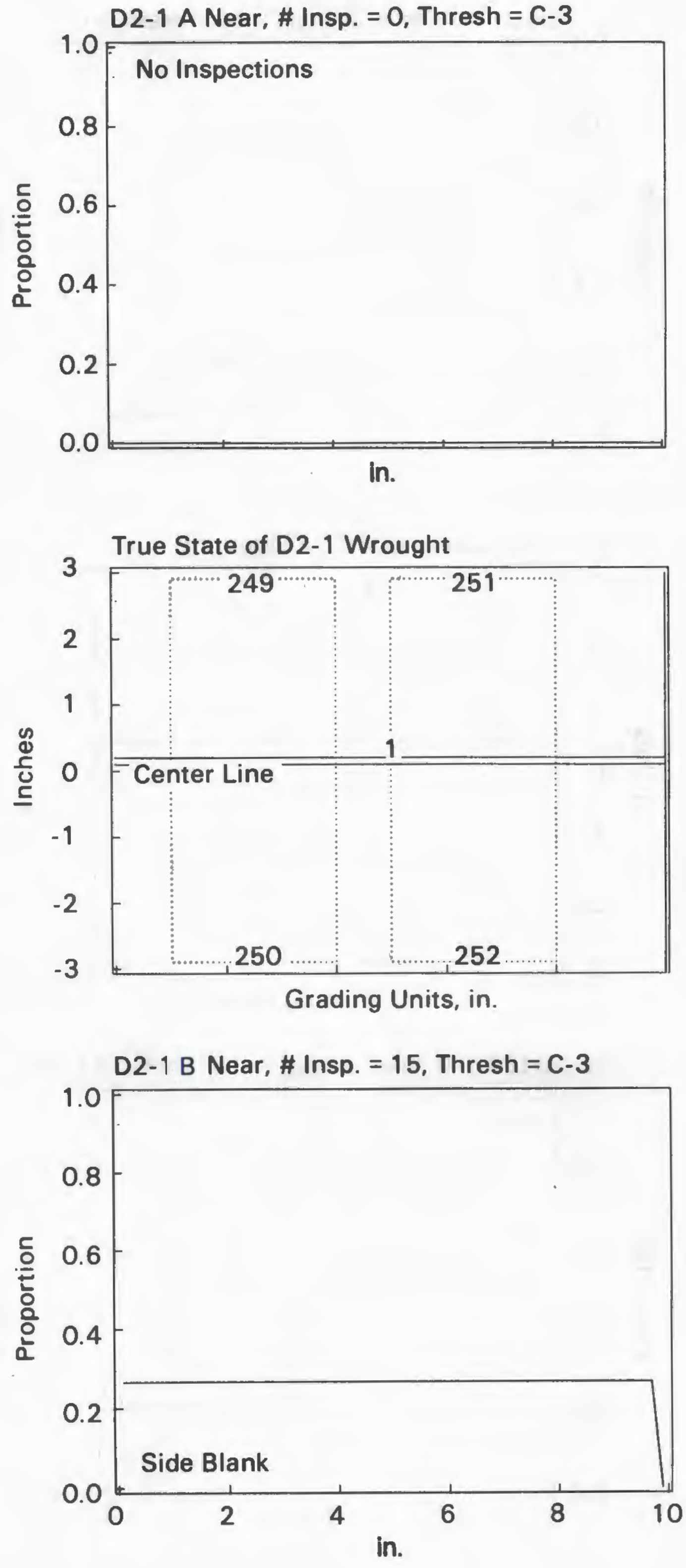


FIGURE E.13. Crack Call Data for Specimen D2-1

(a)

(b)

(c)

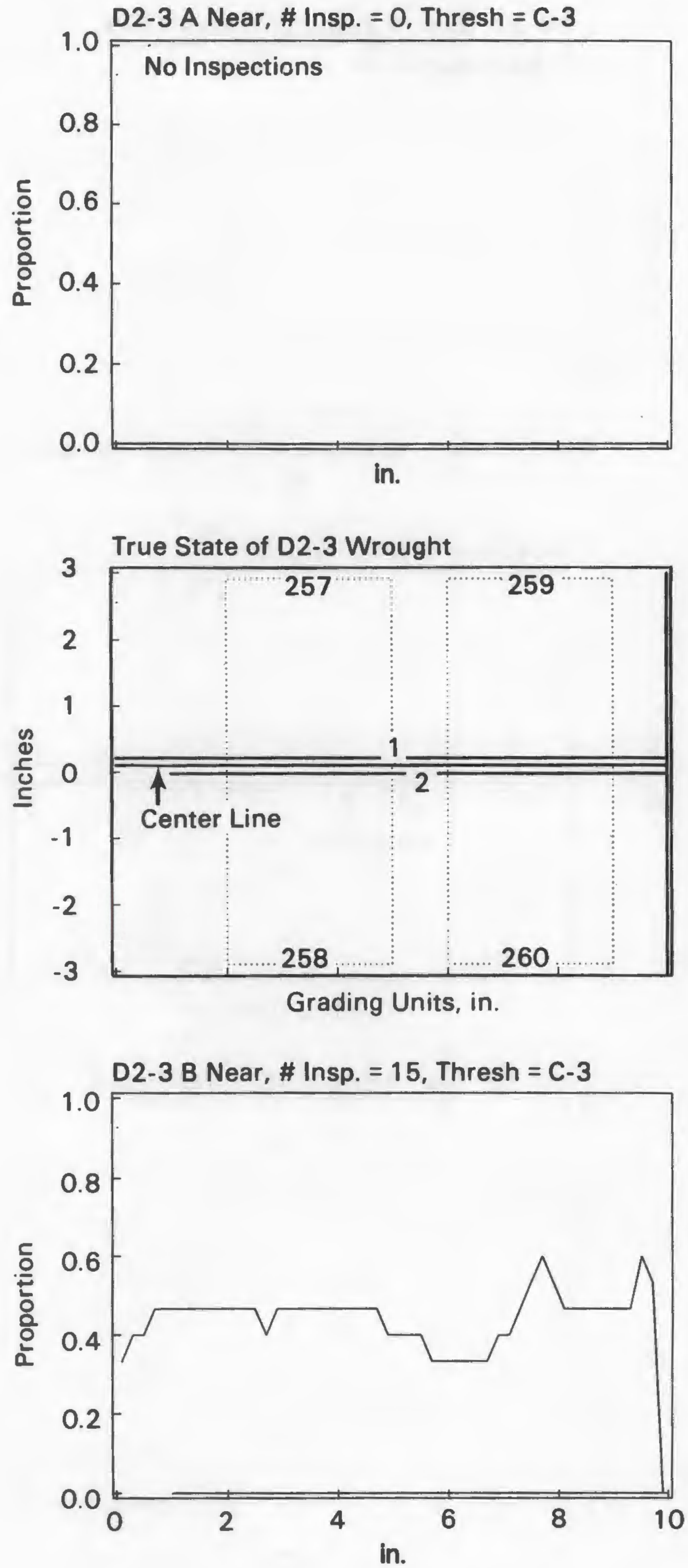


FIGURE E.14. Crack Call Data for Specimen D2-3

(a)

(b)

(c)

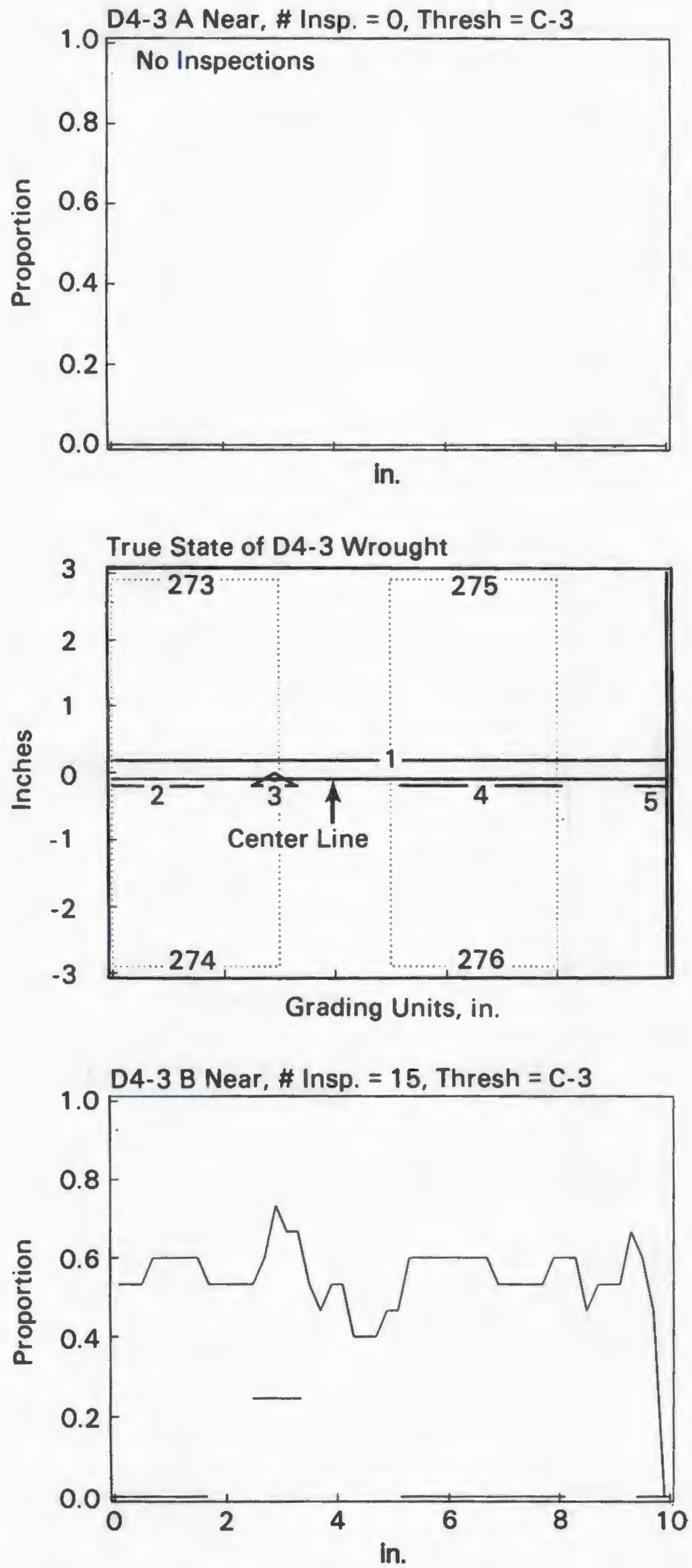


FIGURE E.15. Crack Call Data for Specimen D4-3

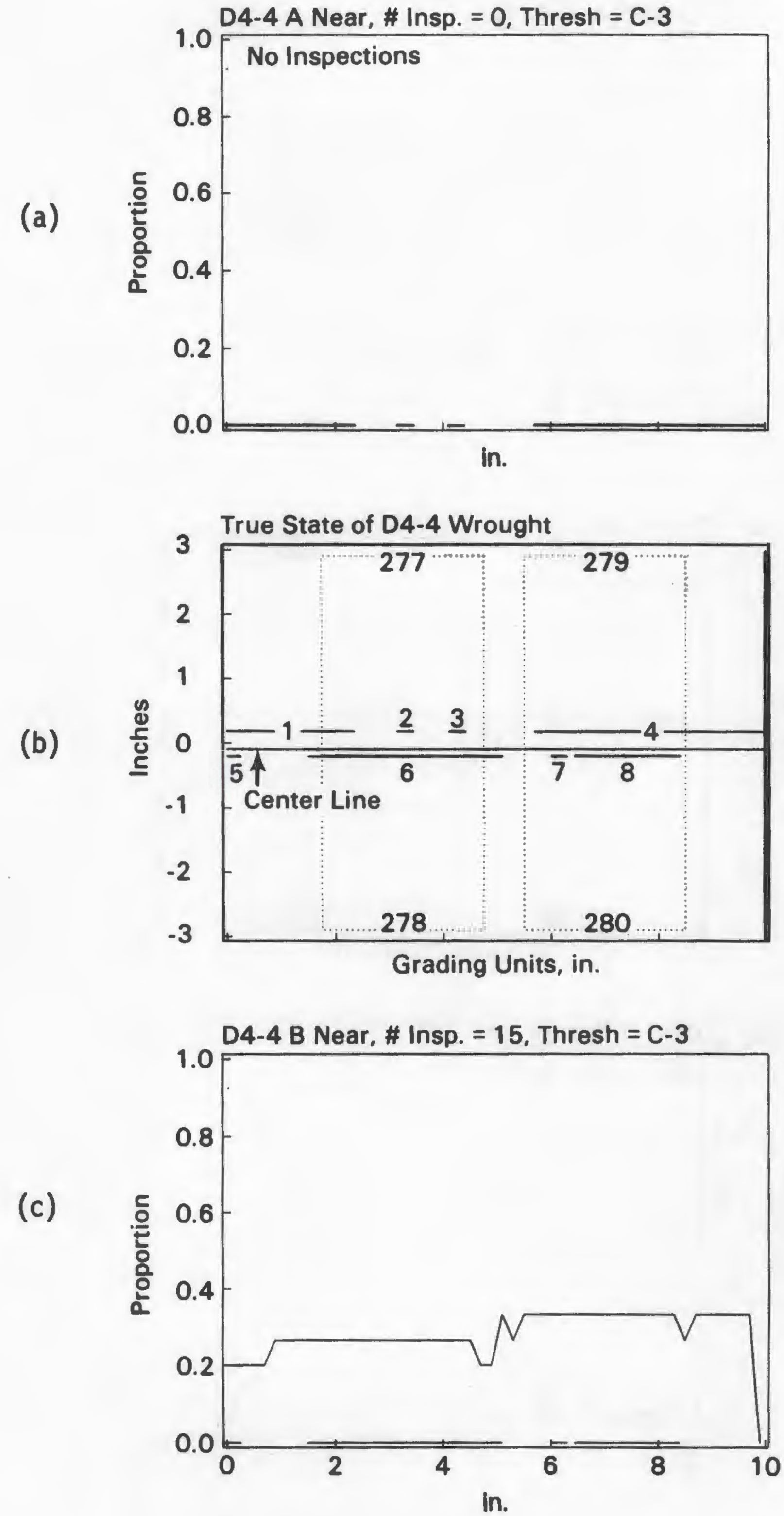


FIGURE E.16. Crack Call Data for Specimen D4-4

(a)

(b)

(c)

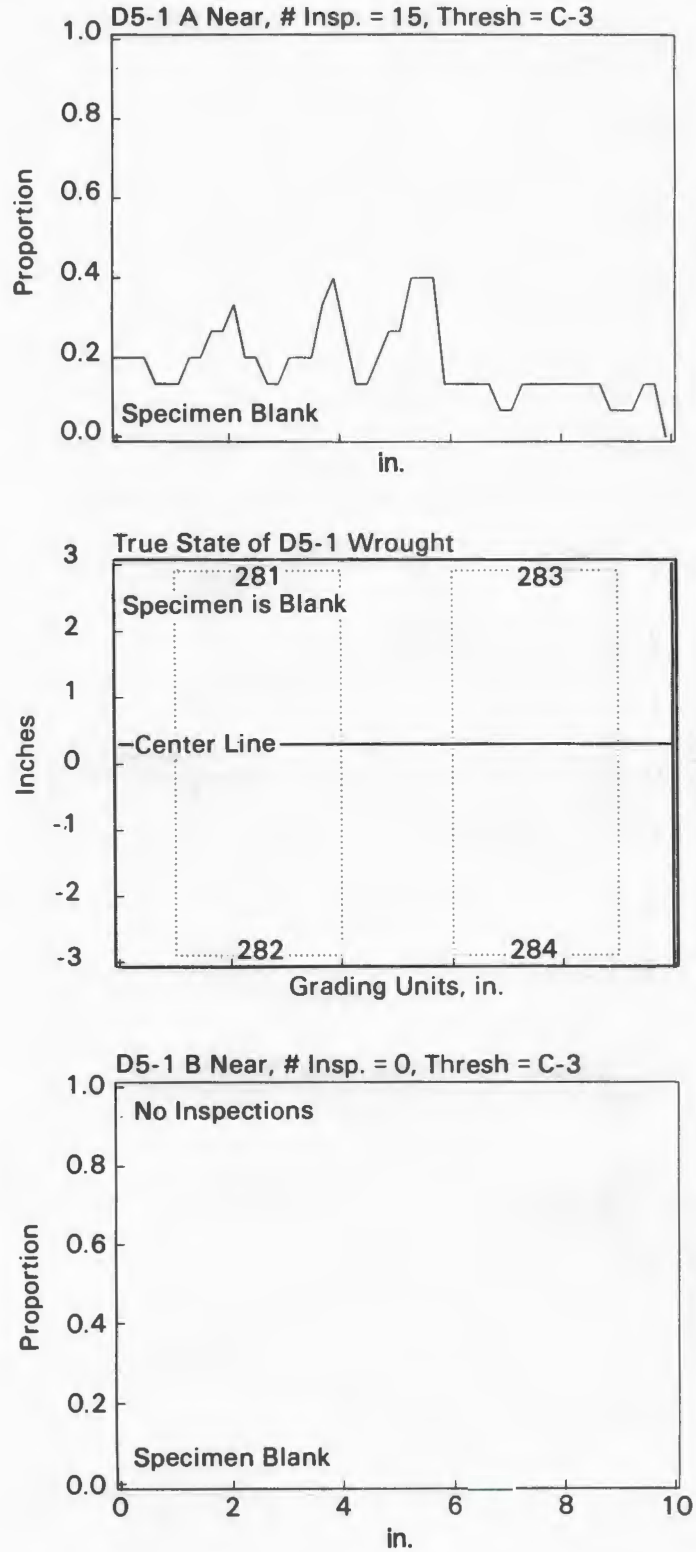


FIGURE E.17. Crack Call Data for Specimen D5-1

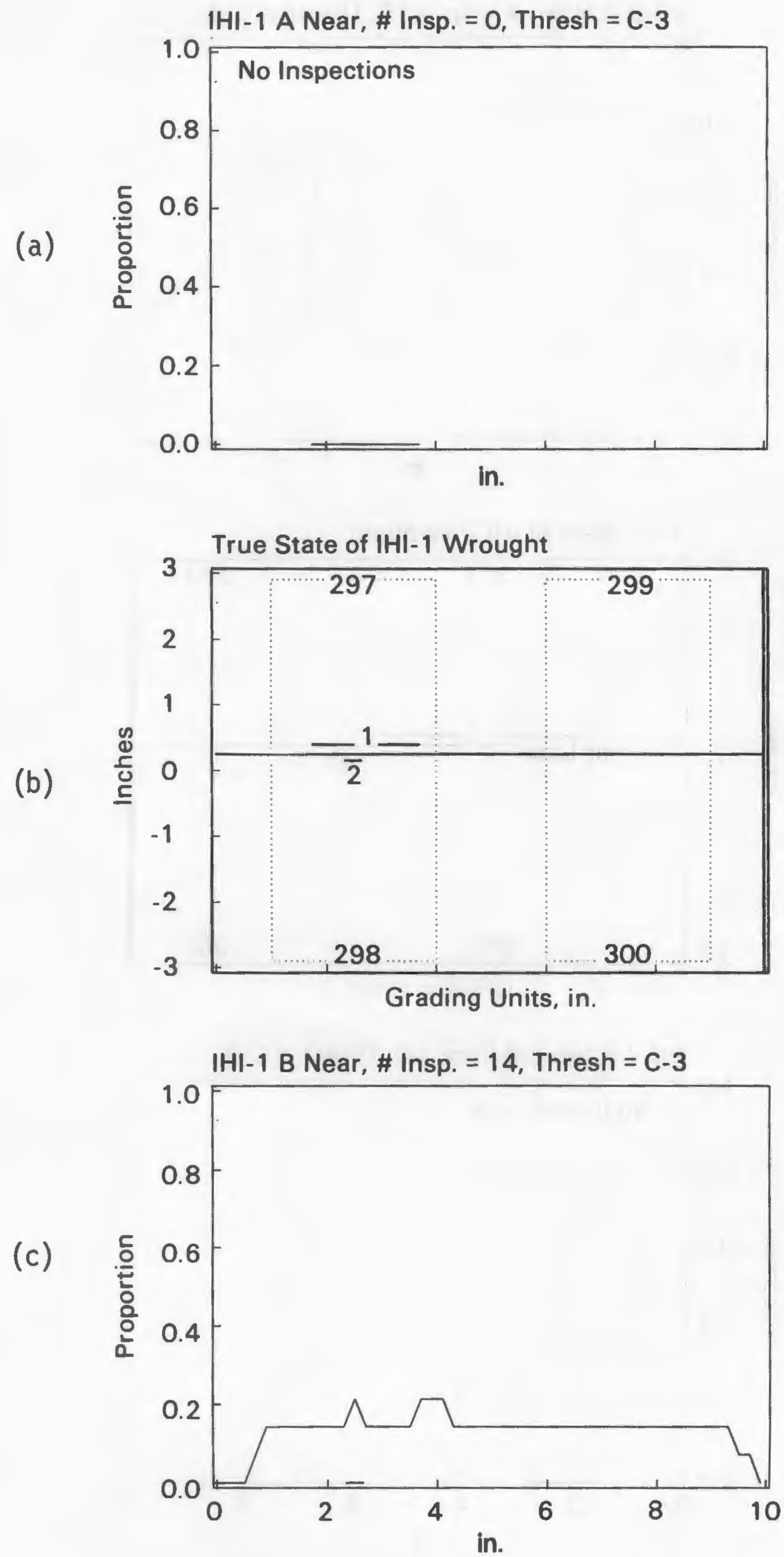


FIGURE E.18. Crack Call Data for Specimen IHI-1

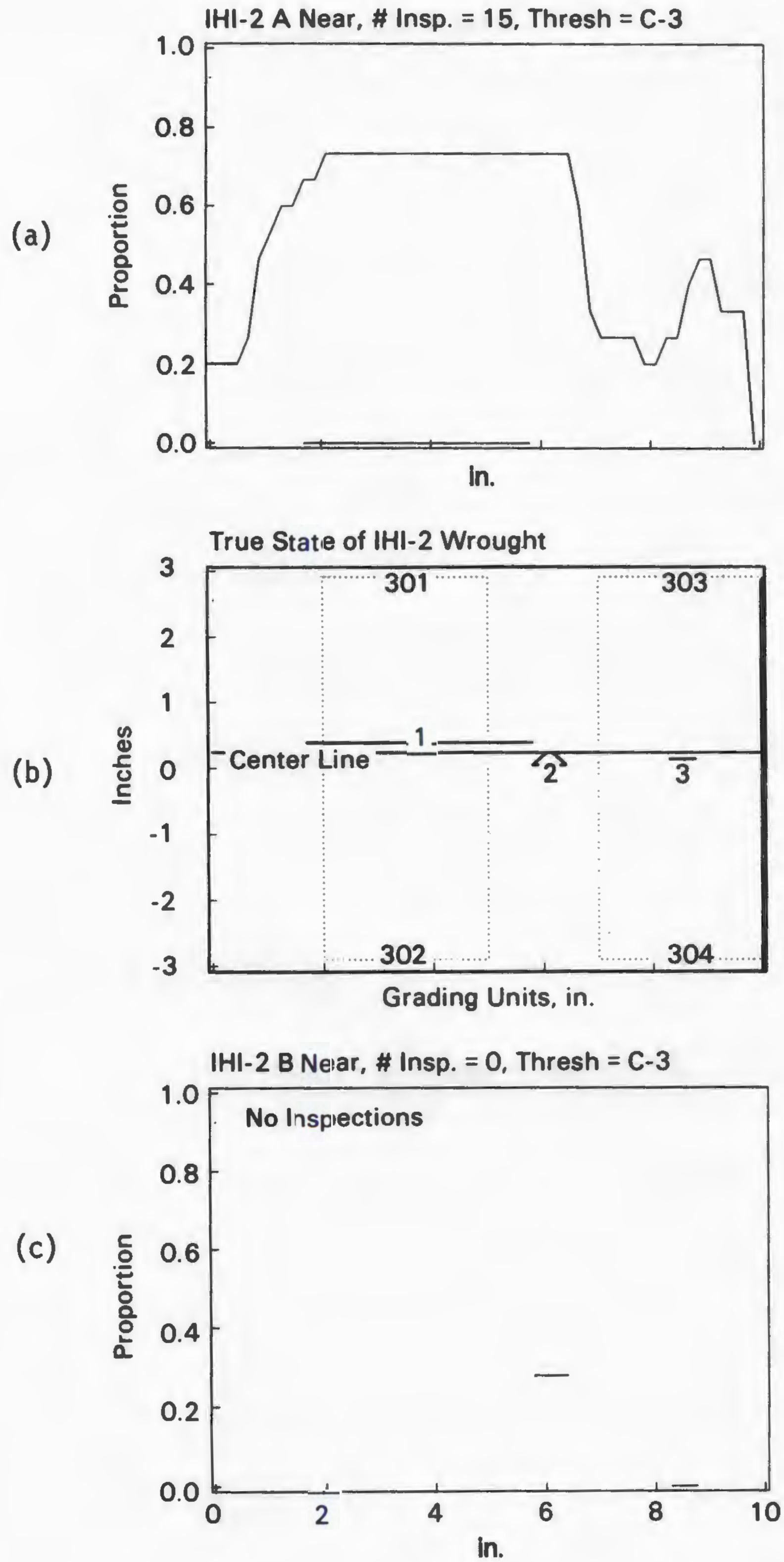


FIGURE E.19. Crack Call Data for Specimen IHI-2

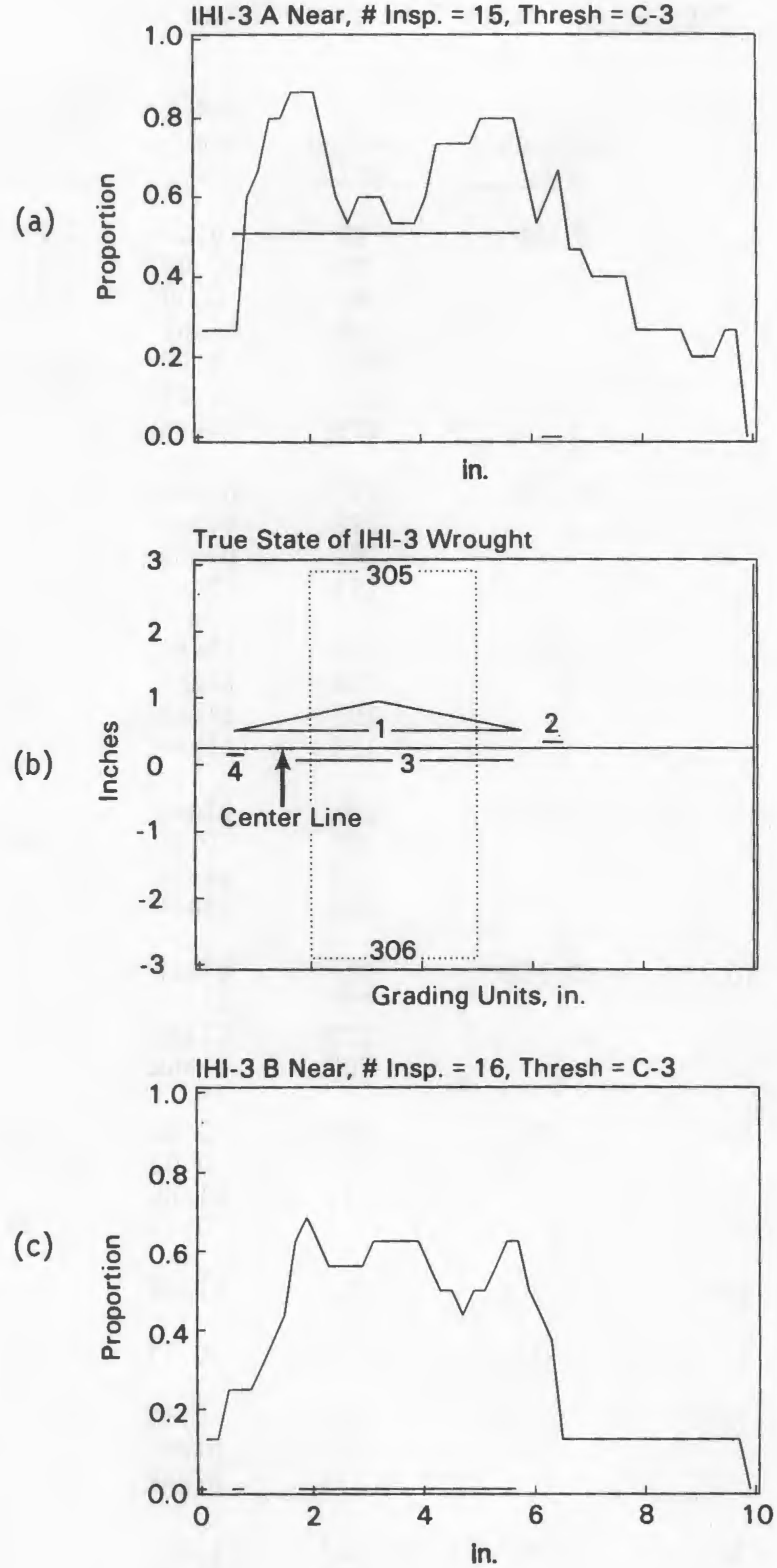


FIGURE E.20. Crack Call Data for Specimen IHI-3

TABLE E.1. IGSCC True-State Data

Specimen	Pipe Diameter, in.	Wall Thickness, in.	Grading Unit	Crack Depth, in.	Crack Length, in.
B217-2	10	0.550	99	0.07	0.25
			99	0.08	0.23
			99	0.09	0.31
			100	blank	
			101	0.07	0.35
			101	0.07	0.31
			102	blank	
D101-1	10	0.550	191	blank	
			192	blank	
			193	blank	
			194	0.01	0.25
D101-2	10	0.550	195	blank	
			196	blank	
			197	blank	
			198	blank	
D101-4	10	0.550	201	blank	
			202	0.10	0.90
			203	blank	
			204	blank	
D102-1	10	0.550	205	blank	
			206	blank	
			207	blank	
			208	blank	
D102-2	10	0.550	209	0.03	0.20
			210	0.02	0.13
			211	blank	
			212	0.12	0.69
D102-3	10	0.550	213	blank	
D102-3	10	0.550	214	0.13	1.04
D102-4	10	0.550	216	0.09	0.39
			217	blank	
			218	blank	
D105-2	10	0.550	223	blank	
			224	blank	
			225	blank	
			226	blank	

TABLE E.1. Contd

Specimen	Pipe Diameter, in.	Wall Thickness, in.	Grading Unit	Crack Depth, in.	Crack Length, in.
D105-4	10	0.550	231	0.06	0.26
			232	blank	
			233	blank	
			234	blank	
D107-1	10	0.550	235	blank	
			236	blank	
			237	blank	
			238	blank	
D107-2	10	0.550	239	0.04	0.67
			240	0.09	0.38
D2-1	12	0.688	249	DNDA	10.00
			250	blank	
			251	DNDA	10.00
			252	blank	
D2-3	12	0.688	257	DNDA	10.00
			258	DNDA	9.00
			259	DNDA	10.00
			260	DNDA	9.00
D4-3	12	0.688	273	DNDA	10.00
			274	DNDA	1.65
			274	0.17	0.85
			275	DNDA	10.00
			276	DNDA	2.95
D4-4	12	0.688	277	DNDA	2.35
			277	DNDA	0.30
			277	DNDA	0.30
			278	DNDA	3.60
			279	DNDA	4.30
			280	DNDA	0.30
			280	DNDA	1.95
D5-1	12	0.688	281	blank	
			282	blank	
			283	blank	
			284	blank	

TABLE E.1. Contd

Specimen	Pipe Diameter, in.	Wall Thickness, in.	Grading Unit	Crack Depth, in.	Crack Length, in.
IHI-1	12	0.866	297	DNDA	1.95
			298	DNDA	0.30
			299	blank	
			300	blank	
IHI-2	12	0.866	301	DNDA	4.10
			302	0.24	0.60
			303	DNDA	0.45
			304	blank	
IHI-3	12	0.866	305	0.44	5.15
			306	DNDA	3.90

*DNDA = did not destructively analyze

DISTRIBUTION

No. of Copies

OFFSITE

2 J. Muscara
NRC/RES
Mail Stop NS 217C

C. Z. Serpan
NRC/RES
Mail Stop NS 217C

L. R. Abramson
NRC/RES
Mail Stop NLS-372

M. R. Hum
NRC/NRR
Mail Stop 9H-15

R. A. Hermann
NRC/NRR
Mail Stop 9H-15

C. Y. Cheng
NRC/NRR
Mail Stop 9H-15

J. P. Durr
NRC/Region I

H. W. Kerch
NRC/Region I

J. R. Strosnider
NRC/Region I

A. R. Herdt
NRC/Region II

J. J. Blake
NRC/Region II

No. of Copies

D. Danielson
NRC/Region III

K. Ward
NRC/Region III

I. Barnes
NRC/Region IV

W. M. McNeil
NRC/Region IV

M. H. Miller
NRC/Region V

R. J. Pate
NRC/Region V

D. S. Kupperman
Materials Science Center
Argonne National Laboratory
9700 S. Cass Avenue
Building 212
Argonne, IL 60439

FOREIGN

50 N. R. McDonald
OECD Nuclear Energy Agency
38 Boulevard Suchet
F-75016 Paris
France
(for distribution to PISC III)

ONSITE

50 Pacific Northwest Laboratory

M. C. Bampton
R. E. Bowey
S. H. Bush
J. D. Deffenbaugh
A. A. Diaz
S. R. Doctor (31)
M. S. Good
E. R. Green
P. G. Heasler
R. L. Hockey
D. K. Lemon
G. J. Posakony
F. A. Simonen
J. C. Spanner
T. T. Taylor
L. G. Van Fleet
T. V. Vo
Technical Report Files (2)
Publishing Coordination

NRC FORM 335 (2-89) NRCM 1102, 3201, 3202

U.S. NUCLEAR REGULATORY COMMISSION

BIBLIOGRAPHIC DATA SHEET

(See instructions on the reverse)

1. REPORT NUMBER (Assigned by NRC. Add Vol., Supp., Rev., and Addendum Numbers, if any.)

NUREG/CR-4908
PNL-6196

2. TITLE AND SUBTITLE

Ultrasonic Inspection Reliability for Intergranular Stress Corrosion Cracks: A Round Robin Study of the Effects of Personnel, Procedures, Equipment, and Crack Characteristics

3. DATE REPORT PUBLISHED

MONTH	YEAR
July	1990

4. FIN OR GRANT NUMBER

B2289

5. AUTHOR(S)

P.G. Heasler, T.T. Taylor, J.C. Spanner, S.R. Doctor, J.D. Deffenbaugh

6. TYPE OF REPORT

Technical

7. PERIOD COVERED *(Inclusive Dates)*

8. PERFORMING ORGANIZATION – NAME AND ADDRESS *(If NRC, provide Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)*

Pacific Northwest Laboratory
Richland, Washington 99352

9. SPONSORING ORGANIZATION – NAME AND ADDRESS *(If NRC, type "Same as above"; if contractor, provide NRC Division, Office or Region, U.S. Nuclear Regulatory Commission and mailing address.)*

Division of Engineering
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, D.C. 20555

10. SUPPLEMENTARY NOTES

11. ABSTRACT *(200 words or less)*

A pipe inspection round robin entitled "Mini-Round Robin" was conducted at Pacific Northwest Laboratory from May 1985 through October 1985. The research was sponsored by the U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research under a program entitled "Evaluation and Improvement of NDE Reliability for Inservice Inspection of Light Water Reactors."

The Mini-Round Robin (MRR) measured the IGSC crack detection and sizing capabilities of inservice inspection (ISI) inspectors that had passed the requirements of IEB 83-02 and the EPRI sizing training course. The MRR data base was compared with an earlier Pipe Inspection Round Robin (PIRR) that had measured effective detection prior to 1982. Comparison of the MRR and PIRR data bases indicated no difference in detection capability was measured for long and short cracks.

In addition to the pipe inspection round robin, a human factors study was conducted in conjunction with the MRR. The most important result of the human factors study is that the Relative Operating Characteristics (ROC) curves provide a better methodology for describing inspector performance than only POD or single-point crack/no crack data.

12. KEY WORDS/DESCRIPTORS *(List words or phrases that will assist researchers in locating the report.)*

nondestructive testing, austenitic steel, ultrasonic inspection, Relative Operating Characteristics curve, ASME Code, performance demonstration, intergranular stress corrosion cracks (IGSCC)

13. AVAILABILITY STATEMENT

Unlimited

14. SECURITY CLASSIFICATION

(This Page) Unclassified

(This Report) Unclassified

15. NUMBER OF PAGES

16. PRICE