

NUREG/CR-5068
PNNL-10475

Piping Inspection Round Robin

Prepared by
P. G. Heasler, S. R. Doctor

Pacific Northwest National Laboratory
Operated by
Battelle Memorial Institute

Prepared for
U.S. Nuclear Regulatory Commission

9604230383 960430
PDR NUREG
CR-5068 R PDR

DF020

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 20555-0001
2. The Superintendent of Documents, U.S. Government Printing Office, P. O. Box 37082, Washington, DC 20402-9328
3. The National Technical Information Service, Springfield, VA 22161-0002

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 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 Government Printing Office: formal NRC staff and contractor reports, NRC-sponsored conference proceedings, international agreement reports, grantee reports, 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 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 Administration, Distribution and Mail Services Section, U.S. Nuclear Regulatory Commission, Washington, DC 20555-0001.

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at the NRC Library, Two White Flint North, 11545 Rockville Pike, Rockville, MD 20852-2738, for 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 10018-3308.

DISCLAIMER NOTICE

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus, product, or process disclosed in this report, or represents that its use by such third party would not infringe privately owned rights.

Piping Inspection Round Robin

Manuscript Completed: January 1996
Date Published: April 1996

Prepared by
P. G. Heasler, S. R. Doctor

Pacific Northwest National Laboratory
Richland, WA 99352

J. Muscara, NRC Project Manager

Prepared for
Division of Engineering Technology
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001
NRC Job Code B2289

Abstract

The piping inspection round robin was conducted in 1981 at the Pacific Northwest National Laboratory (PNNL) to quantify the capability of ultrasonics for inservice inspection and to address some aspects of reliability for this type of nondestructive evaluation (NDE). The research was sponsored by the U.S. Nuclear Regulatory Commission, Office of Research, under a program entitled "Evaluation and Improvement of NDE Reliability for Inservice Inspection of Light Water Reactors."

The round robin measured the crack detection capabilities of seven field inspection teams who employed procedures that met or exceeded the 1977 edition through

the 1978 addenda of the American Society of Mechanical Engineers (ASME) Section XI Code requirements. Three different types of material were employed in the study (cast stainless steel, clad ferritic, and wrought stainless steel), and two different types of flaws were implanted into the specimens (intergranular stress corrosion cracks (IGSCCs) and thermal fatigue cracks (TFCs)). When considering near-side inspection, far-side inspection, and false call rate, the overall performance was found to be best in clad ferritic, less effective in wrought stainless steel and the worst in cast stainless steel. Depth sizing performance showed little correlation with the true crack depths.

Contents

Abstract	iii
Executive Summary	xiii
Acknowledgments	xvii
Glossary of Terms	xix
Previous Reports in the Series	xxi
1.0 Introduction	1.1
2.0 Design Variables and Quantities to be Estimated	2.1
2.1 Detection Statistics	2.1
2.2 ROC Statistics	2.2
2.3 Sizing Statistics	2.2
2.4 Team Error Statistics	2.3
2.5 Design Variables	2.3
3.0 Experimental Design, Data Collection, and Reduction	3.1
3.1 Test Matrices	3.1
3.2 Test Protocol	3.3
3.3 Scoring the Inspection Results	3.3
3.4 Destructive Analysis Results	3.7
4.0 Variables that Influence Detection Performance	4.1
4.1 Effect of Inspection Variables	4.1
4.2 Effect of Material Variables	4.3
4.3 Relative Importance of Variables	4.8
5.0 Relationship Between Detection and Crack Size	5.1
5.1 Detection Curves for Thermal Fatigue Cracks in Cast Stainless Steel	5.4
5.2 Detection Curves for Thermal Fatigue Cracks in Clad Ferritic	5.6
5.3 Detection Curves for Thermal Fatigue Cracks in Wrought Stainless Steel	5.6
5.4 Detection Curves for IGSC Cracks in Wrought Stainless Steel	5.6
5.5 Average Detection Results for Data in Percent Through-Wall	5.8
6.0 Relationships Between Detection Performance and Decision Criteria; ROC Curves	6.1
6.1 ROC Curves for TFC in Cast Stainless Steel	6.1
6.2 ROC Curves for TFC in Clad Ferritic	6.1
6.3 ROC Curves for TFC in Wrought Stainless Steel	6.2
6.4 ROC Curves for IGSCC in Wrought Stainless Steel	6.2
7.0 Sizing Errors	7.1
7.1 Depth Sizing Error	7.1
7.2 Length Sizing Errors	7.4

Contents

8.0 Team Error Statistics	8.1
8.1 Examples of Common Team Errors	8.1
8.2 Team Error Codes	8.1
8.3 Tabulation of Team Errors	8.3
9.0 Examination of False Calls	9.1
9.1 False Call Rates and the Poisson Model	9.1
9.2 The Effect of Location on False Calls	9.3
9.3 False Call Hot Spots	9.4
10.0 Conclusions and Recommendations	10.1
10.1 POD Performance	10.1
10.2 False Call Performance	10.2
10.3 ROC Performance	10.2
10.4 Sizing Performance	10.3
10.5 Team Error Statistics	10.3
11.0 References	11.1
Appendix A Hit-O-Grams for PIRR Specimens	A.1
Appendix B True-State Plots for PIRR Specimens	B.1

Figures

Figure E.1	POD Curve Fit Surrounded by 95% Error Bounds for Near-Side Inspections of Cast Stainless Steel	xv
Figure E.2	POD Curve Fit with Associated 95% Error Bounds for Combined Near- and Far-Side Inspections of Clad Ferritic Material	xv
Figure E.3	POD Curve Fit Surrounded by 95% Error Bounds for Near-Side Inspections of Thermal Fatigue Cracks in Wrought Stainless Steel	xv
Figure E.4	POD Curve Fit Surrounded by 95% Error Bounds for Near-Side Inspections of IGSCC in Wrought Stainless Steel	xv
Figure E.5	Best Team's Length Sizing Performance on Thermal Fatigue Cracks in Clad Ferritic Material	xvi
Figure E.6	Best Team's IGSCC Length Sizing Performance in Wrought Stainless Steel Material	xvi
Figure 2.1	Example of a Relative Operating Characteristic Curve	2.3
Figure 2.2	Example of PIRR Specimens (these two are clad ferritic weldments)	2.5
Figure 2.3	Inspection of a 32 in. dia. (81.3 cm) Cast Stainless Steel Specimen During the Round Robin	2.6
Figure 2.4	Example of an Inspection Conducted under Difficult (simulated "Field") Conditions	2.8
Figure 2.5	Illustrations Showing Near, Far and Both Side Access in a) Cross-sectional View, and b) Plan View.	2.9
Figure 2.6	Defect Size Categories for Cast Stainless Steel Cracks (size categories are indicated by 1, 2, 3, and 4 for 33 in. diameter (83.8 cm), 2-3/8 in. (6 cm) wall)	2.10
Figure 2.7	Defect Size Categories for Clad Ferritic Cracks (size categories are indicated by 1, 2, 3, and 4 for 32 in. (81.3 cm) diameter, 2-3/8 in. (6 cm) wall)	2.11
Figure 2.8	Defect Size Categories for Wrought Stainless Steel Cracks (includes IGSCC and thermal fatigue) for 10-in. (25.4 cm) diameter, either 0.594 in. (1.51 cm) or 0.5 in. (1.27 cm) wall	2.11
Figure 3.1	Data from a Completed Inspection Report Form	3.5
Figure 3.2	Scoring Example Showing the Location of Cracks (T), Crack Calls (C), Non-crack Calls (N), and Grading Units (201, 202, 203, and 204) in Specimen D101-4 With Inspection From the B-side and A-side Masked Off.	3.6
Figure 5.1	Plot of Logistic Function	5.1
Figure 5.2	Logistic Curve Fit to POD Data with 95% Bounds (Cast SS, Near-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length	5.2
Figure 5.3	Logistic Curve Fit to POD Data with 95% Bounds (Cast SS, Far-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length	5.3
Figure 5.4	Logistic Curve Fit to RP Data with 95% Bounds (Cast SS, Near-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length	5.4
Figure 5.5	Logistic Curve Fit to RP Data with 95% Bounds (Cast Stainless Steel, Far-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length	5.11
Figure 5.6	Logistic Curve Fit to POD Data with 95% Bounds (Cast Stainless Steel, Near-Side Inspections, Team #1) a) POD vs. Depth b) POD vs. Length	5.11
Figure 5.7	Logistic Curve Fit to POD Data with 95% Bounds (Cast Stainless Steel, Near-Side Inspections, Team #3) a) POD vs. Depth b) POD vs. Length	5.12
Figure 5.8	Logistic Curve Fit to POD data with 95% Bounds (Cast Stainless Steel, Near-Side Inspections, Team #4) a) POD vs. Depth b) POD vs. Length	5.12
Figure 5.9	Logistic Curve Fit to POD Data with 95% Bounds (Cast Stainless Steel, Near-Side Inspections, Team #5) a) POD vs. Depth b) POD vs. Length	5.13
Figure 5.10	Logistic Curve Fit to POD Data with 95% Bounds (Cast SS, Near-Side Inspections, Team #6) a) POD vs. Depth b) POD vs. Length	5.13
Figure 5.11	Logistic Curve Fit to POD Data with 95% Bounds (Cast SS, Near-Side Inspections, Team #7) a) POD vs. Depth b) POD vs. Length	5.14
Figure 5.12	Logistic Curve Fit to POD Data with 95% Bounds (Clad Ferritic, All Teams) a) POD vs. Depth b) POD vs. Length	5.14

Figure 5.13 Logistic Curve Fit to RP Data with 95% Bounds (Clad Ferritic, All Teams) a) POD vs. Depth b) POD vs. Length	5.15
Figure 5.14 Logistic Curve Fit to POD Data with 95% Bounds (Clad Ferritic, Team #1) a) POD vs. Depth b) POD vs. Length	5.15
Figure 5.15 Logistic Curve Fit to POD Data with 95% Bounds (Clad Ferritic, Team #2) a) POD vs. Depth b) POD vs. Length	5.16
Figure 5.16 Logistic Curve Fit to POD Data with 95% Bounds (Clad Ferritic, Team #3) a) POD vs. Depth b) POD vs. Length	5.16
Figure 5.17 Logistic Curve Fit to POD Data with 95% Bounds (Clad Ferritic, Team #4) a) POD vs. Depth b) POD vs. Length	5.17
Figure 5.18 Logistic Curve Fit to POD Data with 95% Bounds (Clad Ferritic, Team #5) a) POD vs. Depth b) POD vs. Length	5.17
Figure 5.19 Logistic Curve Fit to POD Data with 95% Bounds (Clad Ferritic, Team #6) a) POD vs. Depth b) POD vs. Length	5.18
Figure 5.20 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length	5.18
Figure 5.21 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length	5.19
Figure 5.22 Logistic Curve Fit to RP Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length	5.19
Figure 5.23 Logistic Curve Fit to RP Data with 95% Bounds (Wrought SS with TFC, Far-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length	5.20
Figure 5.24 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, Team #1) a) POD vs. Depth b) POD vs. Length	5.20
Figure 5.25 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, Team #2) a) POD vs. Depth b) POD vs. Length	5.21
Figure 5.26 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, Team #3) a) POD vs. Depth b) POD vs. Length	5.21
Figure 5.27 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, Team #4) a) POD vs. Depth b) POD vs. Length	5.22
Figure 5.28 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, Team #5) a) POD vs. Depth b) POD vs. Length	5.22
Figure 5.29 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, Team #6) a) POD vs. Depth b) POD vs. Length	5.23
Figure 5.30 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with IGSCC, Near-Side Inspec- tions, All Teams) a) POD vs. Depth b) POD vs. Length	5.23
Figure 5.31 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with IGSCC, Far-Side Inspec- tions, All Teams) a) POD vs. Depth b) POD vs. Length	5.24
Figure 5.32 Logistic Curve Fit to RP Data with 95% Bounds (Wrought SS with IGSCC, Near-Side Inspec- tions, All Teams) a) POD vs. Depth b) POD vs. Length	5.24
Figure 5.33 Logistic Curve Fit to RP Data with 95% Bounds (Wrought SS with IGSCC, Far-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length	5.25
Figure 5.34 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with IGSCC, Near-Side Inspec- tions, Team #1) a) POD vs. Depth b) POD vs. Length	5.25
Figure 5.35 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with IGSCC, Near-Side Inspec- tions, Team #2) a) POD vs. Depth b) POD vs. Length	5.26
Figure 5.36 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with IGSCC, Near-Side Inspec- tions, Team #3) a) POD vs. Depth b) POD vs. Length	5.26
Figure 5.37 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with IGSCC, Near-Side Inspec- tions, Team #4) a) POD vs. Depth b) POD vs. Length	5.27
Figure 5.38 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with IGSCC, Near-Side Inspec- tions, Team #5) a) POD vs. Depth b) POD vs. Length	5.27
Figure 5.39 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with IGSCC, Near-Side Inspec- tions, Team #6) a) POD vs. Depth b) POD vs. Length	5.28

Figure 5.40 Logistic Curve Fit to POD Data with 95% Confidence Bounds (Cast Stainless Steel, Near-Side Inspection, All Teams) a) POD vs. Depth b) POD vs. Length	5.28
Figure 5.41 Logistic Curve Fit to POD Data with 95% Confidence Bounds (Wrought Stainless Steel with TFC, Near-Side Inspection, All Teams) a) POD vs. Depth b) POD vs. Length	5.29
Figure 5.42 Logistic Curve Fit to POD Data With 95% Confidence Bounds (Wrought Stainless Steel with IGSCC, Near-Side Inspection, All Teams) a) POD vs. Depth b) POD vs. Length	5.29
Figure 5.43 Logistic Curve Fit to POD Data With 95% Confidence Bounds (Clad Ferritic, Combined Near- and Far-Side Inspection, All Teams) a) POD vs. Depth b) POD vs. Length	5.30
Figure 6.1 ROC Fit in Cast Stainless Steel (Near-Side Inspection, Size 1, Thermal Fatigue Cracks)	6.4
Figure 6.2 ROC Fit in Cast Stainless Steel (Near-Side Inspection, Size 2, Thermal Fatigue Cracks)	6.4
Figure 6.3 ROC Fit in Cast Stainless Steel (Near-Side Inspection, Size 3, Thermal Fatigue Cracks)	6.4
Figure 6.4 ROC Fit in Cast Stainless Steel (Near-Side Inspection, Size 4, Thermal Fatigue Cracks)	6.4
Figure 6.5 ROC Fit in Cast Stainless Steel (Far-Side Inspection, Size 1, Thermal Fatigue Cracks)	6.5
Figure 6.6 ROC Fit in Cast Stainless Steel (Far-Side Inspection, Size 2, Thermal Fatigue Cracks)	6.5
Figure 6.7 ROC Fit in Cast Stainless Steel (Far-Side Inspection, Size 3, Thermal Fatigue Cracks)	6.5
Figure 6.8 ROC Fit in Cast Stainless Steel (Far-Side Inspection, Size 4, Thermal Fatigue Cracks)	6.5
Figure 6.9 ROC Fit in Clad Ferritic Material (Combined Near- and Far-Side Inspection, Size 1, Thermal Fatigue Cracks)	6.6
Figure 6.10 ROC Fit in Clad Ferritic Material (Combined Near- and Far-Side Inspection, Size 2, Thermal Fatigue Cracks)	6.6
Figure 6.11 ROC Fit in Clad Ferritic Material (Combined Near- and Far-Side Inspection, Size 3, Thermal Fatigue Cracks)	6.6
Figure 6.12 ROC Fit in Clad Ferritic Material (Combined Near- and Far-Side Inspection, Size 4, Thermal Fatigue Cracks)	6.6
Figure 6.13 ROC Fit for Thermal Fatigue Cracks in Wrought Stainless Steel (Near-Side Inspection, Size 1 Cracks)	6.7
Figure 6.14 ROC Fit for Thermal Fatigue Cracks in Wrought Stainless Steel (Near-Side Inspection, Size 2 Cracks)	6.7
Figure 6.15 ROC Fit for Thermal Fatigue Cracks in Wrought Stainless Steel (Near-Side Inspection, Size 3 Cracks)	6.7
Figure 6.16 ROC Fit for Thermal Fatigue Cracks in Wrought Stainless Steel (Far-Side Inspection, Size 1 Cracks)	6.7
Figure 6.17 ROC Fit for Thermal Fatigue Cracks in Wrought Stainless Steel (Far-Side Inspection, Size 2 Cracks)	6.8
Figure 6.18 ROC Fit for Thermal Fatigue Cracks in Wrought Stainless Steel (Far-Side Inspection, Size 3 Cracks)	6.8
Figure 6.19 ROC Fit for IGSCC in Wrought Stainless Steel (Near-Side Inspection, Size 1 Cracks)	6.8
Figure 6.20 ROC Fit for IGSCC in Wrought Stainless Steel (Near-Side Inspection, Size 2 Cracks)	6.8
Figure 6.21 ROC Fit for IGSCC in Wrought Stainless Steel (Far-Side Inspection, Size 1 Cracks)	6.9
Figure 6.22 ROC Fit for IGSCC in Wrought Stainless Steel (Far-Side Inspection, Size 2 Cracks)	6.9
Figure 7.1 Plot of All Teams' Near-Side Depth Measurements of TFC in Cast Stainless Steel Material	7.4
Figure 7.2 Plot of All Teams' Near-Side Depth Measurements of TFC in Clad Ferritic Material	7.4
Figure 7.3 Plot of All Teams' Near-Side Depth Measurements in Wrought Stainless Steel with Thermal Fatigue Cracks	7.4
Figure 7.4 Plot of All Teams' Near-Side Depth Measurements in Wrought Stainless Steel with IGSCC	7.5
Figure 7.5 Regression Fit of Near-Side Results of Team #2 for Wrought Stainless Steel with IGSCC	7.5
Figure 7.6 Regression Fit of Near-Side Results of Team #3 for Wrought Stainless Steel with IGSCC	7.5
Figure 7.7 Regression Fit of Near-Side Results of Team #6 for Wrought Stainless Steel with IGSCC	7.5
Figure 7.8 Regression Fit of Length Measurements (Team #1, Clad Ferritic)	7.6
Figure 7.9 Regression Fit of Length Measurements (Team #2, Clad Ferritic)	7.6
Figure 7.10 Regression Fit of Length Measurements (Team #3, Clad Ferritic)	7.6
Figure 7.11 Regression Fit of Length Measurements (Team #4, Clad Ferritic)	7.6
Figure 7.12 Regression Fit of Length Measurements (Team #5, Clad Ferritic)	7.7
Figure 7.13 Regression Fit of Length Measurements (Team #6, Clad Ferritic)	7.7

Figure 7.14 Regression Fit of Length Measurements (Team #1, Wrought Stainless Steel, IGSCC)	7.7
Figure 7.15 Regression Fit of Length Measurements (Team #2, Wrought Stainless Steel, IGSCC)	7.7
Figure 7.16 Regression Fit of Length Measurements (Team #3, Wrought Stainless Steel, IGSCC)	7.8
Figure 7.17 Regression Fit of Length Measurements (Team #4, Wrought Stainless Steel, IGSCC)	7.8
Figure 7.18 Regression Fit of Length Measurements (Team #5, Wrought Stainless Steel, IGSCC)	7.8
Figure 7.19 Regression Fit of Length Measurements (Team #6, Wrought Stainless Steel, IGSCC)	7.8
Figure 8.1 Possible Distribution Model for Team Errors	8.3
Figure 9.1 Indication Plot by Team #6 on Wrought SS Blank Specimen, Grading Units 31 and 32, Showing Clusters of "N" Decisions	9.6
Figure 9.2 Indication Plot by Team #1 on Wrought SS Blank Specimen, Grading Units 41 and 42, Showing Clusters of "N" Decisions	9.6
Figure 9.3 Hit-O-Gram for wrought SS Specimen B213-4 (number of inspections = 12). "Defect" shows the location of the crack and crack length: Crack depth is the height of this line above the horizontal axis. a) A side Results, b) B side Results	9.6
Figure 9.4 Hit-O-Gram for wrought SS Specimen B215-3, (number of inspections = 18).	9.7
Figure 9.5 Hit-O-Gram for wrought SS Specimen B214-2, (number of inspections = 18).	9.7

Tables

Table E.1. Relative Importance of Variables	xiv
Table 2.1. Design Variables	2.4
Table 2.2. Material Types	2.4
Table 2.3. Defect Types	2.5
Table 2.4. Summary of Team Member's Qualifications	2.7
Table 2.5. Inspection Access Conditions	2.7
Table 3.1. Round Robin Test Matrix for a Single Team	3.2
Table 3.2. Structure of Experiments	3.3
Table 3.3. Example of Randomized Inspection List	3.4
Table 3.4. Comparison of POD for Near and Collapsed Grading Units in Clad Ferritic Material	3.7
Table 3.5. Results of Clad Ferritic Specimen Destructive Analysis	3.7
Table 3.6. Destructive Results for Centrifugally Cast Stainless Steel Specimens	3.8
Table 4.1. Summary of Detection Performance for Cast Stainless Steel	4.2
Table 4.2. Summary of Detection Performance for Clad Ferritic Material (using both-side grading units)	4.3
Table 4.3. Summary of Detection Performance in Wrought Stainless Steel Material	4.4
Table 4.4. Significance of Procedure and Environment to POD	4.5
Table 4.5. Detections Collapsed Over Procedure and Environment Variables	4.6
Table 4.6. Summary of the Effect of Access	4.7
Table 4.7. Summary of Significance of Size	4.7
Table 4.8. Summary of Significance of Material	4.8
Table 4.9. Significance of Defect Type in Wrought Stainless Steel (TFC vs. IGSCC)	4.9
Table 4.10. Relative Importance of Variables	4.10
Table 5.1. Table of GOF Statistics for Two POD Curve Fits	5.2
Table 5.2. Summary of POD Logistic Curve Fits in Cast SS	5.5
Table 5.3. Summary of POD Logistic Fits in Clad Ferritic	5.7
Table 5.4. Summary of POD Logistic Fits in Wrought SS with TFC	5.8
Table 5.5. Summary of POD Logistic Fits in Wrought SS with IGSCC	5.9
Table 6.1. Detection Results in Cast Stainless Steel Material for Near-Side Access	6.2
Table 6.2. Detection Results in Cast Stainless Steel Material for Far-Side Access	6.2
Table 6.3. Detection Results in Clad Ferritic Material	6.3
Table 6.4. Detection Results for Near-Side Inspection of Thermal Fatigue Cracks in Wrought SS	6.3
Table 6.5. Detection Results for Far-Side Inspection of Thermal Fatigue Cracks in Wrought SS	6.3
Table 6.6. Detection Results for Near-Side Inspection of IGSCC in Wrought SS Material	6.3
Table 6.7. Detection Results for Far-Side Inspection of IGSCC in Wrought SS Material	6.3
Table 7.1. Summary of Depth Sizing Regressions	7.2
Table 7.2. Summary of Length Sizing Regressions	7.3
Table 8.1. Categorization of Error Codes	8.2
Table 8.2. Team Errors Classified by Error Code and Teams	8.4
Table 8.3. Inservice Inspection Errors Classified by Consequence	8.5
Table 9.1. False Call Rates	9.2
Table 9.2. False Recording Rates (C+N Indications)	9.2
Table 9.3. Comparison of Actual and Calculated FCP	9.3
Table 9.4. Comparison of Actual and Calculated False Recording Probability	9.4
Table 9.5. Effect of Grading Unit Location on FCP and RP	9.5

Executive Summary

The Piping Inspection Round Robin (PIRR) was conducted in 1981 to determine the detection and sizing capabilities of ultrasonic inspection procedures that met or exceeded the 1977 edition through the 1978 addenda of the ASME Section XI Code requirements, particularly as practiced in the field. Seven teams, all of them employed by commercial inservice inspection (ISI) companies, participated in the round robin. An individual team (consisting of Level I, II, and III inspectors--see Table 2.4) conducted approximately 250 inspections on welded pipe consisting of clad ferritic, cast, and wrought stainless steel specimens. Five inspection teams performed all the inspections, one team inspected everything but the cast stainless steel and the other team only inspected the cast stainless steel. An inventory of approximately 80 pipe specimens was assembled for the round robin.

Inspections were conducted under a variety of conditions in order to assess the effect of these conditions on detection performance. Some of the important conditions included in the round robin were:

Procedure:	"As Practiced" vs. "Improved" procedures that meet or exceed ASME requirements at the time of the round robin
Environment:	Laboratory vs. simulated Difficult field conditions
Access:	Near-side vs. Far-side access to defects (i.e., relative to probe, is the defect located on the near-side or far-side of the weld?)
Material:	Cast Stainless Steel vs. Clad Ferritic vs. Wrought Stainless Steel
Crack Type:	Thermal Fatigue vs. Intergranular Stress Corrosion
Crack Size:	Blank, to approximately 50% through-wall in depth and blank, up to approximately 3.5 in. (8.9 cm) in length. (Section 5 of this report describes in detail the relationship of crack depth to crack detection.)

The major objective of the PIRR was to assess the effect of these variables on detection performance. The round robin was designed to accomplish this in several ways.

For example, the importance of each variable can be gauged by a chi-squared statistic that measures the effect the variable (or variables) has on probability of detection (POD). A list of these statistics is given in Table E.1. Table E.1 offers a concise summary of the round robin results. Important variables are those with chi-squared/DOF ratios that are much larger than one.

From this table it is clear that inspection environment and procedure (at least as they were defined in the PIRR) are not important. On the other hand, all other variables have a significant effect on POD, with the two most obvious variables (defect size and material) having the greatest effect.

It should be noted that there was a surprise associated with one of the variables--crack type. We had expected that the thermal fatigue cracks (TFCs) could serve as reasonable replacements for intergranular stress corrosion cracks (IGSCCs) in wrought stainless steel for round-robin testing and other experimental work, and had included both types of cracks to test this hypothesis. The chi-squared statistic in Table E.1 shows that there is a very definite difference between TFC and IGSCC in wrought stainless steel. However, it was found that the TFC were harder to detect than the IGSCC, so they can certainly serve as conservative surrogates to IGSC cracks in tests.

The round-robin data also allowed quantification of the relationship between crack depth and POD. Curves were constructed to plot the relationship of POD to crack depth, using mathematical regression techniques to fit the curve to the experimental data. Figures E.1 and E.2 illustrate the POD curves for cast stainless steel under near-side inspection conditions and clad ferritic piping under combined near- and far-side inspection conditions, respectively. The curves are bracketed by the 95% confidence bounds. The point on the curve associated with a crack depth of 0 actually represents the false call probability. As one can see from the figures, performance for clad ferritic is very good; but in cast stainless steel, it is little better than guessing. (That is, the false call probability is not much different from the POD at the largest crack depth.) The false call rate in the clad ferritic material was the lowest of all the three materials studied. The false call rate in cast stainless steel was four times higher than the rate in clad ferritic material and the false call rate in wrought stainless steel was five times higher than the

Table E.1. Relative Importance of Variables

Variable	Chi-Squared	Degrees of Freedom (DOF)	"Importance" Chi-Squared/DOF
Environment and Procedure	93.8	94	1.0
Crack Type (in wrought stainless steel)	73.6	8	9.2*
Access	260.2	26	10.1*
Material	477.4	42	11.4*
Defect Size	250.8	20	12.5*

*Statistically significant at 0.1% level, which means that these variables are very important and have a large impact on inspection performance.

rate in the clad ferritic material.

In wrought stainless steel, TFC and IGSCC displayed different probability of detection characteristics, as illustrated in Figures E.3 and E.4. As one can see from these curves, TFCs are more difficult to detect than IGSCCs. It should be noted that these curves also represent near-side inspection performance. Far-side inspection performance in wrought stainless steel was similar to the near-side performance in cast stainless steel, as shown in Figure E.1. This is to be expected, since the sound fields in both cases must propagate through coarse-grained anisotropic material.

In general, crack depth measurements performed during the round robin were poor. None of these teams would meet or exceed the flaw depth sizing requirements in the ASME Code Section XI Appendix VIII. The regression fits had large standard deviations, large root mean square errors and small slopes. It should be noted that many of the teams' depth measurements showed no significant correlation with the true sizes. This is to be expected, since all the teams used probe motion to estimate depth. For small defects (relative to the size of the sound field), probe motion measure

ments provide estimates only of the sound field size, and not of the defect size. However, a few of the teams did a better job on sizing the length of cracks. Figures E.5 and E.6 present the best team's length sizing measurements for TFCs in clad ferritic and IGSCCs in wrought stainless steel, respectively.

A relative operating characteristic (ROC) analysis was performed with the data. This study was not designed for an ROC analysis so the curve fits were based on only two points. This work was performed to see if it provided any further insights to the inspection process. The results supported the conclusions based on the other analyses conducted on the data and in some ways made some points easier to interpret because of the graphical presentation of the data.

Extensive efforts were spent in collecting data of the teams in a manner that permitted the identification of errors that are made during the inspection process. This data shows that a team will make a large sizing or location error about 5% of the time. This is not insignificant and points out the need to improve the human factors aspects of ISI in order to increase the reliability of inspections.

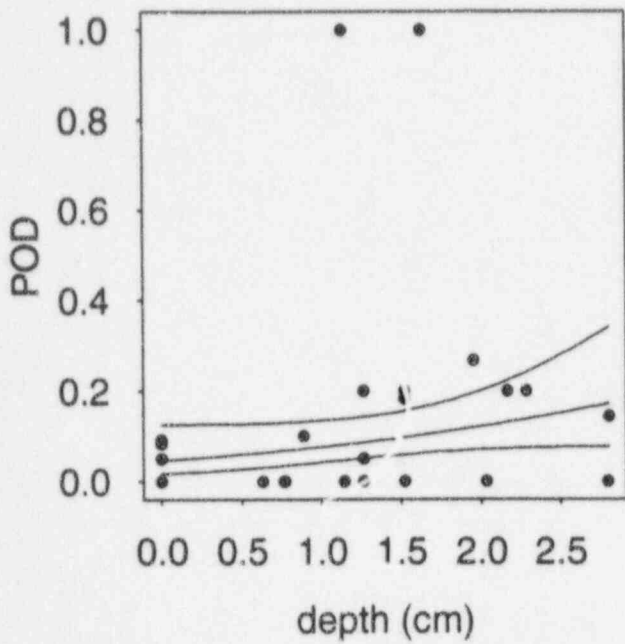


Figure E.1 POD Curve Fit Surrounded by 95% Error Bounds for Near-Side Inspections of Cast Stainless Steel

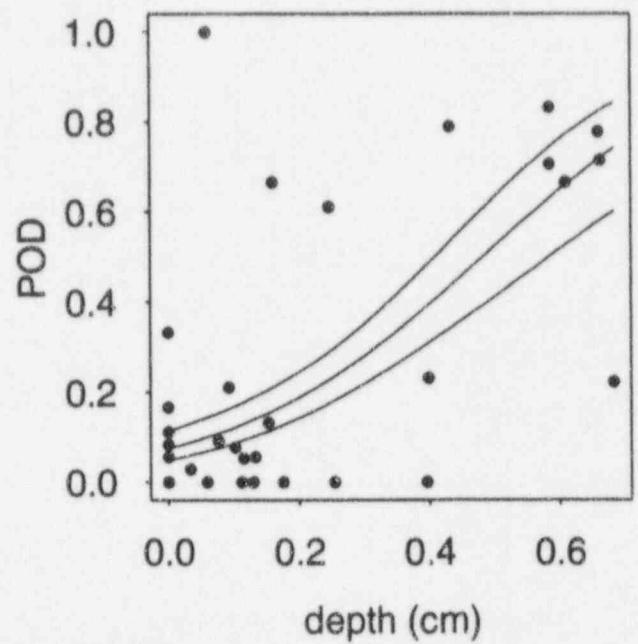


Figure E.3 POD Curve Fit Surrounded by 95% Error Bounds for Near-Side Inspections of Thermal Fatigue Cracks in Wrought Stainless Steel

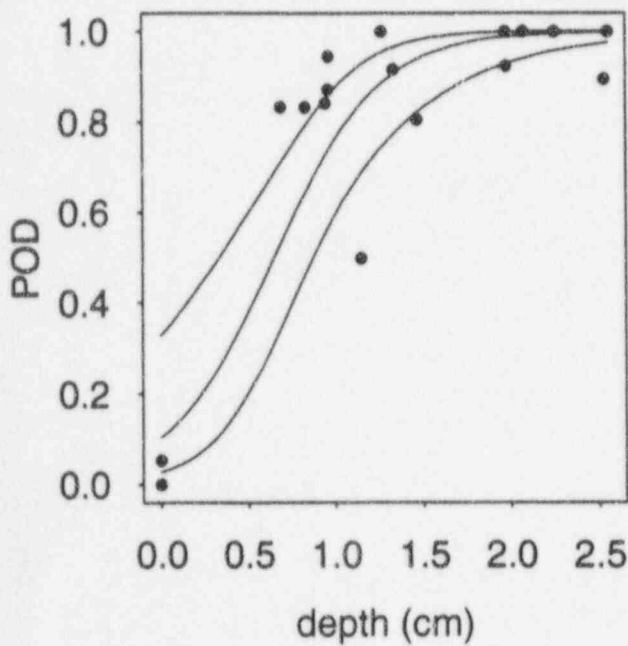


Figure E.2 POD Curve Fit with Associated 95% Error Bounds for Combined Near- and Far-Side Inspections of Clad Ferritic Material

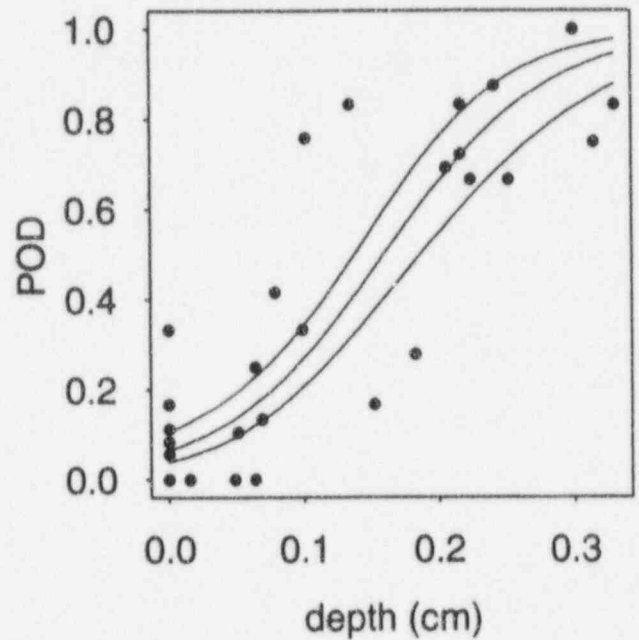


Figure E.4 POD Curve Fit Surrounded by 95% Error Bounds for Near-Side Inspections of IGSCC in Wrought Stainless Steel

Executive Summary

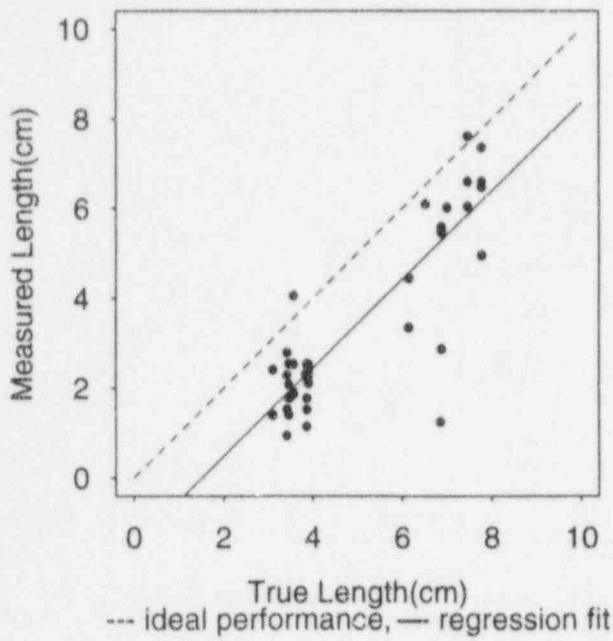


Figure E.5 Best Team's Length Sizing Performance on Thermal Fatigue Cracks in Clad Ferritic Material

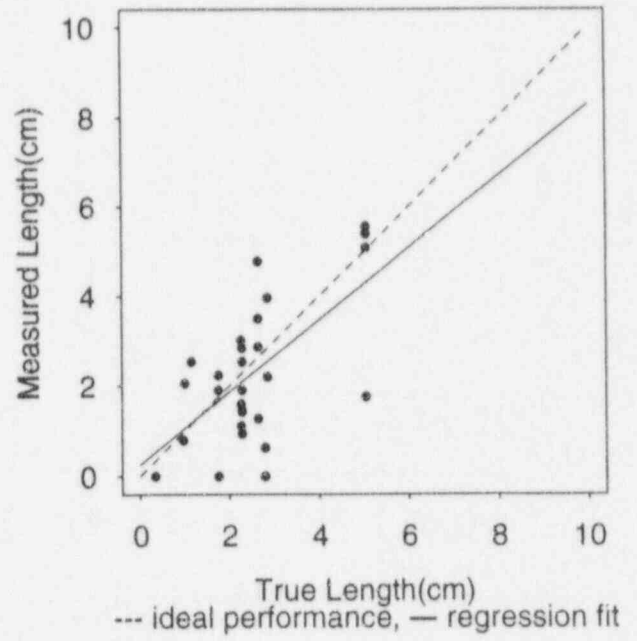


Figure E.6 Best Team's IGSCC Length Sizing Performance in Wrought Stainless Steel Material

Acknowledgments

There have been a number of significant contributions to this study by a large number of staff members at the Pacific Northwest National Laboratory including: Ivar Husa (now at HEDL), Gordon Dudder, Linnea Bickford (now at Savannah River Site), Bob Watson, Tom Fish, Rick Kurtz, and Stan Pitman -- all from the Materials Department; Dan Spanner (now at GTE), Robert Bowey, Howard Hartzog (now at Timken), Tom Taylor [now at Electric Power Research Center (EPRI) NDE Center], and Jack Spanner (now retired) -- from the Energy Division. A special thanks goes to F. Larry Becker (currently at the J. A. Jones Applied Research Center - EPRI NDE Center) who was the PNNL program manager who developed the PIRR design and oversaw the data collection. In addition, special appreciation goes to Greg Selby (also now at the EPRI NDE Center) who was instrumental in conducting the NDE acceptance testing of the specimens, overseeing the

day-to-day data collection activities, and the in-depth review of all data for errors.

The following vendors participated in the PIRR study, and we would like to thank them and their staff who contributed to this study.

- Magnaflux Corporation
- Babcock and Wilcox
- Westinghouse Electric
- Nuclear Energy Services
- Southwest Research Institute
- Lambert, McGill, and Thomas, Inc.
- General Electric

Thanks to Kay Hass and Cathy Stephens for typing and Frank Ryan editing this manuscript, and to Dr. Joe Muscara of the Nuclear Regulatory Commission Research Office for direction and guidance of this work.

Glossary of Terms

C-Indication - A location on the weld identified during inspection as a crack. An indication in this exercise is described by three numbers corresponding to the depth of the crack, axial distance of the indication from the reference scribe line, and the circumferential coordinates that identify each end.

Cast Stainless Steel - Cast stainless steel specimens used in this exercise were fabricated from CF-8A centrifugally cast pipe 32 in. (81.3 cm) in outside diameter with 2-3/8 in. (6 cm) wall thickness.

Clad Ferritic - Clad ferritic specimens used in this exercise were fabricated from A102 mild carbon steel piping 33 in. (83.8 cm) in outside diameter with a 2-3/8 in. (6 cm) wall thickness. Specimens were clad on the I.D. side.

Decision Threshold - When NDE indications are found during inspection, the inspector must decide whether or not the indication represents an actual crack. In this exercise, the inspector was to mark an "N" or "C" with every indication to signify that the indication was either (N)ot associated with a crack or that it represented a (C)rack. These two designations (N and C) are called the decision thresholds.

EDM - Electric Discharge Machined. A type of artificial defect employed in the round robin.

FCP - False Call Probability. The probability that a blank grading unit receives a C-indication.

FRP - False Recording Probability. The probability that a blank grading unit receives any indication (either C or N).

Grading Unit - A fixed length of material used to calculate false call and detection probabilities. In this round robin, the grading units are 3-in. (7.6-cm) long.

IGSCC - Intergranular Stress Corrosion Crack. A type of crack present in some of the wrought stainless steel specimens.

Inspection Access - A design variable for the round robin. There were three different inspection access conditions considered in this exercise: near-side, far-side, and both-side access.

Inspection Environment - A design variable for the round robin. Two different environmental conditions were represented in the round robin: simulated field conditions and laboratory conditions.

ISI - Inservice Inspection

Material/Crack Type - A design variable for the round robin. Four different levels of this variable were present in the round robin: cast stainless steel with thermal fatigue cracks, clad ferritic with thermal fatigue cracks, wrought stainless steel with thermal fatigue cracks, and wrought stainless steel with intergranular stress corrosion cracks.

N-Indication - An ultrasonic signal that is recorded but is not thought to be caused by a crack.

PIRR - Piping Inspection Round Robin

PISC - Programme for the Inspection of Steel Components

POL - Probability of Detection. The probability that a C-indication will be placed in a cracked grading unit during inspection.

Procedure Type - A design variable for the round robin. Two different ultrasonic procedures were employed in the round robin: the teams normal field procedure that met or exceeded the minimum ASME "Code" requirements (up through 1978 addenda) and an "improved" procedure developed by PNNL.

ROC - Relative Operating Characteristic. A method of analysis that describes how detection performance (FCP and POD) change as the inspection decision criteria are varied.

RP - Recording Probability. The probability that any indication (either C or N) will be placed within a grading unit.

Thermal Fatigue Crack (TFC) - A type of crack employed in the round robin.

Wrought Stainless Steel - Identifies specimens fabricated from 10-in.-dia. (25.4 cm) Schedule 80 and 80S, Type 304 wrought stainless steel piping.

Previous Reports in the Series

Doctor, S. R., A. A. Diaz, J. R. Friley, M. S. Good, M. S. Greenwood, P. G. Heasler, R. L. Hockey, R. J. Kurtz, F. A. Simonen, J. C. Spanner, T. T. Taylor, and T. V. Vo. 1993. *Nondestructive Examination (NDE) Reliability for Inservice Inspection of Light Water Reactors*. NUREG/CR-4469, PNL-5711, Vol. 15. Pacific Northwest Laboratory, Richland, Washington.

Doctor, S. R., A. A. Diaz, J. R. Friley, M. S. Good, M. S. Greenwood, P. G. Heasler, R. L. Hockey, R. J. Kurtz, F. A. Simonen, J. C. Spanner, T. T. Taylor, and T. V. Vo. 1992. *Nondestructive Examination (NDE) Reliability for Inservice Inspection of Light Water Reactors*. NUREG/CR-4469, PNL-5711, Vol. 14. Pacific Northwest Laboratory, Richland, Washington.

Green, E. R., S. R. Doctor, R. L. Hockey, and A. A. Diaz. 1992. *Development of Equipment Parameter Tolerances for the Ultrasonic Inspection of Steel Components: Application to Components up to 3 Inches Thick*. NUREG/CR-5817, Vol. 1. Pacific Northwest Laboratory, Richland, Washington.

Doctor, S. R., M. S. Good, P. G. Heasler, R. L. Hockey, F. A. Simonen, J. C. Spanner, T. T. Taylor, and T. V. Vo. 1992. *Nondestructive Examination (NDE) Reliability for Inservice Inspection of Light Water Reactors*. NUREG/CR-4469, PNL-5711, Vol. 13. Pacific Northwest Laboratory, Richland, Washington.

Doctor, S. R., M. S. Good, P. G. Heasler, R. L. Hockey, F. A. Simonen, J. C. Spanner, T. T. Taylor, and T. V. Vo. 1992. *Nondestructive Examination (NDE) Reliability for Inservice Inspection of Light Water Reactors*. NUREG/CR-4469, PNL-5711, Vol. 12. Pacific Northwest Laboratory, Richland, Washington.

Doctor, S. R., M. S. Good, E. R. Green, P. G. Heasler, F. A. Simonen, J. C. Spanner, T. T. Taylor, and T. V. Vo. 1991. *Nondestructive Examination (NDE) Reliability for Inservice Inspection of Light Water Reactors*. NUREG/CR-4469, PNL-5711, Vol. 11. Pacific Northwest Laboratory, Richland, Washington.

Heasler, P. G., T. T. Taylor, J. C. Spanner, S. R. Doctor, and J. D. Deffenbaugh. 1990. *Ultrasonic Inspection Reliability for Intergranular Stress Corrosion Cracks: A Round Robin Study of the Effects of Personnel, Procedures, Equipment and Crack Characteristics*. NUREG/CR-4908. Pacific Northwest Laboratory, Richland, Washington.

Spanner, J. C., S. R. Doctor, T. T. Taylor/PNL and J. Muscara/NRC. 1990. *Qualification Process for Ultrasonic Testing in Nuclear Inservice Inspection Applications*. NUREG/CR-4882, PNL-6179. Pacific Northwest Laboratory, Richland, Washington.

Doctor, S. R., J. D. Deffenbaugh, M. S. Good, E. R. Green, P. G. Heasler, F. A. Simonen, J. C. Spanner, T. T. Taylor, and T. V. Vo. 1990. *Nondestructive Examination (NDE) Reliability for Inservice Inspection of Light Water Reactors*. NUREG/CR-4469, PNL-5711, Vol. 10. Pacific Northwest Laboratory, Richland, Washington.

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. NUREG/CR-4469, PNL-5711, Vol. 9. Pacific Northwest Laboratory, Richland, Washington.

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*. NUREG/CR-4469, PNL-5711, Vol. 8. Pacific Northwest Laboratory, Richland, Washington.

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. 1988. *Nondestructive Examination (NDE) Reliability for Inservice Inspection of Light Water Reactors*. 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*. 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*. 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*. 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*. 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*. NUREG/CR-4469, PNL-5711, Vol. 2. Pacific Northwest Laboratory, Richland, Washington.

Previous Reports in the Series

- 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-3869, PNL-5140. 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

Operators of commercial light-water nuclear power plants are required by Federal Regulations to periodically inspect their primary coolant piping systems. These requirements are accomplished by inservice inspection (ISI) at refueling and other shutdowns, using a variety of nondestructive evaluation (NDE) techniques that are specified in the American Society of Mechanical Engineers (ASME) Boiler and Pressure Vessel Code, Section XI.

Plant operating histories show that the service-induced degradation in piping includes thermal fatigue cracking (TFC) and intergranular stress corrosion cracking (IGSCC). The inservice inspection method mandated by the ASME Code, as most appropriate for the detection of piping cracks, is ultrasonic testing (UT).

Pacific Northwest National Laboratory (PNNL), operated by Battelle Memorial Institute, is engaged in a multi-year study entitled "Evaluation and Improvement of NDE Reliability for Inservice Inspection of Light-Water Reactors," sponsored by the U.S. Nuclear Regulatory Commission (NRC). The program's purpose is to:

- assess the reliability of ISI methods as practiced;
- assess the impact of ISI reliability on piping integrity, using probabilistic fracture mechanics analysis;
- evaluate advanced NDE techniques to improve inspection reliability; and
- improve requirements to ensure more effective inservice inspection.

The assessment of UT crack detection reliability in reactor piping ISI has been accomplished via blind round-robin tests at PNNL. Specifically, the Piping Inspection Round Robin (PIRR) test and its results are the subject of this report. Another round robin was conducted by PNNL to evaluate improvements in inspector performance using some of the same PIRR specimens. This second test is called the Mini-Round Robin (MRR) (Heasler, et al. 1990). The PIRR was conducted from May through October 1981; and the MRR was conducted from June through October 1985. The wrought stainless steel pipe specimens were also used to evaluate advanced inspection techniques (Doctor et al. 1986) and the cast stainless steel specimens were used in an international round-robin exercise in cooperation with the European Committee on the Safety of Nuclear Installations (CSNI), Programme for

the Inspection of Steel Components (FISC) program (Bates et al. 1987).

The PIRR test was designed to generate sufficient data to describe the effects on crack detection performance of seven important inspection parameters (material type, team, defect type, inspection access, defect size, inspection environment, and procedure). With the help of statistical models, the relationships between inspection parameters and performance were quantified and extrapolated to include conditions not measured in the test.

Inspection performance under a given set of conditions cannot be coherently summarized by a single quantity. In this report, we chose to summarize detection performance with probability of detection (the probability that a defective unit of material is classified as defective) and false call probability (the probability that a "good" unit of material is classified as defective). The other aspect of inspection performance, that of sizing error, was evaluated with the aid of a regression model.

The important elements in obtaining an accurate measurement of ISI performance are the nature and quantity of the pipe specimens and the defects used. PNNL has developed techniques for inexpensive and controllable production of thermal fatigue and IGSC cracks in segments of welded nuclear reactor piping. The round-robin test used approximately 80 pipe specimens, each containing about 8 in. (20.3 cm) of circumferential pipe-to-pipe butt weld. Some specimens were left uncracked, but the majority contained one or two cracks and a few had machined notches. The three types of piping included in this test were: 32-in.-dia. (81.3-cm), 2-3/8-in. (6-cm) wall, ferritic steel with stainless steel cladding on the inside surface; 33-in.-dia. (83.8-cm), 2-3/8-in. (6-cm) wall, centrifugally cast stainless steel (CCSS) from two different heats of ASTM A-351 Grade CF-8A (which is a cast 304 material); and 10-in.-dia. (25.4-cm) Schedule 80, 0.594-in. (1.51 cm) wall, and 80S, 0.500-in. (1.27 cm) wall, rolled and welded Type 304 stainless steel.

Three-man inspection teams from six commercial ISI vendors participated in the round robin, while a seventh team participated exclusively in the cast stainless steel portion of the round robin (Taylor 1984). Each team consisted of Level I, II, and III examiners (see Table 2.4), selected to represent neither the most nor the

1.0 Introduction

least levels of ISI experience available. Thus, the results should be representative of the average industry capability at that time, and of current teams that employ procedures similar to the ones employed in the PIRR.

Each team spent three weeks at PNNL, working on average eleven hours per day and six days per week, to complete a carefully designed test matrix of 253 separate inspections per team. Each specimen inspection area consisted of approximately 8 in. (20.3 cm) of circumferential weldment, so each team inspected 169 ft. (51.5 m) of weld. A variety of inspection conditions were simulated; but in almost all cases, specimens were masked to permit inspection access from only one side of the weld. A PNNL observer monitored the test at all times, ensuring that test protocols were followed. The observer was also responsible for the operation of a four-channel strip chart recorder, used to record transducer position and UT data for the duration of each inspection.

The specimens were presented to the teams in random order. The large number and uniform appearance of the specimens, together with the random order of inspection, precluded guesswork as to the contents of upcoming specimens. Even the PNNL observer was unaware of the contents of the specimens, because each one arrived for the inspection already masked and mounted in a holding jig. These measures ensured the "blindness" of the test.

The inspection results of all teams were manually reviewed to remove those errors that might affect the scoring procedure, but which would not have occurred in the field. These were the type of errors that would normally be found in a standard ISI review. The actual scoring of the results was performed by computer in order to follow a consistent set of decision rules to evaluate the results.

This report is organized in ten sections. The first section is an introduction as to why the work was conducted. Section 2 contains a description of the performance parameters to be measured and the inspection variables selected for study. Section 3 contains a detailed explanation of the design of the PIRR, the data to be collected, the analysis methodology, and an overview of the destructive assay. Section 4 is an analysis of the importance of each of the inspection variables studied. Section 5 examines the relationship between the POD and the defect depth as well as defect length. Section 6 examines the relationship between POD and the decision criteria, through use of relative operating characteristic (ROC) analysis. Section 7 evaluates the team errors associated with defect depth and length estimates. Section 8 is a detailed look at all of the team errors that occurred in the data, with emphasis on how these errors impacted defect detection. Section 9 investigates the distribution of false calls and identifies the important variables that impact false call rates. Section 10 contains conclusions and recommendations drawn from this study.

2.0 Design Variables and Quantities to be Estimated

This section describes the statistical analysis methods used to quantify inspection performance. Because inspection performance is influenced by the materials and the conditions under which inspections are performed, another important set of variables in the round-robin analysis are the design variables that define these conditions. These design variables are described in this section.

2.1 Detection Statistics

The fundamental objective of any nondestructive examination is to find defects reliably and size them accurately. The resulting data may be used in a fracture mechanics analysis to determine the integrity of components: i.e., those which are "fit" for service; those which can remain in service but require further monitoring; and those which must be repaired, replaced, or removed from service. In particular, nondestructive examinations of nuclear power plant piping must distinguish those weldments that contain cracks from those that do not. Cracks are the most significant inservice degradation mode and were the only defect type studied in the PIRR.

Consequently, two statistics of fundamental importance are:

POD: Probability of Detection; i.e., the frequency with which an inspection system (personnel, equipment, and procedure) correctly classifies a specified unit of defective weld material as a crack

and correspondingly,

FCP: False Call Probability; i.e., the frequency with which an inspection system incorrectly classifies a specified unit of good (blank) weld material as defective.

Together, these two statistics define the crack-discriminating capability of any combination of inspector, equipment, and procedure. For an effective inspection, the POD score will be much larger than the FCP score. For an ineffective inspection, the two scores may be equal. Also, a "perfect" inspection (one that never misclassifies any material) will have a POD score of 1 and an FCP score of 0.

For meaningful estimates of FCP and POD, certain important features of the test must be specified:

- the size and shape of the unit of material used to define FCP and POD,
- the criteria used to define detection and/or classification, and
- the test environment under which the inspections will be conducted.

As one increases the length or width of the unit of material under test (this unit of material is called the "grading unit" in this report), both FCP and POD will typically increase. Under ideal circumstances, a simple mathematical relationship exists between these two quantities and grading unit size, so results from one grading unit size can be extrapolated to other sizes. Most importantly, such a relationship allows experimental results obtained from small round-robin specimens to be extrapolated without bias to entire pipe weldments that are inspected in the field.

Unfortunately, in-field inspection procedures do not produce results that fit these simple mathematical relationships. It is therefore important to select a grading unit that is relevant to in-field performance. All grading units employed within this study were standardized to a grading unit size of 3 in. (7.6 cm) (except for an analysis in Section 9, whose purpose was to investigate the effect of grading unit size). A grading unit size of 3 in. (7.6 cm) is large enough to contain almost all of the largest cracks in the round robin.

Blank and defective grading units were externally identical in this study and were not marked on the pipe, so the inspectors did not know their locations. Furthermore, the rules used to define POD were exactly the same as those used to define FCP; blank and defective material were "scored" in exactly the same way. This feature allows FCP and POD statistics to be directly compared.

Two additional detection statistics closely related to POD and FCP were also employed in this study. These statistics describe the frequency with which defect indications [both crack defects and spurious metallurgical or geometric (non-crack) effects] were recorded in a grading unit. We call these frequencies the Recording

2.0 Design Variables

Probability (RP) and False Recording Probability (FRP) and define them as follows:

RP: The frequency with which an inspection system produces a recordable (defect) indication within a specified unit of defective weld material.

FRP: The frequency with which an inspection system produces a recordable (defect) indication within a specified unit of good (blank) weld material.

The ordered pair of statistics (FRP,RP) can be thought of as (FCP, POD) calculated with different decision criteria.

2.2 ROC Statistics

ROC statistics are used to determine the manner in which detection performance changes as the decision criteria employed in the inspection procedure are varied. The (FCP, POD) performance of a well-calibrated inspection system can be represented as a single point on a Relative Operating Characteristics (ROC) diagram, as illustrated in Figure 2.1. As the decision criteria are varied, this point should trace out a curve, known as the ROC curve. This curve provides a very clear description of the compromises between false calls and detections that an inspection system is forced to make.

The ROC method was applied to other studies at PNNL after the PIRR had been conducted. Because of its demonstrated usefulness, it was decided to apply this approach to the PIRR data to see if further useful information could be extracted from this study. Because the PIRR test was not designed to be analyzed using the ROC method, a full ROC curve with many points could not be developed. Instead, we developed a very limited one with only four points--two of which are (0,0) and (1,1).

In UT methods, two decision criteria significantly affect the procedure. First, the distance-amplitude curve (DAC) used to define the recording threshold for a signal can be varied. At the time the PIRR was being conducted, the ASME Code (1977 edition through the 1978 addenda) prescribed a minimum requirement (50% DAC), but a specific inspection procedure may

use a recording threshold that is more stringent (i.e., 20% DAC).

Second, the inspector may be more or less stringent in classifying recorded indications as "cracks." In most UT procedures, this classification process is quite dependent on the experience of the inspectors and is, therefore, most likely to show variation from one inspector to another.

In the PIRR data, we had the opportunity to use probability of detection and recording probability statistics to define an ROC curve. [In other words, a curve that passes through the ordered pairs (FCP, POD) and (FCP, RP) can be defined.] A curve defined in this manner is determined exclusively by the effect that an inspector's crack classification scheme has on detection performance; therefore, the ROC curves presented in this report do not attempt to describe the effect of a change in DAC recording level.

The shape of an inspection system's ROC curve tells a great deal about its potential capabilities. A "good" ROC curve gets close to the point FCP=0, POD=1 while a poor curve lies close to the diagonal line shown in Figure 2.1. In this study, the ROC curve was determined by regressing a curve with a set parametric form onto the two estimated (FCP, POD) and (FRP,RP) points. Because the curve was fit to only two points in the ROC space, the curve must have a simple shape. The shape chosen is symmetric about the line $POD = 1 - FCP$, and contains only one unknown parameter. The equation for the ROC curve is

$$POD = \text{logit} (A + \text{logit}^{-1} (FCP)) \quad (2.1)$$

where A is the unknown parameter.

2.3 Sizing Statistics

To compare actual crack sizes and locations with inspection results, it was necessary to identify which ultrasonic indications are associated with a particular crack. It is not easy to make this identification because the association may not be unique or one-to-one; several ultrasonic indications may be associated with a single crack, and vice-versa. There is no generally accepted algorithm to determine which cracks are associated with

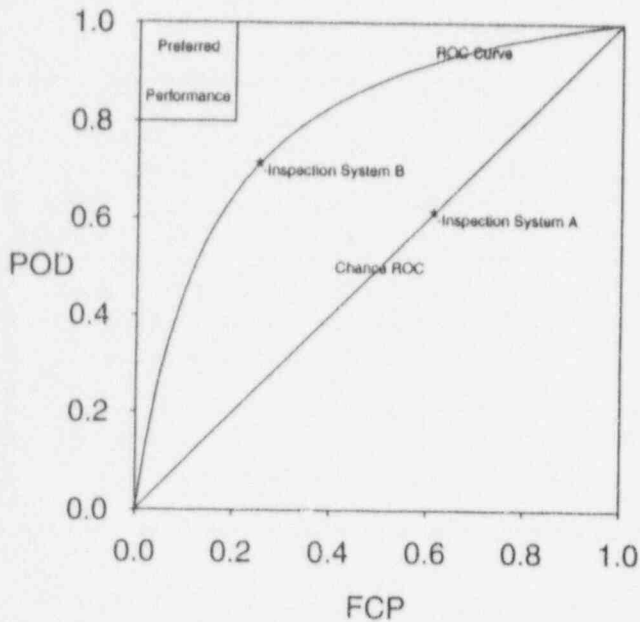


Figure 2.1 Example of a Relative Operating Characteristic Curve

which ultrasonic indications.

In general, the scoring scheme will associate the crack indication with the actual crack nearest to it. A crack indication is an ultrasonic indication that has been explicitly labeled as a crack by the inspection team. When more than one crack indication intersects an actual crack, then the indication with the largest depth (or if the depths are equal, the longest length) was associated with the crack. This scheme always associates a unique crack indication with each crack. The scheme is slightly more complicated than the description above implies, because it must also deal with missing data (e.g., dimensions of the indication).

The relationship between true and indicated crack size was analyzed with a linear regression model. The indicated crack depth (ICD) was assumed to be related to the true depth according to a regression model of the form:

$$ICD = A + B \times (\text{True Depth}) + \text{Error} \quad (2.2)$$

where A and B are parameters determined by the regression. Therefore, the crack sizing ability of an

inspection system is described by the parameters A and B (measurement/bias), and the standard deviation of the measurement error (measurement variability).

2.4 Team Error Statistics

After the round robin was completed and all data were computerized, the computerized records were checked against the team's inspection strip charts. The four channels recorded from each scan included axial probe position, circumferential probe position, indication range, and indication amplitude. The data were checked to find data recording errors and, more importantly, to determine causes for missed cracks.

All identified errors the teams committed during the inspection were tabulated, and identified causes for missed cracks were annotated to the inspection results in the computer, so that the team error rates could be calculated. These error rates can be used to determine the potential improvement in (FCP, POD) performance if certain causes of error were eliminated.

2.5 Design Variables

A major objective of the PIRR test was to determine detection effectiveness [i.e., (FCP, POD) estimates] for procedures that meet or exceed ASME Code requirements, in effect during the early 1980s, when they are employed under typical field conditions. Unfortunately, FCP and POD cannot be expected to be constant under typical field conditions; the phrase "typical field conditions" does not describe a single set of inspection conditions, but rather a spectrum of conditions. There is very little reason to believe that detection performance will remain constant over all the different sets of conditions this phrase encompasses.

One of the first steps undertaken in the planning stages of the Piping Inspection Round Robin was to explicitly enumerate all the different sets of inspection conditions that might be encompassed by the phrase "Code procedure inspections conducted under typical field conditions." After all conditions were enumerated, a subset of the list was selected for use in the PIRR. The variables that define these inspection and material conditions are called the "design variables" of the test.

2.0 Design Variables

Seven design variables were selected to uniquely define an inspection condition. While more variables could be added to this list, these seven were considered to be the most important, and should account for most of the variation in the test results. The seven variables are described in Table 2.1.

Table 2.1. Design Variables

Design Variable	Example of a Set of Inspection Conditions Defined by the Variables
Material Type	10-in. (25.4-cm), Schedule 80S, Type 304 stainless steel
Defect Type	Thermal fatigue cracks (TFC)
Inspection Team	Team "1"
Inspection Environment	Difficult (inspector dressed in radiation clothing, weld located in an awkward-to-reach position)
Access	Near side (defect is located on the accessible side of the weld centerline)
Procedure Type	Code (procedure as practiced by the team in the field)
Defect Size	10% through-wall crack

We therefore consider the phrase "a set of inspection conditions" to refer to a list of settings for the seven variables listed in Table 2.1.

Three different possible choices, options, or "levels" were defined for the variable "material type" and are listed in Table 2.2.

The centrifugally cast stainless steel and clad ferritic steel specimens represent the main coolant piping of pressurized water reactors. The 10-in. (25.4-cm) wrought stainless steel specimens represent primary coolant piping of boiling water reactors.

The weld specimen dimensions were chosen so that a specimen could be easily transported by a single technician, as shown in Figures 2.2 and 2.3. The specimens

Table 2.2. Material Types

Material Types	Description
Cast Stainless Steel	Centrifugally cast stainless steel with a 32 in. (81.3 cm) OD and 2-3/8 in. (6 cm) wall
Clad Ferritic	A-102 mild carbon steel with a stainless steel ID cladding; 33 in. (83.8 cm) OD and 2-3/8 in. (6 cm) wall
Wrought Stainless Steel	10 in. (25.4 cm), Schedule 80 and 80S, Type 304 wrought stainless steel

were made by welders qualified to ASME Code. The welds were made under shop conditions, but are typical of field practice. The weld crowns were ground flush and blended with the parent pipe. Reference marks were placed on each pipe specimen to provide a means of locating all indications.

The counterbore configuration chosen for the specimens represents a conservative (i.e., difficult to inspect) condition. The manufactured counterbores conformed to configurations as reported by Morris et al. (1982). Flush-ground crowns are not necessarily conservative, but represented some Class I welds, and allowed us to thoroughly evaluate single-side access conditions with no inspection constraint imposed by the crown. Some Level III inspectors (see Table 2.4) claimed they had never inspected an unground weld before in the field.

Some specimens had no counterbore; others had the counterbore transition near the weld fusion line. The root zones had drop-through and suck-back. The centrifugally cast stainless steel specimens were predominantly of either equiaxed or columnar microstructure.

Three types of defects were considered for the round-robin test, as shown in Table 2.3.

Although thermal fatigue cracks have been found in some reactor components, this type of defect was included because it is conservative in terms of ultrasonic testing for other types of cracking. In centrifugally cast stainless steel, the only reported cracks are TFC. The reliable detection of an ultrasonically conservative crack

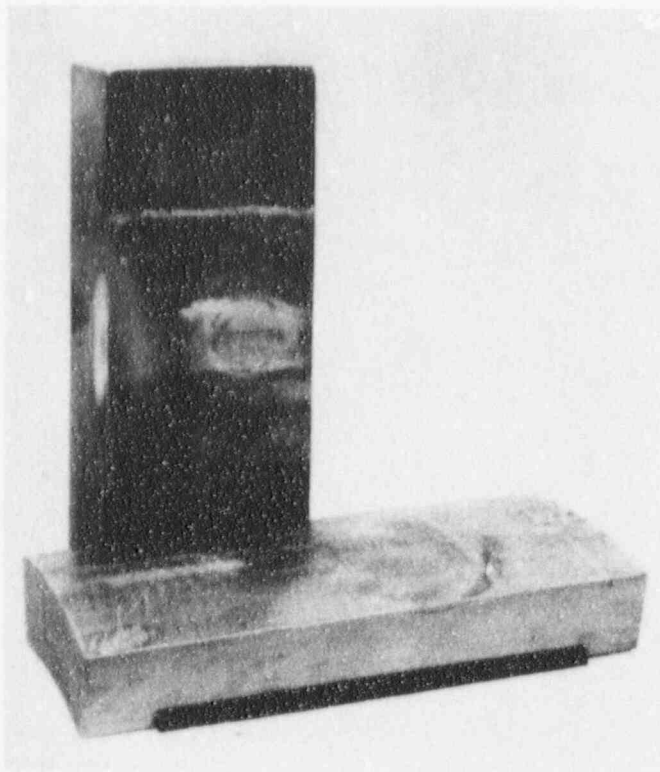


Figure 2.2 Example of PIRR Specimens (these two are clad ferritic weldments)

Table 2.3. Defect Types

Defect Type	Description
Thermal Fatigue	Thermal fatigue crack
IGSCC	Intergranular stress corrosion crack
EDM	Artificial notches cut into the weld area, using electro-discharge machining

gives assurance that other kinds of defects could be detected in this material. Another useful feature of these cracks is that they can be economically produced according to pre-defined specifications. Without this capability, it would have been much more difficult to obtain useful data.

The tightness of TFCs makes them difficult to detect, as compared to most mechanical fatigue cracks. The crack faces are very close together, causing them to touch in spots, and therefore, transmit some of the incident ultrasonic energy, which reduces the ultrasonic energy reflected back to the sensor. The tightness of the TFCs simulate the crack opening under compressive

stresses that may occur during cold shutdown, when inservice inspections are performed. The defect-making process also results in cracks that are relatively short compared to their depth (less than 5:1 aspect ratio in most cases), which are more difficult to detect.

Because IGSC cracking also occurs in large-diameter piping, this type of crack was also included in the PIRR test, but was not examined as extensively as the TFC.

Finally, a type of commonly used artificial defect (EDM notch) was included in the test. Eight EDM notches were machined into one wrought stainless steel specimen. This type of defect was included so that team performance on real and artificial defects could be contrasted. The artificial defects were also included so that a class of "easy" defects would be present in the test. The results with these defects would serve as a minimum criterion for inspection effectiveness; if a team could not find the EDM notches, their poor performance could not be blamed on any peculiarities of the thermal fatigue or IGSC cracks manufactured for the PIRR.

Seven inspection teams participated in the PIRR test; each team came from a different commercial ISI organization. The seven ISI organizations that participated performed the majority of the inspections conducted in the U.S. at the time of the PIRR. Five inspection teams performed all of the inspections, one team inspected everything but the cast stainless steel, and one team inspected only the cast stainless steel. The inspection teams that were sent by the ISI organizations were considered to be average or better than average, but not the most highly qualified or experienced teams available. To participate in the PIRR, ISI team members provided histories of their nuclear reactor inspection experience and UT experience (see Table 2.4). It is evident from these data that the Level II and III participants had extensive experience in both preservice (PSI) and inservice (ISI) inspection.

Two levels were defined for the inspection environment variable. They were "Difficult" (simulated field) and "Laboratory" environments. These two levels were meant to describe the two extremes in a team's working environment. In the "laboratory" environment, the pipe specimen was conveniently located and the inspection team could either stand up or sit down while performing the inspection. The "difficult" environment required

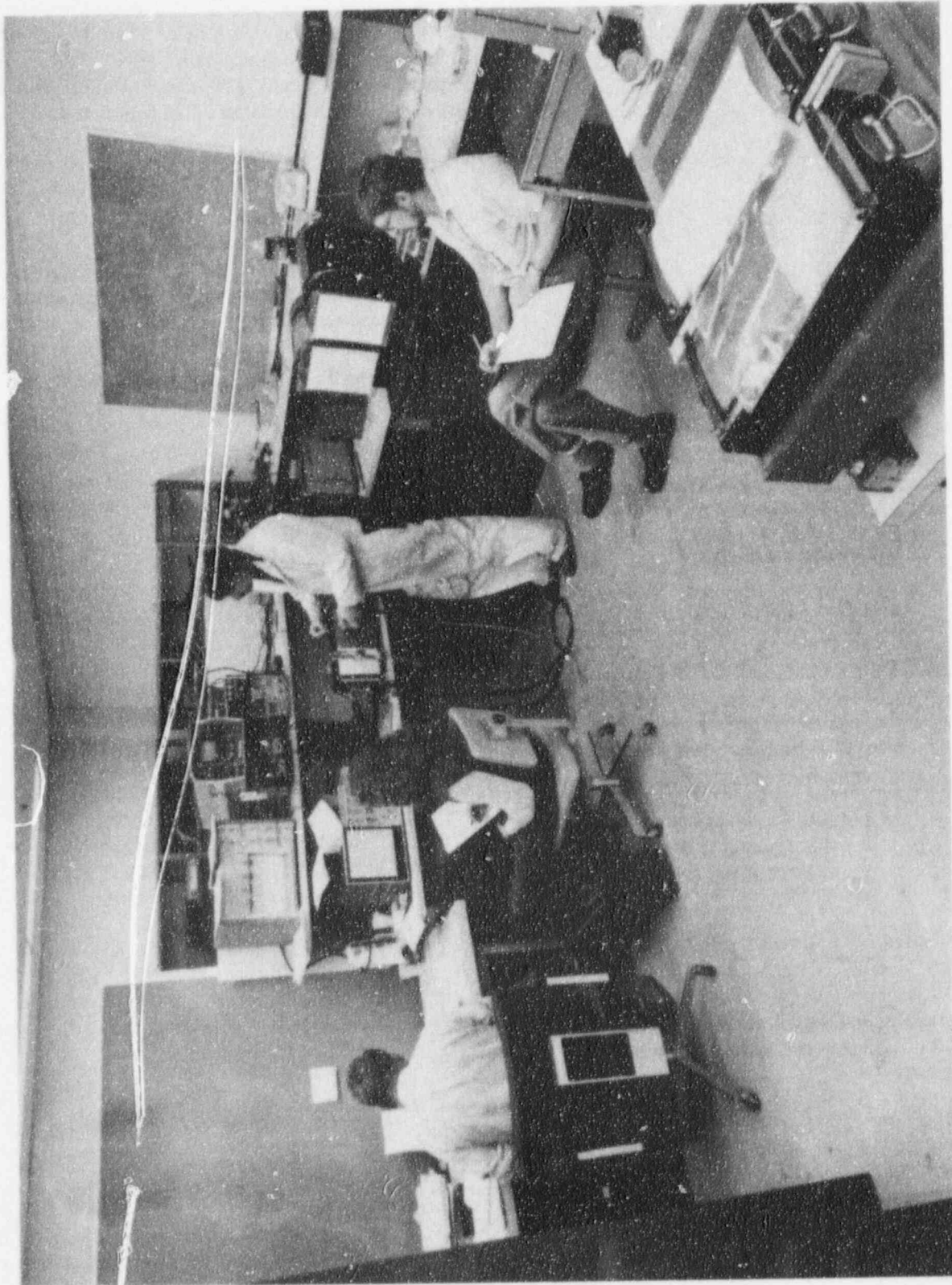


Figure 2.3 Inspection of a 32 in. dia. (81.3 cm) Cast Stainless Steel Specimen During the Round Robin

Table 2.4. Summary of Team Member's Qualifications

Team Member Classification	UT Experience, yrs.		Number of PSIs and ISI	
	Average	Range	Average	Range
ASNT Level III	10.2	4 to 23	28	7 to 62
ASNT Level II	7.4	2.5 to 13	16.7	2 to 57
ASNT Level I	1.1	0.5 to 2.5	2	0 to 5

the inspector to wear radiation clothing (including gloves) and access to the specimen was awkward. Also, the inspector was not easily able to see the test specimen and the UT instrument simultaneously, as shown in Figure 2.4.

In other respects, both environments simulated field conditions whenever possible. The inspectors worked eleven-hour days, six days a week for three weeks. All the testing was performed at PNNL located in Richland, Washington, which was away from home for all teams. The teams had 30 minutes to collect the data for each pipe specimen. Of course, the pipe specimens were randomly presented to the teams, so that sequential data patterns did not exist.

Three types of inspection access were identified, as listed in Table 2.5, and illustrated in Figure 2.5.

Table 2.5. Inspection Access Conditions

Inspection Access	Description
Near	Probe access allowed only on the side of the weld that contained the defect (defect between probe and weld)
Far	Probe access allowed only on the far side of the weld from the defect (weld between probe and defect)
Both	Probe access allowed on both sides of the weld

The pipe specimens were mounted in a jig that prevented access to the inside surface of the pipe. A mask was

placed on each specimen, which also limited access to the outside surface and to one side of the weld. This limited access condition is typical of field inspections, where many of the welds are from pipe-to-component. A typical component would be a valve, pump, or elbow. If a defect is located on the far-side of the weld from the probe, its UT response amplitude could be drastically reduced by adverse sound-field scattering due to the coarse-grained material properties of the weld. Almost all welds were inspected under either near- or far-side access conditions; both-side access was rarely provided.

It should be noted that both-side access conditions are expected to produce results almost identical to near-side access conditions; in both situations, the signal reflected from the defect can be maximized. Therefore, only a few inspections were performed under both-side access conditions. These were used simply to serve as confirmation that there is no great difference between near- and both-side conditions.

It is important to note also that the teams were not aware of which inspections were near-side and which were far-side. All near- and far-side specimens were mounted identically and could not be distinguished from each other by the team. Furthermore, the inspections were randomly presented to the teams, who could not use the inspection sequence to determine which side of the weld the crack might be on. Only the technician mounting the specimen in the jig (not the test observer) had access to the near/far-side information.

Two variants of the ASME Code procedure were employed by the teams in the PIRR test. The first of these was simply called the "Code" procedure, and refers to the procedure the team typically used in the



Figure 2.4 Example of an Inspection Conducted under Difficult (Simulated "Field") Conditions

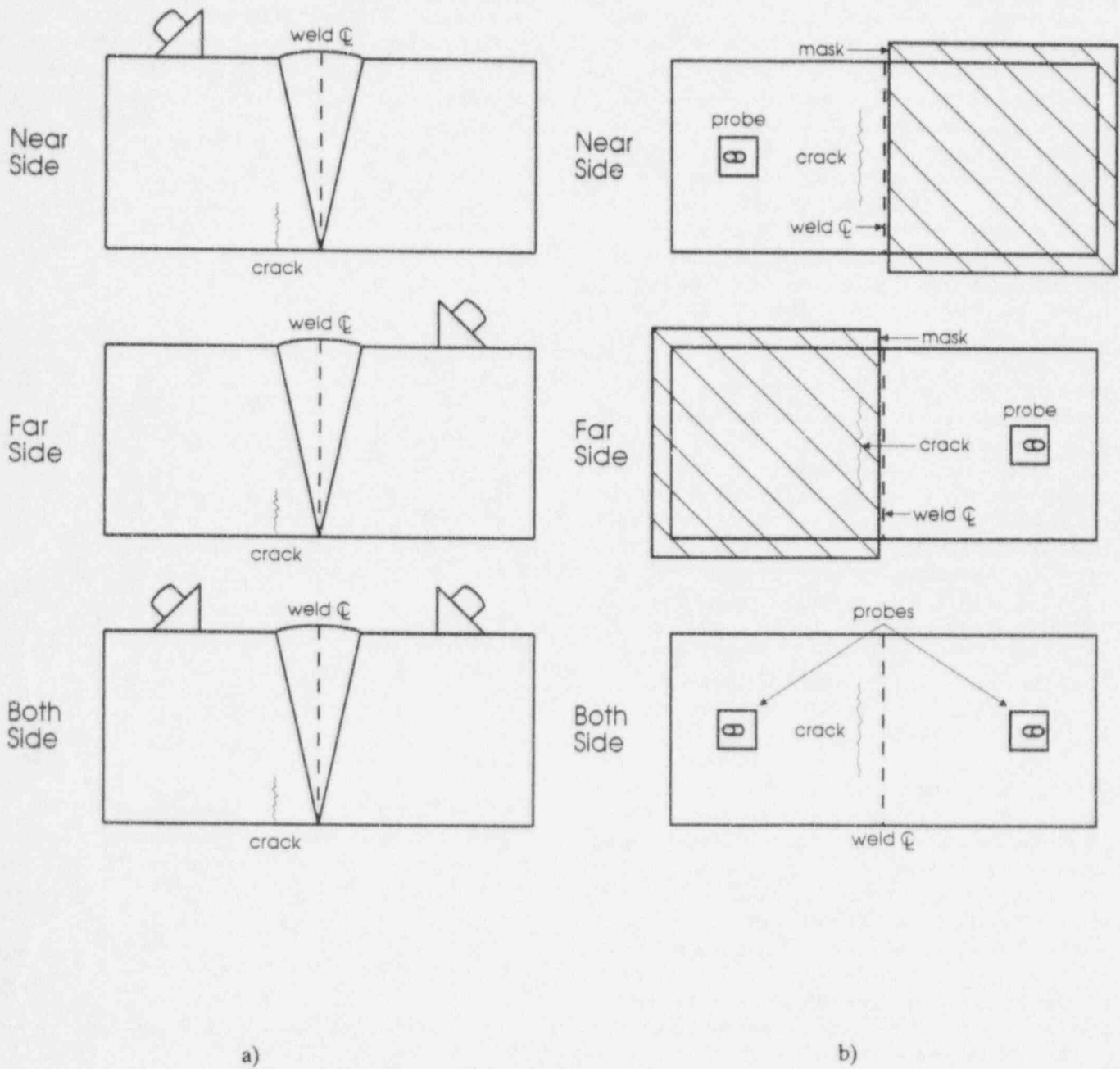


Figure 2.5 Illustrations Showing Near, Far and Both Side Access in a) Cross-sectional View, and b) Plan View.

2.0 Design Variables

field. The Code procedure a team employed conforms to all requirements specified in Section XI of the ASME Code (1977 edition through the 1978 addenda); but it may have been more thorough than the minimum Code requirements.

The second variant of the Code procedure was designated the "Improved" procedure. This improved procedure was developed by the Pacific Northwest National Laboratory staff based on extensive laboratory testing for each combination of flaw type and material. All teams were required to use the identical improved procedure. This procedure also conforms to the ASME Code, but it contains additional requirements on inspection technique, based primarily on Code Case N-335. These requirements were expected to improve the performance of the inspection teams. The "Improved" procedure also allowed us to observe the difference between teams when they were using identical procedures.

In planning the Pipe Inspection Round Robin, it was assumed that "replicate" cracks could be manufactured in the specimens. Replicate cracks are cracks of the same size and shape. These replicate cracks were to be used to quantify crack-to-crack variability. Several of the planned methods of analysis assumed that cracks could be clustered into four groups of replicates. The destructive examination following the PIRR revealed depth and length sizes as shown in Figures 2.6 through 2.8. As one can see from the figures, we did not achieve well-defined crack clusters as originally intended. Only clad material displayed well-defined crack clusters; and even in this material, the planned number of clusters (four) was not achieved.

Because of this problem, the analysis strategy had to be altered. The crack size analysis presented in Section 5 (POD Analysis) treats crack size as a continuous variable. However, two other analyses, the contingency table analysis presented in Section 5 and the ROC analysis in Section 6, still required that cracks be classified into discrete categories by size (which is much easier to do when the cracks are clustered). Figures 2.6 through 2.8 identify the size categories used in these analyses. Size category 1 is meant to signify a "small" crack, while category 4 signifies a "large" crack.

In the case of the cast stainless steel specimens, Figure 2.6 four size categories have been defined on the plot.

Size category 1 represents small cracks [depth < 30% and length < 1.5 in. (3.8 cm)]; category 2 represents long, shallow cracks; category 3 represents deep cracks of moderate length; and category 4 represents the largest and deepest cracks. In the case of clad ferritic, Figure 2.7, defects were divided into four size categories according to depth. However, since depth and length were strongly correlated in this material, the results would have been almost the same if length were used. In the case of wrought stainless steel, Figure 2.8, defects were divided into three categories on the basis of depth.

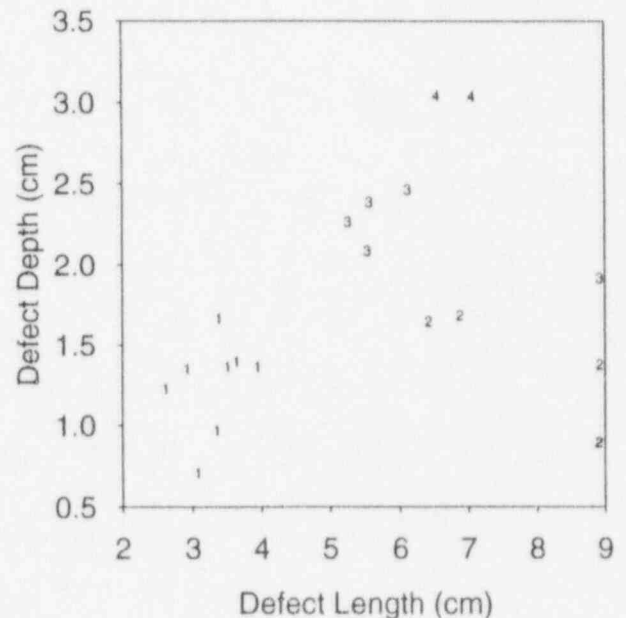


Figure 2.6 Defect Size Categories for Cast Stainless Steel Cracks (size categories are indicated by 1, 2, 3, and 4 for 33 in. diameter (83.8 cm), 2-3/8 in. (6 cm) wall)

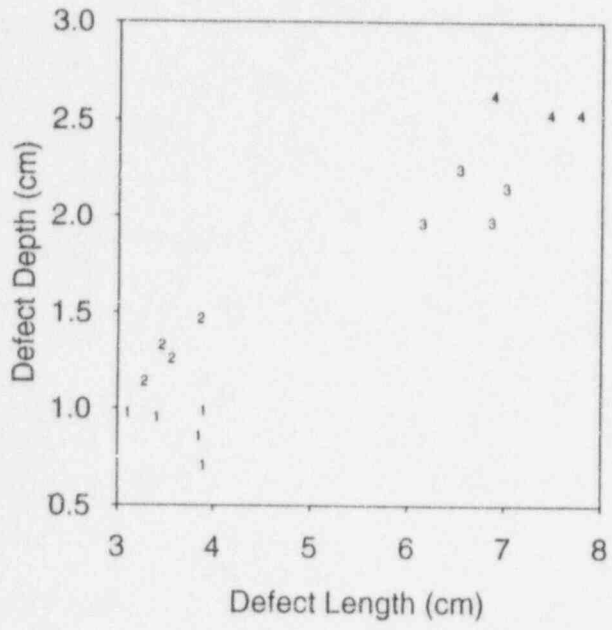


Figure 2.7 Defect Size Categories for Clad Ferritic Cracks (size categories are indicated by 1, 2, 3, and 4 for 32 in. (81.3 cm) diameter, 2-3/8 in. (6 cm) wall)

2.0 Design Variables

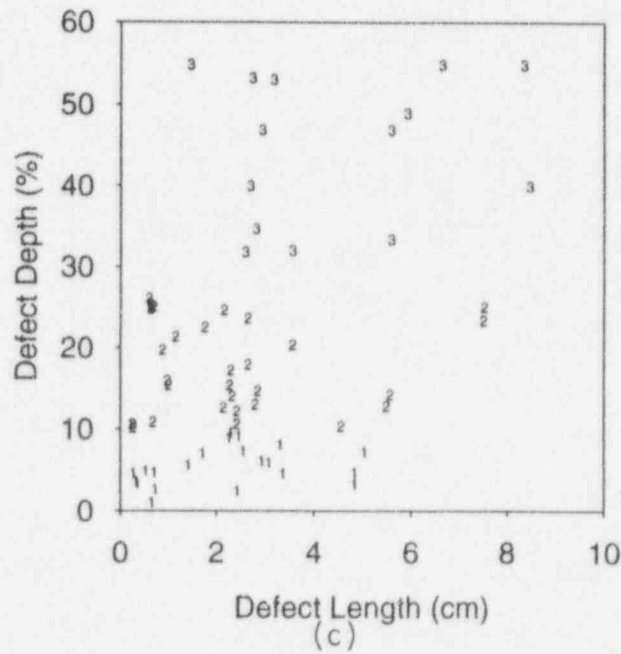
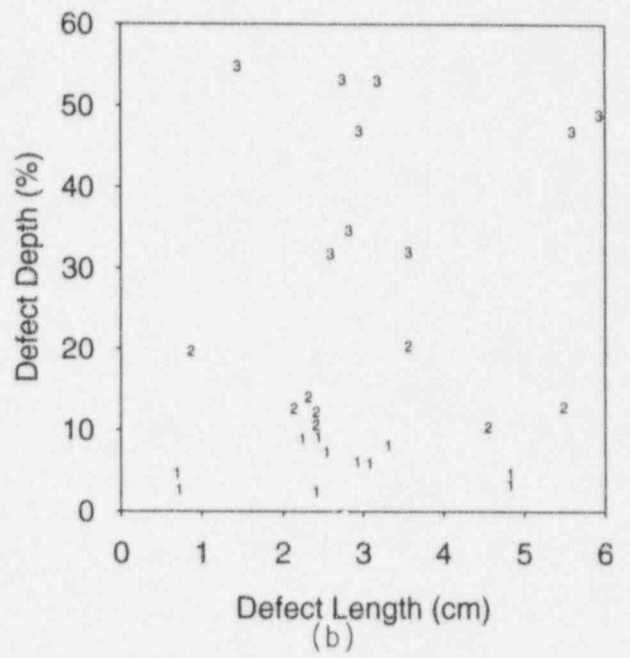
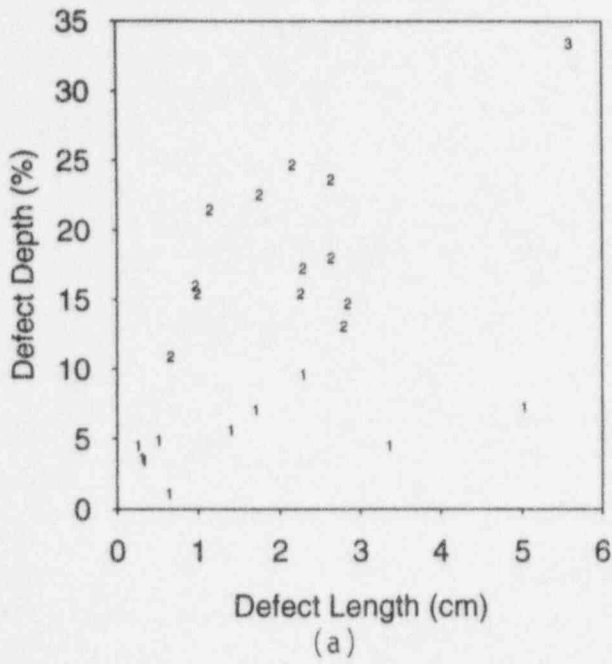


Figure 2.8 Defect Size Categories for Wrought Stainless Steel Cracks for 10 in. (25.4 cm) diameter, either 0.594 in. (1.51 cm) or 0.5 in. (1.27 cm) wall. a) IGSCC, b) TFC, and c) IGSCC and TFC

3.0 Experimental Design, Data Collection, and Reduction

This section describes the PIRR test design and resulting data. The test matrices (i.e., schedule of inspections and specimens) are defined in terms of the seven "design" variables described in the previous section. The test protocol that was followed is also described, as well as the data reduction methods employed. (We call the data reduction "scoring" because it produces detection statistics from the raw data.)

3.1 Test Matrices

The seven design variables described in Section 2.5 define a potential group of 3240 candidate inspection conditions that the PIRR test might examine. That is:

$$3240 = \begin{array}{l} (3 \text{ materials}) \times (3 \text{ defect types}) \times \\ (6 \text{ teams}) \times (3 \text{ access conditions}) \times \\ (2 \text{ environments}) \times (2 \text{ procedures}) \times \\ (5 \text{ defect sizes, including blanks}) \end{array}$$

It is, of course, not economically feasible to determine inspection performance for all 3240 inspection conditions, particularly since accurate POD statistics require from 12 to 25 inspections per condition. The set of candidate inspection conditions was therefore restricted to those conditions that most frequently occur during field inspections. Points in the inspection condition space were also chosen so that rough extrapolations could be made to regions that were not measured.

Table 3.1 summarizes the inspection conditions that were measured in the PIRR. The cells of the test matrix presented in Table 3.1 identify "inspection experiments" of different sizes that were performed under the designated inspection conditions. The different experiment sizes or types are labeled as A, B, C, and D. An individual inspection experiment is designed to produce the information necessary to estimate a single set of POD curves. The general strategy used in the construction of the test matrix was to conduct a large inspection experiment (Type A or B), for important sets of inspection conditions and a smaller experiment (Type C or D) for the less important sets.

The test matrix of Table 3.1 describes the inspections that one team performed in the PIRR. Each of the participating teams performed exactly the same schedule of inspections. At the bottom of Table 3.1 is an accounting of the total number of inspections a single team performed in the PIRR. Most teams were able to

complete the 253 scheduled inspections in the allocated time, so a few extra inspections were added. These included inspections with both-side access and with EDM notches. Because there were six teams participating in the PIRR, more than 1500 inspections were conducted in total.

Experiments of four sizes or types were used in the PIRR test. The schedule of inspections to be performed by a single team on each specimen is shown in Table 3.2 for the four different types of experiment. The largest-sized inspection experiment (type A) required each team to perform 25 inspections, and these inspections were distributed among specimens containing five different sizes of cracks.

In Table 3.2, the numbers in each cell indicate the number of inspections (1 or 2) performed on each specimen. The smaller-sized experiments (B, C, and D) eliminated some of the replicate measurements contained in experiment A and also examined fewer crack size categories. Experiment B used only three replicate specimens per crack size, and the number of size categories was reduced from five to four. An experiment of type C required no "pure" replicate inspections to be performed. Finally, experiment D prescribed inspections for only one crack size. This size of experiment produced only one point on a Probability of Detection curve and is, therefore, suited only for relatively crude comparisons. Basically, experiments of size D were inserted at inspection condition points that were thought to be optimum for ultrasonic inspection. Comparing the crude statistics with the main body of results provided estimates of the best performances to be expected from ultrasonic testing.

Which of the five crack sizes were examined in each experiment depended on the material type, but the "target" crack sizes were, in terms of through-wall percentages: 0% = size 0, 10% = size 1, 20% = size 2, 30% = size 3, and 40% = size 4. (A specimen with a 0% through-wall crack is a blank specimen.) This experimental design, therefore, assumed a capability to produce cracks of predetermined depth. As illustrated in Section 2.5 (see Figures 2.6 through 2.8), this assumption was in error; it was not possible to control crack sizes with sufficient accuracy.

Consequently, the crack size categories listed above can only be considered to be the "target" sizes used in the

3.0 Experimental Design

Table 3.1. Round Robin Test Matrix for a Single Team

Inspection Environment and Access					
UT Procedure	Material/ Crack Type	Laboratory		Difficult	
		Near	Far	Near	Far
Code	Wrought SS TFC	B	B	A	C
	Wrought SS IGSCC	D	//////////////// ////////////////	A	C
	CCSS TFC	D	//////////////// ////////////////	A	C
	Clad Ferritic TFC	D	//////////////// ////////////////	A	C
Improved	Wrought SS TFC	//////////////// ////////////////	//////////////// ////////////////	B	//////////////// ////////////////
	Wrought SS IGSCC	//////////////// ////////////////	//////////////// ////////////////	B	//////////////// ////////////////
	CCSS TFC	//////////////// ////////////////	//////////////// ////////////////	B	//////////////// ////////////////
	Clad Ferritic TFC	//////////////// ////////////////	//////////////// ////////////////	B	//////////////// ////////////////

Inspection Experiment Type	Number of Inspections Per Team	Total Number of Experiments in Matrix	Total Inspections Per Team
A	25	4	100
B	16	6	96
C	12	4	48
D	3	3	9
Total			253

Wrought SS = wrought stainless steel
 TFC = thermal fatigue crack
 CCSS = centrifugally cast stainless steel
 IGSCC = intergranular stress corrosion crack

production of cracked specimens. After the test was completed and the results of the destructive analysis were available, the cracks were re-assigned into the size categories that are described in Section 2.5.

It should be emphasized that the design matrices presented in Tables 3.1 and 3.2 describe the logical organization of the PIRR (i.e., the way the results were organized after the test was done). They DO NOT, however, describe the order in which the inspections were performed. The inspections described in Table 3.1

Table 3.2. Structure of Experiments

Experiment Type	Specimen Description																			
	Blank				Crack Size															
	0				1			2			3			4						
Replicate #		Replicate #		Replicate #		Replicate #		Replicate #		Replicate #		Replicate #		Replicate #						
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4					
A	2	1	1	1	2	1	1	1	2	1	1	1	2	1	1	1	2	1	1	1
B	2	1	1		2	1	1						2	1	1		2	1	1	
C	1	1	1		1	1	1						1	1	1		1	1	1	
D									1	1	1									

were randomly ordered in blocks of eight inspections (in order to avoid an excessive amount of recalibration, eight inspections in a block had to use the same material type). Therefore, each team was presented with a large array of different materials and environments every day. Once the experiment was underway, it was quite obvious that randomization introduced a very important dimension to the PIRR test. The random alternation between inspection conditions is one very important source of realism in this study that distinguishes it from many previous studies.

3.2 Test Protocol

Two people were responsible for conducting the PIRR: an observer who continuously monitored the inspections, and a technician who mounted the specimens in their holding jigs. The technician had a randomized inspection list that contained all the inspections the team was to perform and the order in which they were to be performed. Each inspection in the list was indexed by the INSPECTION #, which was used to uniquely identify all inspections in the experiment. It was also used to index the inspection data in the computer data base and in the raw data files. A typical portion of the inspection list is illustrated in Table 3.3.

The technician made certain that the inspection condition variables were set properly for each inspection. The critical variables in this respect are ENVIRONMENT, SPECIMEN-CODE, and WELD-SIDE¹. To reduce errors, a technician other than the one who prepared the specimens was required to unload the

specimens and to check that the proper test conditions had been provided to the inspection team. The observer was responsible for the proper completion of all the data forms and had to assemble these forms into an INSPECTION FOLDER and file them in the RAW DATA FILE.

An inspection generates a great deal of information, but all the information in the inspection folder was not processed. The important inspection information, which was input into the computer for analysis, resided on the INSPECTION REPORT FORM. An example of a completed inspection report form is illustrated in Figure 3.1.

3.3 Scoring the Inspection Results

It is never a straightforward task to associate inspection results with the true state of the material. We call this process "scoring" the inspection results. The basic outputs for our scoring procedure are detection statistics and indicated crack dimensions.

Fundamental to our scoring procedure is the concept of a "grading unit," which allows straightforward calculation of detection performance. Grading units are

¹ All welds have two sides and these were permanently marked in an arbitrary way as either A or B on the end of each specimen in a manner so that the inspector could not see the marking.

3.0 Experimental Design

Table 3.3. Example of Randomized Inspection List

Inspection No.	Reference No.	Material	Environment	Access	Weld Size	Crack Size	Specimen Code	Specimen Repetition	Inspection Repetition
.
456	5	SS	DIFF	NEAR	B	1	B211-2	1	1
457	32	SS	DIFF	NEAR	B	4	D101-4	1	2
458	14	SS	DIFF	NEAR	A	4	B213-3	1	2
.
.

Reference No. = Inspection #
 SS = Stainless Steel
 DIFF = Difficult

fixed lengths of weld material that are to be classified by inspection. Grading units are not marked on the specimen, and the inspector is not aware of the number of grading units on a specimen, their location, or size during the inspection. During the inspection, the inspector simply marks down indications on a map of the weld and categorizes them as N (not a crack) or C (crack).

During the scoring procedure, the "grading units" are identified on the weld map, as illustrated in Figure 3.2. The results for each grading unit are summarized by examining all the indications that lie within the grading unit, and categorize them as follows:

- M: (Not Classified) No (N or C) indications were in grading unit
- N: (No Crack) At least one N-indication was present in grading unit but no C-indication
- C: (Crack) At least one C-indication was present in grading unit

These detection categories were then used to calculate POD and FCP statistics. For example, to calculate the POD that team A achieved when using decision

threshold C, one would form the quotient:

$$POD = \frac{\text{Number of Cracked Grading Units Classified as C by Team A}}{\text{Total Number of Cracked Grading Units Inspected by Team A}}$$

and FCP for the team would be defined by:

$$FCP = \frac{\text{Number of Blank Grading Units Classified as C by Team A}}{\text{Total Number of Blank Grading Units Inspected by Team A}}$$

This same scheme was used to define recording probability and false recording probability:

$$RP = \frac{\text{Number of Cracked Grading Units Classified as C or N}}{\text{Total Number of Cracked Grading Units Inspected}}$$

$$FRP = \frac{\text{Number of Blank Grading Units Classified as C or N}}{\text{Total Number of Blank Grading Units Inspected}}$$

BATTELLE
INSPECTION REPORT FORM

To be filled out by Observer

1. INSPECTION <u>544</u>	2. TEAM <u>#6</u>
3. ENVIRONMENT: (Lab, <u>Diff</u>)	4. WELD-SIDE: (A, <u>B</u>)
5. INSP. PROCEDURE: (<u>Code</u> , Imp)	6. SPECIMEN CODE: <u>D101-3</u>
7. INSPECTION REPL. <u>/</u>	8. OBSERVER _____

	HOUR*		DATE			11. CALIBRATION
	BEGIN	END	DAY	MONTH	YEAR	
9. INSPEC.	<u>1757</u>	<u>1935</u>	<u>13</u>	<u>7</u>	<u>81</u>	<u>190001</u>
10. ANALYSIS	<u>1230</u>	<u>1337</u>	<u>14</u>	<u>7</u>	<u>81</u>	

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

INDICATIONS

	dB RESP	L ₀ (in)	L ₁ (in)	L ₂ (in)	W ₀ (in) - Far + Near	Depth %T-Wall	DISPOSITION C(rack) N(o crack)
16.	<u>-11</u>	<u>8 1/2</u>	<u>8 1/4</u>	<u>8 5/8</u>	<u>0</u>	<u>-</u>	<u>N</u>
17.	<u>-7</u>	<u>7 1/2</u>	<u>7 3/8</u>	<u>8</u>	<u>0</u>	<u>-</u>	<u>N</u>
18.	<u>-7</u>	<u>7 3/16</u>	<u>7</u>	<u>7 3/8</u>	<u>0</u>	<u>-</u>	<u>N</u>
19.	<u>-2</u>	<u>3</u>	<u>2 5/8</u>	<u>3 3/8</u>	<u>3/8</u>	<u>10</u>	<u>C</u>
20.	<u>8</u>	<u>3 1/16</u>	<u>3</u>	<u>3 7/8</u>	<u>3/8</u>	<u>10</u>	<u>C</u>
21.							
22.							
23.							
24.							
25.							

60°

12. _____ 13. _____ 14. _____
Level I Level II Level III

Figure 3.1 Data from a Completed Inspection Report Form

3.0 Experimental Design

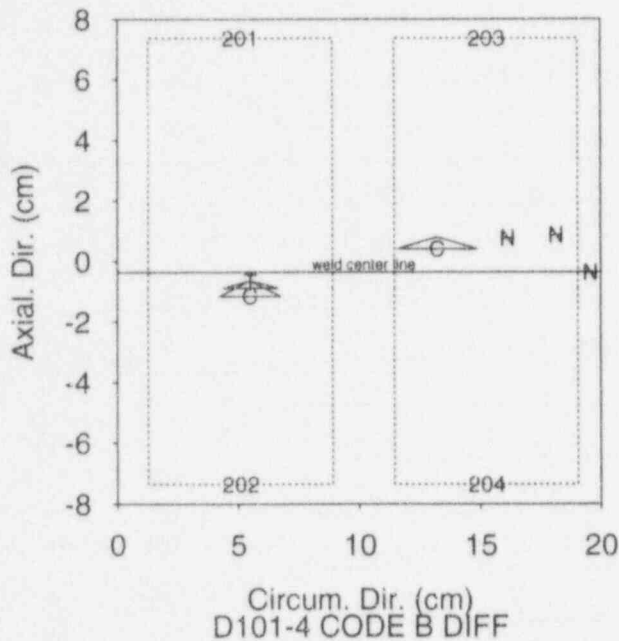


Figure 3.2 Scoring Example Showing the Location of Cracks (T), Crack Calls (C), Non-crack Calls (N), and Grading Units (201, 202, 203, and 204) in Specimen D101-4 With Inspection From the B-side and A-side Masked Off.

The grading units employed for the bulk of the analysis are like those displayed in Figure 3.2. The length of these grading units is 3 in. (7.6 cm), which is long enough to accommodate all but the longest cracks in the PIRR. It is generally bad practice to use grading units that cannot accommodate whole cracks. The trade-off is that many grading units are needed to assess FCP and the larger the grading unit, the more pipe that must be inspected; but because our specimen set was already fixed, it was more important to have larger grading units to accommodate the larger cracks, even though fewer blank grading units would be available to assess FCP. The grading unit size chosen is still small enough to allow us to fit more than one unit on a specimen, with some room to spare between them.

It is wise not to attempt to crowd too many grading units on a single specimen. Small grading units, spaced closely together, produce detection statistics with a complex correlated structure that can cause great problems with the standard statistical analysis procedures (all computed confidence bounds and error estimates will be too optimistic). In the PIRR analysis, we at-

tempted to separate grading units with at least 1 in. (2.54 cm) of defect-free material. Frequently we employed only one grading unit per specimen weld side, as contrasted with the four grading units shown in Figure 3.2.

In order to measure near/far-side effects, separate grading units were placed on each side of the weld centerline, as illustrated in Figure 3.2. Because the inspection illustrated in Figure 3.2 was from the "A" side of the weld (i.e., the probe was located on the A side), grading units 201 and 203 were inspected from the near-side, while grading units 202 and 204 were inspected from the far-side.

A problem with the grading-unit approach was discovered with respect to clad ferritic material, and this required us to make an important change in the experimental design and analysis. In the clad ferritic material, many cracks were placed almost on the weld centerline, so that a small axial location error could easily place indications outside of the correct grading unit. Because of this problem, it was not possible to determine near/far-side effects in this material, and grading units on adjacent sides of the welds were "collapsed" (combined) together to form a single unit.

Table 3.4 illustrates the consequences of this problem. In size category 3 where several cracks were close to the weld centerline, the POD increases from 63% to 97% when grading units were extended across the weld centerline and the access variable was dropped. This is a dramatic change in performance that cannot be ignored, so "collapsed" grading units were employed in all analyses of clad ferritic material. It was fortunate that this problem occurred only in the clad material. Because the material is easy to inspect, it is the one material for which the access condition is relatively unimportant and can be ignored.

POD analysis with collapsed grading units was also conducted on the other materials in the PIRR (CCSS and wrought stainless steel) to determine if the same problem existed in those materials. No such effect was observed, and therefore separate near/far-side grading units were retained. When reviewing the results in this report, the reader must remember that the near- and far-side inspections have been combined for clad ferritic material.

Table 3.4. Comparison of POD for Near and Collapsed Grading Units in Clad Ferritic Material

Crack Size	Average Depth	Near-Side Grading Units	Collapsed (Near + Far) Grading Units
1	15%	60%	88%
2	22%	64%	87%
3	35%	63%	97%
4	43%	75%	97%

In the majority of cases, the grading units contained only one of the following:

1. no defects (they are blank);
2. one crack, completely contained within the grading unit; or
3. one large crack, surrounded by a cluster of smaller cracks.

The cases in which more complicated crack geometries were present were few. Consequently, it is fairly easy to summarize the "true state" of a grading unit. The crack type, depth, and length associated with the grading unit is that of the largest crack in the unit.

3.4 Destructive Analysis Results

Destructive analysis of the test specimens was carried out in a number of phases. The cracks were either cross sectioned in sequential steps, broken apart, or bent in a fixture and sized with an optical microscope looking down the opened crack (this last procedure was validated by cross sectioning some cracks to verify their depths).

The UT inspectability of the clad ferritic specimens was very good, and they therefore provided excellent crack-tip signals to estimate the depths of the thermal fatigue cracks that they contained. There were 16 cracks and four blanks used in the clad ferritic portion of the study. Three of the larger cracks were selected for destruction, because they were the ones that would have the greatest impact on detection statistics if there were errors in the estimated depth of the crack tip. Table

3.5 compares the destructive test results with the depths estimated by the crack-tip diffraction method. Based on the very good agreement shown by these results, it was decided that the cracks were characterized sufficiently well by the nondestructive measurements, and that no further destructive tests were needed.

Table 3.5. Results of Clad Ferritic Specimen Destructive Analysis

Specimen Number	Destructively Measured Depth	Crack-Tip Depth
B605	0.996" (2.53 cm)	1.005" (2.55 cm)
B609	0.769" (1.95 cm)	0.778" (1.98 cm)
B617	0.446" (1.13 cm)	0.380" (0.97 cm)

On the other hand, preliminary UT measurements on the wrought stainless steel specimens did not provide accurate depths for the cracks, so all the depths had to be determined by destructive analysis. Documentation of the destructive testing of the IGSCC can be found in Heasler et al. 1990. Furthermore, the destructive analysis showed that all thermal fatigue cracks grown near inside surface discontinuities (i.e., counterbore transitions or weld fusion lines within the crack-growing nozzle) were surrounded by "bonus" cracks. These supplemental cracks were generally not significant, being between 2% and 5% of wall thickness in depth. Because all the bonus cracks fit into the grading unit surrounding the major crack, this caused no difficulty in the analysis.

Bonus cracks were also found for some of the IGSC cracks in the wrought stainless steel specimens. These bonus cracks were also very shallow and short, and should not have any effect on the results, because the grading units were adjusted to accommodate them.

In the case of the centrifugally cast stainless steel specimens, the preliminary UT measurements were also found to be in error. However, the complete set of specimens were not destructively analyzed, because it was felt that they would be useful for other studies. Furthermore, it was obvious that inspection was inadequate, so a knowledge of the true depths would not add

3.0 Experimental Design

much to the study. The destructive test results for cast stainless steel are presented in Table 3.6. The cracks tended to be smaller than the intended depth based on growth rate calculations. One of the blank specimens was also destructively tested, and no defects were found.

Table 3.6. Destructive Results for Centrifugally Cast Stainless Steel Specimens

Specimen	UT-Estimated Depth	Destructive Depth
B523	30%	15.6%
B521	30%	28%

4.0 Variables that Influence Detection Performance

The round-robin experiment was designed to determine the effect of important inspection and material variables on detection performance. In this section, we calculate the statistical significance of the design variables on detection performance, using contingency tables [see Bishop (1976), for an introduction to contingency tables]. Based on this analysis, the variables are ranked from the most important to the least. Two of the variables were found to have very little effect on detection performance and are consequently dropped from analyses in the following sections of this report.

The design variables evaluated in this section are: the Inspection Variables

1. Procedure: (Code versus Improved)
2. Environment: (Laboratory versus Difficult) conditions
3. Access: (Near versus Far) side access to material

and the Material Variables

4. Material/Crack Type: (Cast SS/TFC, Clad Ferritic/TFC, Wrought SS/TFC, Wrought SS/IGSCC, Wrought SS/EDM)
5. Crack Size: (Blank, Size 1, Size 2, Size 3, Size 4)

One important design variable was intentionally omitted from consideration in this section--the team variable. The effect of the team variable is considered in Section 8. Omission of the team variable should not cause significant problems with the analyses in this section, because all teams in the round robin performed the same schedule of inspections, and are therefore representative samples of all teams performing inspections. Nevertheless, it must be remembered that the conclusions in this section deal with average team detection performance, and a particular team could (and probably did) behave somewhat differently.

Tables 4.1 through 4.3 present summaries of detection data obtained for each material in the piping inspection round robin. These contingency tables present the basic data used to determine whether or not the listed variables significantly affect detection. In these tables, POD and RP detection statistics are presented, as well as the number of inspections performed. For example, Table 4.1 indicates that procedures meeting ASME Code requirements in operation under field conditions

yielded a POD of 2% for inspections of size 1 cracks from the near-side direction.

These three tables display the effects that the design variables have on detection performance (i.e., POD and RP), although some of the patterns present in the tables may be due to random variation. For example, if the detection statistics in the first three columns displayed little variation from one column to the next, this would lead one to postulate that the procedure and environment variables do not affect POD or RP, at least for near-side inspections. Postulates such as these can be more formally evaluated using a chi-squared test. In this section, chi-squared tests are used to determine the significance of any identified patterns in the tables. The resulting chi-squared tests and tables offer the reader one of the most direct overviews of the major round-robin results.

4.1 Effect of Inspection Variables

Application of the chi-squared test to the statistics in Tables 4.1 through 4.3 confirmed that procedure and environment have no significant effect on detection performance. This implies that the POD statistics listed under the headings "Code/Difficult," "Code/Lab," and "Improved/Difficult" in Tables 4.1 to 4.3 are essentially the same, except for experimental error. Table 4.4 presents the relevant chi-squared statistics.

Table 4.4 resembles the earlier tables, except that it is collapsed over the variables to be tested. Each cell in the table contains a chi-squared statistic and its associated degrees of freedom. The values in the cell show the effect these variables have on detection performance.

A chi-squared statistic is formally compared with its degrees of freedom through the use of a chi-squared table (Bishop 1976). This table contains critical values that determine when a chi-squared statistic is significant. For example, a chi-squared statistic with 4 degrees of freedom is significant at the 1% level if it is larger than the critical value of 13.3. Because the critical value is always greater than the associated degrees of freedom, a quick "rule of thumb" test of significance can be performed by comparing the chi-squared statistic with the number of degrees of freedom.

4.0 Variables that Influence Detection Performance

Table 4.1. Summary of Detection Performance for Cast Stainless Steel

Defect Size	Near Side Access			Far Side Access		
	Code/ Difficult	Code/ Lab	Improved/ Difficult	Code/ Difficult	Code/ Lab	Improved/ Difficult
Blank						
FCP	0.06	0.0	0.05	0.14	0.0	0.20
FRP	0.11	0.0	0.10	0.26	0.0	0.20
# Insp.	35	16	20	35	16	20
Size 1						
POD	0.02	0.0	0.05	0.17	0.0	0.0
RP	0.12	0.0	0.10	0.25	0.0	0.0
# Insp.	41	10	39	12	3	1
Size 2						
POD	0.25	0.18	--	0.22	0.60	--
RP	0.38	0.35	--	0.22	0.60	--
# Insp.	8	17	0	9	5	0
Size 3						
POD	0.19	0.36	0.20	0.0	0.0	--
RP	0.24	0.36	0.20	0.0	0.0	--
# Insp.	21	11	5	9	4	0
Size 4						
POD	0.15	0.0	0.07	0.25	0.0	--
RP	0.31	0.0	0.07	0.38	0.0	--
# Insp.	13	3	15	8	2	0

Insp. = number of inspections

A chi-squared value that is small relative to its degrees of freedom indicates that the variables for which chi-squared values were computed do not affect detection performance, under the identified cell conditions. For example, the chi-squared value (2.3) in the cell identified by "Size 1" cracks, cast stainless steel material, and near-side access is small with respect to its degrees of freedom (4), which implies that procedure and inspection environment do not influence detection performance under these conditions. The conclusion that "2.3" is small results from a comparison with the critical value of 13.3.

In the right-hand column of Table 4.4, the chi-squared statistics and degrees of freedom are added up for each

material. These sums provide an overall test of significance for each material. The overall chi-squared statistics for both cast SS and wrought SS are small (relative to the criteria values in Bishop's table), indicating that the procedure and environment variables are not important in these materials. The overall statistic for clad ferritic is significant at the 10% level, but not at 1%. An examination of Table 4.4 reveals that the chief contribution to this large statistic is related to the detection of size 2 cracks. An examination of Table 4.2 shows that the improved procedure might be better than the field procedures meeting ASME Code requirements at detecting these medium-sized cracks, and this has caused the large chi-squared statistic. In other words, there is weak evidence that the improved proce-

Table 4.2. Summary of Detection Performance for Clad Ferritic Material (using both-side grading units)

Defect Size	Code/ Difficult	Code/ Lab	Improved/ Difficult
Blank			
FCP	0.02	--	0.0
FRP	0.10	--	0.0
# Insp.	48	0	24
Size 1			
POD	0.84	--	0.96
RP	0.89	--	0.96
# Insp.	55	0	25
Size 2			
POD	0.79	0.89	1.00
RP	0.85	1.00	1.00
# Insp.	48	18	24
Size 3			
POD	0.97	--	--
RP	1.00	--	--
# Insp.	31	0	0
Size 4			
POD	0.95	--	1.00
RP	0.98	--	1.00
# Insp.	42	0	24

Insp. = number of inspections

procedure does a better job at detecting medium-sized cracks in clad ferritic material.

The largest chi-squared statistics (those significant at the 1% level) have been marked with an asterisk in the table. One of these statistics is for the clad material and has already been discussed. The other two large statistics both occur for blank material under far-side access conditions. Examinations of Tables 4.1 and 4.3 do not reveal a simple answer for these two large values. In the case of cast stainless steel material (Table 4.1), there seems to be a difference between difficult and laboratory conditions. The laboratory inspections make no false calls, but the difficult inspections make 14% to 26%. In the case of the wrought stainless steel material (Table 4.3), both procedure (many team proce-

dures used a fixed DAC level for crack detection and the improved procedure required evaluation of any signal regardless of amplitude if the signal has "crack-like properties") and environment may have an effect.

Even though a few statistics in Table 4.4 are large, the overall message is that the procedure and environment variables have surprisingly little effect on POD and RP. Consequently, in the rest of this report, we will eliminate these two variables from the analysis and work with the collapsed data set illustrated in Table 4.5. This provides more observations per cell for the ensuing analyses, and reduces experimental error.

The chi-squared statistics derived from Table 4.5, which measure the effect of the variable "access" on detection performance, are presented in Table 4.6. The major conclusions to be drawn from these chi-squared statistics are fairly straightforward. There is a big difference between near- and far-side access for all types of defects present in wrought stainless steel. In cast stainless steel material, the effect of access is much weaker and not significant. This last result is not unexpected since the structure of the weld and cast SS material are both coarse-grained, anisotropic materials that are difficult to inspect with ultrasonics. Furthermore, the POD values for cast SS are small for all conditions making the assessment of inspection variables on performance very difficult. The overall chi-squared statistic for cast SS is only significant at the 15% level--an indication of weak influence. Notice that no one chi-squared statistic in the cast SS column seems to be particularly large.

4.2 Effect of Material Variables

Table 4.7 presents the chi-squared statistics that measure the effect of the variable defect size. These chi-squared statistics are also computed from Table 4.5. It is important to note that Table 4.7 also answers a question that is of central importance to inspection effectiveness: Can inspection discriminate between blank material and defective material? Defect size is relevant to this question because blanks are one category of defect size (blanks contain a defect of size 0).

The chi-squared statistics associated with each material are highly significant, indicating that defects can be found in each material and that defect size is important. Defect size is clearly most important in clad ferritic,

4.0 Variables that Influence Detection Performance

Table 4.3. Summary of Detection Performance in Wrought Stainless Steel Material

Defect Size/Type	Near Side Access			Far Side Access		
	Code/Difficult	Code/Lab	Improved/Difficult	Code/Difficult	Code/Lab	Improved/Difficult
Blank						
FCP	0.03	0.02	0.05	0.09	0.17	0.16
FRP	0.19	0.15	0.11	0.43	0.50	0.34
# Insp.	160	94	74	160	94	74
Size 1/TFC						
POD	0.16	0.20	0.11	0.15	0.16	0.29
RP	0.27	0.30	0.32	0.51	0.47	0.55
# Insp.	49	50	19	72	49	31
Size 2/TFC						
POD	0.17	0.27	0.40	0.07	0.17	0.0
RP	0.22	0.39	0.48	0.33	0.43	0.0
# Insp.	36	33	25	43	30	5
Size 3/TFC						
POD	0.55	0.61	0.69	0.21	0.23	0.0
RP	0.68	0.79	0.83	0.58	0.61	0.0
# Insp.	66	33	29	24	31	1
Size 1/IGSCC						
POD	0.32	0.33	0.56	0.08	0.08	0.06
RP	0.52	0.50	0.62	0.40	0.25	0.28
# Insp.	95	12	34	86	12	47
Size 2/IGSCC						
POD	0.68	0.50	0.72	0.16	--	0.15
RP	0.79	0.67	0.77	0.43	--	0.38
# Insp.	84	12	47	67	0	13
Size 2/EDM						
POD	0.51	0.57	0.53	0.21	0.18	0.30
RP	0.63	0.63	0.59	0.52	0.44	0.45
# Insp.	49	54	17	42	45	20
Size 3/EDM						
POD	--	0.88	--	--	0.38	--
RP	--	0.88	--	--	0.63	--
# Insp.	0	24	0	0	16	0

Insp. = number of inspections

Table 4.4. Significance of Procedure and Environment to POD

Material/ Access		Blank	Size 1	Size 2	Size 3	Size 4	Total
Cast SS	Chi DOF						22.9 32
Near	Chi DOF	1.9 4	2.3 4	0.2 2	1.9 4	4.2 4	
Far	Chi DOF	8.0* 4	1.2 4	2.0 2	0.0 2	1.1 2	
Clad Ferritic	Chi DOF						15.8 10
Both	Chi DOF	2.7 2	2.6 2	9.3* 4	-- --	1.2 2	
Wrought SS for all three crack types	Chi DOF						55.1 52
Wrought SS/TFC							
Near	Chi DOF	6.0 4	2.4 4	5.6 4	3.1 4	-- --	
Far	Chi DOF	9.9* 4	3.4 4	4.8 4	1.5 4	-- --	
Wrought SS/IGSCC							
Near	Chi DOF	-- --	7.7 4	3.6 4	-- --	-- --	
Far	Chi DOF	-- --	2.6 4	0.1 2	-- --	-- --	
Wrought SS/EDM							
Near	Chi DOF	-- --	-- --	1.8 4	-- --	-- --	
Far	Chi DOF	-- --	-- --	2.6 4	-- --	-- --	

* indicates a statistic significant at 1% level (see Table E.1 for explanation)

Chi = chi-squared statistic

DOF = degrees of freedom

with wrought SS next, followed by cast SS. From Table 4.7, one can also see that the access condition makes a big difference in the effect of defect size on POD.

With far-side access, defect size is not significant (at the 1% level) in cast SS or wrought SS. With near-side access, size is significant for all three defect types listed

4.0 Variables that Influence Detection Performance

Table 4.5. Detections Collapsed Over Procedure and Environment Variables

Access		Blank	Size 1	Size 2	Size 3	Size 4
Cast Stainless Steel						
Near	POD	0.04	0.03	0.20	0.24	0.10
	RP	0.08	0.10	0.36	0.27	0.16
	# Insp.	71	90	25	37	31
Far	POD	0.13	0.13	0.36	0.0	0.20
	RP	0.18	0.19	0.36	0.0	0.30
	# Insp.	71	16	14	13	10
Clad Ferritic						
Both	POD	0.01	0.88	0.87	0.97	0.97
	RP	0.07	0.91	0.92	1.00	0.98
	# Insp.	72	80	90	31	66
Wrought SS/TFC						
Near	POD	0.03	0.17	0.27	0.59	NA
	RP	0.16	0.29	0.35	0.74	NA
	# Insp.	328	118	94	128	0
Far	POD	0.13	0.18	0.10	0.21	NA
	RP	0.43	0.51	0.35	0.59	NA
	# Insp.	328	152	78	56	0
Wrought SS/IGSCC						
Near	POD	NA	0.38	0.68	NA	NA
	RP	NA	0.54	0.77	NA	NA
	# Insp.	0	141	143	0	0
Far	POD	NA	0.08	0.16	NA	NA
	RP	NA	0.34	0.43	NA	NA
	# Insp.	0	145	80	0	0
Wrought SS/EDM						
Near	POD	NA	NA	0.54	0.88	NA
	RP	NA	NA	0.63	0.88	NA
	# Insp.	0	0	120	24	0
Far	POD	NA	NA	0.21	0.38	NA
	RP	NA	NA	0.48	0.63	NA
	# Insp.	0	0	107	16	0

Insp. = number of inspections

under wrought SS, for clad ferritic, and for cast SS.

Tables 4.8 and 4.9 present the significance of the variables material and crack type, respectively, to detection

Table 4.6. Summary of the Effect of Access

Size		Cast Stainless Steel	Clad Ferritic	Wrought Stainless Steel		
				TFC	IGSCC	EDM
Blank	Chi	3.5	--	58.4*	--	--
	DOF	2	--	2	--	--
Size 1	Chi	2.5	--	17.3*	37.3*	--
	DOF	2	--	2	2	--
Size 2	Chi	3.1	--	12.8*	54.7*	29.1*
	DOF	2	--	2	2	2
Size 3	Chi	4.4	--	23.9*	--	12.2*
	DOF	2	--	2	--	2
Size 4	Chi	1.0	--	--	--	--
	DOF	2	--	--	--	--
Total	Chi	14.5	--	112.4*	92.0*	41.3*
	DOF	10	--	8	4	4

* indicates statistic is significant at 1% level

performance. Table 4.8 demonstrates that detection in the three types of material is inherently different; i.e., the material is highly significant to detection performance. Even the inspection of blank material (see the first two rows in Table 4.8) yields significant statistics; the false call probabilities for these three types of material are not even the same (which is also apparent in Tables 4.1 through 4.3).

Table 4.9 reports on the effect of crack type on detection performance. Two important types of cracks were present in wrought material: thermal fatigue (TFC) and IGSCC. The chi-squared statistics in Table 4.9 show that there are significant differences between TFC and IGSCC in all cases except the far-side access inspection of medium-sized cracks (size 2). Because no IGSCC cracks larger than medium size existed, it is not possible to compare detection performance for larger

Table 4.7. Summary of Significance of Size

Access		Cast Stainless Steel	Clad Ferritic	Wrought Stainless Steel		
				Blank and TFC	Blank and IGSCC	Blank and EDM
Near	Chi	25.5*	246.4*	197.3*	26.1*	9.5*
	DOF	8	8	6	2	2
Far	Chi	9.9	--	11.6	4.2	2.1
	DOF	8	--	6	2	2
Total	Chi	35.4	246.4	208.9	30.3	11.6
	DOF	16	8	12	4	4

* indicates statistic is significant at 1% level

4.0 Variables that Influence Detection Performance

Table 4.8. Summary of Significance of Material

Size		Near	Far	Total
Blank	Chi DOF	7.5 4	18.8* 2	26.3* 6
Size 1	Chi DOF	170.6* 6	15.1* 4	185.7* 10
Size 2	Chi DOF	96.9* 8	11.1 6	108.0* 14
Size 3	Chi DOF	63.2* 6	17.9* 4	81.1* 10
Size 4	Chi DOF	76.3* 2	-- --	76.3* 2
Total	Chi DOF	414.5* 26	62.9* 16	477.4* 42

* significant at 1% level

sizes. If one refers to Table 4.5, it is possible to determine exactly how crack type seems to influence POD. It appears that the misfit is caused because POD is much higher for IGSC cracks than for TFC. For example, near-side POD is 38% for small IGSCC, but only 17% for small TFC.

4.3 Relative Importance of Variables

Table 4.10 orders the variables discussed in this section according to their importance. "Importance" has been defined as the chi-squared value divided by degrees of freedom. Two variables (inspection environment and procedure) are not significant. All other variables are. As one can see from Table 4.10, all significant variables have about the same magnitude of importance.

The inspection variable of "environment" can be understood that it should not influence an inspection if the inspection team conscientiously follows their procedure.

Highly motivated inspection teams should be able to deal with the annoyance of physically awkward positions which was the primary difference between the two environment conditions (laboratory versus difficult). In the case of the procedure, extensive laboratory testing had shown that the improved procedure should have

been more effective than the procedures that most of the teams employed in the field. Most of the teams procedures were based on a DAC level that an indication needed to exceed in order for it to be recorded and evaluated. The improved procedures required a team to evaluate all signals regardless of amplitude that exhibited "crack-like properties." Thus, the improved procedures required new skills and evaluations of signals that were not normally evaluated. It is thought that the most probable explanation for the lack of improvement for the improved procedures is that the time spent in training the inspectors to follow the new procedures was not adequate and that the inspectors were, therefore, not able to effectively apply the improved procedures.

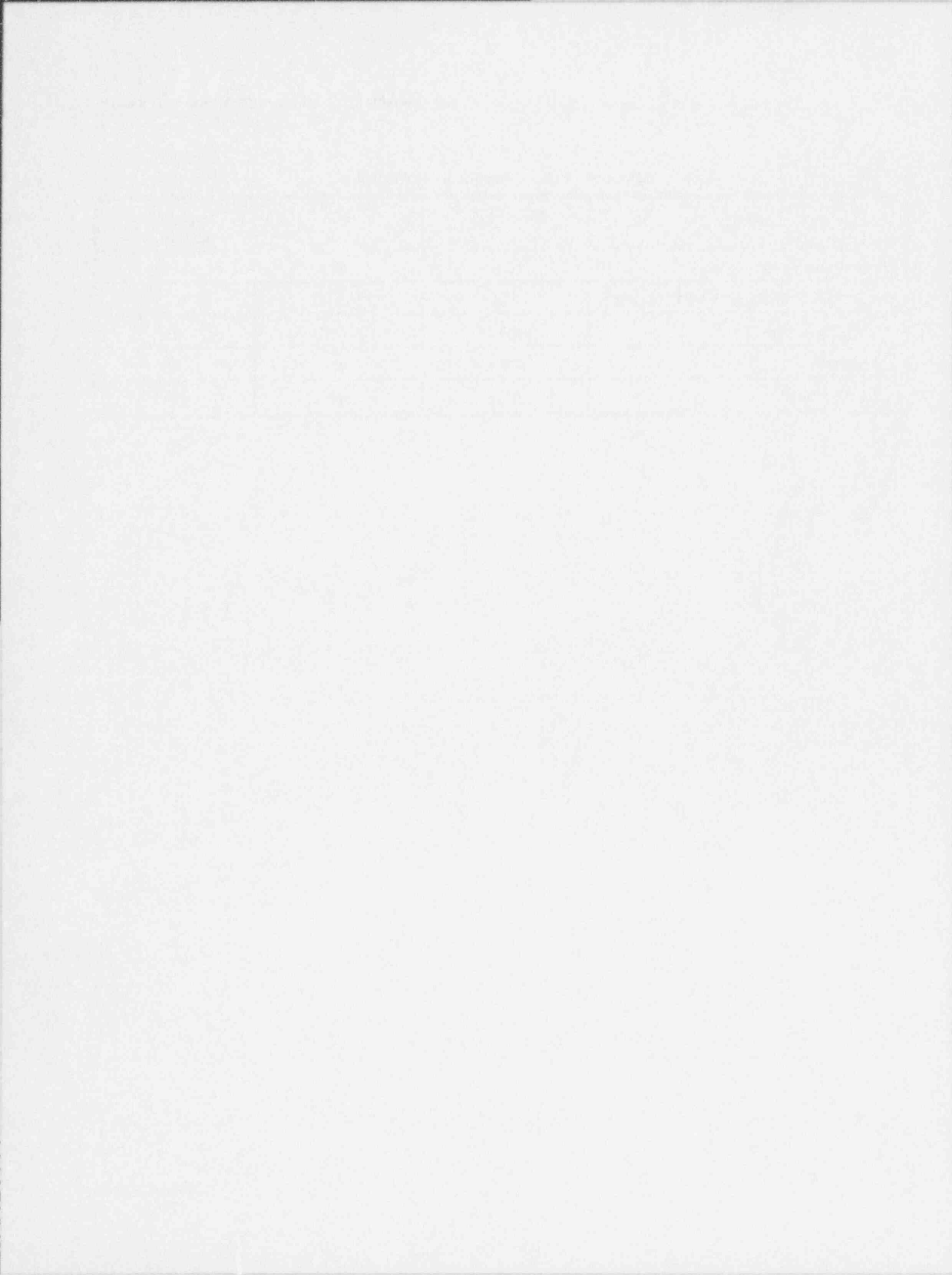
Table 4.9. Significance of Defect Type in Wrought Stainless Steel (TFC vs. IGSCC)

Size		Near	Far	Total
Size 1	Chi DOF	17.6* 2	10.7* 2	28.3* 4
Size 2	Chi DOF	43.8* 2	1.5 2	45.3* 4
Total	Chi DOF	61.4* 4	12.2 4	73.6* 8

4.0 Variables that Influence Detection Performance

Table 4.10. Relative Importance of Variables

Variable	Chi-squared	DOF	Importance Chi/DOF
Environment and Procedure	93.8	94	1.0
Crack Type in Wrought (IGSCC vs. TFC)	73.6	8	9.2
Access (Near, Far)	260.2	26	10.0
Material	477.4	42	11.4
Defect Size	532.6	44	12.1



5.0 Relationship Between Detection and Crack Size

This section examines the relationship between POD and defect size in detail. The specific objective of this section is to determine a quantitative relationship between POD and crack size, which is accomplished with quantal response regression (Agresti, 1990). The general type of relationship assumed to exist between crack size and POD is described by a logistic function, which has the form;

$$POD(s) = \text{logit}(\beta_1 + \beta_2 s) \quad (5.1)$$

where

$$\text{logit}(\beta_1 + \beta_2 s) = (1 + \exp(-\beta_1 - \beta_2 s))^{-1} \quad (5.2)$$

(β_1, β_2) are unknown parameters to be determined by the regression fit, and s is the crack size. The relationship postulated above can be derived from a "signal/response" model (Swets, 1983). Generally, the model assumes that the inspection signals are normally distributed, with the mean of the distribution depending on crack size. In the above formulation, the normal cumulative distribution function has been replaced by a logistic function. (When a normal distribution function is employed, the model is called a Probit model). A logistic function has been substituted for the normal because it is simpler to work with and both curves have very similar shapes.

A plot of the logistic function (Equation 5.1) is presented in Figure 5.1. Other investigators have employed a curve of this form to model probability of detection, see for example (Berens, 1984). The model can also be extended to include more than one independent variable. Such extended models allow the effect of several variables on POD to be quantified simultaneously.

Inspection results from all teams were employed and the logistic model was fit to POD statistics. Near-side inspections were used for all materials except clad ferritic, for which "both-side" access inspections were used.

Crack size may be quantified in several different ways. The size of a crack may be described by its length, depth, or some function of both, such as area. When cracks with complicated geometry are considered (such as IGSCC), even more complicated quantities might be proposed.

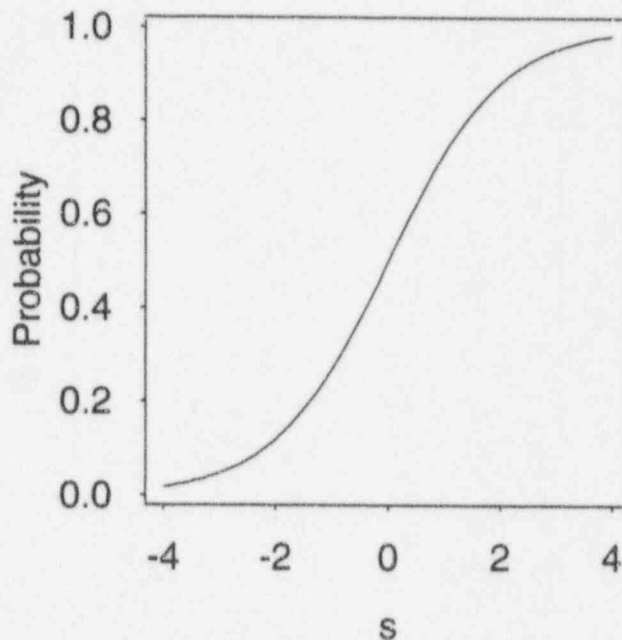


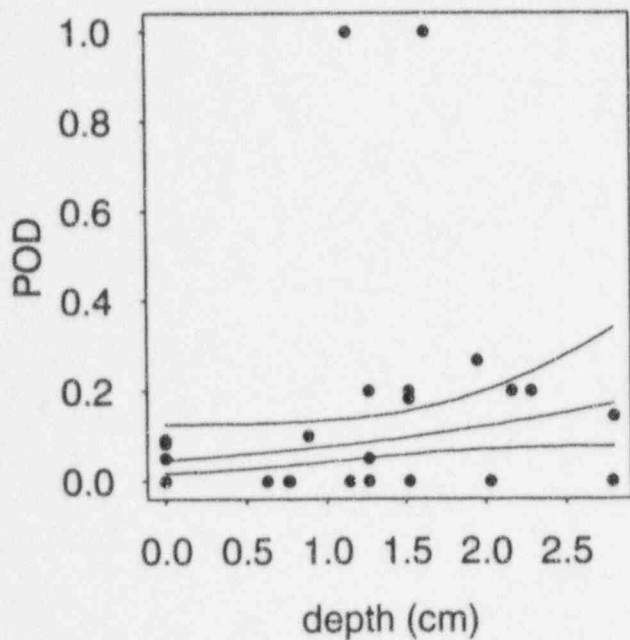
Figure 5.1 Plot of Logistic Function

During the planning stages of the study, crack depth was identified as the single dimension most relevant to pipe failure. Consequently, the Piping Inspection Round-Robin specimens were designed to allow the relationship between detection performance and crack depth to be evaluated. Crack depth was chosen because fracture mechanics calculations indicated that this parameter was most closely associated with structural integrity.

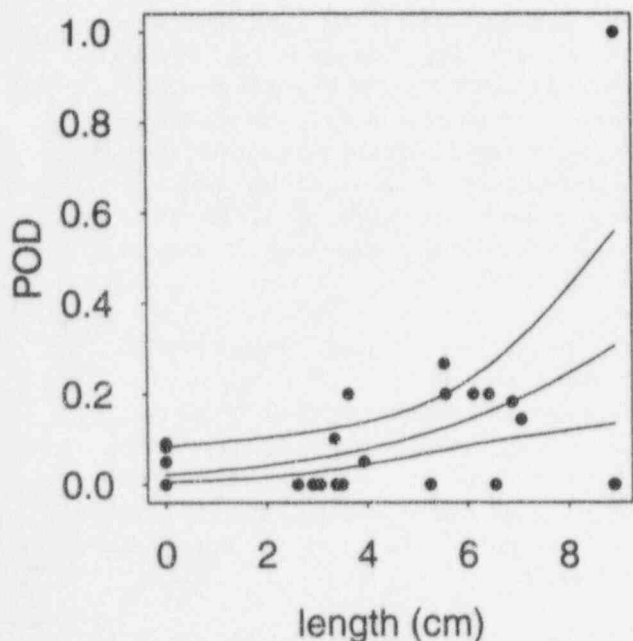
However, when preliminary results became available, questions were raised about the effect on detection of other crack dimensions (e.g., length, cross-sectional area). Even though crack depth is most significant from a failure point of view, it is important to determine whether other size parameters are more directly related to POD and RP. In this section, this question is also investigated.

Table 5.1 presents a summary of goodness-of-fit (GOF) statistics for two logistic models; one based only on crack depth, and the other based only on crack length. A large GOF statistic indicates that either the POD function does not fit the data, or substantial variability exists in the data (most likely crack-to-crack or

5.0 Relationship Between Detection and Crack Size



(a)



(b)

Figure 5.2 Logistic Curve Fit to POD Data with 95% Bounds (Cast SS, Near-Side Inspections, All Teams)
a) POD vs. Depth b) POD vs. Length

team-to-team variability). For a model that fits the data properly, the GOF statistic should be comparable in magnitude to its degrees of freedom (DOF). To perform a formal goodness-of-fit test, the GOF statistic should be compared to a critical value obtained from a chi-squared table.

Table 5.1. Table of GOF Statistics for Two POD Curve Fits

Material		Variables in Model	
		Depth	Length
Cast SS	GOF	38.3	29.9
	DOF	31	31
Clad Ferritic	GOF	55.9	40.2
	DOF	18	18
Wrought SS/TFC	GOF	177.5	223.5
	DOF	50	50
Wrought SS/IGSCC	GOF	112.3	158.8
	DOF	47	47

GOF = goodness of fit
DOF = degrees of freedom

The GOF statistics in Table 5.1 indicate that:

- Either length or depth would be an acceptable metric for TFCs in cast stainless steel, although crack length is slightly better.
- Length is a better metric than depth in clad ferritic.
- Depth is a better metric than length for both TFC and IGSCC in wrought stainless steel. Length is not a good size metric to use for wrought SS.

The fact that crack depth is not more significant in cast stainless steel is undoubtedly influenced in part by the fact that the detection rate for cracks in this material is very low. Another factor is that because cast SS materials produce a large amount of coherent acoustic noise, one of the discriminants that inspectors use is to look for signals that have lengths which exceed that

5.0 Relationship Between Detection and Crack Size

associated with the noise; thus possibly providing greater correlation with crack length. As will be seen in Section 5.1, there was no differences in performance for depth or length for four (4) of the teams, but the remaining two (2) teams did have a higher POD for the longer cracks. However, the measurement errors (95% confidence bounds) are very large and could account for the results found. For the clad ferritic cracks, there is no good technical explanation for crack length being more significant than crack depth.

The following sections present the results of fitting POD vs. crack data to Equation (5.1), for:

- TFC in cast SS (Section 5.1, Table 5.2, Figures 5.2-5.11)
- TFC in clad ferritic (Section 5.2, Table 5.3, Figures 5.12-5.19)
- TFC in wrought SS (Section 5.3, Table 5.4, Figures 5.20-5.29)
- IGSCC in wrought SS (Section 5.4, Table 5.5, Figures 5.30-5.39)

In each section, a table summarizes the curve fits for a) POD vs. depth and b) POD vs. length. Results are shown for various inspection conditions and for various teams. The table lists the fit parameters β_1 and β_2 for Equation (5.1), along with their standard deviations SD. Also tabulated are the chi-squared goodness-of-fit (GOF) and degrees-of-freedom (DOF) criteria. The fitted curves for all listed cases are shown, in order, at the end of this section. Each figure shows the POD vs. depth regression as Part a), and the POD vs. length regression as Part b).

Of course, POD regression can also be used to describe the relationship between recording probability and crack size. Logistic regression was used to produce recording probability curves for the different inspected materials. The recording probability curve will always be higher than the associated POD curve, and is useful for checking the inspector's internal calibration.

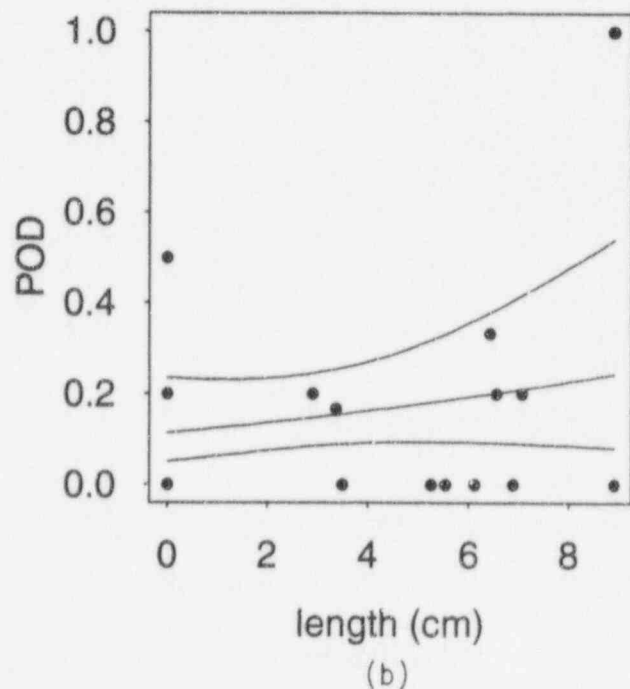
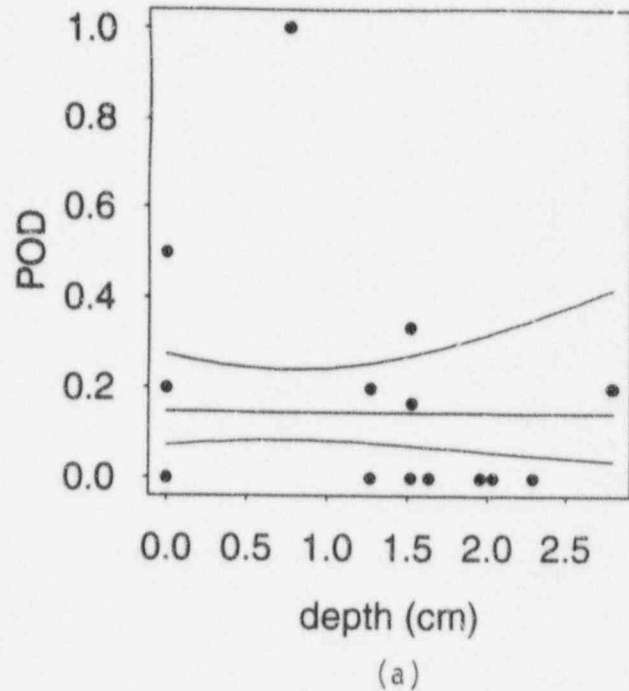
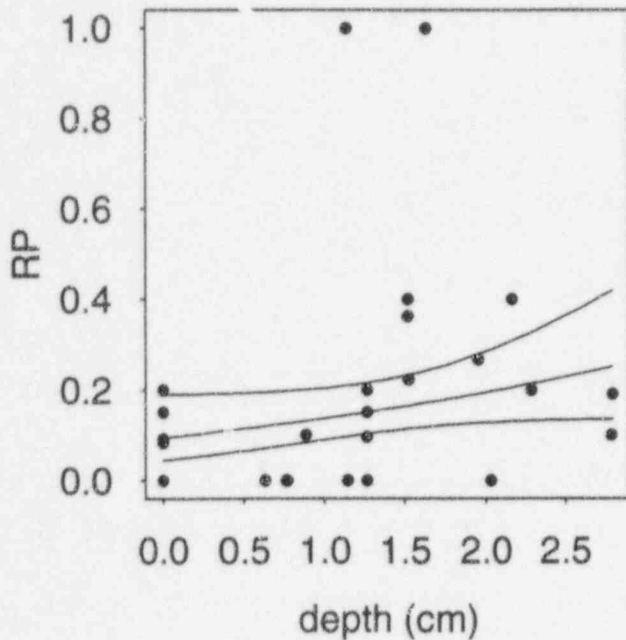
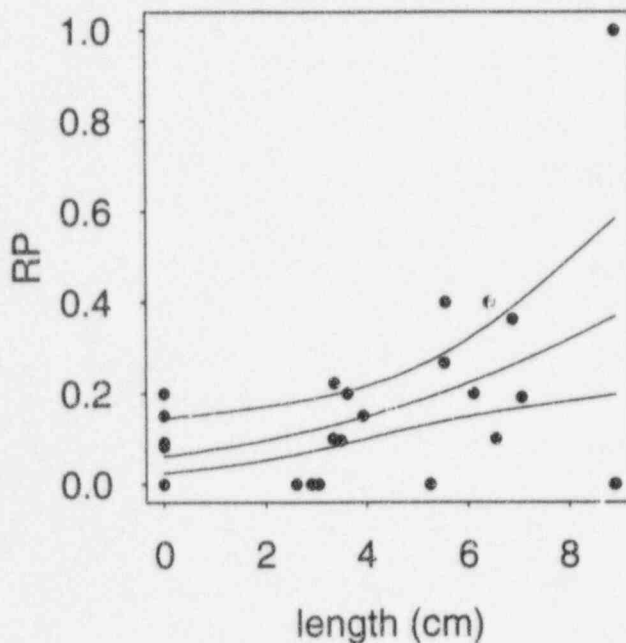


Figure 5.3 Logistic Curve Fit to POD Data with 95% Bounds (Cast SS, Far-Side Inspections, All Teams)
a) POD vs. Depth b) POD vs. Length

5.0 Relationship Between Detection and Crack Size



(a)



(b)

Figure 5.4 Logistic Curve Fit to RP Data with 95% Bounds (Cast SS, Near-Side Inspections, All Teams)
a) POD vs. Depth b) POD vs. Length

5.1 Detection Curves for Thermal Fatigue Cracks in Cast Stainless Steel

Cast stainless steel is one of the most difficult materials to inspect reliably. All the POD curves fit to the cast SS results confirm this assertion. None of the POD curves show a significant relationship between POD and crack size. Since the blank grading units (0 crack size) are included in these POD regressions, this means that the inspectors cannot effectively distinguish between blank and cracked material.

POD curves were computed for several sets of conditions; POD curves were fitted to each individual team's results. Also, POD curves were fitted to aggregated near-side and far-side inspections. Finally, POD curves were computed using crack depth and length as the independent variable.

It is important to note that no far-side detection curves are available for individual teams. Individual teams did not conduct enough far-side inspections to allow such curves to be calculated.

The regression procedure produces estimates for the unknown model parameters and also gives an uncertainty for the estimates. This information can be used to surround the POD curve with confidence bounds. The basic results of the POD regressions are given in Table 5.2.

The results of these fits are also displayed graphically in Figures 5.2 through 5.11. The logistic curves are surrounded by 95% confidence bounds and the raw POD (or RP) points used in the fit are also illustrated on the plots. Each POD point describes the detection results on an individual crack (or blank grading unit), so the variation around the POD curve represents crack-to-crack variability.

It should be noted that the false call statistics associated with blank grading units are displayed at size = 0 in these plots. These statistics are included in the logistic regression, so the detection curves reflect the effects of false calls.

From these plots, it is apparent that it is very difficult to detect cracks in cast SS; none of the plotted curves are particularly good. In fact, because the curves are

Table 5.2 Summary of POD Logistic Curve Fits in Cast SS

Access	Detection Statistic	Team	β_1	Std. Dev. β_1	β_2 (1/cm)	Std. Dev. β_2 (1/cm)	GOF	DOF
POD vs. Depth Regressions								
near	POD	all	-3.0	0.44	0.518	0.244	38.6	31
far	POD	all	-1.8	0.32	-0.007	0.264	43.0	24
near	RP	all	-2.3	0.33	0.421	0.193	37.3	31
far	RP	all	-1.4	0.29	-0.037	0.236	50.5	24
near	POD	#1	-2.4	0.86	0.199	0.520	15.1	21
near	POD	#3	-50.8	8.23	0.085	5.30	0.1	21
near	POD	#4	-50.8	11.64	0.019	6.86	0.04	12
near	POD	#5	-2.4	0.80	0.507	0.469	17.9	21
near	POD	#6	-3.0	0.93	0.881	0.491	23.3	21
near	POD	#7	-3.4	1.14	0.650	0.681	21.0	30
POD vs. Length Regressions								
near	POD	all	-3.8	0.58	0.336	0.102	29.9	31
far	POD	all	-2.1	0.36	0.104	0.079	41.4	24
near	RP	all	-2.7	0.39	0.246	0.075	29.7	31
far	RP	all	-1.6	0.30	0.057	0.071	50.0	24
near	POD	#1	-2.2	0.85	0.028	0.189	15.2	21
near	POD	#3	-50.8	8.47	-0.030	1.94	0.1	21
near	POD	#4	-50.9	12.03	0.016	2.65	0.04	12
near	POD	#5	-2.7	0.92	0.254	0.185	17.0	21
near	POD	#6	-4.9	1.66	0.704	0.287	16.8	21
near	POD	#7	-77.6	7.24	8.717	0.819	8.0	30

essentially flat, we can conclude that far-side crack-detection performance in this material is no better than chance. In this material, there is no large difference between the two decision criteria; the "N" decision

criteria used to compute Recording Probability produces a slightly higher curve than the POD curve associated with the "C" decision criteria, but it is also flatter.

5.0 Relationship Between Detection and Crack Size

The near-side detection curve results show that there may be a slightly better (but still poor) ability to detect cracks in this material (see the POD and RP curves plotted against length). These results are somewhat confounded because Team 7 inspected several long (3.5 in. (8.9 cm) and deep cracks that none of the other teams inspected. Team 7 performed very well on them since they detected both of them. This performance had a strong influence on the slope of the POD curves as shown in Figure 5.11. If the results for these cracks for team 7 are removed then there is little evidence to support depth or length being correlated with POD.

The individual team plots demonstrate substantial differences between teams. Some team's POD curves are flat, indicating that their performance is no better than guessing. In a few extreme cases, teams did not successfully detect a single crack. Team No. 6 produced the best results of all the teams. This team had a false call probability of 0 while displaying a 42% POD for the longest cracks.

5.2 Detection Curves for Thermal Fatigue Cracks in Clad Ferritic

Inspection performance in clad ferritic was dramatically better than that demonstrated in the cast stainless steel material. All cracks in this material displayed relatively high probabilities for detection, while the false call probability was low. Table 5.3 displays detection curve fits for the important inspection conditions investigated in clad ferritic.

Figures 5.12 through 5.19 present plots of the fitted detection curves for the listed inspection conditions and individual teams. From Figures 5.12 and 5.13, we see that there is no dramatic difference due to decision criteria. Use of the "N" decision criterion (Recording Probability) seems to raise the lower part of the curve by about 5% over the "C" decision criterion (Probability of Detection). An interesting "outlier" crack exists at length = 2.8 in. (7.1 cm) in these plots. According to the POD curve, a crack of this size should have a POD of 99.7%, but this crack displays a POD of only 90%. Since this crack was inspected many times, the difference between 90 and 99.7% is significant. This is a good example of crack-to-crack variability.

Individual team plots demonstrate good ability for all teams. Even the worst team (No. 2) demonstrates a 95% POD for cracks of length 2.8 in. (7.1 cm). The best team, (No. 6) produced perfect results in this material; all cracks detected and a false call probability of 0.

5.3 Detection Curves for Thermal Fatigue Cracks in Wrought Stainless Steel

Thermal fatigue cracks in wrought stainless steel are much easier to detect than TFCs in cast stainless steel material, and far-side cracks are equally difficult in wrought and cast stainless steel.

The results of the logistic fits for the various inspection conditions are given in Table 5.4 while Figures 5.20 through 5.29 are the regression curves. The fits to far-side conditions (see Figures 5.21 and 5.23) demonstrate that inspection is ineffective from the far-side. A comparison of the two decision criteria (for near-side access) indicates that the "C" decision criteria (POD Figure 5.20) may be better than the "N" criteria (RP Figure 5.22). Using the "C" criteria, false calls are approximately half that of the "N" criteria, while detectability of large cracks are virtually the same (POD = RP = 75%) at a depth size of 0.25 in. (0.6 cm). Thermal fatigue cracks in wrought SS display the largest crack-to-crack variability of any of the crack/material combinations, as Figure 5.20 illustrates. Five outlier cracks (substantially outside of confidence bounds) exist on this POD plot and one blank grading unit (see size = 0) also seems to differ significantly from the average.

However, less team-to-team variability seems to exist in this material than the others. Only one team (No. 3) displays results that differ greatly from the average. This team performs more poorly than the others, and is also the same team that had the worst performance in the clad ferritic material.

5.4 Detection Curves for IGSC Cracks in Wrought Stainless Steel

As experienced in the thermal fatigue data fits, we find far-side performance for IGSCC (shown in Figure 5.31)

Table 5.3 Summary of POD Logistic Fits in Clad Ferritic

Access	Detection Statistics	Team	β_1	Std. Dev. β_1	β_2 (1/cm)	Std. Dev. β_2 (1/cm)	GOF	DOF
POD vs. Depth Regressions								
both	POD	all	-2.2	0.46	3.346	0.472	57.4	18
both	RP	all	-2.0	0.43	3.858	0.512	35.8	18
both	POD	#1	-2.5	1.04	31.929	2.95	2.6	18
both	POD	#2	-3.0	1.21	4.331	1.26	12.3	18
both	POD	#3	-1.7	0.64	1.732	0.512	27.4	18
both	POD	#4	-1.7	0.95	3.031	1.02	198.2	18
both	POD	#5	-3.8	1.55	4.803	1.54	12.8	18
both	POD	#6	-18.2	7.81	42.677	5.12	0.1	18
POD vs. Length Regressions								
both	POD	all	-2.8	0.64	1.260	0.193	40.2	18
both	RP	all	-2.4	0.49	1.299	0.161	27.0	18
both	POD	#1	-2.5	1.04	10.315	0.984	2.6	18
both	POD	#2	-10.3	4.81	3.465	1.39	6.9	18
both	POD	#3	-2.2	0.78	0.709	0.201	23.9	18
both	POD	#4	-2.2	1.21	1.102	0.382	16.1	18
both	POD	#5	-7.1	4.38	2.402	1.26	12.3	18
both	POD	#6	-23.3	8.01	14.528	1.74	0.1	18

that is no better than guessing. However, in contrast to the thermal fatigue curves presented in the last section, we find that the curves obtained for IGSCC are much steeper. In fact, a comparison of the fitted parameters for wrought SS displayed in Table 5.5 with those presented in Table 5.4 show that:

- the β_1 parameter is about 0.30 units smaller for the IGSCC as compared to the TFC values for near-side fits (this parameter determines FCP),
- the β_2 parameter for IGSCC is approximately three times larger than the TFC values for

near-side fits.

Therefore, an IGSCC crack of depth size 0.25 in. (0.6 cm) (the size of the deepest thermal fatigue cracks examined) would experience a POD of 99.95 while the corresponding thermal fatigue crack will experience a POD of only 75%.

Figures 5.30 through 5.39 display the POD logistic curve fits to the data. The plots show that less crack-to-crack variability seems to be present in the IGSCC cracks than that found in the thermal fatigue cracks. In Figure 5.30, for example, only three "outlier" cracks are

5.0 Relationship Between Detection and Crack Size

Table 5.4 Summary of POD Logistic Fits in Wrought SS with TFC

Access	Detection Statistics	Team	β_1	Std. Dev. β_1	β_2 (1/cm)	Std. Dev. β_2 (1/cm)	GOF	DOF
POD vs. Depth Regressions								
near	POD	all	-2.50	0.174	5.197	0.472	178	50
far	POD	all	-1.85	0.135	0.984	0.591	82	47
near	RP	all	-1.51	0.124	4.606	0.472	135	50
far	RP	all	-0.28	0.097	0.906	0.512	91	47
near	POD	#1	-2.76	0.441	4.646	1.10	55	43
near	POD	#2	-2.90	0.451	8.189	1.50	53	45
near	POD	#3	-3.29	0.547	3.780	1.26	41	46
near	POD	#4	-1.56	0.295	4.724	1.06	74	41
near	POD	#5	-2.97	0.469	8.071	1.50	50	43
near	POD	#6	-2.43	0.385	4.843	1.10	57	45
POD vs. Length Regressions								
near	POD	all	-2.54	0.20	0.640	0.075	224	50
far	POD	all	-1.80	0.15	0.035	0.067	85	47
near	RP	all	-1.46	0.14	0.474	0.063	193	50
far	RP	all	-0.23	0.11	0.028	0.051	94	47
near	POD	#1	-2.93	0.49	0.653	0.173	57	43
near	POD	#2	-3.02	0.53	0.982	0.205	67	45
near	POD	#3	-3.12	0.56	0.395	0.189	45	46
near	POD	#4	-1.48	0.31	0.492	0.138	85	41
near	POD	#5	-2.83	0.49	0.867	0.189	69	43
near	POD	#6	-2.81	0.47	0.750	0.177	55	45

present. The individual team results show a pattern for IGSCC that is exactly the same as that experienced for TFC; all teams, with the exception of team No. 3, behave very similarly.

5.5 Average Detection Results for Data in Percent Through-Wall

It is straight forward to compare the POD performance in clad ferritic with that in cast stainless steel. The

Table 5.5 Summary of POD Logistic Fits in Wrought SS with IGSCC

Access	Detection Statistics	Team	β_1	Std. Dev. β_1	β_2 (1/cm)	Std. Dev. β_2 (1/cm)	GOF	DOF
POD vs. Depth Regression								
near	POD	all	-2.67	0.20	16.709	1.38	112	47
far	POD	all	-2.02	0.15	0.685	1.34	60	44
near	RP	all	-1.49	0.14	13.791	1.30	127	47
far	RP	all	-0.37	0.10	-0.291	0.945	78	44
near	POD	#1	-3.08	0.53	21.531	3.90	37	43
near	POD	#7	-2.89	0.48	23.480	4.29	40	44
near	POD	#2	-3.52	0.66	10.315	3.46	26	42
near	POD	#3	-1.64	0.34	16.807	3.39	66	41
near	POD	#4	-3.11	0.54	18.976	3.54	39	43
near	POD	#5	-3.21	0.56	21.713	3.94	28	45
POD vs. Length Regressions								
near	POD	all	-2.39	0.19	1.093	0.106	159	47
far	POD	all	-1.99	0.15	0.014	0.126	60	44
near	RP	all	-1.46	0.14	1.044	0.098	126	47
far	RP	all	-0.37	0.10	-0.026	0.087	78	44
near	POD	#1	-2.24	0.40	1.051	0.220	62	43
near	POD	#7	-3.17	0.55	2.107	0.366	30	44
near	POD	#2	-3.42	0.61	0.676	0.220	25	42
near	POD	#3	-1.80	0.35	1.528	0.291	57	41
near	POD	#4	-2.71	0.48	1.214	0.248	49	43
near	POD	#5	-2.39	0.42	1.101	0.228	52	45

crack type (TFCs), the range of crack sizes, and the wall thickness (60 mm) for the two materials were the same. However, the results for the cast stainless steel are for near-side access and the results for the clad ferritic are combined for both near- and far-side access.

It is also straight forward to compare the performance of the TFCs and the IGSCCs in wrought stainless steel because the material type, wall thickness, crack sizes and access conditions are basically the same. However, some difficulty comes in making a comparison of POD performance between that in clad ferritic and the cast

5.0 Relationship Between Detection and Crack Size

stainless steel with that found in wrought stainless steel.

- First, the cracks in the CSS and the clad ferritic were from 6 mm deep to 25 mm deep. In contrast the cracks in the wrought stainless steel ranged from less than 1 mm up to about 6.9 mm deep. Thus, there is virtually no overlap of the flaw sizes used to estimate POD performance.
- The flaw sizes for the wrought stainless steel are concentrated in the steep transition zone of the POD curve whereas, the flaws in the clad ferritic are concentrated in the upper portion of the curve where it is fairly flat.
- The clad and the cast material have a wall thickness of 60 mm while the wrought stainless steel had wall thicknesses of 12 to about 15.2 mm.
- The POD performance results for the clad ferritic are combined for both near- and far-side access and those for the wrought stainless steel are for near-side only.

Therefore, one must be careful in making a comparison between these materials since differences in performance may be the result of differences in these conditions and not related to the POD effectiveness for a common set of conditions.

All of the results so far presented in this section have been in absolute crack depth. It can be useful to have the data also in a relative form such as a percentage of the through-wall for particular applications. This may be one way to compensate for the difference in wall thickness and flaw depth range. Therefore, it was decided to present a series of plots for all teams, near-side access (except for clad ferritic which is a combination of near- and far-side), and each material and flaw combination.

The trends for this data are generally, consistent with those found when using the absolute data.

Figure 5.40 shows the POD logistic curve fit for cast stainless steel material containing TFC. Figure 5.41 shows the POD logistic curve fit for wrought stainless steel containing TFC. Figure 5.42 shows the POD logistic curve fit for wrought stainless steel containing

IGSCC. Figure 5.43 shows the POD logistic curve fit for clad ferritic steel containing TFC.

It is apparent that the cast stainless steel is the most difficult to inspect achieving a 10% POD for a crack 50% through-wall. Then the wrought stainless steel containing TFC is substantially better achieving a 70% POD for a flaw 50% through-wall. The results for the clad ferritic steel containing TFC and the wrought stainless steel containing IGSCC are nearly identical and substantially better than the other two material/flaw combinations.

5.0 Relationship Between Detection and Crack Size

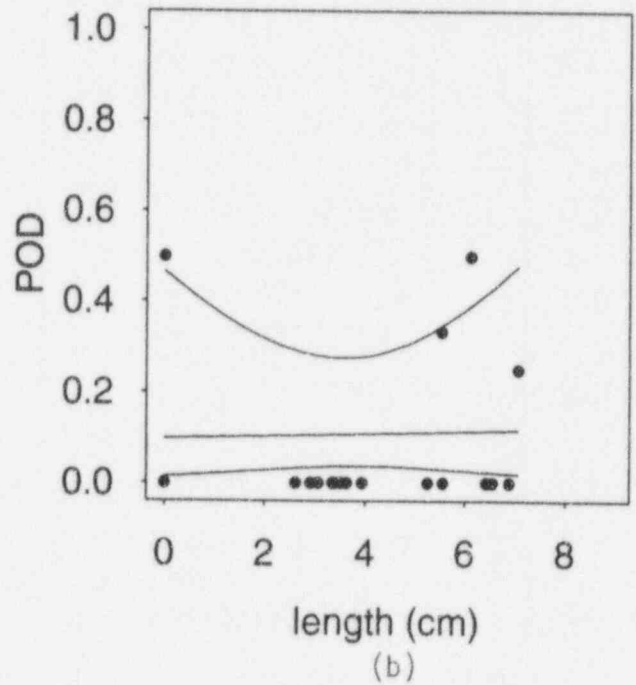
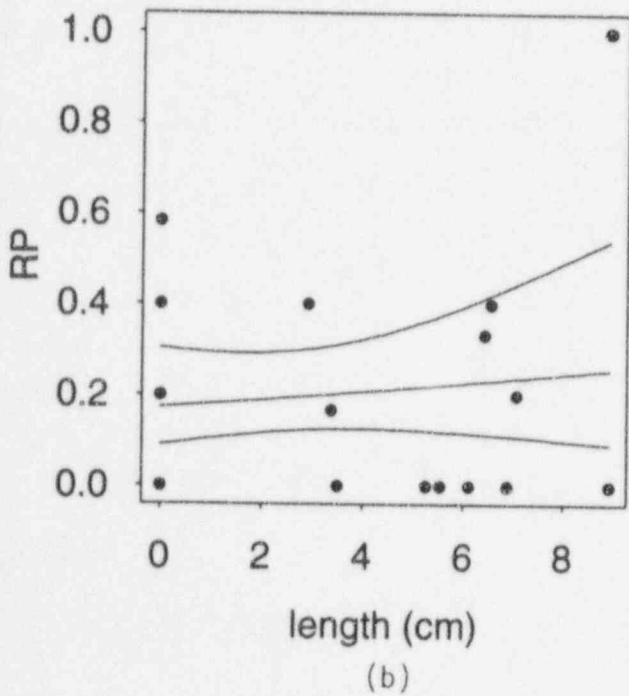
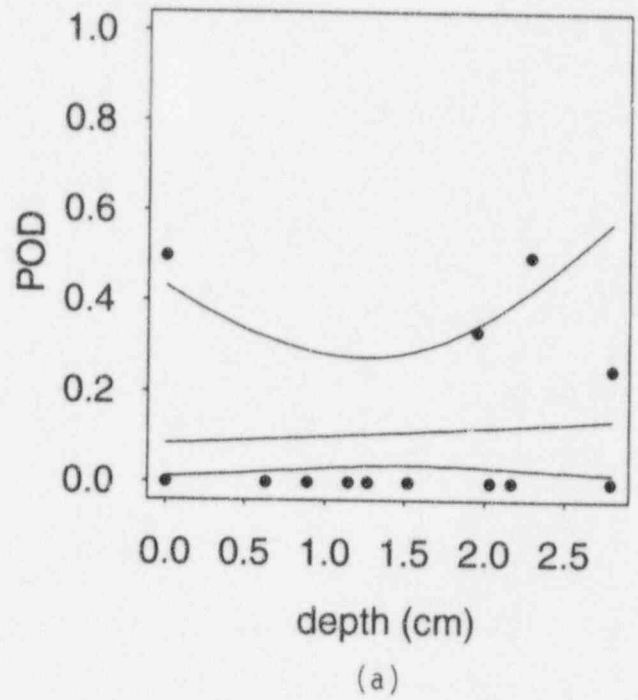
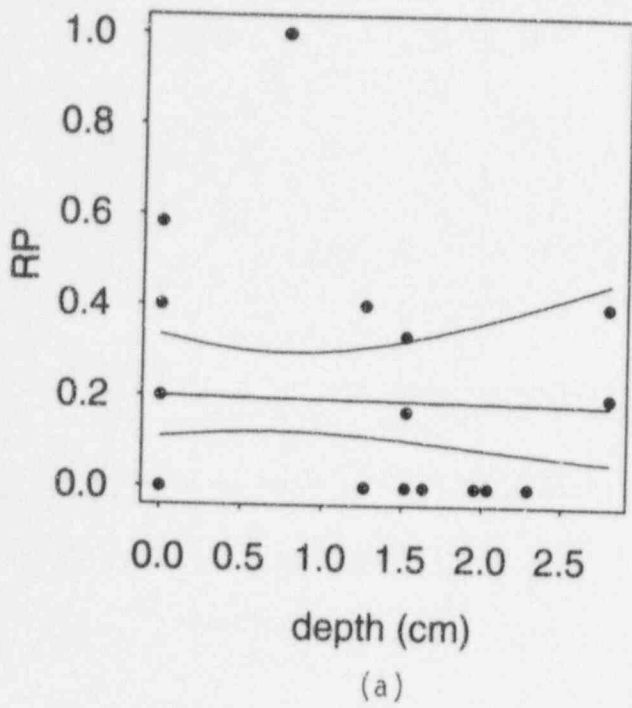
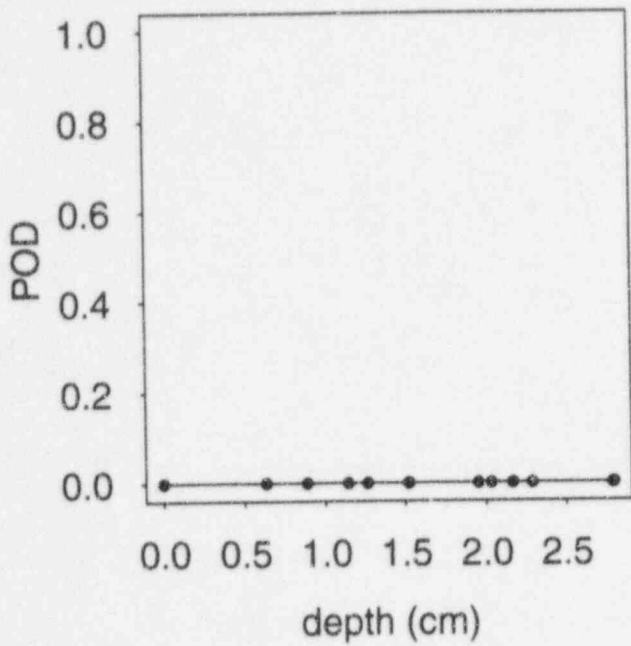


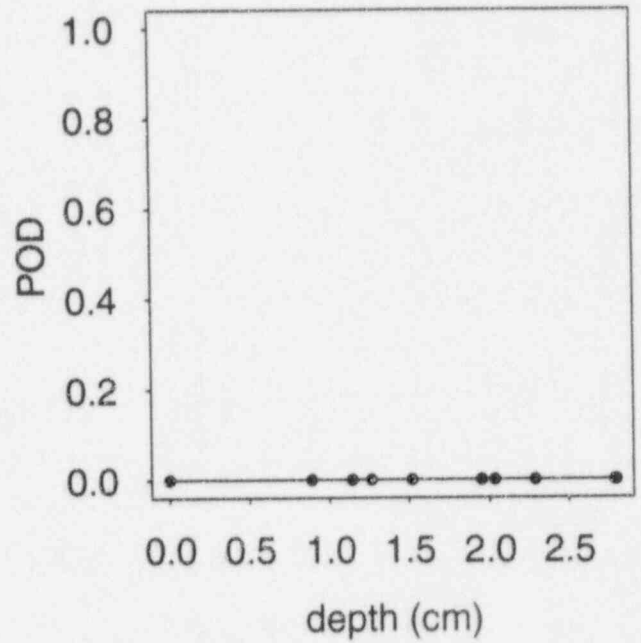
Figure 5.5 Logistic Curve Fit to RP Data with 95% Bounds (Cast Stainless Steel, Far-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length

Figure 5.6 Logistic Curve Fit to POD Data with 95% Bounds (Cast Stainless Steel, Near-Side Inspections, Team #1) a) POD vs. Depth b) POD vs. Length

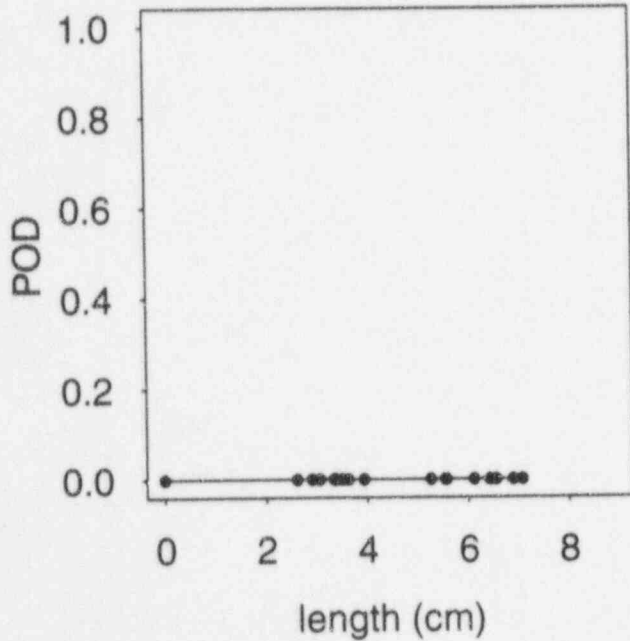
5.0 Relationship Between Detection and Crack Size



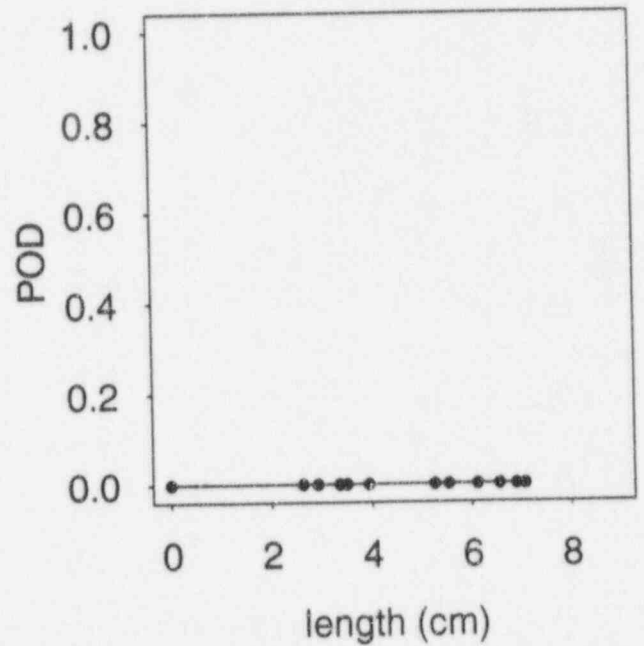
(a)



(a)



(b)

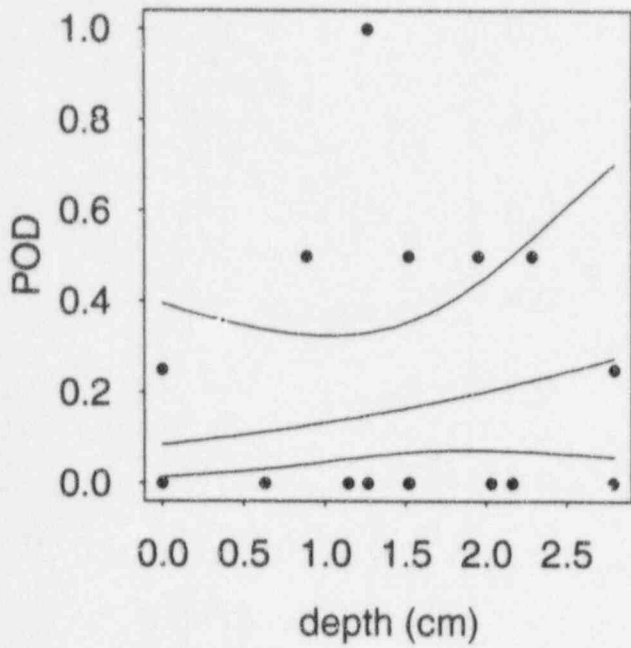


(b)

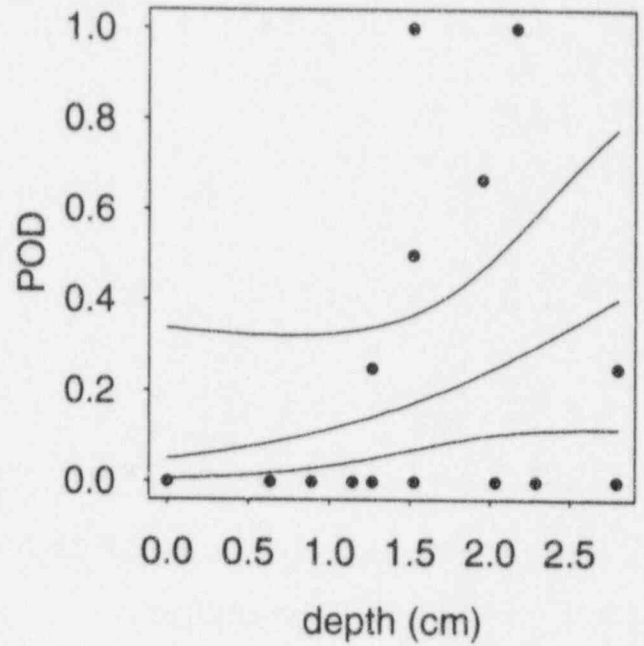
Figure 5.7 Logistic Curve Fit to POD Data with 95% Bounds (Cast Stainless Steel, Near-Side Inspections, Team #3) a) POD vs. Depth b) POD vs. Length

Figure 5.8 Logistic Curve Fit to POD data with 95% Bounds (Cast Stainless Steel, Near-Side Inspections, Team #4) a) POD vs. Depth b) POD vs. Length

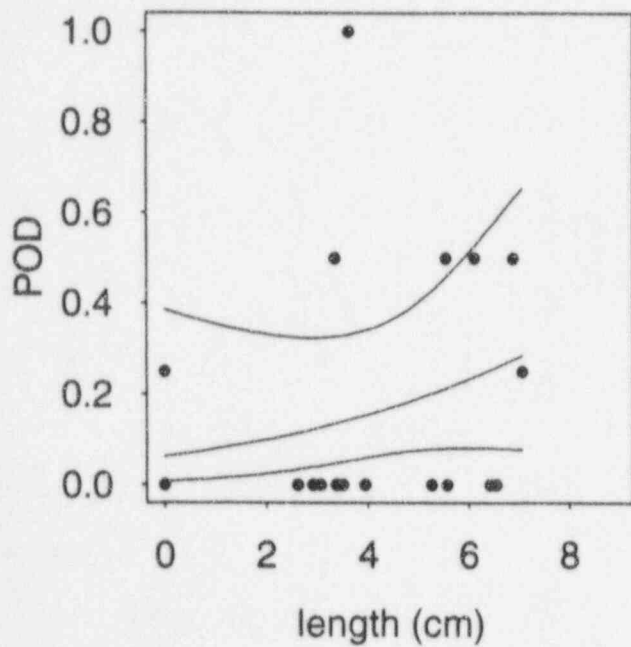
5.0 Relationship Between Detection and Crack Size



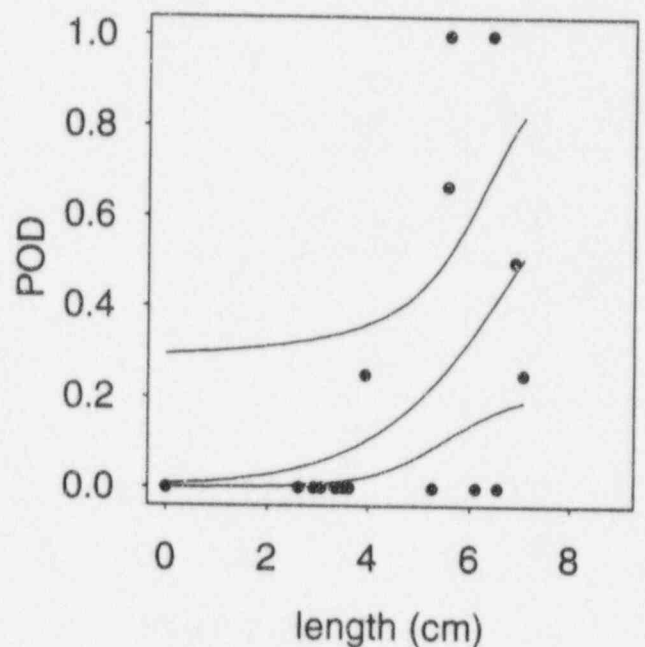
(a)



(a)



(b)



(b)

Figure 5.9 Logistic Curve Fit to POD Data with 95% Bounds (Cast Stainless Steel, Near-Side Inspections, Team #5) a) POD vs. Depth b) POD vs. Length

Figure 5.10 Logistic Curve Fit to POD Data with 95% Bounds (Cast SS, Near-Side Inspections, Team #6) a) POD vs. Depth b) POD vs. Length

5.0 Relationship Between Detection and Crack Size

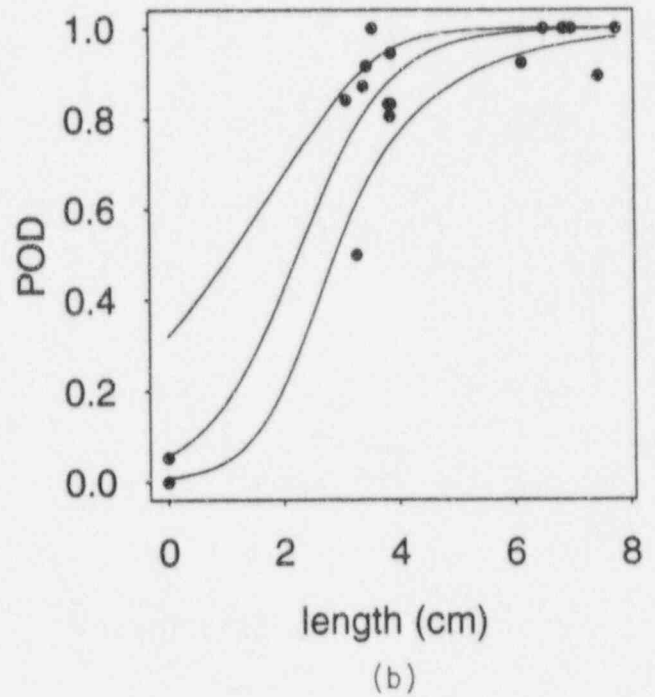
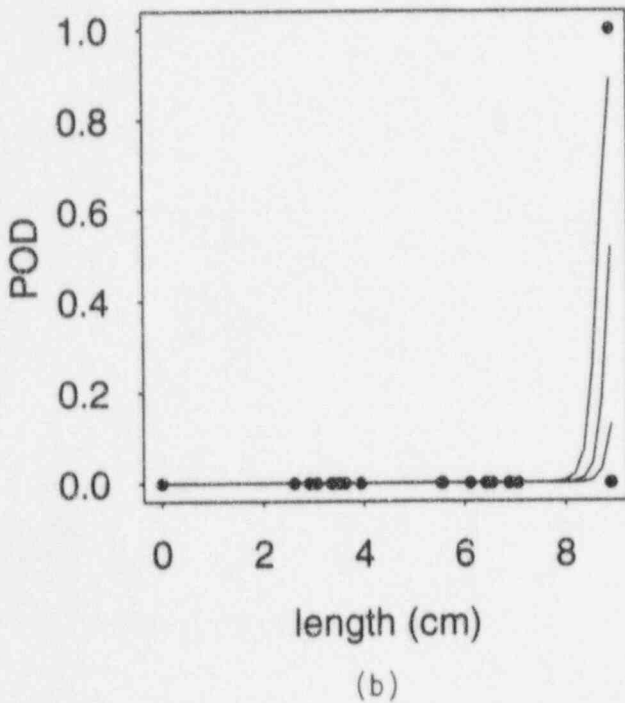
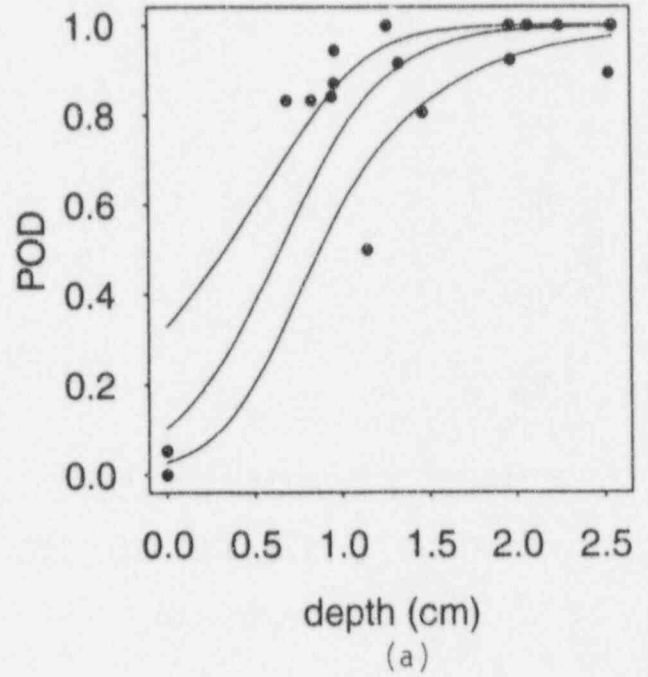
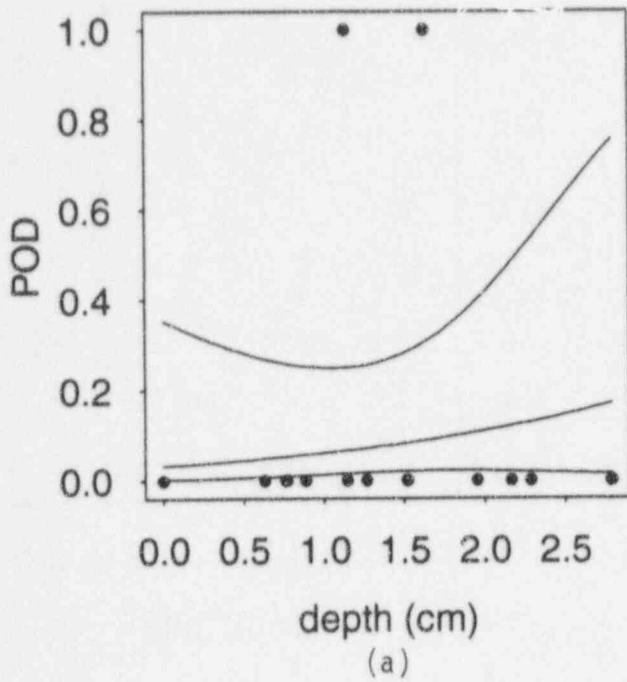


Figure 5.11 Logistic Curve Fit to POD Data with 95% Bounds (Cast SS, Near-Side Inspections, Team #7)
 a) POD vs. Depth b) POD vs. Length

Figure 5.12 Logistic Curve Fit to POD Data with 95% Bounds (Clad Ferritic, All Teams) a) POD vs. Depth
 b) POD vs. Length

5.0 Relationship Between Detection and Crack Size

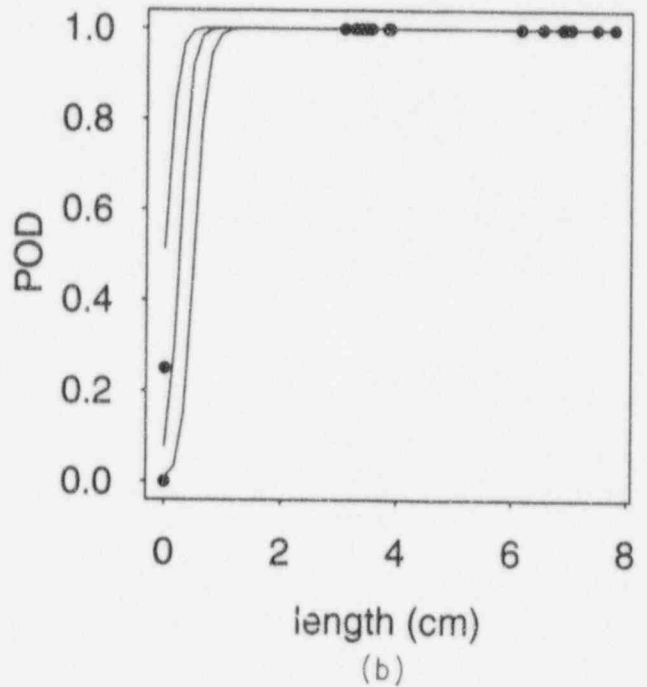
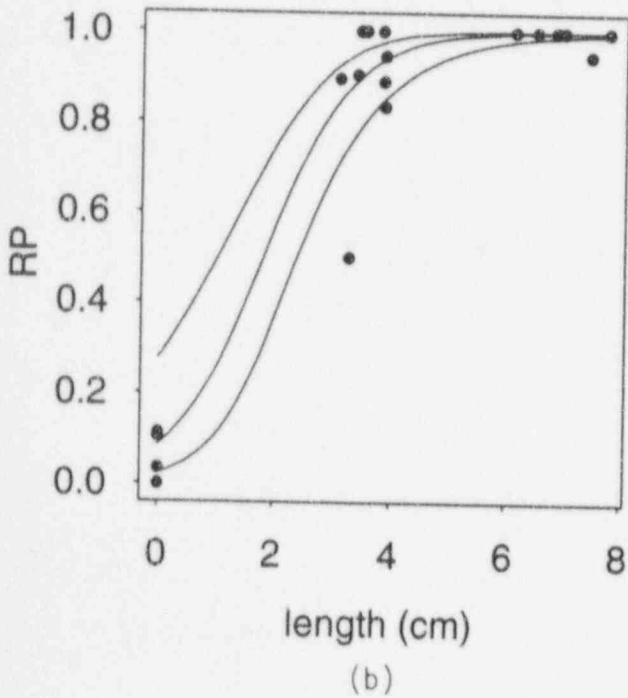
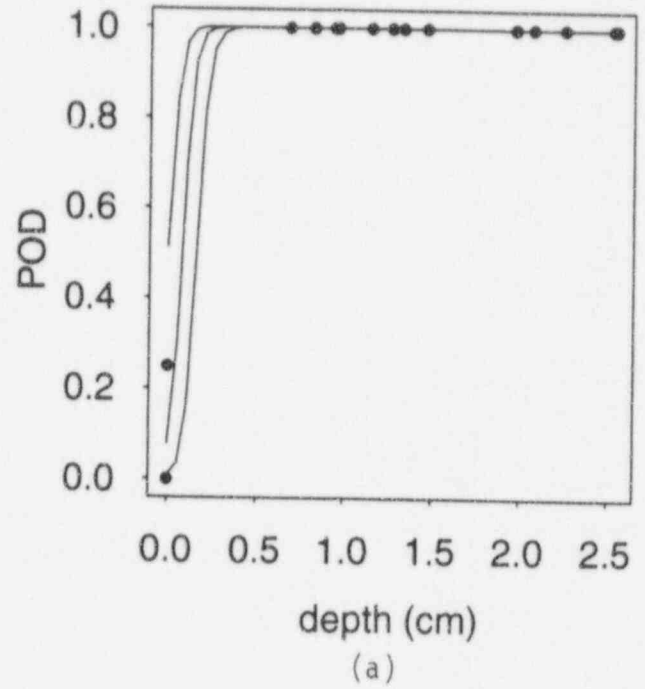
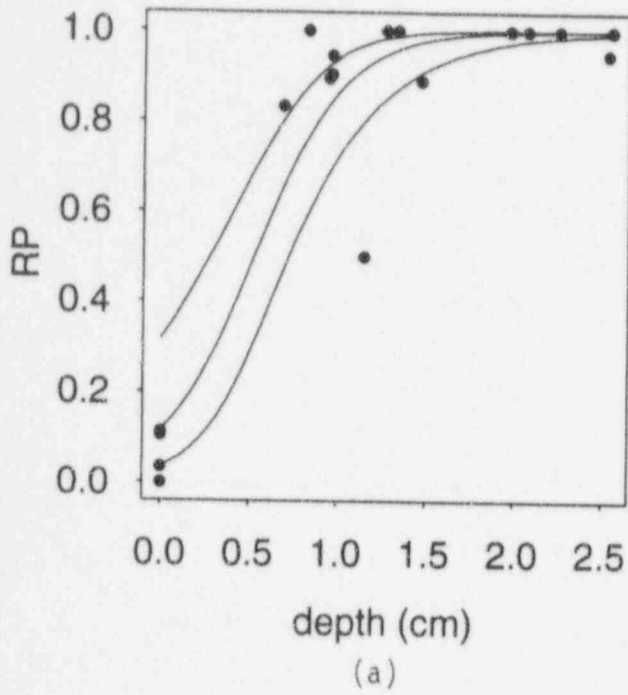


Figure 5.13 Logistic Curve Fit to RP Data with 95% Bounds (Clad Ferritic, All Teams) a) POD vs. Depth b) POD vs. Length

Figure 5.14 Logistic Curve Fit to POD Data with 95% Bounds (Clad Ferritic, Team #1) a) POD vs. Depth b) POD vs. Length

5.0 Relationship Between Detection and Crack Size

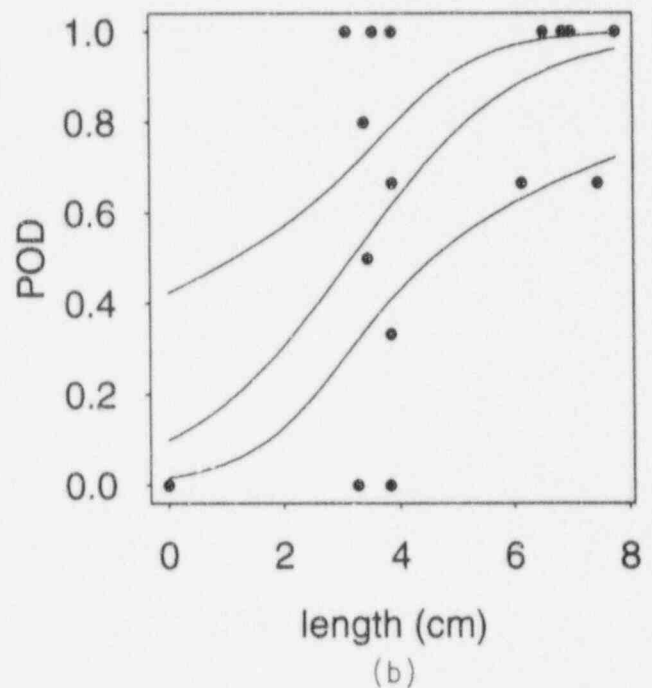
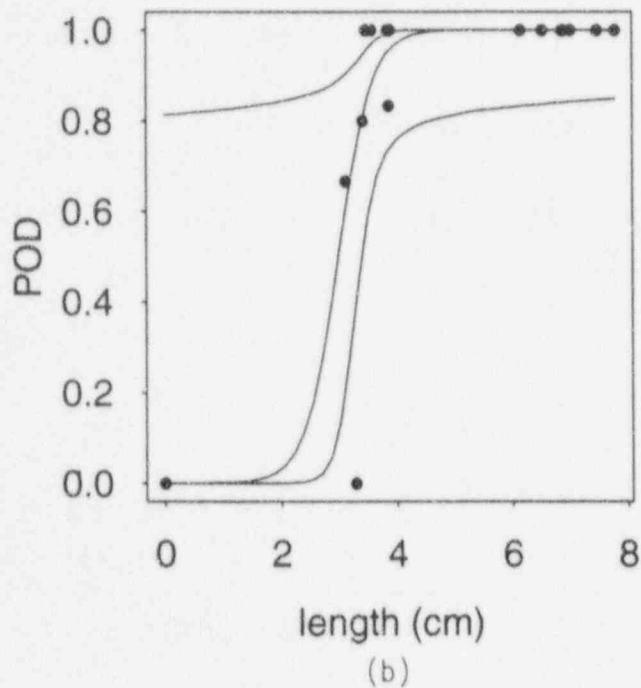
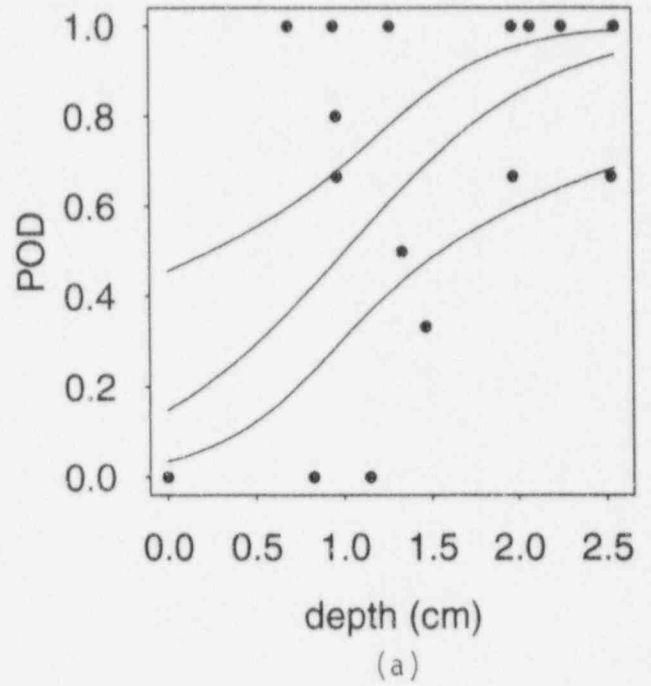
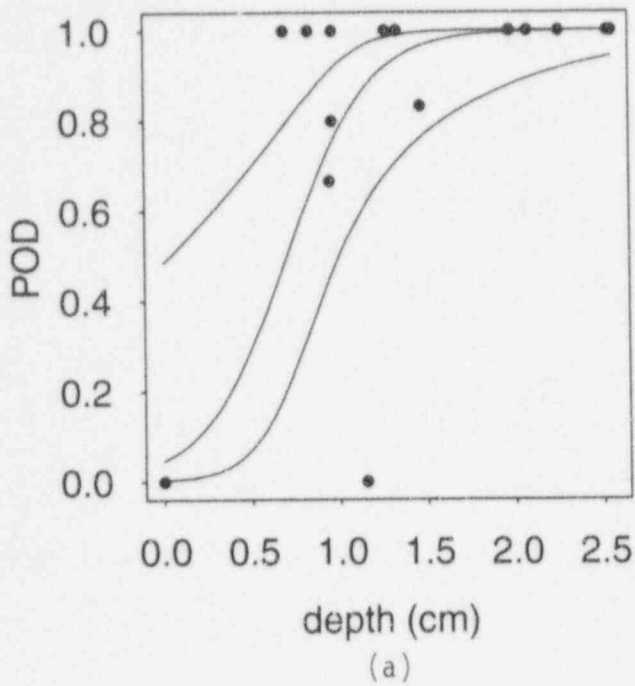


Figure 5.15 Logistic Curve Fit to POD Data with 95% Bounds (Clad Ferritic, Team #2) a) POD vs. Depth b) POD vs. Length

Figure 5.16 Logistic Curve Fit to POD Data with 95% Bounds (Clad Ferritic, Team #3) a) POD vs. Depth b) POD vs. Length

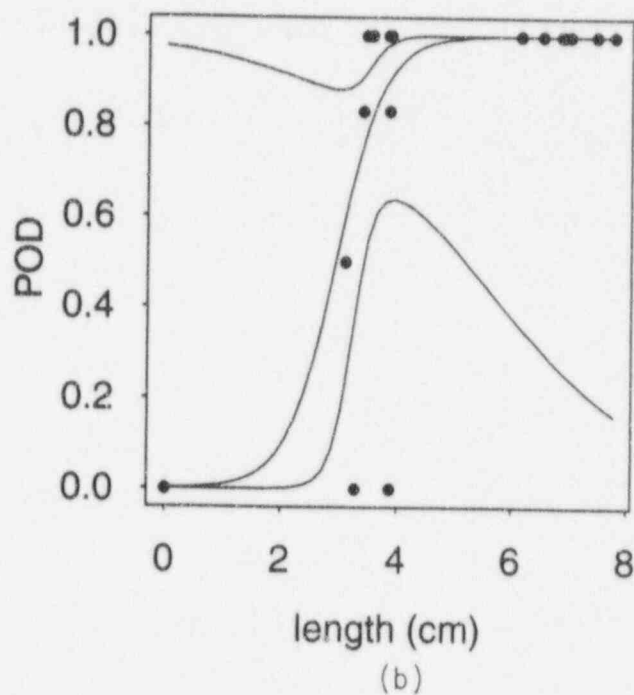
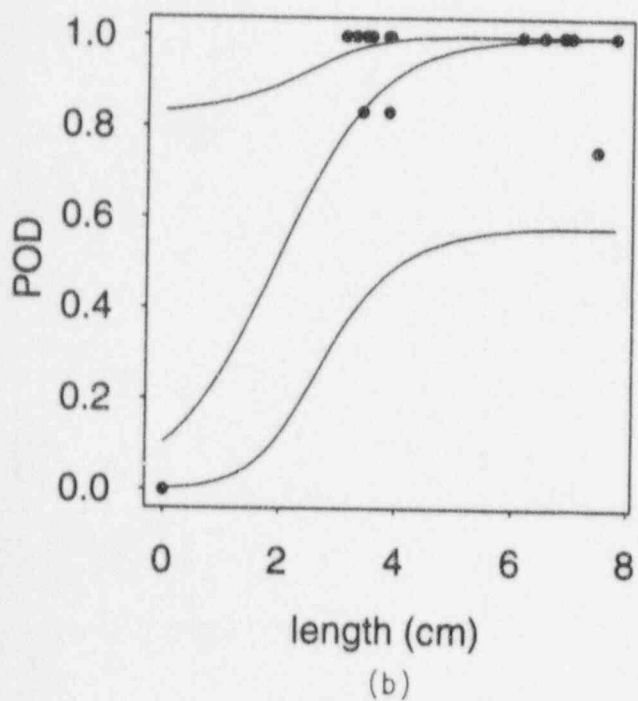
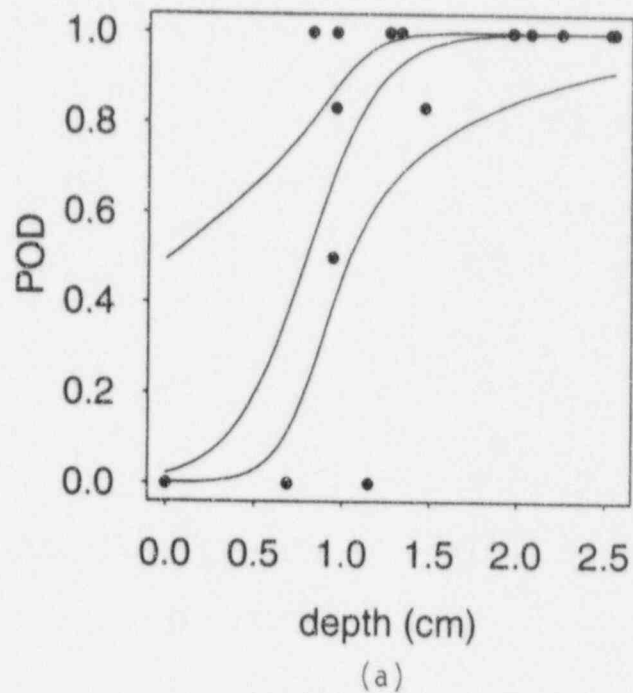
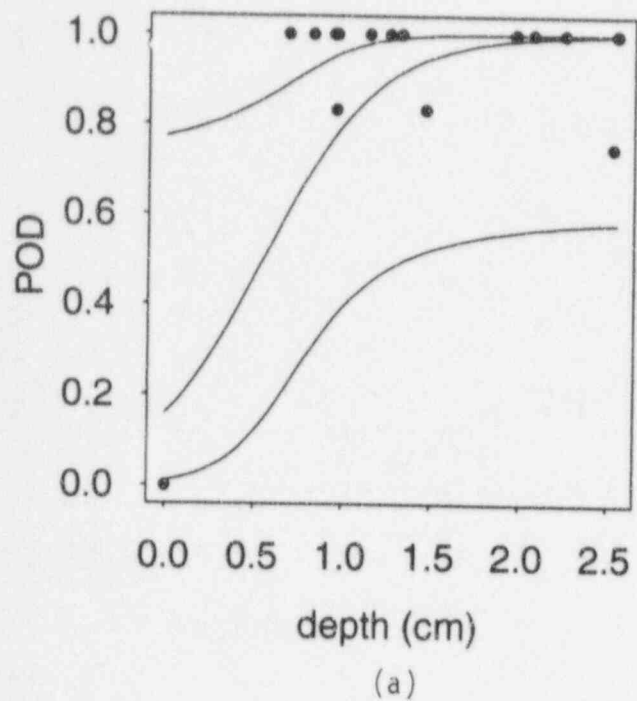
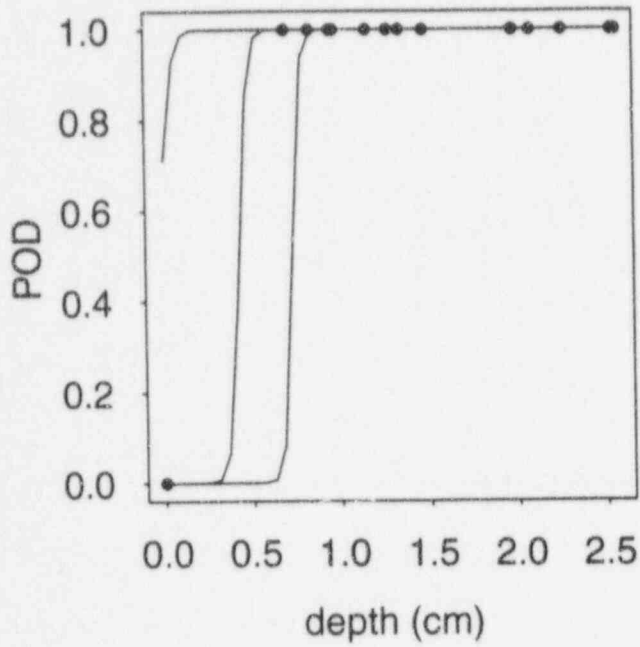


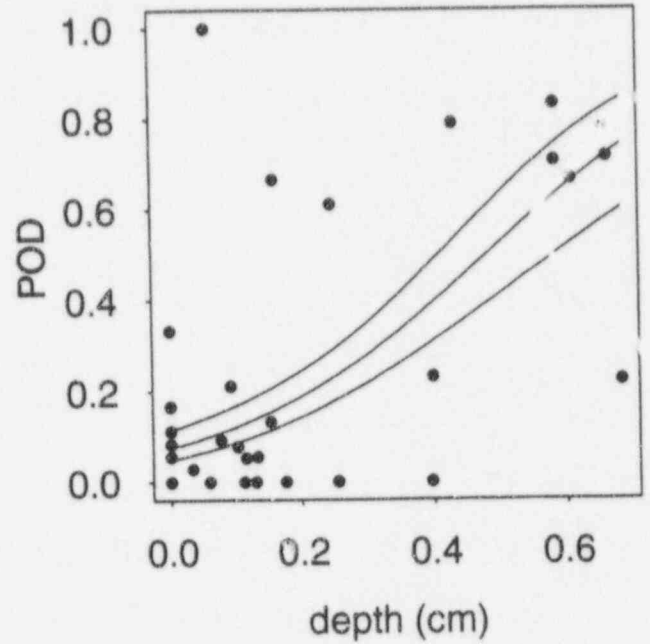
Figure 5.17 Logistic Curve Fit to POD Data with 95% Bounds (Clad Ferritic, Team #4) a) POD vs. Depth b) POD vs. Length

Figure 5.18 Logistic Curve Fit to POD Data with 95% Bounds (Clad Ferritic, Team #5) a) POD vs. Depth b) POD vs. Length

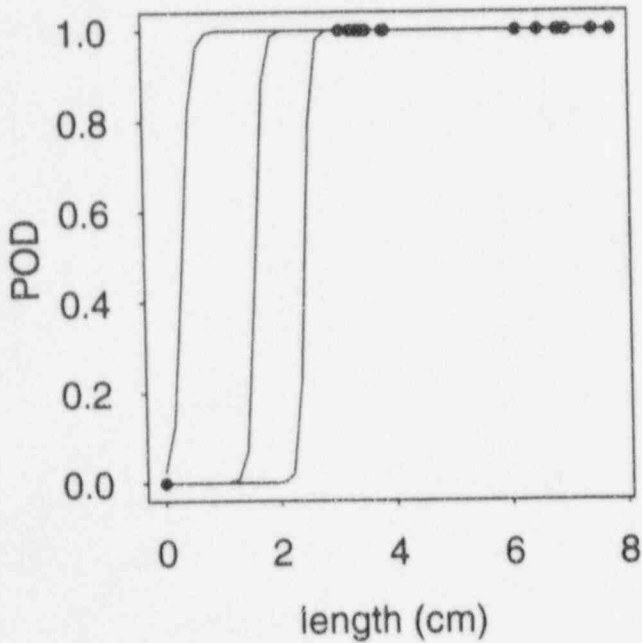
5.0 Relationship Between Detection and Crack Size



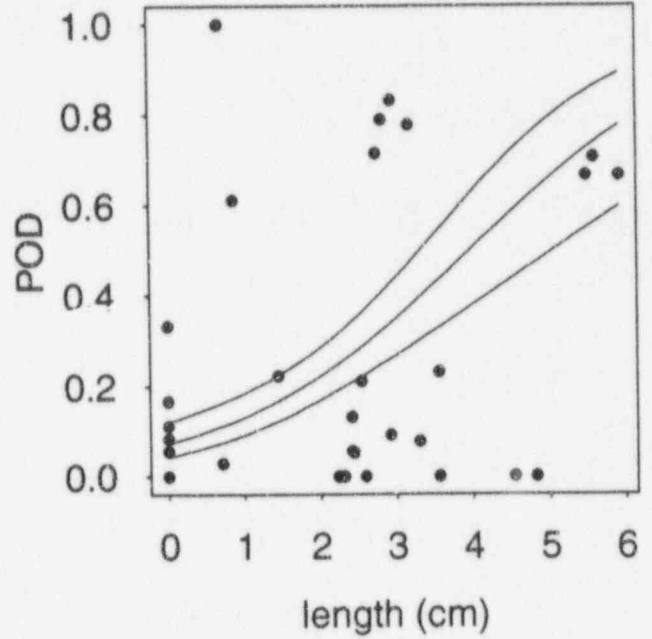
(a)



(a)



(b)



(b)

Figure 5.19 Logistic Curve Fit to POD Data with 95% Bounds (Clad Ferritic, Team #6) a) POD vs. Depth b) POD vs. Length

Figure 5.20 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length

5.0 Relationship Between Detection and Crack Size

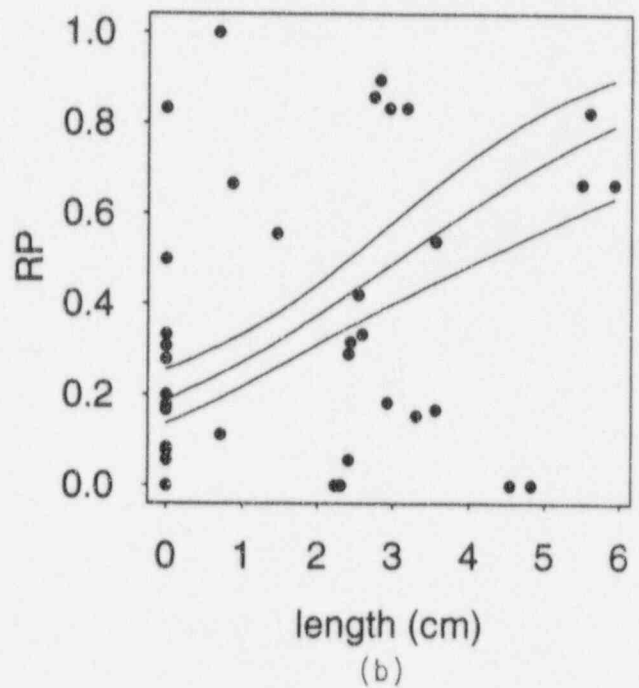
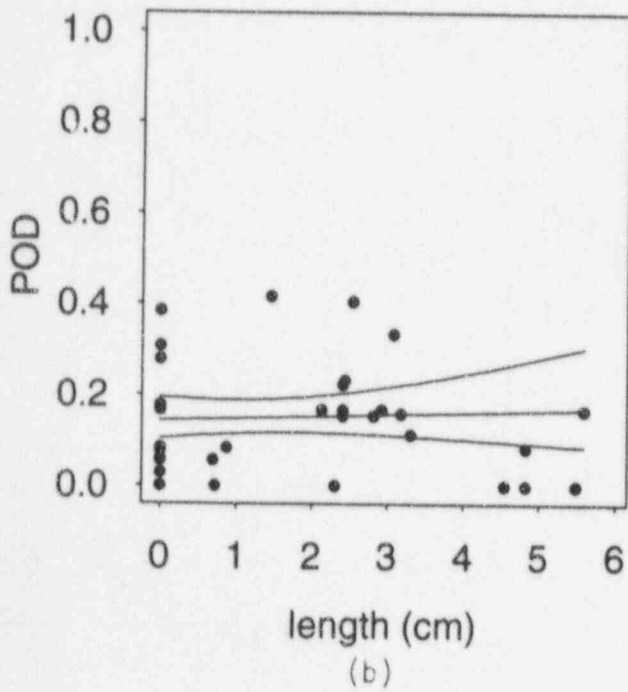
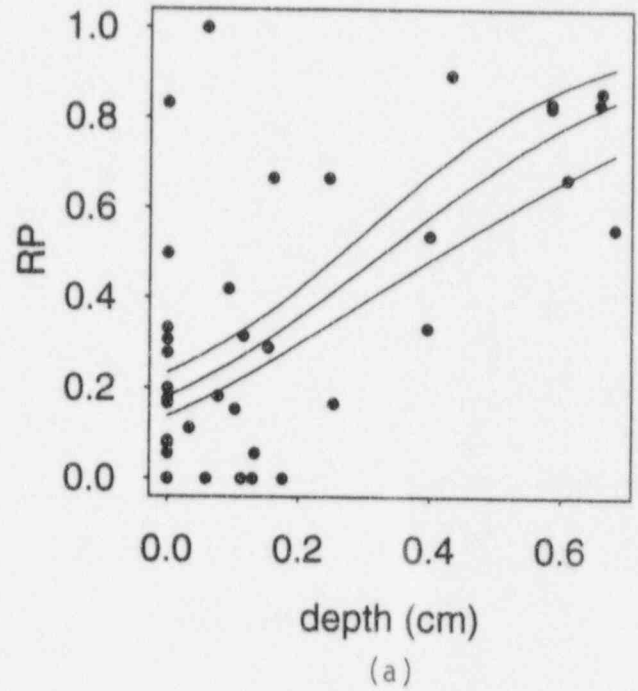
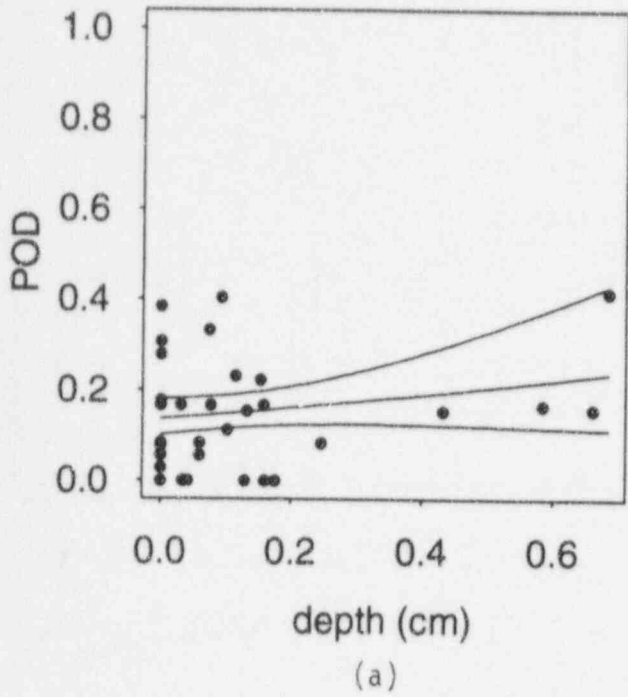


Figure 5.21 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with TFC, Far-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length

Figure 5.22 Logistic Curve Fit to RP Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length

5.0 Relationship Between Detection and Crack Size

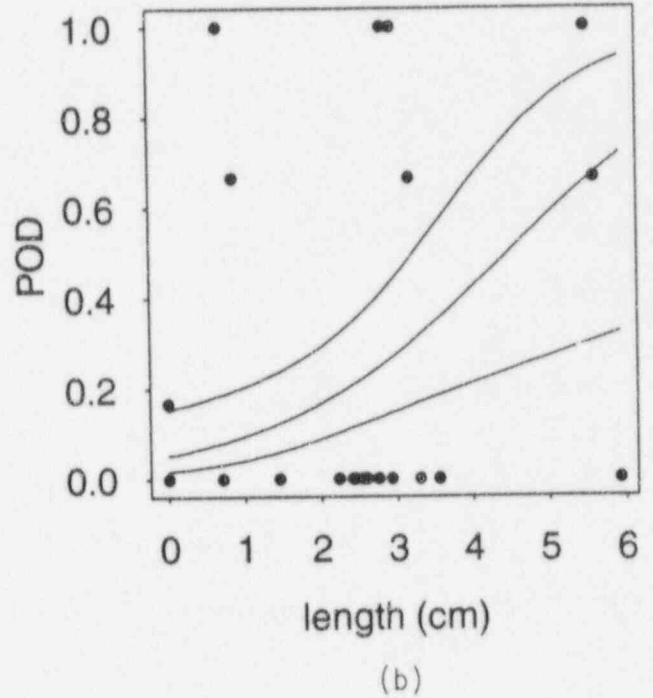
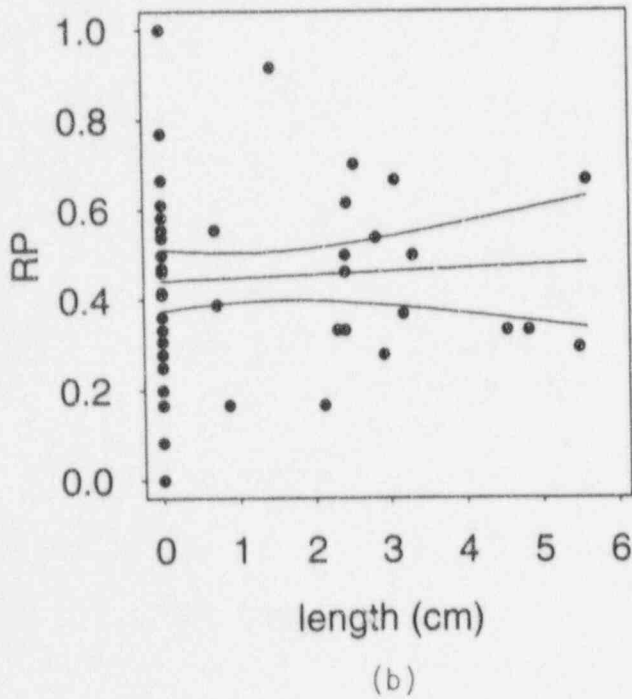
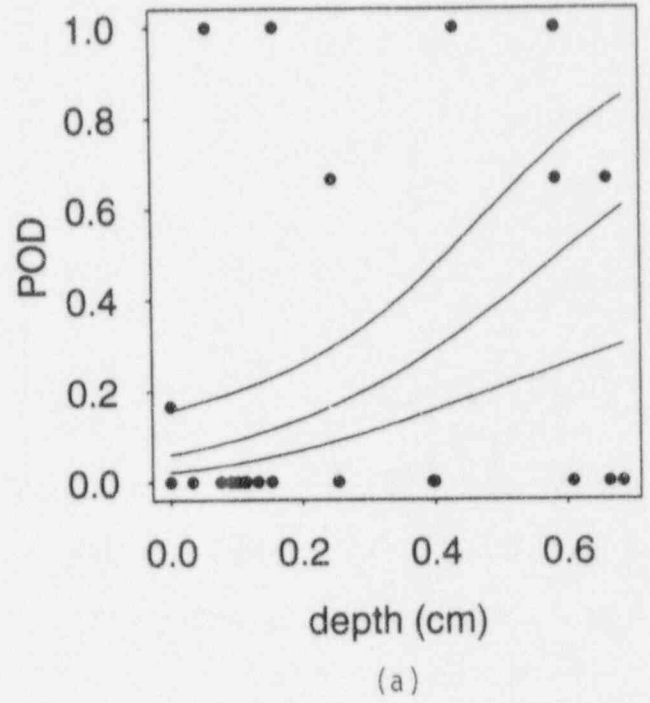
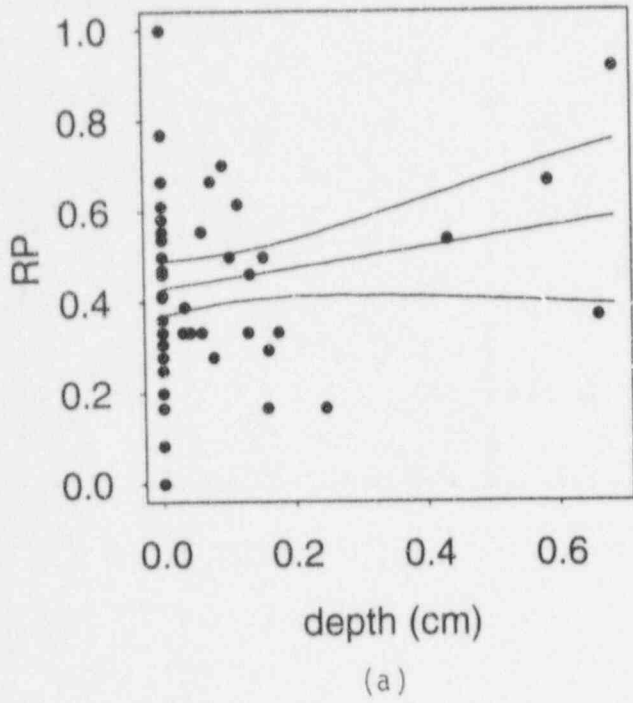
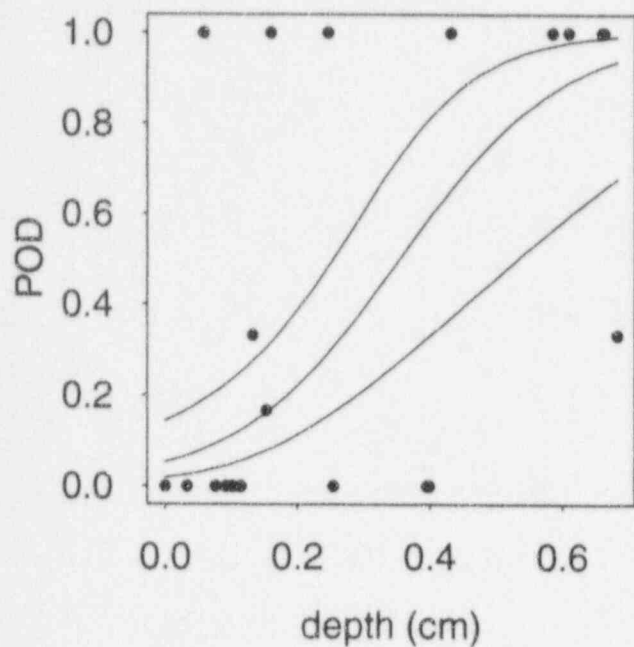
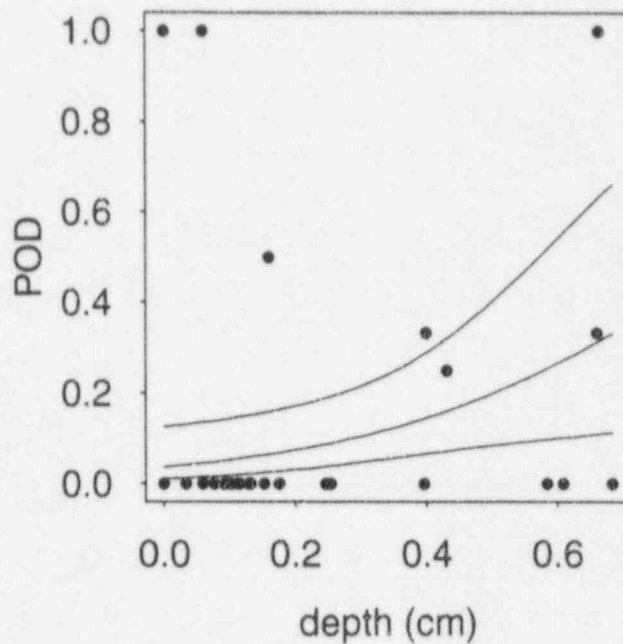


Figure 5.23 Logistic Curve Fit to RP Data with 95% Bounds (Wrought SS with TFC, Far-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length

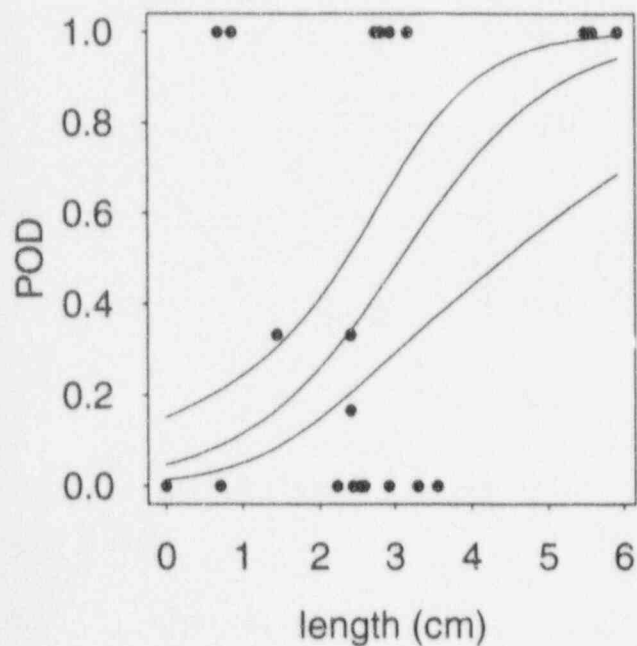
Figure 5.24 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, Team #1) a) POD vs. Depth b) POD vs. Length



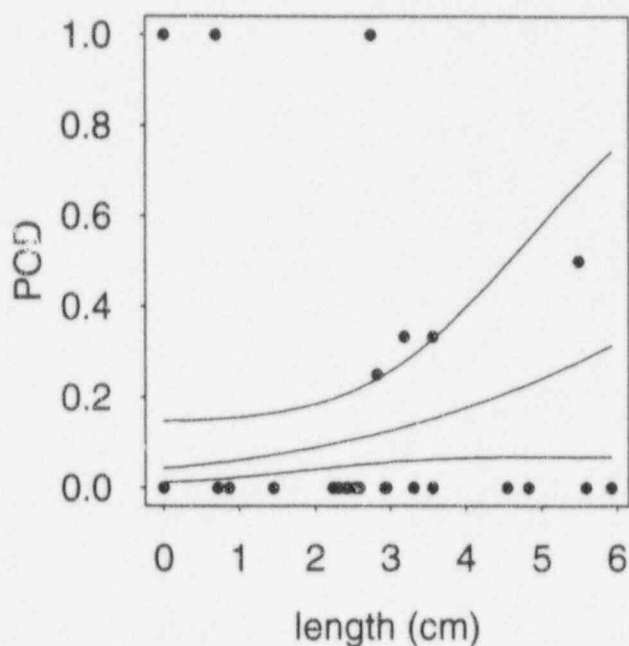
(a)



(a)



(b)

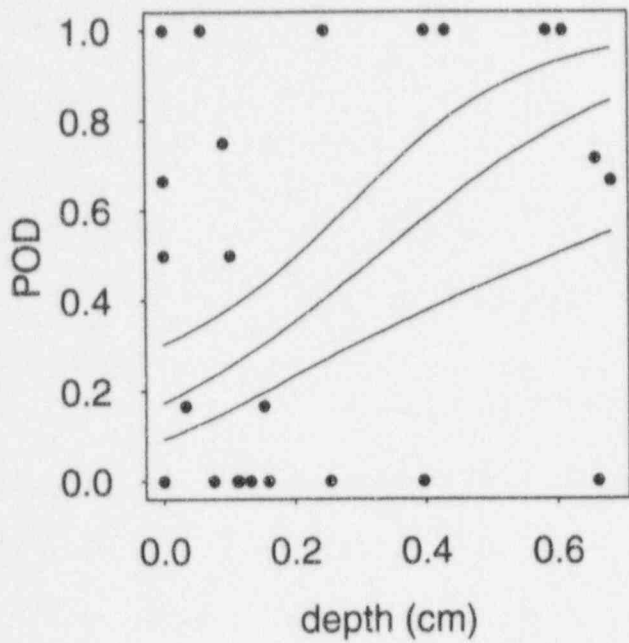


(b)

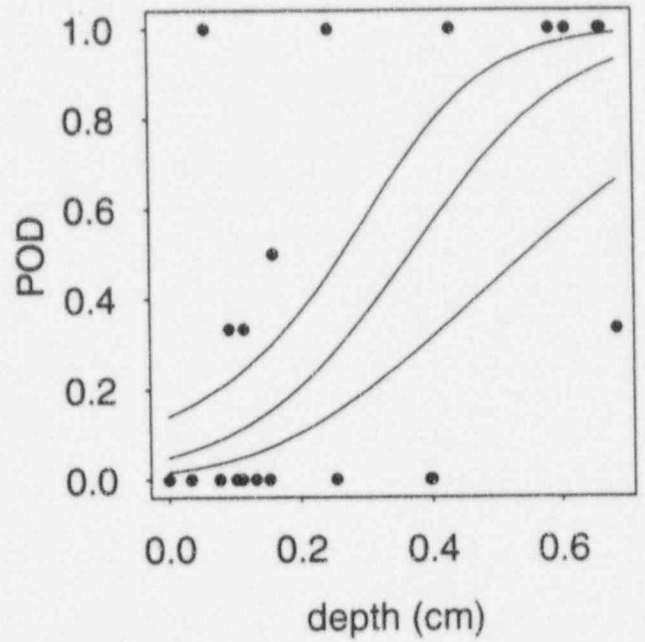
Figure 5.25 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, Team #2) a) POD vs. Depth b) POD vs. Length

Figure 5.26 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, Team #3) a) POD vs. Depth b) POD vs. Length

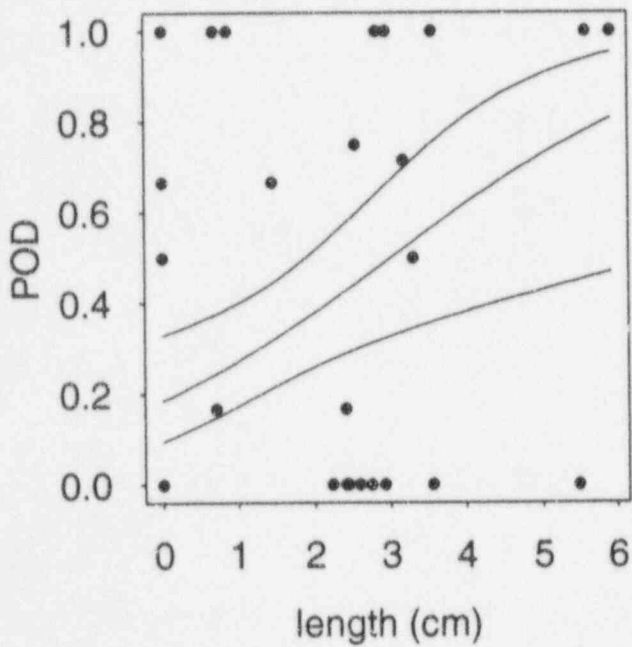
5.0 Relationship Between Detection and Crack Size



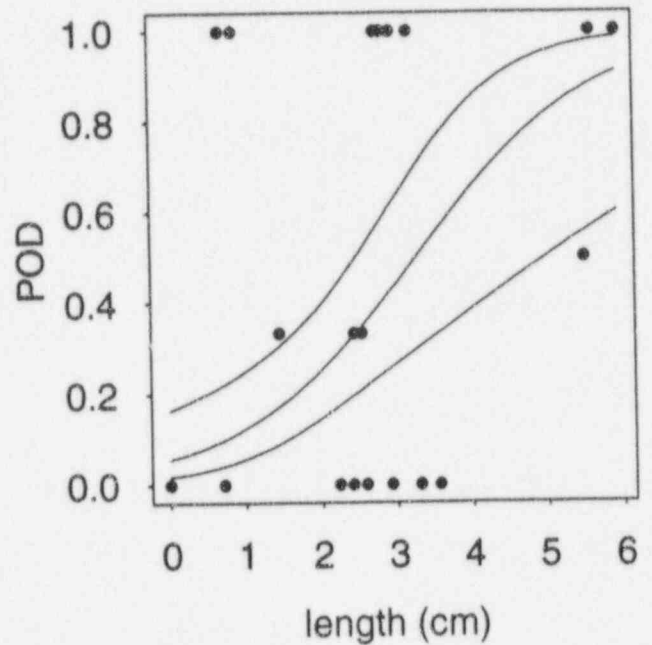
(a)



(a)



(b)



(b)

Figure 5.27 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, Team #4) a) POD vs. Depth b) POD vs. Length

Figure 5.28 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, Team #5) a) POD vs. Depth b) POD vs. Length

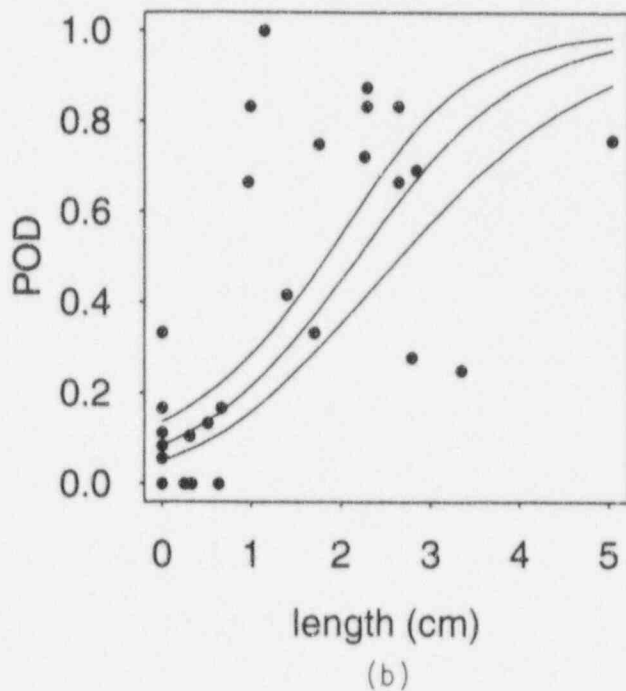
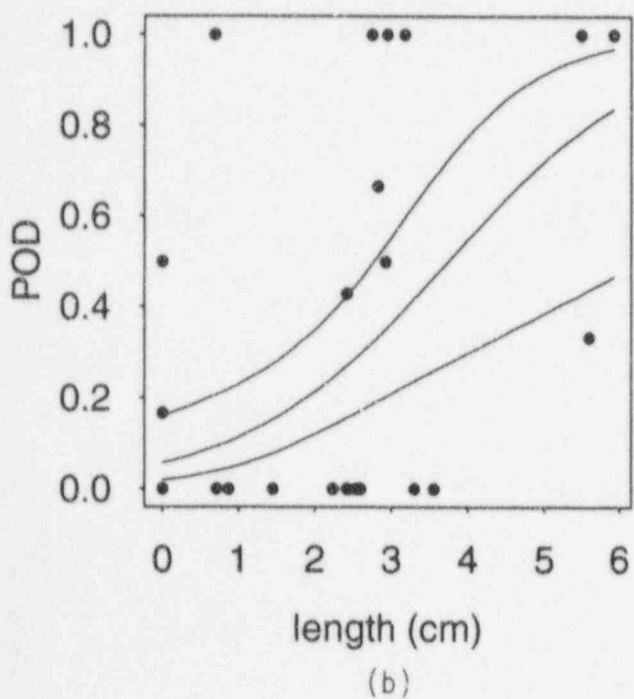
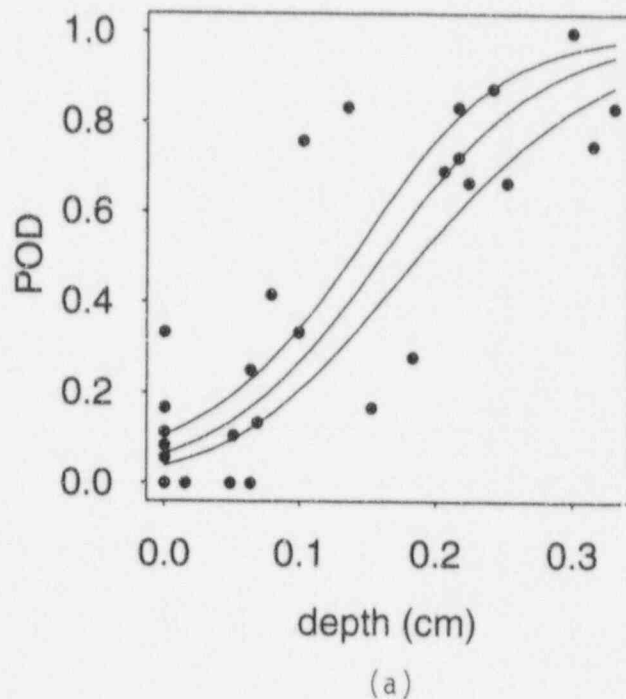
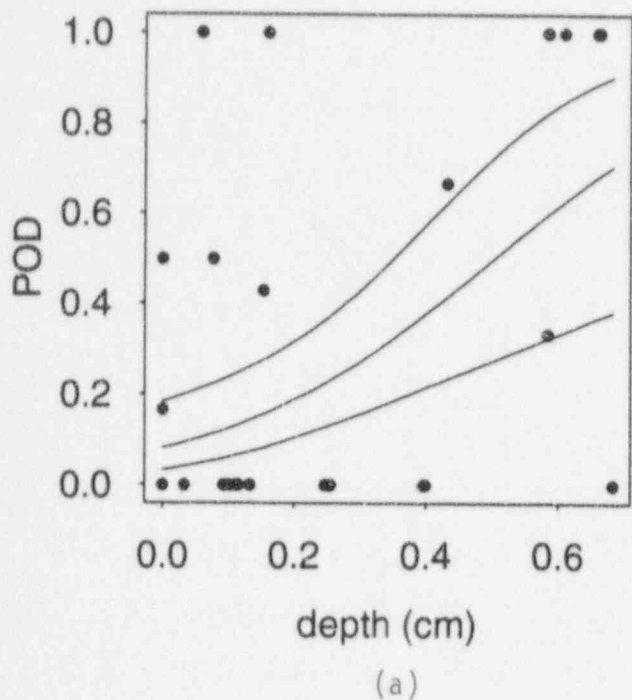


Figure 5.29 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with TFC, Near-Side Inspections, Team #6) a) POD vs. Depth b) POD vs. Length

Figure 5.30 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with IGSCC, Near-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length

5.0 Relationship Between Detection and Crack Size

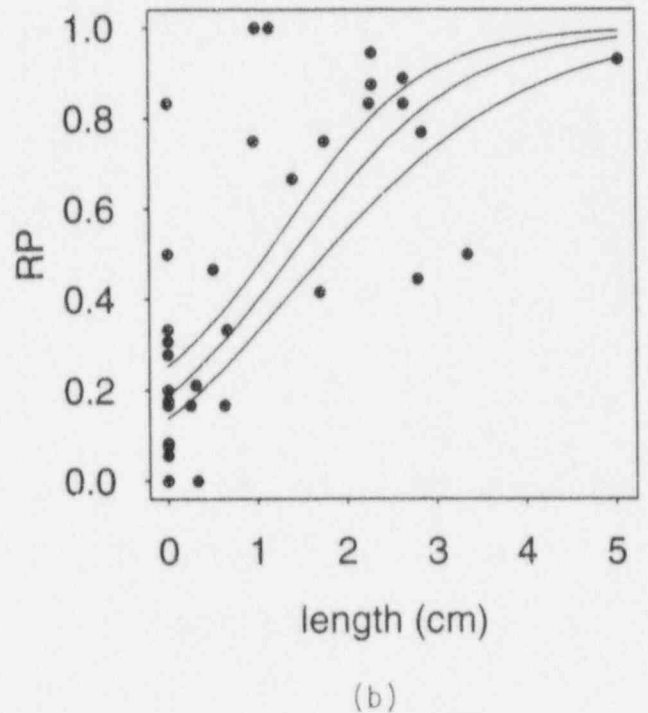
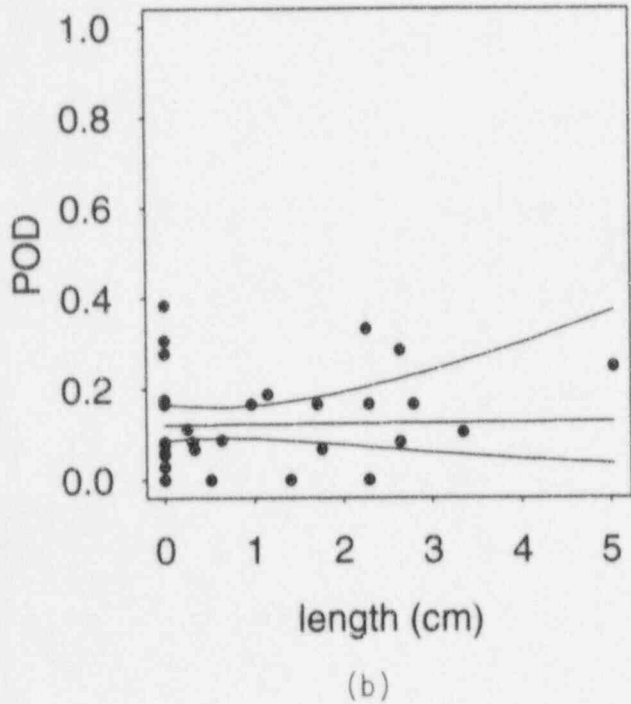
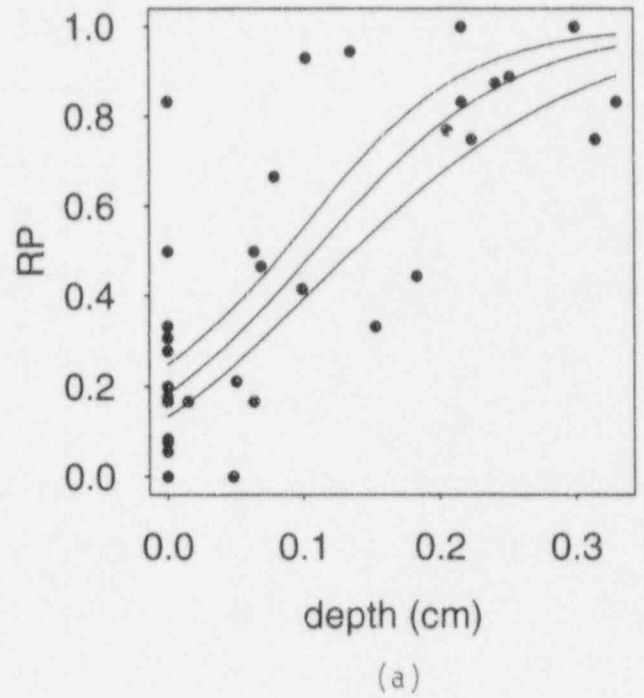
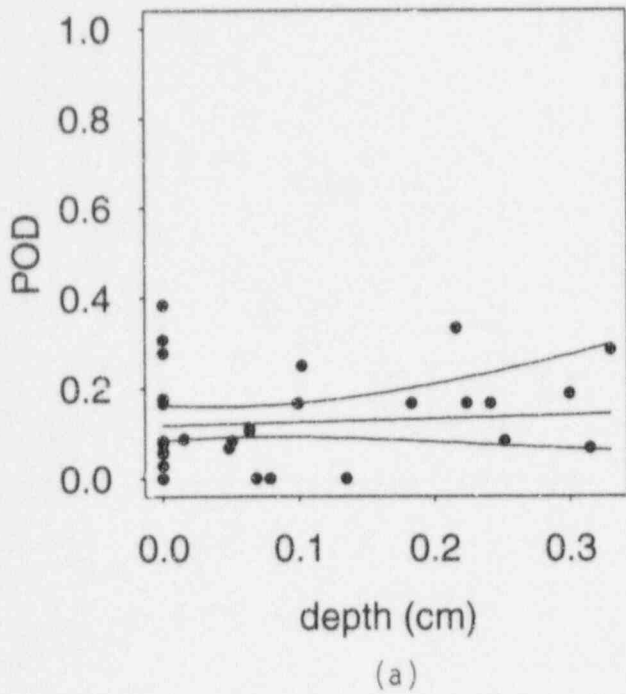


Figure 5.31 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with IGSCC, Far-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length

Figure 5.32 Logistic Curve Fit to RP Data with 95% Bounds (Wrought SS with IGSCC, Near-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length

5.0 Relationship Between Detection and Crack Size

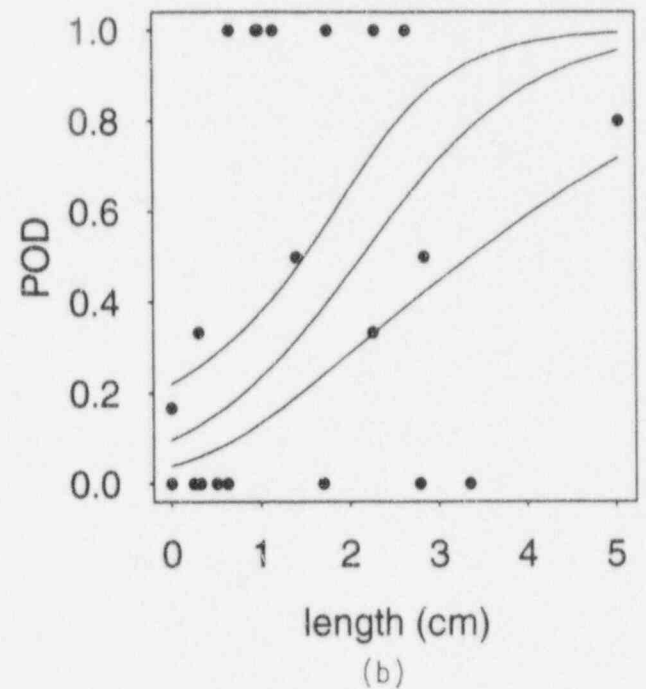
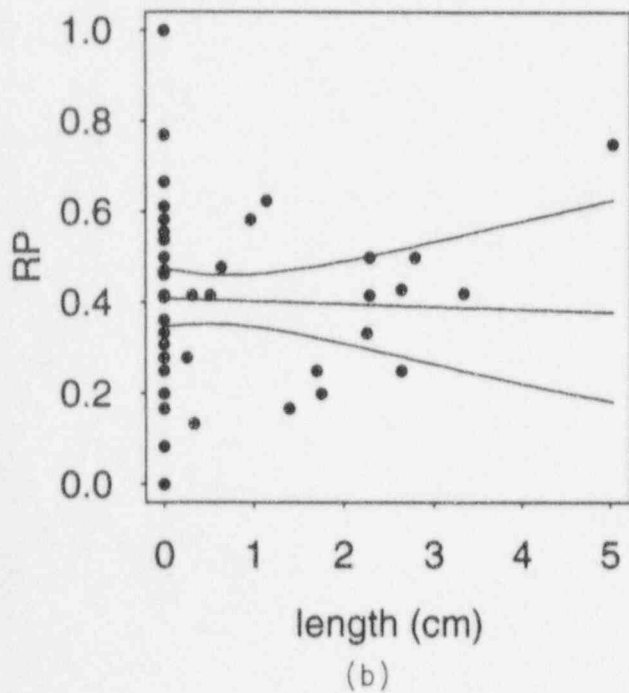
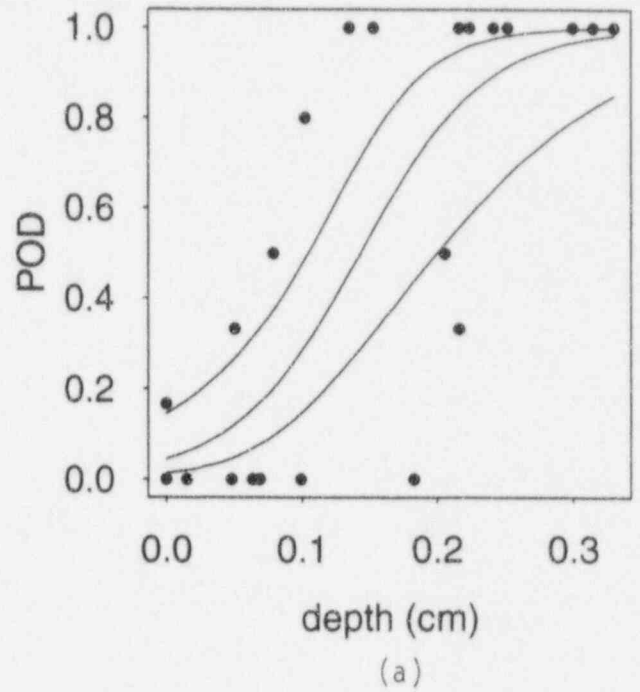
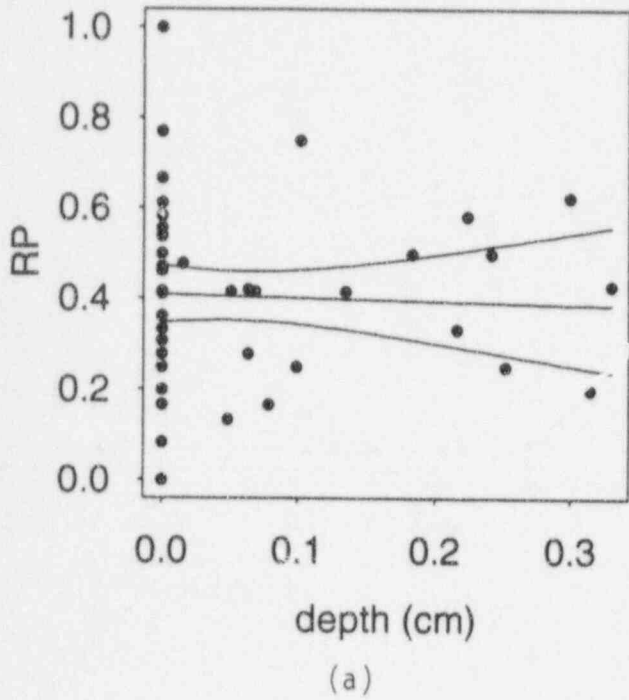


Figure 5.33 Logistic Curve Fit to RP Data with 95% Bounds (Wrought SS with IGSCC, Far-Side Inspections, All Teams) a) POD vs. Depth b) POD vs. Length

Figure 5.34 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with IGSCC, Near-Side Inspections, Team #1) a) POD vs. Depth b) POD vs. Length

5.0 Relationship Between Detection and Crack Size

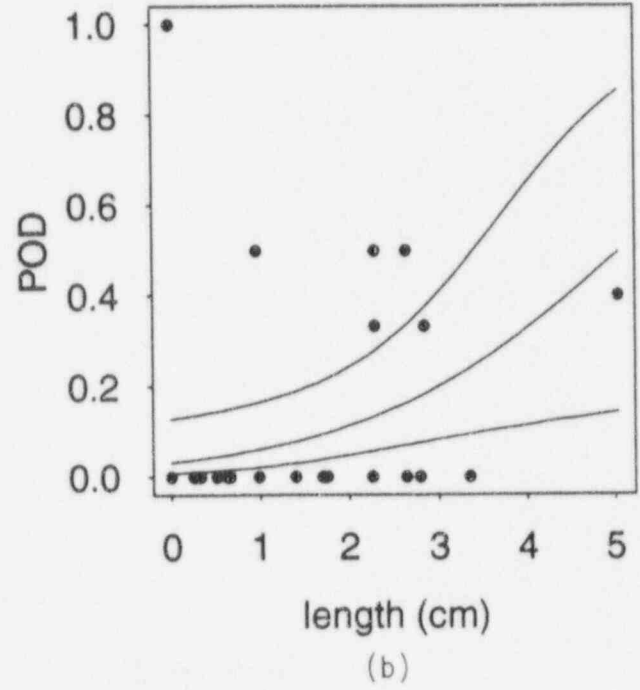
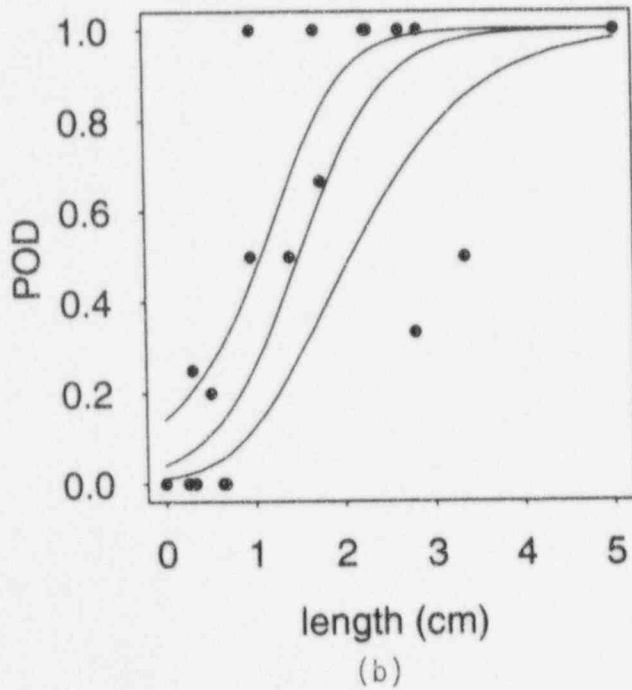
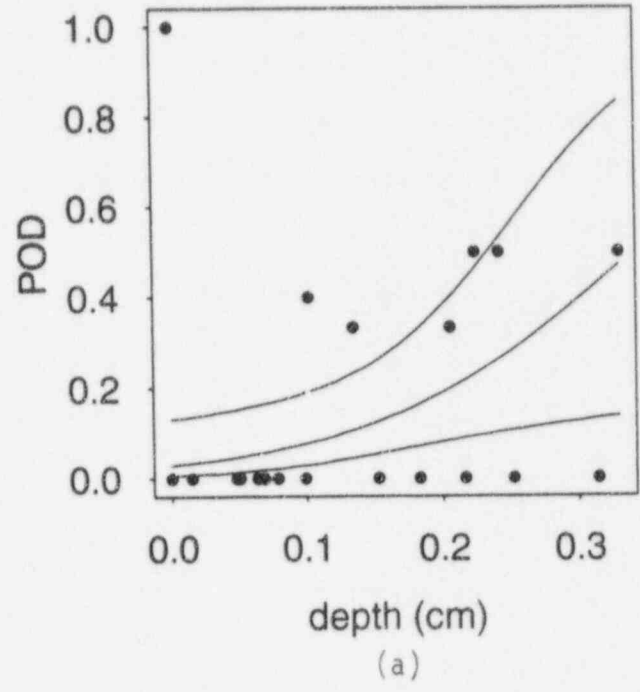
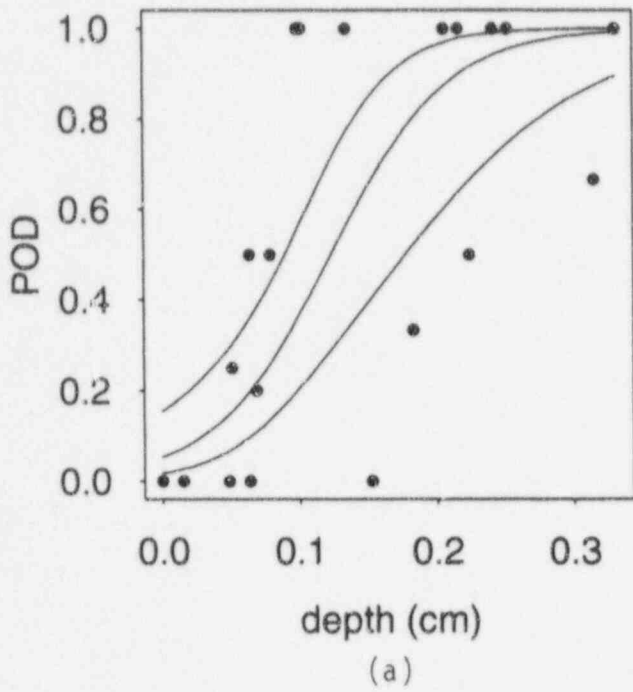


Figure 5.35 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with IGSCC, Near-Side Inspections, Team #2) a) POD vs. Depth b) POD vs. Length

Figure 5.36 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with IGSCC, Near-Side Inspections, Team #3) a) POD vs. Depth b) POD vs. Length

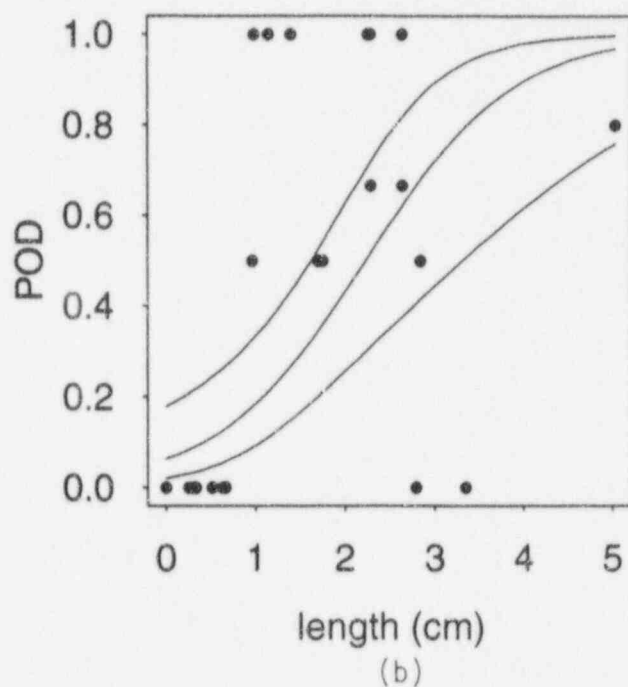
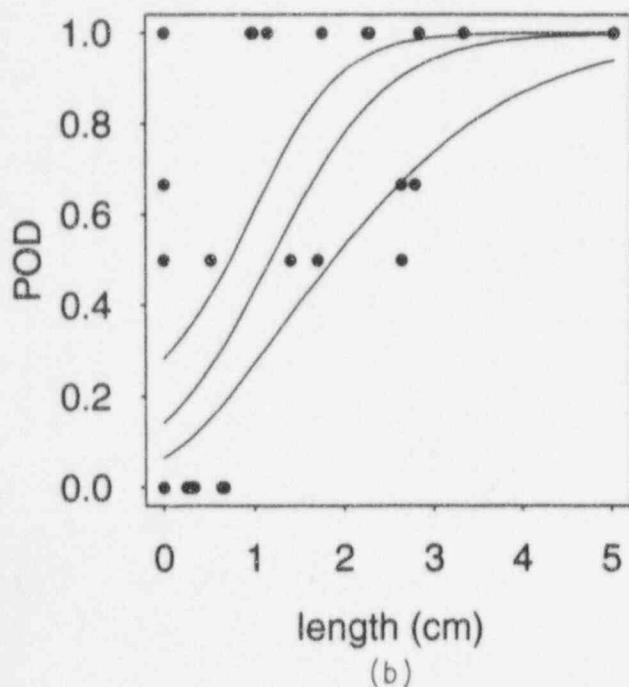
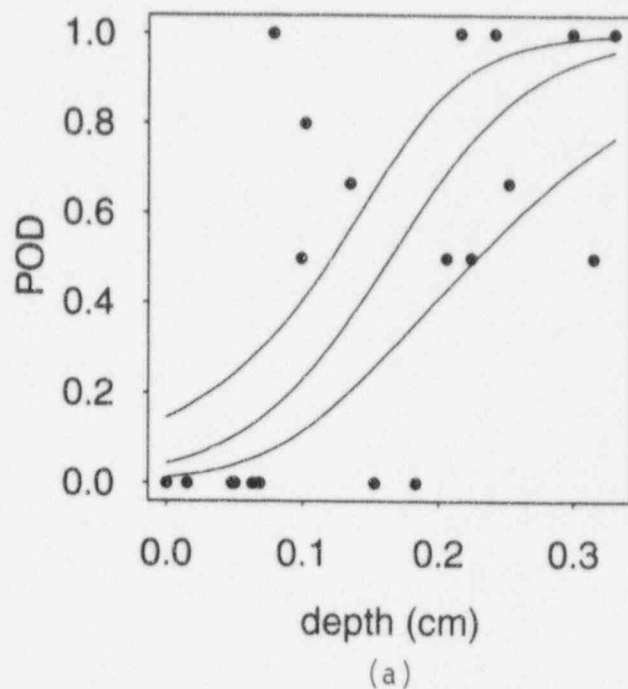
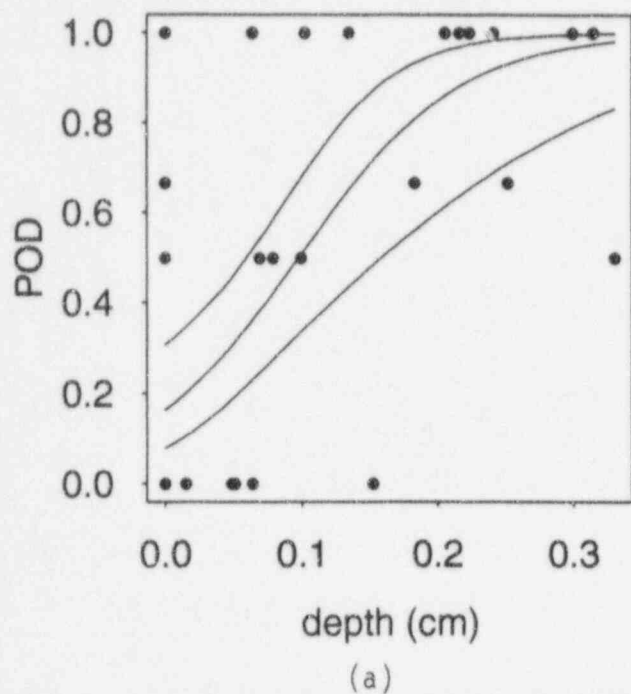
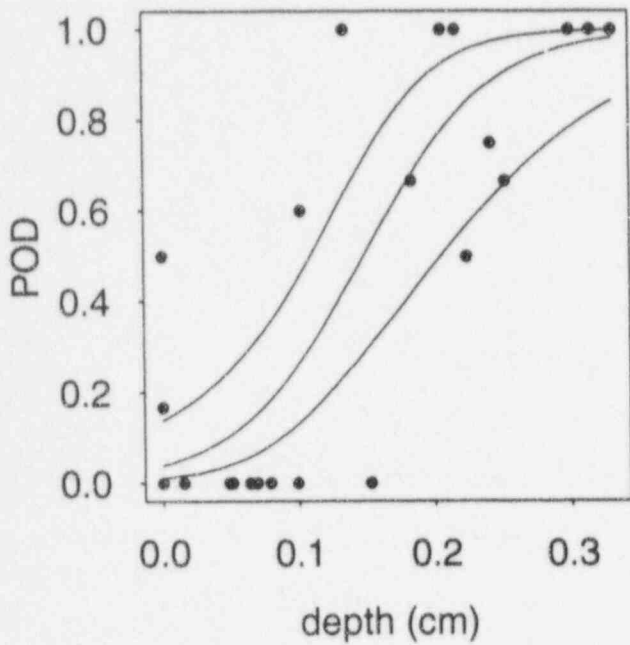


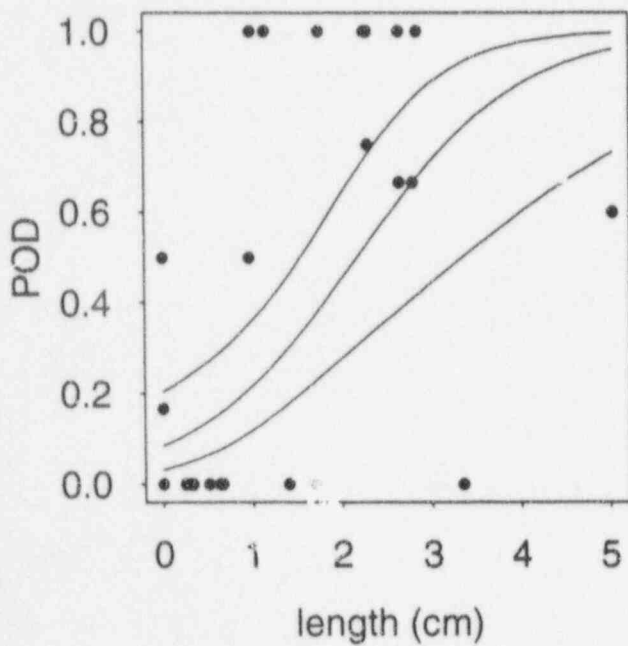
Figure 5.37 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with IGSCC, Near-Side Inspections, Team #4) a) POD vs. Depth b) POD vs. Length

Figure 5.38 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with IGSCC, Near-Side Inspections, Team #5) a) POD vs. Depth b) POD vs. Length

5.0 Relationship Between Detection and Crack Size



(a)



(b)

Figure 5.39 Logistic Curve Fit to POD Data with 95% Bounds (Wrought SS with IGSCC, Near-Side Inspections, Team #6) a) POD vs. Depth b) POD vs. Length

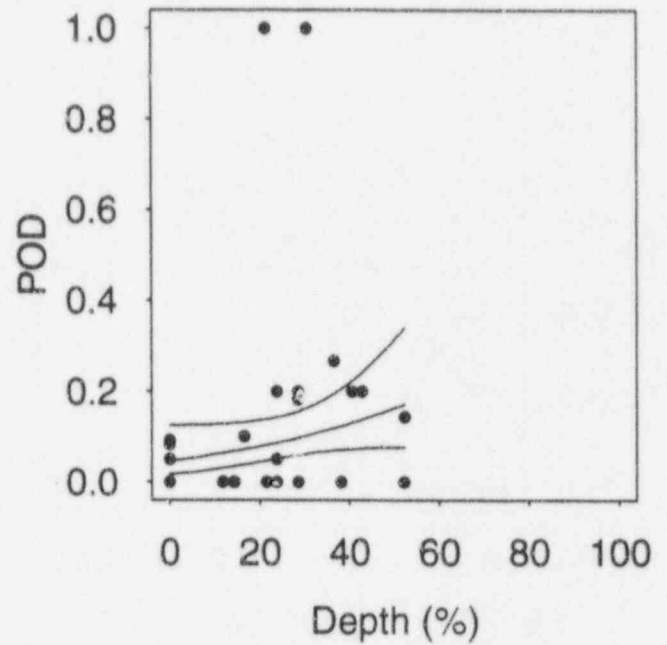


Figure 5.40 Logistic Curve Fit to POD vs. % Through-Wall Depth Data with 95% Confidence Bounds (Cast Stainless Steel, Near-Side Inspection, All Teams)

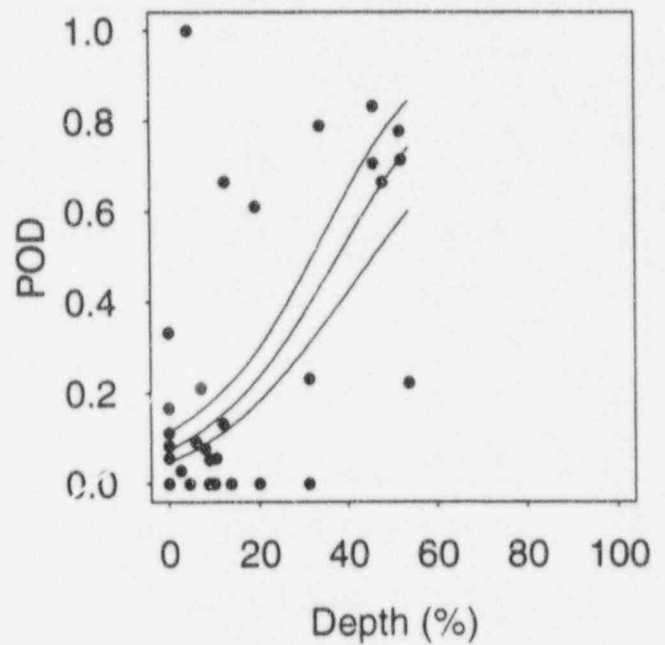


Figure 5.41 Logistic Curve Fit to POD vs. % Through-Wall Depth Data with 95% Confidence Bounds (Wrought Stainless Steel with TFC, Near-Side Inspection, All Teams)

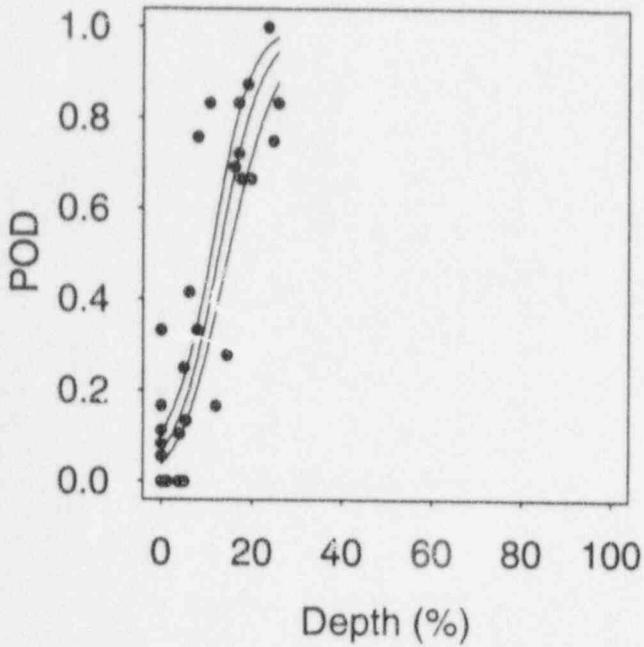


Figure 5.42 Logistic Curve Fit to POD vs. % Through-Wall Depth Data With 95% Confidence Bounds (Wrought Stainless Steel with IGSCC, Near-Side Inspection, All Teams)

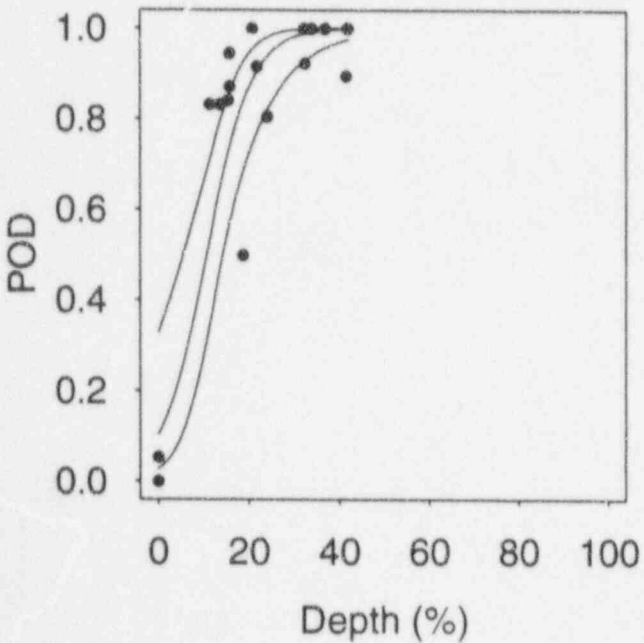


Figure 5.43 Logistic Curve Fit to POD vs. % Through-Wall Depth Data With 95% Confidence Bounds (Clad Ferritic, Combined Near- and Far-Side Inspection, All Teams)

6.0 Relationships Between Detection Performance and Decision Criteria; ROC Curves

In this section, relative operating characteristic (ROC) curves are used to describe the relationship between detection performance and decision criteria. Although, the PIRR was not designed to employ ROC analysis, it is being utilized to see if it will provide some additional insights into inspection performance. The ROC curve attempts to describe how (FCP,POD) performance changes as the detection decision threshold is altered by the inspector. Because the inspector employed a two-point decision scale to report results, C means the indication is definitely a crack and N¹ means the indication is not a crack, only two points were available to fit a curve in the ROC space besides (0,0) and (1,1). These two points are in fact represented by the ordered pairs (FCP,POD) and (FRP, RP). Therefore, we employed a simple family of ROC curves with only one unknown parameter as given by the equation:

$$POD = \text{logit}(A + \text{logit}^{-1}(FCP)) \quad (6.1)$$

where A is the unknown parameter to be determined by the curve-fitting procedure. This formula results in an ROC curve that is always symmetric about the $POD = 1 - FCP$ diagonal line.

In the following sections, ROC curves are presented for each of the four important material/crack type combinations employed in the experiment. All ROC curves are surrounded by 95% confidence bounds, and the data points used to fit the curves are also illustrated in the plots as "C" or "N". The ROC fits are made as a function of crack size groupings and these are the groupings defined in Section 2 and plotted in Figures 2.5, 2.6 and 2.7.

¹Some teams always reported any signal that exceeded their recording level. Other teams made decisions based on geometry and signal properties as to whether to record a given signal. Thus, some data may not reflect the true recording probability of the procedure.

6.1 ROC Curves for TFC in Cast Stainless Steel

Tables 6.1 and 6.2, present the detection data for the ROC curve fits in cast stainless steel. Separate ROC curves have been calculated for different access conditions and different sizes of cracks. All cracks employed in cast stainless steel were thermal fatigue cracks.

The "C" and "N" values for a given crack size have quite similar values which means that the inspectors, if they recorded an indication, would most likely classify it as a crack.

The detection data presented in the two tables were fit to ROC curves and the results are displayed in Figures 6.1 through 6.8. From the figures, it is evident that:

1. Far-side performance was poor (not much better than guessing) for all sizes of cracks.
2. The inspections seemed to be better than guessing for near-side inspections of medium and large cracks.
3. The plotted points associated with C and N decision criteria were not optimally located to allow very accurate estimates of the ROC curves, particularly for the near-side ROC curves.

6.2 ROC Curves for TFC in Clad Ferritic

Table 6.3 presents detection statistics for clad ferritic material. The table employs inspection results from all teams, and for both near- and far-side access. This material was easiest to inspect, as indicated by the computed ROC curves in Figures 6.9 through 6.12.

In fact, the results associated with the two largest size categories (sizes 3 and 4) were so good (values are about 1.0) that it was difficult to fit ROC curves to them. These results trace out a curve that is nearly a "step function" at the origin. Because the family of curves (equation 6.1) we have chosen to model ROC is continuous, there are difficulties in finding a member of this family that fits this data (a step function). The effects of these difficulties are displayed in the

6.0 Relationships

Table 6.1. Detection Results in Cast Stainless Steel Material for Near-Side Access

Decision Threshold	Crack Size				
	Blank	Size 1	Size 2	Size 3	Size 4
C (POD)	0.04	0.03	0.20	0.24	0.10
N (RP)	0.08	0.10	0.36	0.27	0.16
No. of Inspections	71	90	25	37	31

Table 6.2. Detection Results in Cast Stainless Steel Material for Far-Side Access

Decision Threshold	Crack Size				
	Blank	Size 1	Size 2	Size 3	Size 4
C (POD)	0.13	0.13	0.36	0.0	0.20
N (RP)	0.18	0.19	0.36	0.0	0.30
No. of Inspections	71	16	14	13	10

95%-confidence bounds, which are larger than they should be.

6.3 ROC Curves for TFC in Wrought Stainless Steel

The ROC curves for thermal fatigue cracks in wrought stainless steel show a large difference between near-side and far-side access conditions, as shown in Tables 6.4 and 6.5 respectively. Note that the detection statistics in the far-side table are comparable to the false call statistics, while this is definitely not the case for near-side access.

The ROC curves fit to the data in Figures 6.13 through 6.18 show that far-side inspections are little better than results produced by guessing, even for deep cracks. In the near-side inspections, we see an improvement in ROC curve shape as one moves from shallow cracks to deep cracks (compare Figures 6.13 through 6.15).

6.4 ROC Curves for IGSCC in Wrought Stainless Steel

No "deep" IGSC cracks were present in the test specimens, so the complement of crack sizes is not as large as that exhibited for the thermal fatigue cracks. As with the thermal fatigue cracks in wrought SS, a dramatic difference was apparent between near-side and far-side access as shown in Figures 6.19 through 6.22. Also as with thermal fatigue cracks in wrought stainless steel, one can conclude that far-side inspections are ineffective.

However, the near-side ROC results seem to be much better than the corresponding results on thermal fatigue cracks (compare Figure 6.20 to 6.14). This effect is also evident in the POD data presented in Table 6.6. Using a "C" decision threshold, the POD associated with Size 2 IGSC cracks is 68%, while the value associated with Size 2 thermal fatigue cracks is only 27%. Finally, Table 6.7 presents the detection results for far-side inspections, which lead to the same conclusions obtained from the thermal fatigue crack tables; detection performance from the far-side is no better than results produced by guessing.

Table 6.3. Detection Results in Clad Ferritic Material

Decision Threshold	Crack Size				
	Blank	Size 1	Size 2	Size 3	Size 4
C (POD)	0.01	0.88	0.87	0.97	0.97
N (RP)	0.07	0.91	0.92	1.00	0.98
No. of Inspections	72	80	90	31	66

Table 6.4. Detection Results for Near-Side Inspection of Thermal Fatigue Cracks in Wrought SS

Decision Threshold	Crack Size			
	Blank	Size 1	Size 2	Size 3
C (POD)	0.03	0.17	0.27	0.59
N (RP)	0.16	0.29	0.35	0.74
No. of Inspections	328	118	94	128

Table 6.6. Detection Results for Near-Side Inspection of IGSCC in Wrought SS Material

Decision Threshold	Crack Size		
	Blank	Size 1	Size 2
C (POD)	0.03	0.38	0.68
N (RP)	0.16	0.54	0.77
No. of Inspections	328	141	143

Table 6.5. Detection Results for Far-Side Inspection of Thermal Fatigue Cracks in Wrought SS

Decision Threshold	Crack Size			
	Blank	Size 1	Size 2	Size 3
C (POD)	0.13	0.18	0.10	0.21
N (RP)	0.43	0.51	0.35	0.59
No. of Inspections	328	152	78	56

Table 6.7. Detection Results for Far-Side Inspection of IGSCC in Wrought SS Material

Decision Threshold	Crack Size		
	Blank	Size 1	Size 2
C (POD)	0.13	0.08	0.16
N (RP)	0.43	0.34	0.43
No. of Inspections	328	145	80

6.0 Relationships

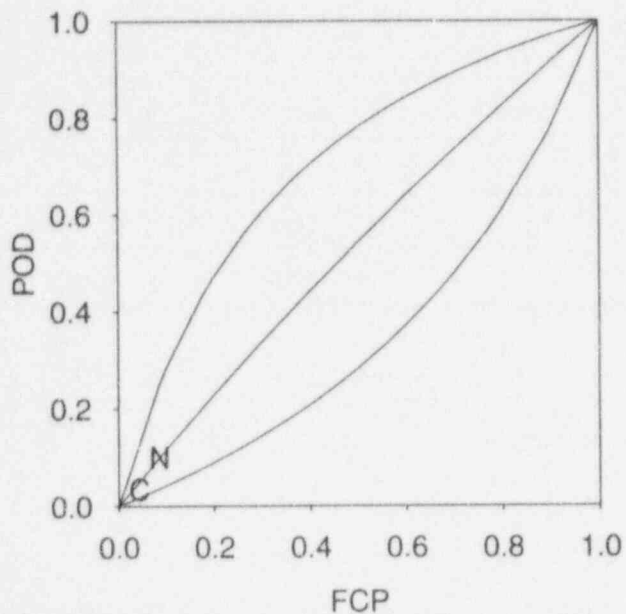


Figure 6.1 ROC Fit in Cast Stainless Steel (Near-Side Inspection, Size 1, Thermal Fatigue Cracks)

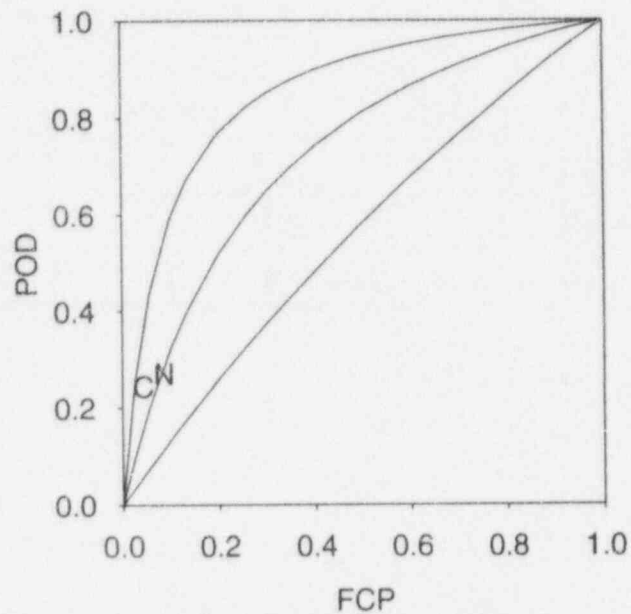


Figure 6.3 ROC Fit in Cast Stainless Steel (Near-Side Inspection, Size 3, Thermal Fatigue Cracks)

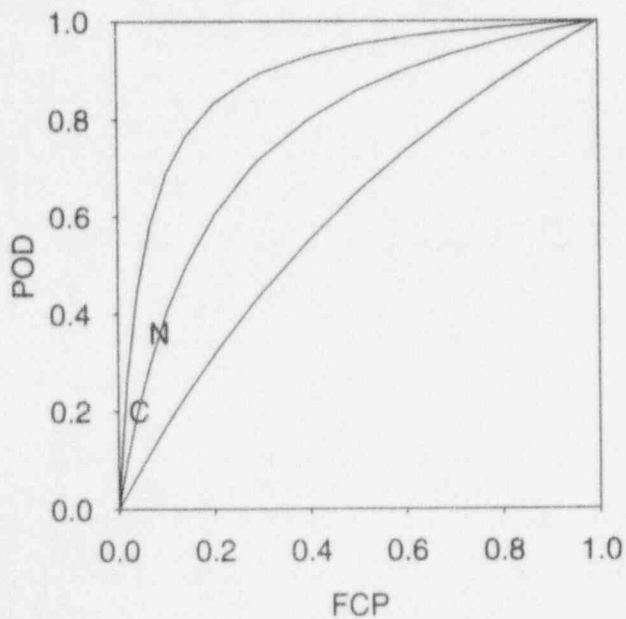


Figure 6.2 ROC Fit in Cast Stainless Steel (Near-Side Inspection, Size 2, Thermal Fatigue Cracks)

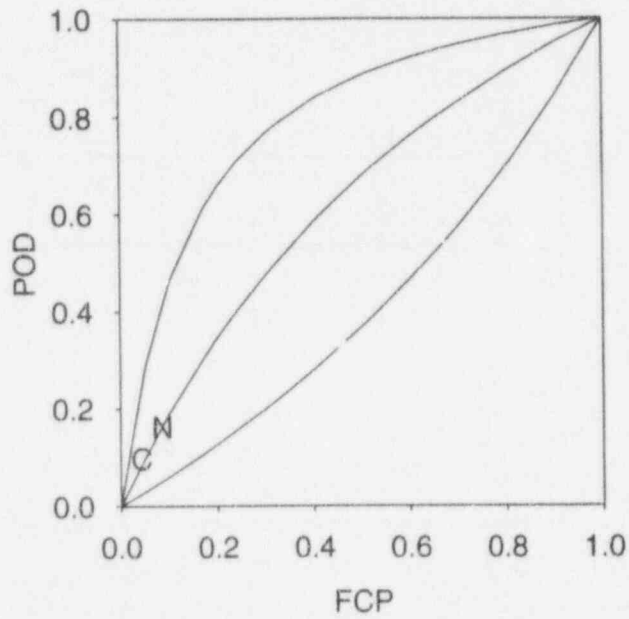


Figure 6.4 ROC Fit in Cast Stainless Steel (Near-Side Inspection, Size 4, Thermal Fatigue Cracks)

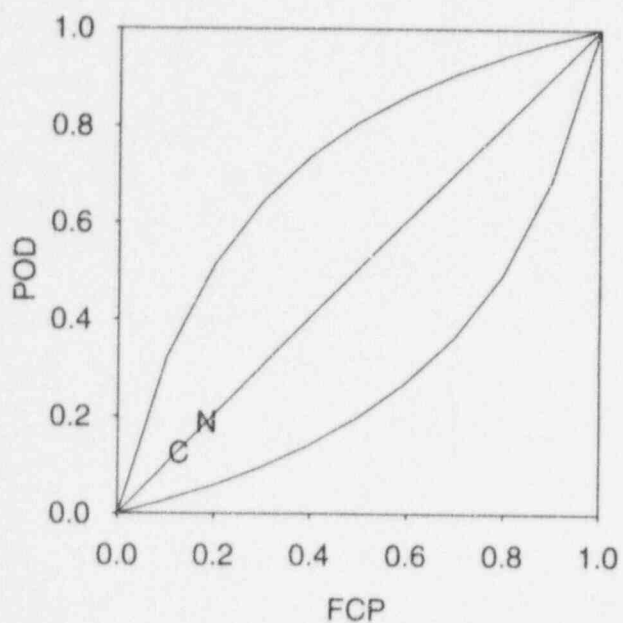


Figure 6.5 ROC Fit in Cast Stainless Steel (Far-Side Inspection, Size 1, Thermal Fatigue Cracks)

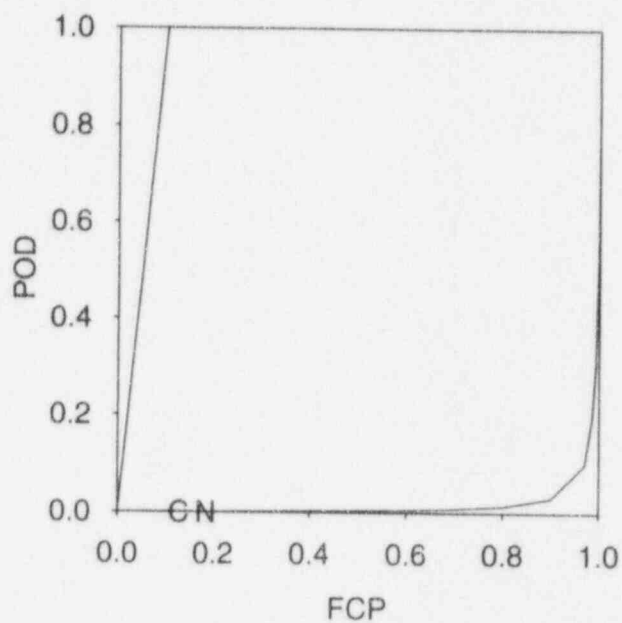


Figure 6.7 ROC Fit in Cast Stainless Steel (Far-Side Inspection, Size 3, Thermal Fatigue Cracks)

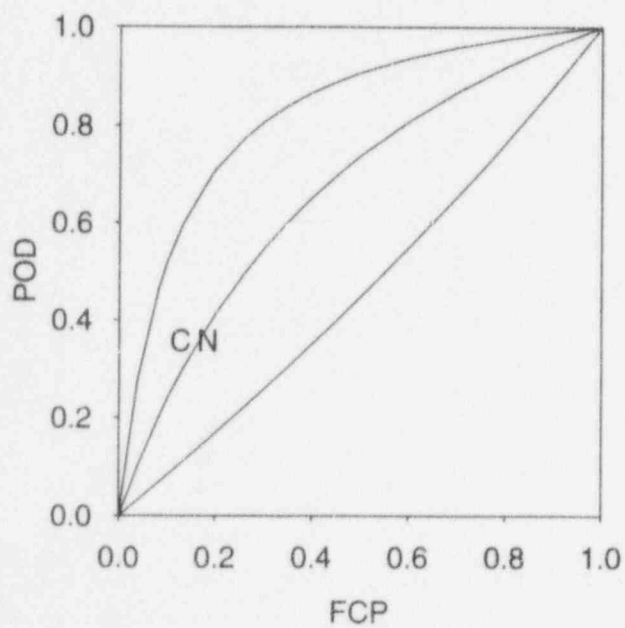


Figure 6.6 ROC Fit in Cast Stainless Steel (Far-Side Inspection, Size 2, Thermal Fatigue Cracks)

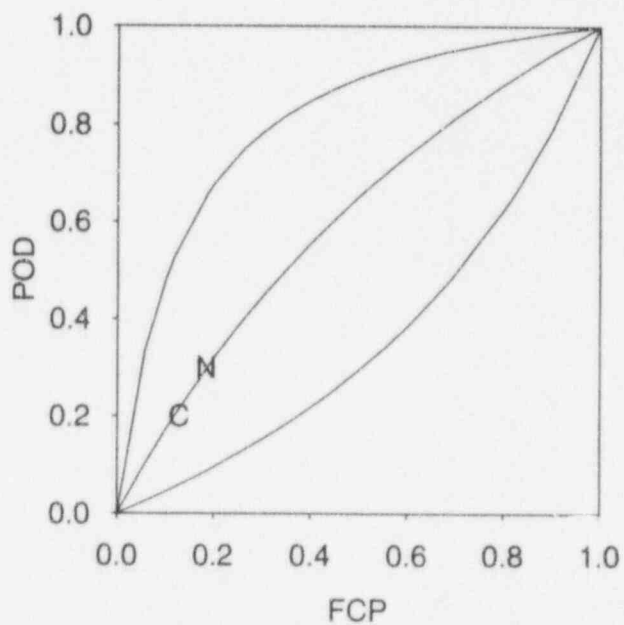


Figure 6.8 ROC Fit in Cast Stainless Steel (Far-Side Inspection, Size 4, Thermal Fatigue Cracks)

6.0 Relationships

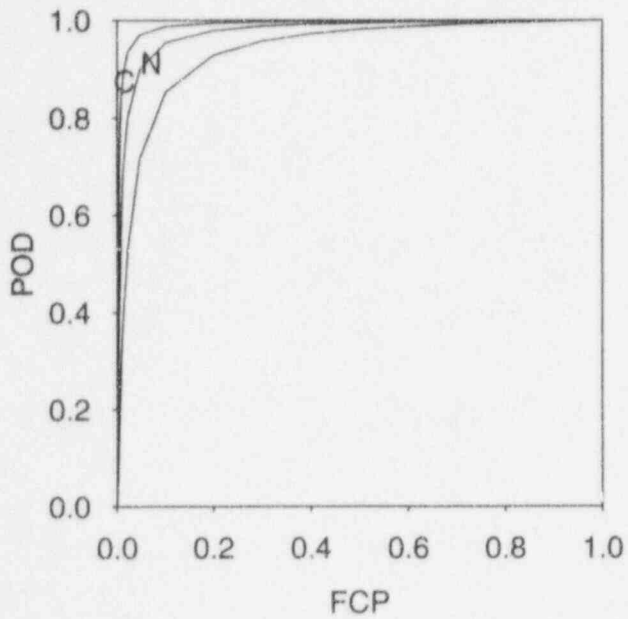


Figure 6.9 ROC Fit in Clad Ferritic Material (Combined Near- and Far-Side Inspection, Size 1, Thermal Fatigue Cracks)

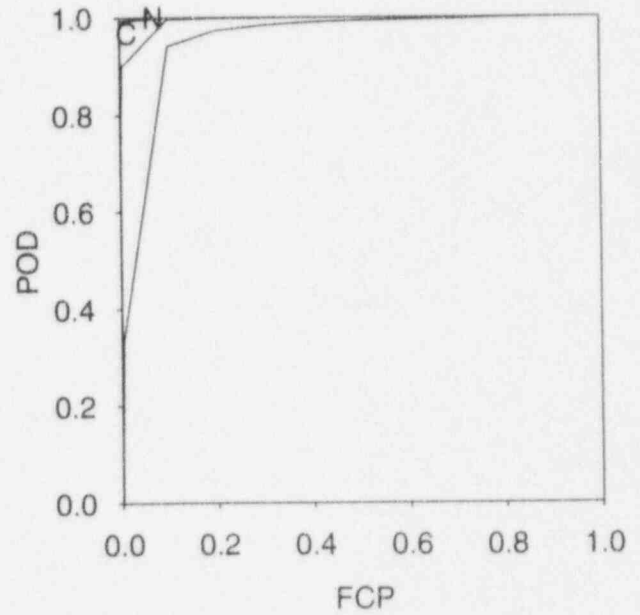


Figure 6.11 ROC Fit in Clad Ferritic Material (Combined Near- and Far-Side Inspection, Size 3, Thermal Fatigue Cracks)

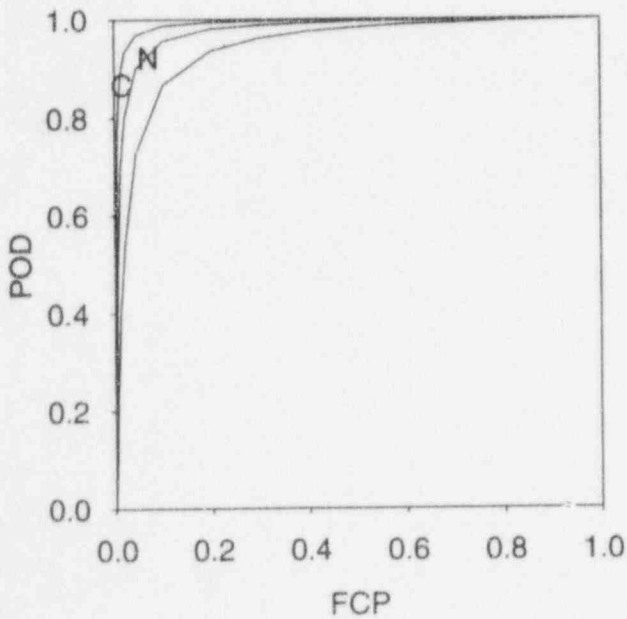


Figure 6.10 ROC Fit in Clad Ferritic Material (Combined Near- and Far-Side Inspection, Size 2, Thermal Fatigue Cracks)

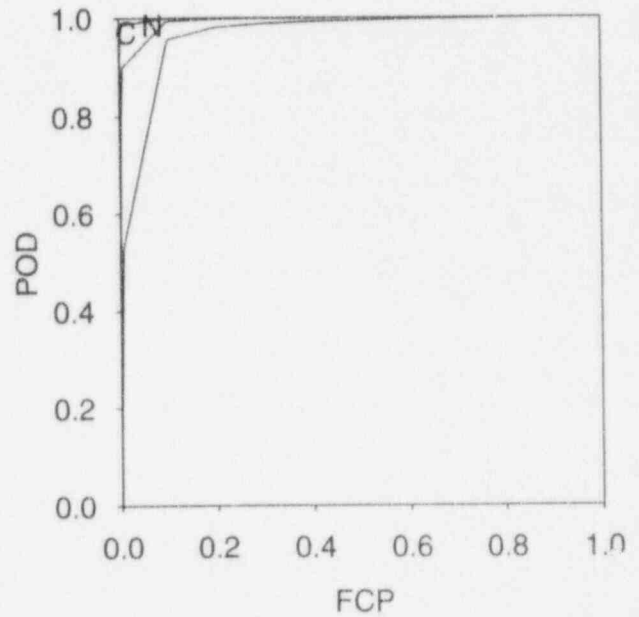


Figure 6.12 ROC Fit in Clad Ferritic Material (Combined Near- and Far-Side Inspection, Size 4, Thermal Fatigue Cracks)

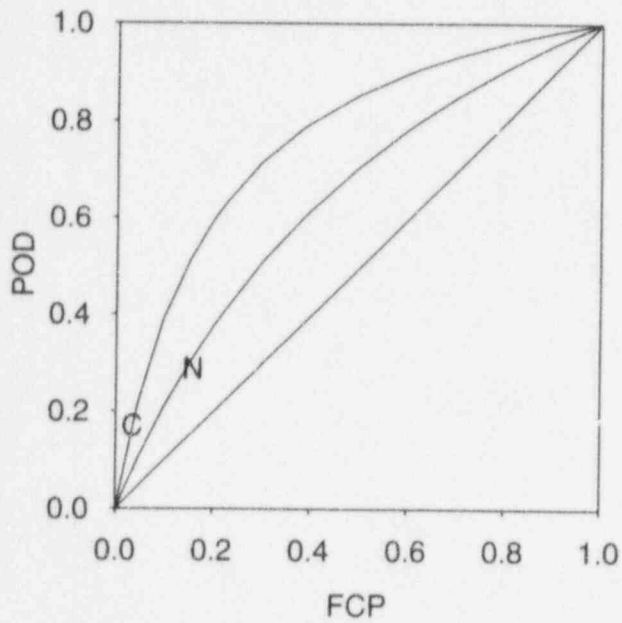


Figure 6.13 ROC Fit for Thermal Fatigue Cracks in Wrought Stainless Steel (Near-Side Inspection, Size 1 Cracks)

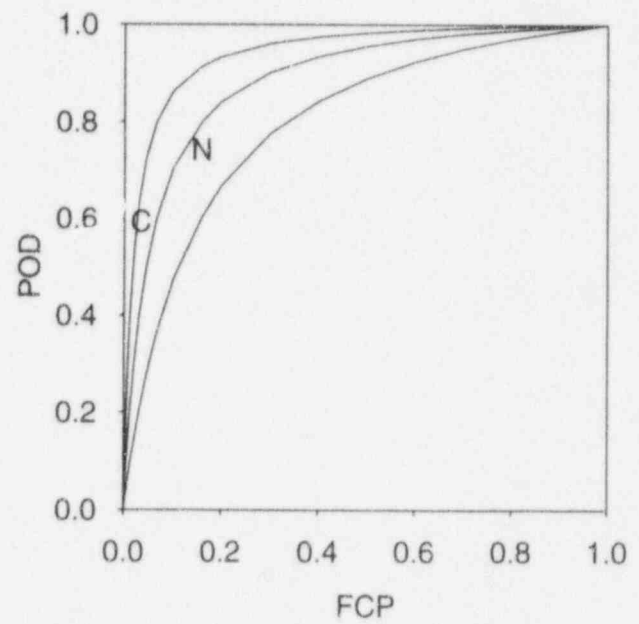


Figure 6.15 ROC Fit for Thermal Fatigue Cracks in Wrought Stainless Steel (Near-Side Inspection, Size 3 Cracks)

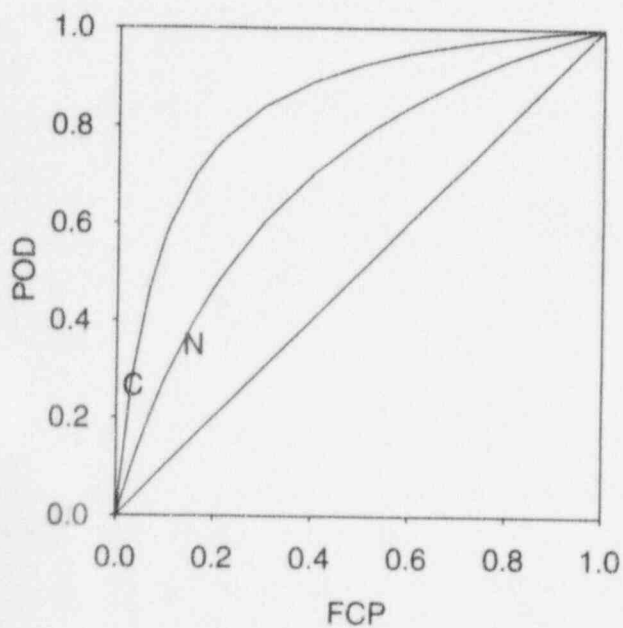


Figure 6.14 ROC Fit for Thermal Fatigue Cracks in Wrought Stainless Steel (Near-Side Inspection, Size 2 Cracks)

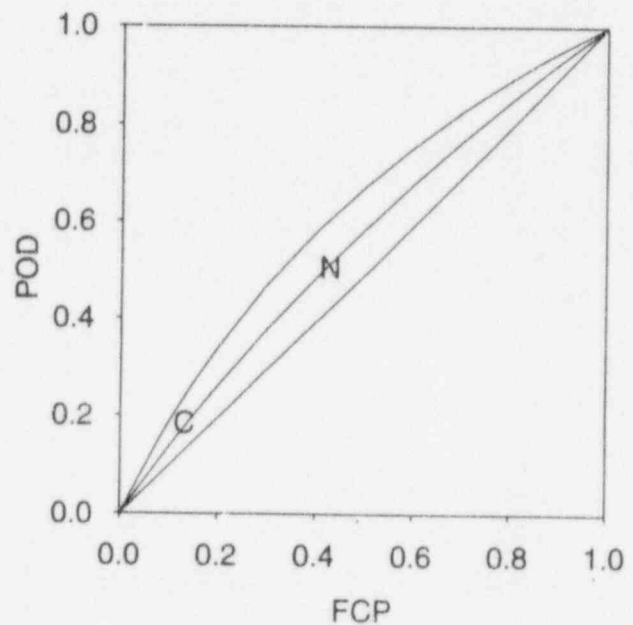


Figure 6.16 ROC Fit for Thermal Fatigue Cracks in Wrought Stainless Steel (Far-Side Inspection, Size 1 Cracks)

6.0 Relationships

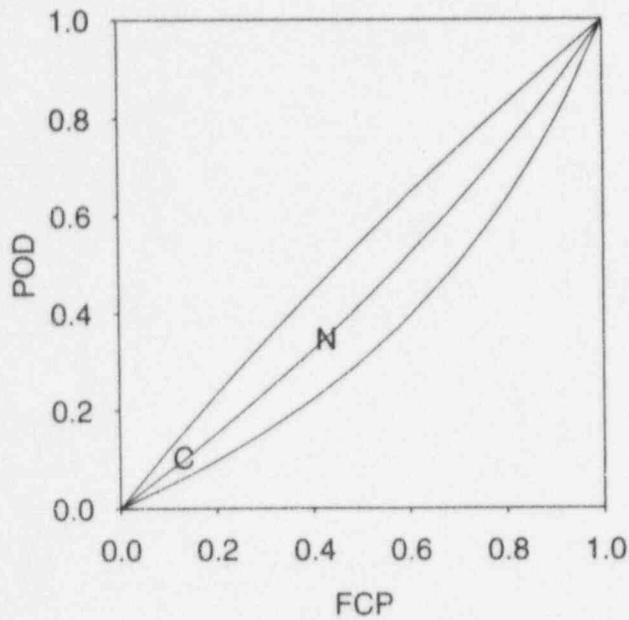


Figure 6.17 ROC Fit for Thermal Fatigue Cracks in Wrought Stainless Steel (Far-Side Inspection, Size 2 Cracks)

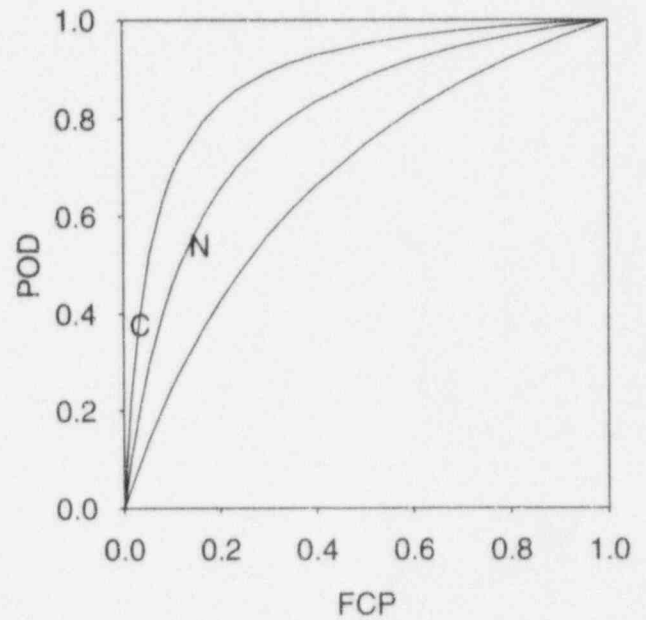


Figure 6.19 ROC Fit for IGSCC in Wrought Stainless Steel (Near-Side Inspection, Size 1 Cracks)

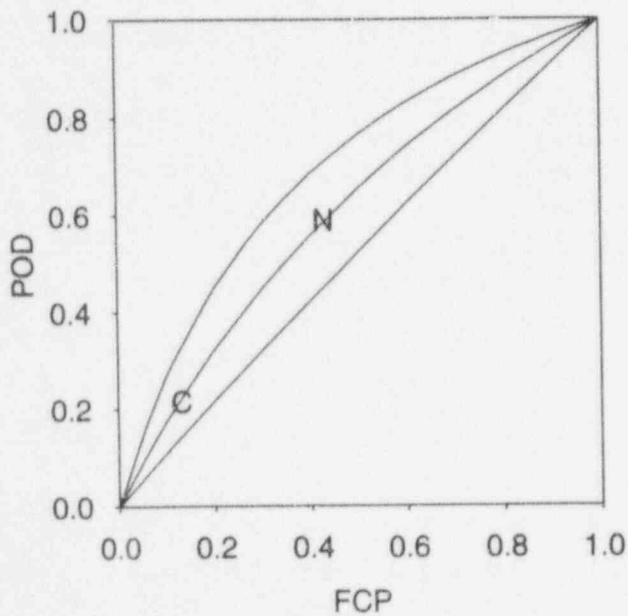


Figure 6.18 ROC Fit for Thermal Fatigue Cracks in Wrought Stainless Steel (Far-Side Inspection, Size 3 Cracks)

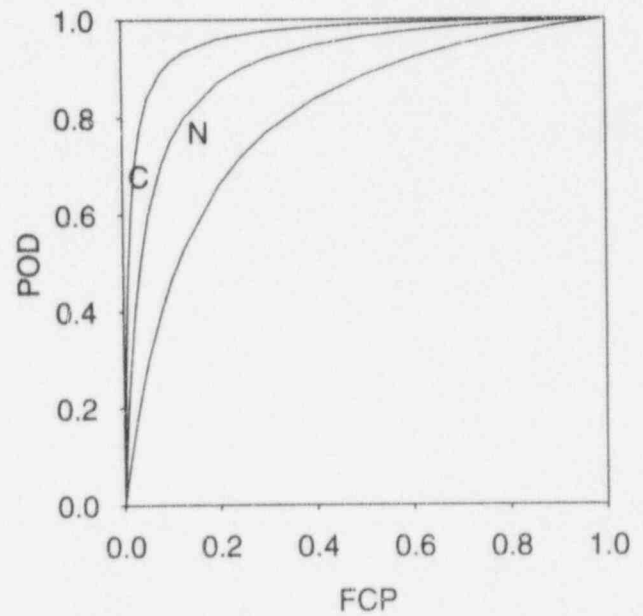


Figure 6.20 ROC Fit for IGSCC in Wrought Stainless Steel (Near-Side Inspection, Size 2 Cracks)

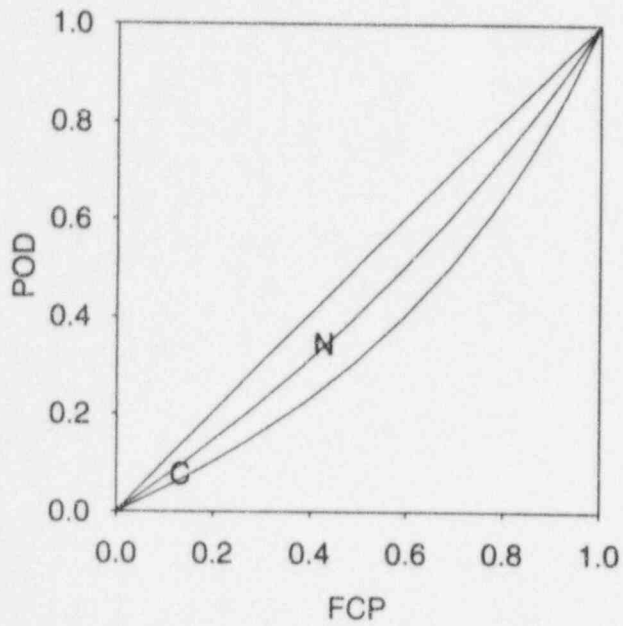


Figure 6.21 ROC Fit for IGSCC in Wrought Stainless Steel (Far-Side Inspection, Size 1 Cracks)

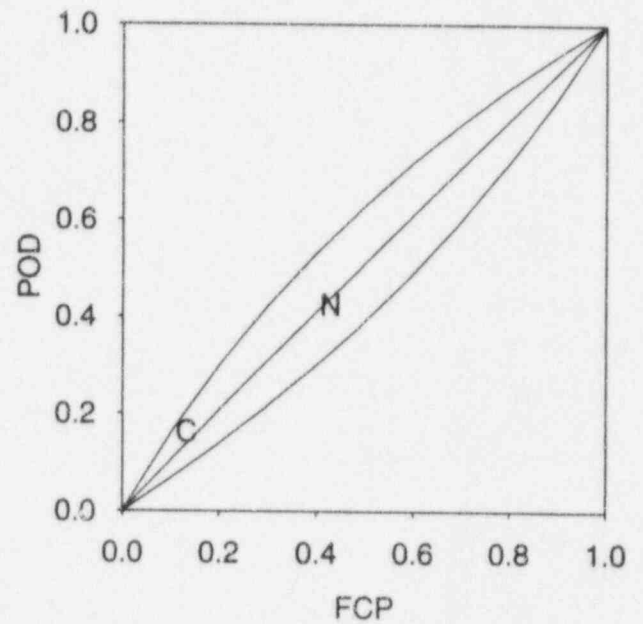


Figure 6.22 ROC Fit for IGSCC in Wrought Stainless Steel (Far-Side Inspection, Size 2 Cracks)

7.0 Sizing Errors

This section evaluates the crack sizing errors (in depth and length) attributable to the various inspection teams. In this section, we consider only those sizing errors associated with cracks that were detected and classified as cracks. That is, only the depth and length measurements reported for "C" indications are used in this analysis. Because cracks that have received a "C" should be the easiest to size, this perspective should yield an optimistic picture of sizing capabilities. The sizing results are influenced by the fact that the number of size measurements produced by any particular team depend on the number of cracks that team detected. Those teams that did poorly in the detection phase of the round robin reported few legitimate crack sizes, and their sizing performance could not be quantified accurately. Grading units were established to associate unique depth and length measurements with each crack (or crack cluster). When more than one indication was recorded within a grading unit, the maximum reported depth and length were associated with the crack in that grading unit.

Linear regression was employed to analyze sizing errors in both depth and length. It was assumed that the measured and true sizes are related by a regression model of the formula:

$$M_i = A + B \times T_i + E_i \quad (7.1)$$

where M_i represents the measured size associated with grading unit i and T_i represents the true size of the crack in the grading unit. A and B are the regression parameters and E_i represents measurement error. Ideal performance occurs when $A=0$, $B=1$, and $E_i=0$. To be able to compare two different regression fits, and to order a set of fits from worst to best, we utilize the root mean square error (RMSE), a statistic that summarizes the three deviations of the regression parameters from their respective ideals. The root mean square statistic can be directly defined in terms of the original sizing data (M_i, T_i), $i = 1 \dots n$ by the formula:

$$RMSE^2 = \frac{\sum_i (M_i - T_i)^2}{n} \quad (7.2)$$

7.1 Depth Sizing Error

In Table 7.1 not all teams are listed for each combination of material and flaw. Some teams had never inspected CSS before and declared this a no test. Some teams only participated in a subset of the study and others declared that they could detect cracks in CSS but could not depth size any detected cracks. Table 7.1 summarizes the fitted regression results for depth sizing. Individual regression lines were fitted to each combination of team and material. To present the most favorable case, only near-side inspection results were used for the analysis. The table shows that the depth sizing results were very poor, and there is no strong evidence to indicate that depth sizing was effective in any of the materials. In cast stainless steel and clad ferritic materials, the average regression slopes are close to zero, indicating no relationship between the measured and true depths. This was expected because all of the teams used probe motion measurements to determine crack depth. This measurement procedure estimates the size of the sound field rather than the size of the defect, except when the defects are much larger than the sound field. For the transducers used by the teams in the PIRR, the cracks were smaller than the sound field and therefore difficult to size.

A visual overview of the sizing results is given in Figures 7.1 through 7.4, which plot the measured versus true depths for all materials. These plots confirm the results displayed in Table 7.1; depth sizing capabilities are poor.

In the wrought material, there is weak evidence of some positive depth sizing capability for IGSCC. The regression fits for three teams (#2, #3, and #6) display slopes (B) that exceed their standard deviations ($Sd(B)$). The regression fits for these three teams are plotted in Figures 7.5 through 7.7. These results represent the best depth-sizing performance seen in the PIRR.

It is interesting to note that the IGSCC sizing capability seems to exceed the TFC sizing capability. This corresponds to the detection results reported in Section 5; the IGSCC defects were also easier to detect than the TFC defects.

7.0 Sizing Errors

Table 7.1. Summary of Depth Sizing Regressions

Material	Team	Intercept				Slope		Standard Error		RMSE		No. of Observations
		A		sd(A)		B	sd(B)	(in)	(mm)	(in)	(mm)	
		(in.)	(mm)	(in.)	(mm)							
CSS	1	0.28	7.1	0.13	3.3	0.25	0.14	0.04	1.0	0.46	11.7	4
	5	0.40	10.2	0.21	5.3	0.05	0.29	0.18	4.6	0.37	9.4	7
	6	0.54	13.7	0.28	7.1	-0.30	0.38	0.26	6.6	0.53	13.5	12
	7	0.59	15.0	0.18	4.6	-0.36	0.38	0.14	3.6	0.24	6.1	6
Mean		0.45	11.4							0.4	10.2	
CF	1	0.30	7.6	0.09	2.3	0.30	0.13	0.22	5.6	0.31	7.9	43
	2	0.17	4.3	0.05	1.3	0.06	0.07	0.11	2.8	0.51	13.0	40
	3	0.45	11.4	0.11	2.8	0.05	0.15	0.23	5.8	0.38	9.7	32
	4	0.57	14.5	0.12	3.0	-0.17	0.18	0.29	7.4	0.45	11.4	43
	5	0.30	7.6	0.05	1.3	0.03	0.08	0.12	3.0	0.43	10.9	40
	6	0.11	2.8	0.03	0.76	0.01	0.04	0.07	1.8	0.57	14.5	44
Mean		0.32	8.1							0.44	11.2	
Wr SS/ TFC	1	0.06	1.5	0.02	0.51	0.15	0.11	0.04	1.0	0.11	2.8	16
	2	0.08	2.0	0.01	0.25	0.02	0.07	0.04	1.0	0.12	3.0	30
	3	0.17	4.3	0.06	1.5	0.22	0.33	0.13	3.3	0.16	4.1	15
	4	0.13	3.3	0.01	0.25	-0.08	0.09	0.06	1.5	0.12	3.0	51
	5	0.14	3.6	0.02	0.51	-0.14	0.10	0.05	1.3	0.13	3.3	31
	6	0.08	2.0	0.02	0.51	0.27	0.10	0.05	1.3	0.09	2.3	24
Mean		0.11	2.8							0.12	3.1	
Wr SS/ IGSCC	1	0.08	2.0	0.02	0.51	0.10	0.23	0.04	1.0	0.05	1.3	27
	2	0.08	2.0	0.02	0.51	0.31	0.24	0.04	1.0	0.06	1.5	34
	3	0.04	1.0	0.10	2.5	2.71	1.32	0.16	4.1	0.22	5.6	13
	4	0.14	3.6	0.02	0.51	-0.01	0.28	0.06	1.5	0.09	2.3	42
	5	0.16	4.1	0.03	0.76	-0.23	0.30	0.06	1.5	0.10	2.5	29
	6	0.07	4.3	0.03	0.76	0.44	0.33	0.05	1.3	0.06	1.5	28
Mean		0.10	2.5							0.097	2.5	
Overall Mean		0.22	5.6	0.07	1.8	0.17	0.24	0.11	2.8	0.25	6.4	27.8

- CSS = cast stainless steel
- CF = clad ferritic
- Wr SS/TFC = wrought stainless steel/thermal fatigue cracks
- Wr SS/IGSCC = wrought stainless steel/IGSCC

Table 7.2. Summary of Length Sizing Regressions

Material	Team	Intercept				Slope		Standard Error		RMSE		No. of Observations
		A		sd(A)		B	sd(B)	(in.)	(mm)	(in.)	(mm)	
		(in.)	(mm)	(in.)	(mm)							
CSS	1	0.59	15.0	5.37	136.4	0.22	2.10	1.08	27.4	1.60	40.6	4
	5	0.62	15.7	0.39	9.9	-0.02	0.17	0.25	6.4	1.72	43.7	7
	6	4.08	104	1.50	38.1	-1.08	0.67	1.22	31.0	1.64	41.7	12
	7	113	2870	122	3099	-31.75	34.8	0.40	10.2	1.93	49.0	6
Mean		29.62	751			-8.2				1.72	43.8	
CF	1	0.97	24.6	0.97	24.6	0.88	0.47	2.15	54.6	2.23	56.6	43
	2	-0.58	14.7	0.20	5.1	0.98	0.10	0.42	10.7	0.74	18.8	40
	3	1.47	37.3	0.54	13.7	0.41	0.25	1.03	26.2	1.12	28.4	32
	4	0.60	15.2	0.32	8.1	0.70	0.16	0.72	18.3	0.74	18.8	43
	5	0.45	11.4	0.27	6.9	0.35	0.13	0.59	15.0	1.10	27.9	40
	6	-0.07	1.8	0.29	7.4	0.84	0.14	0.66	16.8	0.75	19.0	44
Mean		0.47	12.0			0.70				1.11	28.3	
Wr SS/ TFC	1	0.63	16.0	0.12	3.0	0.04	0.09	0.24	6.1	0.83	21.1	17
	2	0.40	10.2	0.11	2.8	0.14	0.08	0.28	7.1	0.53	13.5	30
	3	2.55	64.8	1.24	31.5	-0.20	1.08	1.87	47.5	2.23	56.6	15
	4	1.38	35.1	0.30	7.6	-0.27	0.25	0.94	23.9	1.14	29.0	51
	5	0.59	15.0	0.10	2.5	0.06	0.08	0.26	6.6	0.76	19.3	31
	6	1.01	25.7	0.29	7.4	-0.09	0.22	0.56	14.2	0.83	21.1	25
Mean		1.09	27.8			-0.05				1.05	26.8	
Wr SS/ IGSCC	1	0.30	7.6	0.16	4.1	0.72	0.15	0.39	9.9	0.40	10.2	27
	2	0.13	3.3	0.16	4.1	0.76	0.15	0.41	10.4	0.42	10.7	34
	3	1.87	47.5	0.64	16.3	0.51	0.57	1.12	28.4	1.75	44.4	13
	4	0.77	19.6	0.36	9.1	0.36	0.34	1.10	27.9	1.14	29.0	42
	5	0.45	11.4	0.14	3.6	0.40	0.13	0.35	8.9	0.46	11.7	29
	6	0.10	2.5	0.22	5.6	0.80	0.20	0.49	12.4	0.49	12.4	29
Mean		0.60	15.3			0.59				0.78	19.7	
Overall Mean		5.97	151.6	6.16	156.5	-1.15	1.92	0.75	19.0	1.13	28.7	27.9

CSS = cast stainless steel

CF = clad ferritic

Wr SS/TFC = wrought stainless steel/thermal fatigue cracks

Wr SS/IGSCC = wrought stainless steel/IGSCC

7.2 Length Sizing Errors

Table 7.2 presents the regression fits that relate true crack length to measured length. For length sizing, we find poor performance in two materials, cast stainless steel and wrought stainless steel, with TFC. In fact, for these two combinations of material and crack-type, there is no evidence that length sizing is effective at all. The regression slopes are essentially zero.

However, in the other two combinations of material and crack-type, the results are positive for some teams. Those regression fits that display slopes significantly different from zero (at the 97.5% confidence level, which means that we are 97.5% confident that the slopes are nonzero) are marked with an asterisk in Table 7.2. Five different teams display positive results for clad ferritic with TFC defects and for wrought SS with IGSCC defects. The regression fits for all six teams are plotted in Figures 7.8 through 7.19. In these figures, two teams stand out from the others in sizing performance; #2 and #6 (see Figures 7.9, 7.13, 7.15, and 7.19).

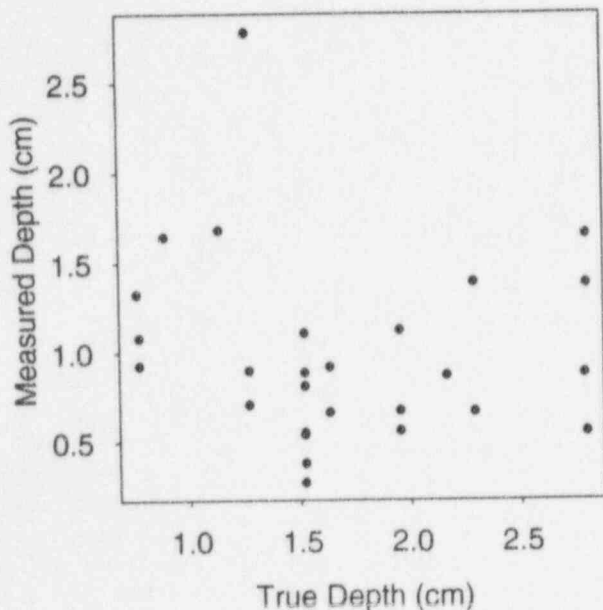


Figure 7.1 Plot of All Teams' Near-Side Depth Measurements of TFC in Cast Stainless Steel Material

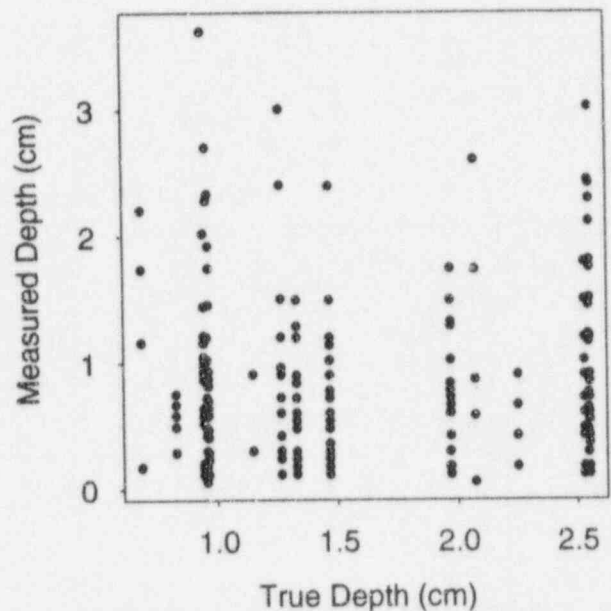


Figure 7.2 Plot of All Teams' Near-Side Depth Measurements of TFC in Clad Ferritic Material

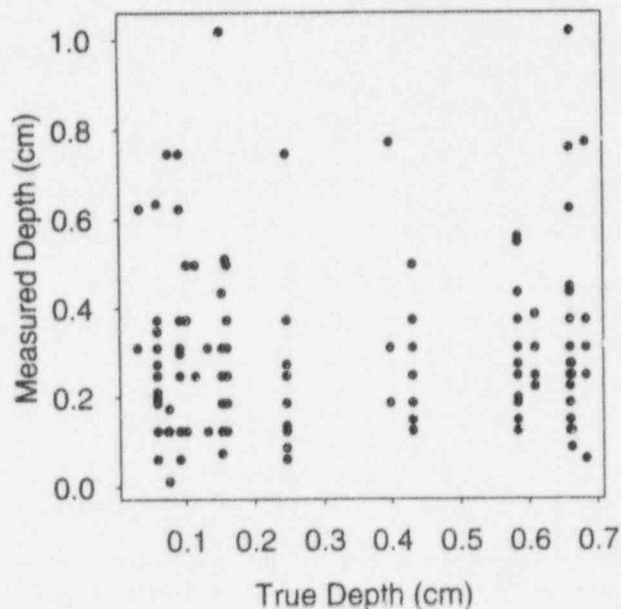


Figure 7.3 Plot of All Teams' Near-Side Depth Measurements in Wrought Stainless Steel with Thermal Fatigue Cracks

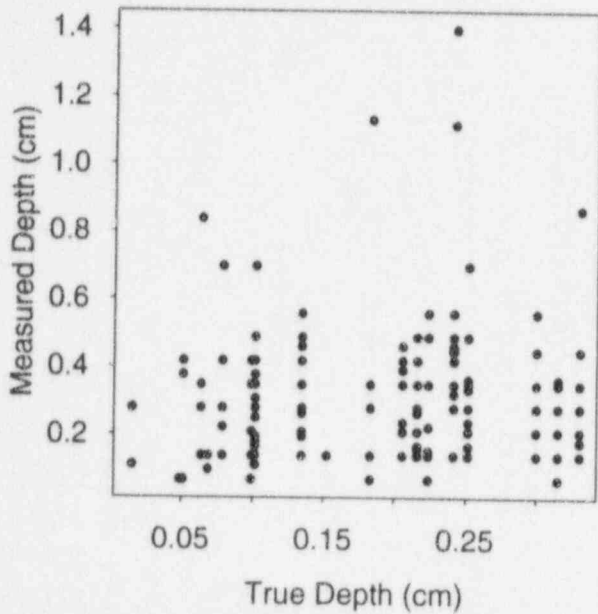


Figure 7.4 Plot of All Teams' Near-Side Depth Measurements in Wrought Stainless Steel with IGSCC

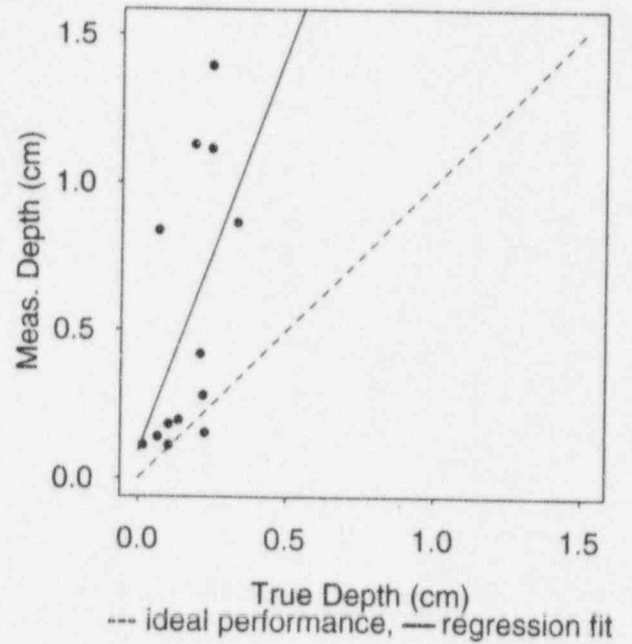


Figure 7.6 Regression Fit of Near-Side Results of Team #3 for Wrought Stainless Steel with IGSCC

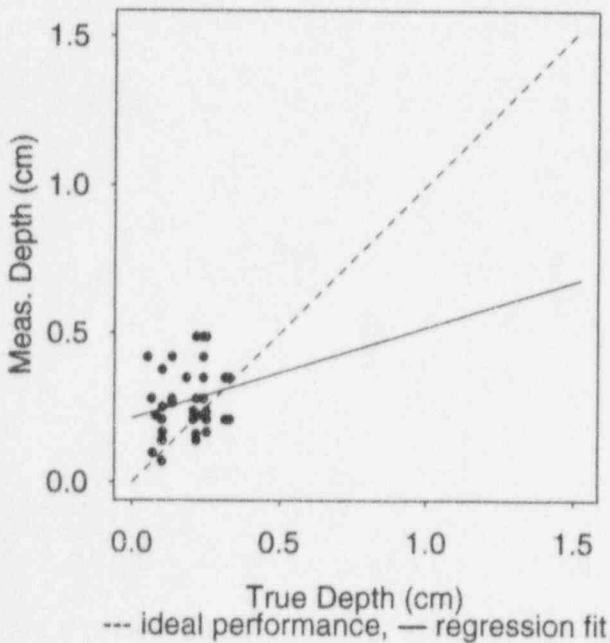


Figure 7.5 Regression Fit of Near-Side Results of Team #2 for Wrought Stainless Steel with IGSCC

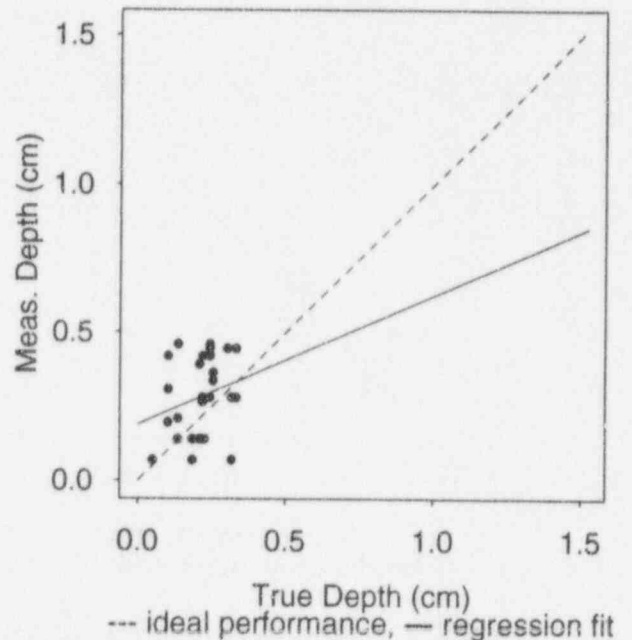


Figure 7.7 Regression Fit of Near-Side Results of Team #6 for Wrought Stainless Steel with IGSCC

7.0 Sizing Errors

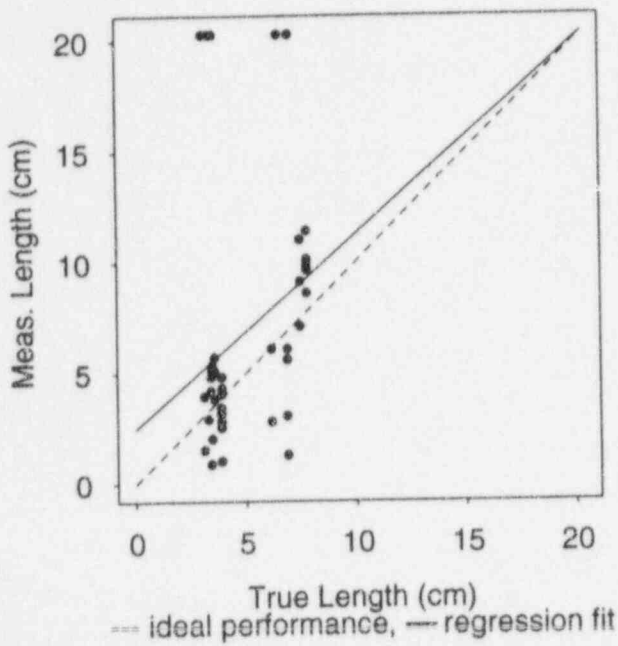


Figure 7.8 Regression Fit of Length Measurements (Team #1, Clad Ferritic)

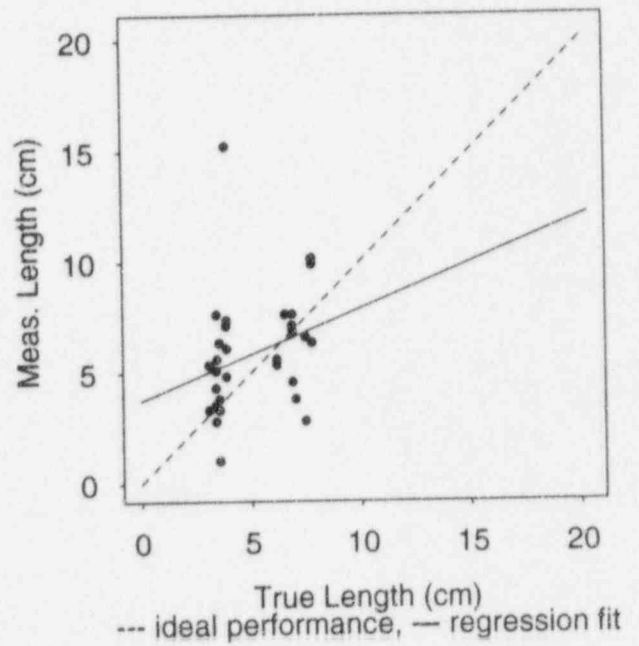


Figure 7.10 Regression Fit of Length Measurements (Team #3, Clad Ferritic)

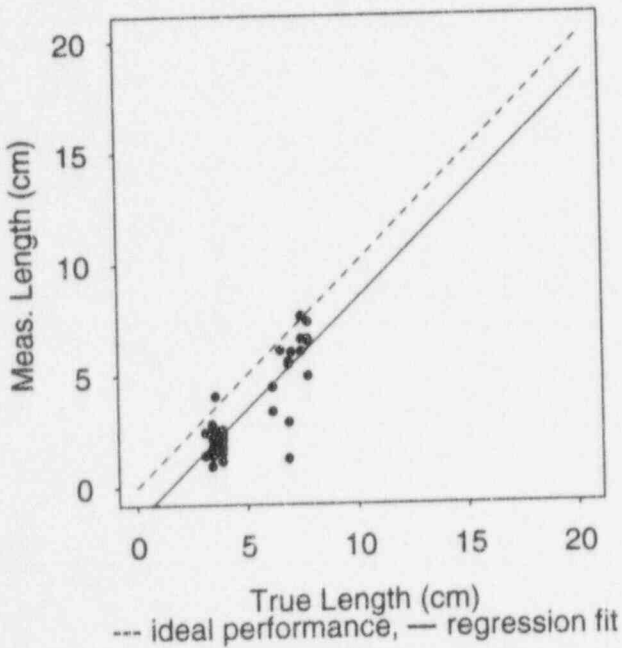


Figure 7.9 Regression Fit of Length Measurements (Team #2, Clad Ferritic)

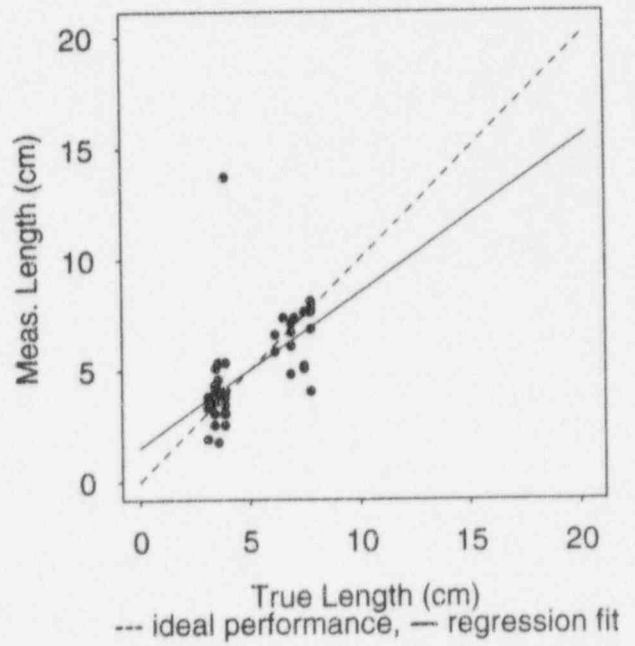


Figure 7.11 Regression Fit of Length Measurements (Team #4, Clad Ferritic)

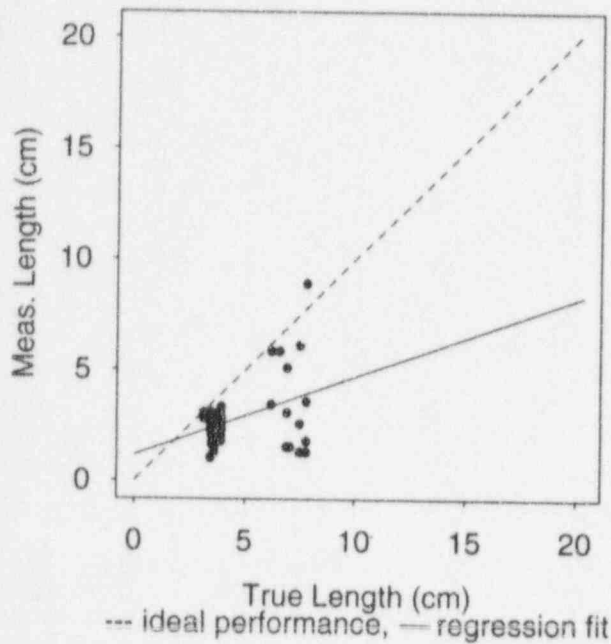


Figure 7.12 Regression Fit of Length Measurements (Team #5, Clad Ferritic)

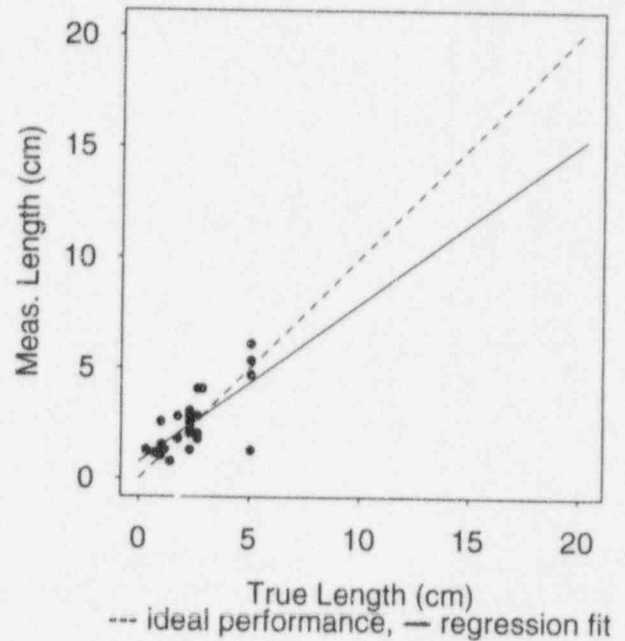


Figure 7.14 Regression Fit of Length Measurements (Team #1, Wrought Stainless Steel, IGSCC)

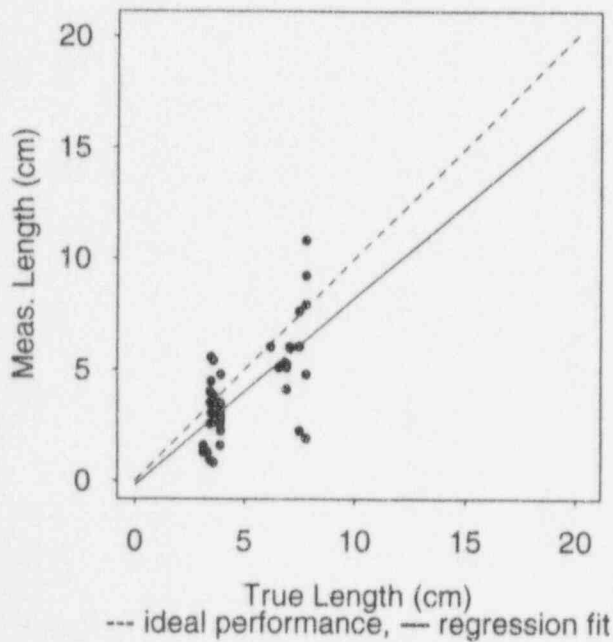


Figure 7.13 Regression Fit of Length Measurements (Team #6, Clad Ferritic)

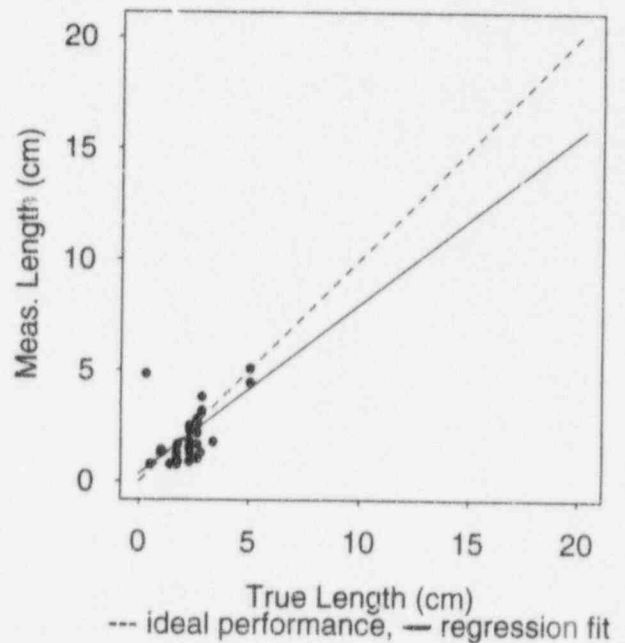


Figure 7.15 Regression Fit of Length Measurements (Team #2, Wrought Stainless Steel, IGSCC)

7.0 Sizing Errors

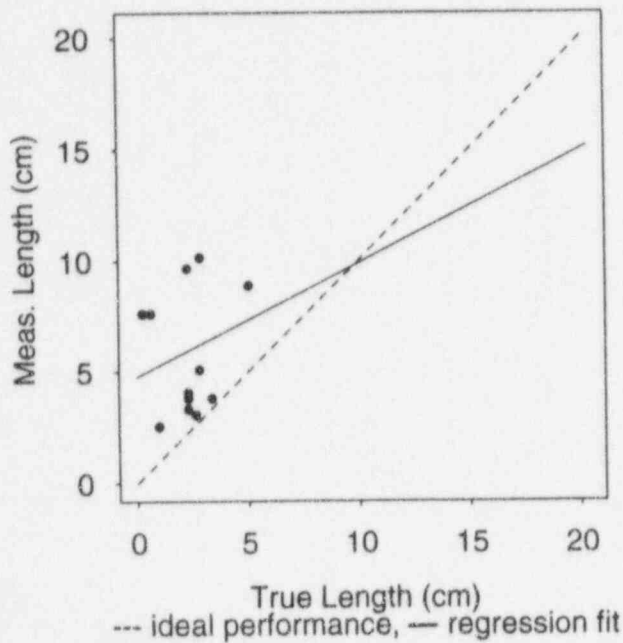


Figure 7.16 Regression Fit of Length Measurements
(Team #3, Wrought Stainless Steel, IGSCC)

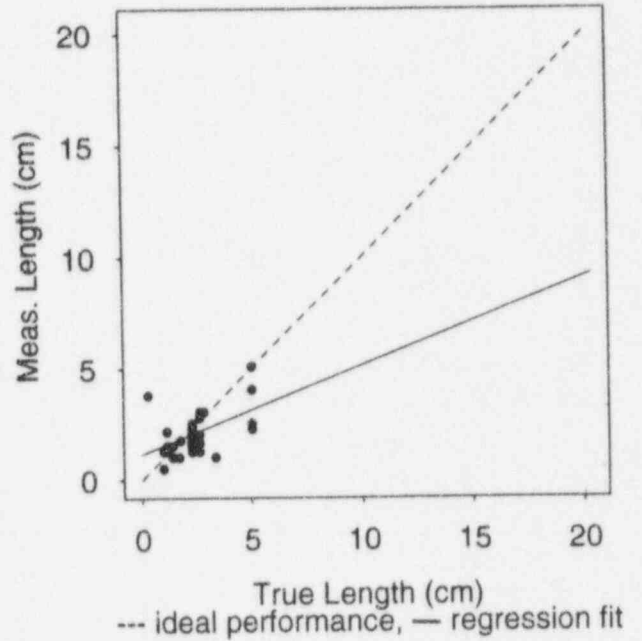


Figure 7.18 Regression Fit of Length Measurements
(Team #5, Wrought Stainless Steel, IGSCC)

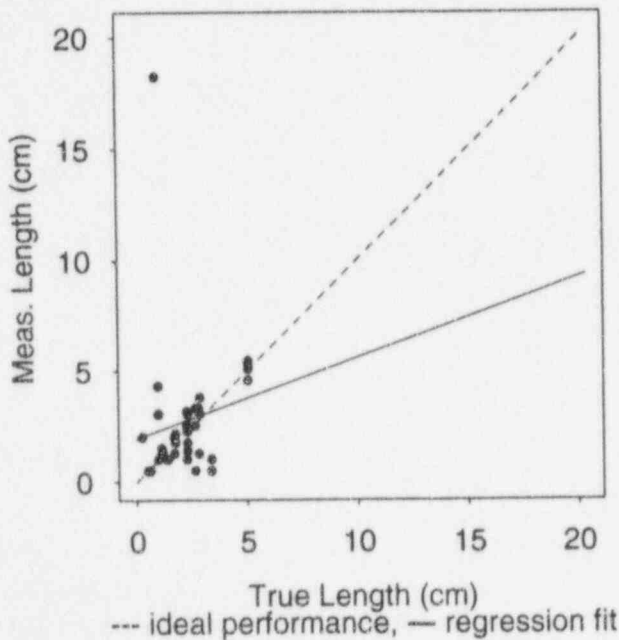


Figure 7.17 Regression Fit of Length Measurements
(Team #4, Wrought Stainless Steel, IGSCC)

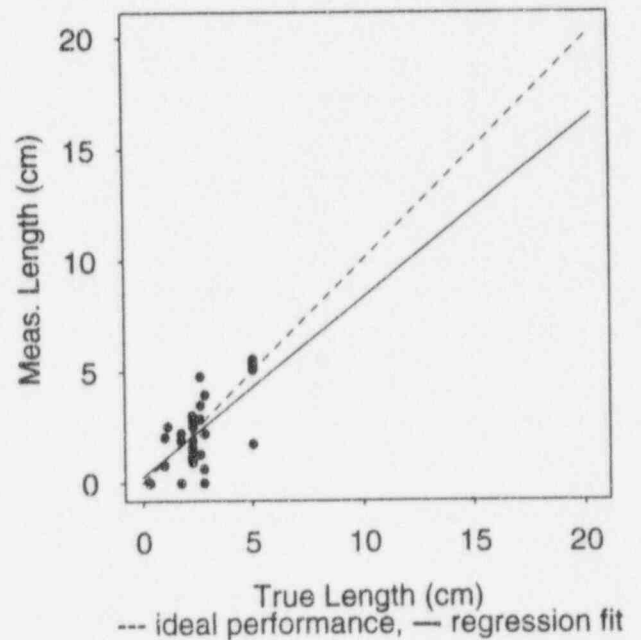


Figure 7.19 Regression Fit of Length Measurements
(Team #6, Wrought Stainless Steel, IGSCC)

8.0 Team Error Statistics

Thousands of data entries were recorded by the ISI teams during the round robin and errors were inevitable. Some of these errors would have changed the scoring of inspection results if entered verbatim into the computer; for example, a transcription error in reporting its position might have caused a crack indication to fall outside a grading unit.

The errors were generally classified into two major categories, according to what caused the error. "Test protocol" errors were due to the team's misunderstanding some aspect of the test protocol. "Inservice inspection" errors were those that would have occurred during an actual inservice inspection. In the inspection data set, test protocol errors were simply corrected and were not counted against the team. In contrast, inservice inspection errors were not corrected; so these errors are reflected in the detection and sizing statistics discussed in previous sections.

The principal objective of this section is to identify the important inservice inspection errors and tabulate their frequency. The effect of these inservice inspection errors on probability of detection and sizing capability is also discussed. A manual review of all inspection reports was conducted to identify and tabulate all such errors. Also during this review, all identified test protocol errors were corrected in the raw data files.

Each inspection folder contained the raw inspection data for a team, including calibration records, raw data sheets generated by the Level I and II examiners, indication plotting sheets, and four-channel strip chart recordings of the output of the Search Unit Tracking And Ranging System (SUTARS). This data allowed us to verify the transducer axial position, circumferential position, time-of-flight, and signal amplitude during all inspections. In most cases, it was possible to trace a team's actions and logic from the detection of a signal through the final step of data entry onto the inspection report form.

At the same time, the strip chart records were searched for the presence of signals from defects for which no indication was reported. From the amplitudes of these unrecorded indications, conclusions can be drawn as to the increased sensitivity necessary to enhance detection performance.

8.1 Examples of Common Team Errors

Errors in plotting the indications and in reading off the plotted axial locations of the ultrasonic reflectors were frequent, no doubt due to the repetitiveness of the chore. In these cases, the information contained in the raw data sheets was sufficient to calculate the correct position values.

Plotting errors can be systematic: in one case, a Level III inspector had consistent slight errors in axial location of defects for about half of the team inspections of clad ferritic pipe. It was noted that the plotted sound field central ray beam angle was slightly off 45° in all the plots for the mislocated defects. We then realized that the Level III had used a Pacific Northwest National Laboratory drafting table as a graphics aid for these plots, and was apparently not proficient in its use. This is a typical test protocol error.

Another team reported the axial locations of all indications relative to the weld centerline (instead of to the scribe line); we obviously failed to make our reference system clear to them. If not corrected, this team's results would have been scored as hopelessly poor. This is another test protocol error.

Errors in distance measurement on the pipe surface were not infrequent. Using the four-channel strip chart records, several errors of precisely one inch (2.54 cm) were noted, and these were ascribed to mistakes in reading the ruler. Occasionally, the clarity of the chart recordings also allowed detection of small position measurement errors, probably due to the awkwardness of manipulating the couplant-coated flexible rule with gloved fingers. These are examples of inservice inspection errors.

8.2 Team Error Codes

This section presents a listing and short description of the error codes developed during the manual review of inspection data. Some codes represent errors in data acquisition and handling; others represent incorrect decision-making. This list was compiled as the review progressed and new error types were discovered. In a few cases, the errors were unique or nearly unique, so their incidence was minimal. The error codes are:

8.0 Team Error Statistics

Measurement Errors

- M1: Small error in distance measurement, deduced from strip chart and raw data sheets.
- M2: Value of defect length L exceeded specimen dimensions; on 10 in. (25.4 cm) pipes, some indications extended into adjacent quadrants; on heavy-wall pipe, some teams used L1=0, L2=8 to describe full-width indications, though most specimens were not 8 in. (20.3 cm) wide.
- M3: 1 in. (2.54 cm) error in distance measurement; major divisions of ruler were misread or miscounted--to be distinguished from inaccurate placement of the ruler (M1).
- M4: Strip chart shows defect detection at two discrete values of axial position (a "double-peak indication"); the wrong peak was chosen for recording.
- M5: Measured location from weld centerline was entered with the wrong sign.

Plotting Errors

- P1: Minor plotting inaccuracy leading to a small error.
- P2: A correctly made plot was misread, producing an error in reflector axial location.
- P3: Major plotting error; includes 1 in. (2.54 cm) errors in plotting transducer positions (see M3 above), plotting transducer positions from the wrong reference point, using a 60° beam plot for a 45° inspection, etc.
- P4: Axial position of indication was reported relative to the measured location of the weld centerline, instead of to the scribe line provided.

Logical Decision Errors

- LD1: Multiple defects were lumped together by the Level I and II team members.
- LD2: Multiple defects were lumped together by the Level III team member.

Scanning Errors

- S: Strip chart record shows that transducer was never in position to detect defect (inadequate scan pattern).

These error codes can be classified in several ways. The classifications most relevant for analysis purposes are the causes and consequences of the error. Accordingly, the error codes were arranged in six categories, as illustrated in Table 8.1.

Table 8.1. Categorization of Error Codes

Consequence of Error	Cause of Error	
	ISI	Test Protocol
Small Sizing/ Location Error	M1, P1	M2
Large Sizing/ Location Error	M4, P2, LD1, LD2, M3, P3, M5	P4
Crack Missed	S	

Those error codes classified under "small sizing/location error" do not have a large effect on the inservice inspection POD, but they do increase the variability in the sizing errors by a modest amount. On the other hand, "large sizing/location error" might cause enough confusion in the inspection results to cast doubt on the very existence of defects, and could therefore affect POD.

From a statistical point of view, these two sets of inservice inspection errors have fundamentally different effects on the sizing/location error distributions, as illustrated in Figure 8.1. The small inservice inspection errors are well within the main body of the distribution and therefore can be accounted for under a normal distribution. However, the large inservice inspection errors are well out in the tail, and tend to put abnormal bumps on the distribution. Basically, small sizing/location errors are considered to be no greater than 0.2 in. (0.5 cm), while the large errors average around 1 in. (2.54 cm) in magnitude.

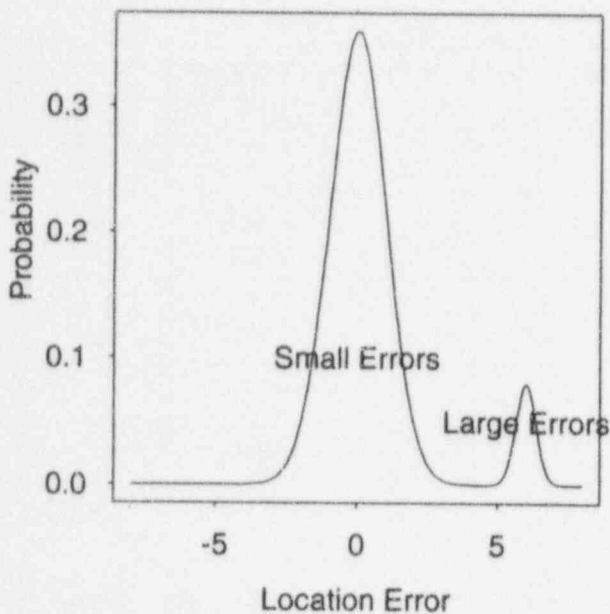


Figure 8.1 Possible Distribution Model for Team Errors

Finally, the third consequence listed in Table 8.1 is a defect being entirely missed. Only one type of inservice inspection error resulted in this consequence: during some of the inspections, the inspector did not scan the weld as tightly as the procedure called for. On the other hand, it should be emphasized that some cracks were missed because the signal obtained from them did not exceed the recording level specified in the procedure. In those cases, the team performed according to the specified procedure, so no inservice inspection error occurred.

8.3 Tabulation of Team Errors

Table 8.2 summarizes the team error data by error code and team. The total number of inspections completed by each team is shown at the bottom of the table. The main body of the table lists the number of errors committed (count) and the error rate (errors/inspection). In general, these error rates are grouped around the 3/100 to the 1/1000 range, although an error rate of 0.68 was observed for error code P4.

There are significant differences among the team error rates listed in Table 8.2; it is obvious that teams do not

commit the same errors. This fact has important consequences for ISI inspection reliability. Many of the errors are due to misunderstandings of data recording, measurement, plotting, or test protocol procedure.

Even though test protocol errors are not directly relevant to an inservice inspection environment, these errors can provide some insightful indirect information. A close examination of the test protocol error codes revealed that all these errors are due to team misunderstandings about the designated coordinate axis or the data recording procedures.

It must be emphasized that neither the coordinate system nor the data recording procedures were particularly complicated or unusual. They were, in fact, the type of procedures a power company might impose on a team so that past and current inspection results could be compared unambiguously, particularly results originating from different teams. The test protocol error rates provide a measure of a teams' ability to follow standardized data reporting procedures, after receiving clear instructions. Table 8.2 indicates that five out of six teams had fundamental misunderstandings about the coordinate axis and data recording (Team #2 did not make these errors).

Table 8.3 displays the "inservice inspection" error rates by consequence. This table also displays important differences among teams. For the sake of simplicity, we focused our attention on the average team error rates given in the last column. Perhaps the most important number in the table is the 0.019 rate for a team missing a defect because of an inservice inspection error. This statistic indicates that probability of detection may be no higher than 0.98, even for large cracks. This is a very important deficiency in present manual procedures which might be rectified with a well-constructed automated procedure. More attention to human factors should also help to lower this important error rate.

It is also important to note that the chance of some large sizing or location error occurring during an inservice inspection is 0.048. This error rate makes it more difficult to use present inservice inspection results to monitor crack growth over time, because there is almost a 5% chance of a gross sizing or location error on any individual measurement.

8.0 Team Error Statistics

Table 8.2. Team Errors Classified by Error Code and Teams

Error Code		Team						Total
		#1	#2	#3	#4	#5	#6	
M1	count	5	5	15	4	5	0	34
	rate	0.024	0.031	0.072	0.022	0.024	0.0	0.029
M2	count	8	0	3	0	1	1	13
	rate	0.039	0.0	0.014	0.0	0.005	0.005	0.011
M3	count	0	1	0	0	0	0	1
	rate	0.00	0.006	0.0	0.0	0.0	0.0	0.001
M4	count	0	0	0	1	0	0	1
	rate	0.0	0.0	0.0	0.005	0.0	0.0	0.001
M5	count	0	0	0	1	0	0	1
	rate	0.0	0.0	0.0	0.005	0.0	0.0	0.001
P1	count	1	2	2	9	2	0	16
	rate	0.005	0.013	0.010	0.049	0.010	0.0	0.014
P2	count	0	0	6	6	12	10	34
	rate	0.0	0.0	0.029	0.033	0.057	0.049	0.029
P3	count	2	0	0	5	0	1	8
	rate	0.010	0.0	0.0	0.027	0.0	0.005	0.007
P4	count	0	0	0	125	0	1	126
	rate	0.0	0.0	0.0	0.683	0.0	0.005	0.108
LD1	count	1	0	1	2	0	0	4
	rate	0.005	0.0	0.005	0.011	0.0	0.0	0.003
LD2	count	0	0	1	3	3	0	7
	rate	0.0	0.0	0.005	0.016	0.014	0.0	0.006
S	count	1	3	8	4	6	0	22
	rate	0.005	0.019	0.038	0.022	0.029	0.0	0.019
Total count		18	11	36	160	29	13	267
Total Error rate		0.087	0.069	0.172	0.874	0.139	0.063	0.228
No. of Inspections		206	160	209	183	209	205	1172

Table 8.3. Inservice Inspection Errors Classified by Consequence

Consequence		Team						Total
		#1	#2	#3	#4	#5	#6	
Small Sizing/ Location Error	count	6	7	17	13	7	0	50
	rate	0.029	0.044	0.081	0.071	0.033	0.0	0.043
Large Sizing/ Location Error	count	3	1	8	18	15	11	56
	rate	0.015	0.006	0.038	0.098	0.072	0.054	0.048
Crack Missed	count	1	3	8	4	6	0	22
	rate	0.005	0.019	0.038	0.022	0.029	0.0	0.019
Total	count	10	11	33	35	28	11	128
Total Error	rate	0.049	0.069	0.158	0.191	0.134	0.054	0.109
No. of Inspections		206	160	209	183	209	205	1172

9.0 Examination of False Calls

In this section, we investigate the distribution of false calls within inspected material and identify important variables that affect false calls. A proper understanding of false calls is necessary to assess the economic consequences of inspection and also to construct relevant grading units. Determination of a model for false calls is important because it allows the false call information obtained in a round-robin test to be extrapolated to the extended lengths of weld material that are examined in an actual field examination. A simple distributional model for false calls is compared to the results obtained with the grading-unit methodology used in the round-robin test.

A false call results from the decision criteria, and relates to a C- or N-indication that is located in blank (uncracked) weld material. Such false indications are usually the result of random guessing and misinterpreted geometrical and/or metallurgical reflectors. In this discussion, a distinction is made between calls in blank grading units that are labeled as C (definitely a crack) and those that are labeled as N, which are less serious. Consequently, we will employ a nomenclature that distinguishes between these two types of calls. The terms false call probability and false call rate will be used to refer to C-indications in blank material, while the terms false recording probability and rate will refer to both C- and N-indications in blank material.

9.1 False Call Rates and the Poisson Model

In homogeneous material, false calls should occur randomly, with no one length of weld more likely to receive a false call than another. Under this assumption, the number of false calls would be described by a Poisson model. In a Poisson model, the number of false calls $N(s)$ that will occur within a length s of blank weld will be distributed according to:

$$Pr\{N(s) = k\} = e^{-r*s} \frac{(r*s)^k}{k!} \quad (9.1)$$

where the parameter r is called the false call rate and k is the number of flaws in internal s . The false call rate represents the average number of false calls that can be expected in one unit length of material and is typically estimated by dividing the total number of false calls by the total length of uncracked material inspected.

Tables 9.1 and 9.2 present estimated false call and false recording rates for various conditions in the round robin. The estimates in these tables were computed using data solely from inspections of completely blank specimens (i.e., no weld material from specimens containing a crack was employed). In addition to the false call rates, the table contains several associated statistics such as the total length of weld inspected, and the average indication length.

It is important to note that all false call rates are probably higher than they would be in actual in-field inspections. However, because this round robin included many blank specimens and the teams knew blanks were present, these statistics may represent reasonable estimates for nuclear piping.

Another important observation is that the material with the highest probability of detection also has the lowest false call rate. The clad ferritic material experienced only two false calls in 1016 inches (25.8 m) of material, producing an overall false call rate of 0.002 false calls per inch (0.001 per cm), or about one false call for every 42 feet (12.8 m) of inspected weld.

Once the parameter r has been determined from data, Equation 9.1 can be employed to determine the probability that a specified number of false calls might occur in an arbitrary length of weld material. Because FCP is simply the probability that one or more indications intersect with a blank grading unit (GU), FCP should be determined by the formula:

$$FCP(GU \text{ of length } s) = 1 - e^{-r*(u*s)} \quad (9.2)$$

where u represents the average length of an indication. The indication length s occurs in the formula because indications outside the grading unit can still intersect with the grading unit, if they are long enough.

Is the Poisson model reasonable for false calls? There are many "goodness-of-fit" tests that might be used to determine whether the Poisson model is adequate. These tests usually attempt to check for some sort of non-random pattern in false call locations. Such a pattern may appear in the form of anomalous "clusters" or as a spacing of false calls that is too regular.

9.0 Examination of False Calls

Table 9.1. False Call Rates

Material	Total Inspection Length, in. (m)	Access	Average Indication Length, in. (cm)	Number of "C" Calls	False Call Rate*
Cast SS	477 (12.1 m)	Near	0.95 (2.41 cm)	4	0.008 (.003)
		Far	2.12 (5.38 cm)	14	0.029 (.011)
Clad Ferritic	508 (12.9 m)	Near	0.60 (1.52 cm)	1	0.002 (.001)
		Far	3.50 (8.9 cm)	1	0.002 (.001)
Wrought SS	1378 (35.0 m)	Near	0.95 (2.41 cm)	14	0.010 (.004)
		Far	0.98 (2.49 cm)	60	0.044 (.017)

* Units are false calls per inch or (per cm).

However, one simple test of adequacy of the Poisson model would be to apply Equation 9.2 to the grading unit sizes used in the round robin and compare the calculated FCPs to those estimated from actual data. If the numbers match up, this will provide evidence that the Poisson model is adequate for calculating false call probability; a very important task for such a probability model.

To test Equation 9.2, false call and false recording probabilities for four sizes of grading units were computed from the data as illustrated in Tables 9.3 and 9.4. These sizes include the standard 3 in. (7.62 cm) grading unit as well as three other sizes, which were used only in this section of the report to provide a better test of Equation 9.2. The grading unit sizes range from 1 in. (2.54 cm) to 8 in. (20.3 cm) in length, which effectively includes the entire specimen. Adjacent to the actual probabilities listed in Tables 9.3 and 9.4, which were

computed from grading unit data, probabilities calculated from Equation 9.2 are presented. These calculated probabilities make use of the rates and average indication lengths presented in Tables 9.1 and 9.2.

Generally speaking, the actual and calculated false call probabilities agree much more closely than do the recording probabilities. In fact, except for possibly the clad ferritic values, the results in Table 9.3 tend to support the assertion that Equation 9.2 adequately describes the effect of size on false call probability.

The situation is different for the false recording probabilities listed in Table 9.4, however. In almost every case, the calculated probability exceeds the actual, often by as much as 100%. It is apparent that Equation 9.2 is producing answers that are too large.

Table 9.2. False Recording Rates (C+N Indications)

Material	Total Inspection Length, in. (m)	Access	Average Indication Length, in. (cm)	Number of "C + N" Calls	False Call Rate*
Cast SS	477 (12.1 m)	Near	1.46 (3.71 cm)	9	0.019 (.007)
		Far	2.45 (6.22 cm)	22	0.046 (.018)
Clad Ferritic	508 (12.9 m)	Near	2.03 (5.16 cm)	7	0.014 (.005)
		Far	4.02 (10.2 cm)	6	0.012 (.005)
Wrought SS	1378 (35.0 m)	Near	1.52 (3.86 cm)	108	0.078 (.031)
		Far	2.20 (5.59 cm)	253	0.184 (.072)

* Units are false calls per inch or (per cm).

Table 9.3. Comparison of Actual and Calculated FCP

Material	Access		Grading Unit Size			
			1 in. (2.54 cm)	2 in. (5.1 cm)	3 in. (7.6 cm)	8 in. (20.3 cm)
Cast SS	Near	Actual	0.014	0.028	0.042	0.056
		Calc.	0.016	0.024	0.033	0.072
	Far	Actual	0.099	0.113	0.127	0.141
		Calc.	0.087	0.114	0.139	0.257
Clad Ferritic	Near	Actual	0.014	0.014	0.014	0.014
		Calc.	0.003	0.005	0.007	0.017
	Far	Actual	0.0	0.0	0.0	0.014
		Calc.	0.009	0.011	0.013	0.022
Wrought SS	Near	Actual	0.012	0.024	0.034	0.073
		Calc.	0.020	0.030	0.039	0.087
	Far	Actual	0.079	0.101	0.131	0.244
		Calc.	0.083	0.122	0.159	0.324

An examination of the indication plots for several teams identified one of the causes of this over-estimation. Indications, particularly N-indications, are not placed randomly in blank material, but in clusters. These clusters of indications are probably associated with some large geometric reflector or with metallurgical conditions. A typical example of this phenomenon is illustrated in Figures 9.1 and 9.2, where the inspector has plotted definite clusters of N-indications that are probably caused by the weld root.

Since N-indications are considered much less important than C-indications, the inspectors did not take the time to examine these two regions in detail. If they had, they might very well have combined all the indications in each cluster into a single indication that would have accurately described the geometric/metallurgical reflector. This would have reduced the false call rate considerably. Because the inspectors were much more careful in the evaluation of C-indications, the results in Table 9.3 (for FCP) are much better than those in Table 9.4 (for false RP).

9.2 The Effect of Location on False Calls

False call and false recording probability are influenced by many other factors besides grading unit size. Another

important factor is the location of the grading unit. Table 9.5 presents the false call probability for blank grading units located at three different sites within the specimens. The three different sites are:

- *Isolated Grading Units:* These grading units are located on a specimen that is completely blank.
- *Same-Side Contaminated Grading Units:* These grading units are located adjacent to a crack. The crack is on the same side of the weld as the grading unit.
- *Opposite-Side Contaminated Grading Units:* These grading units are located directly opposite a crack on the other side of the weld.

The reason we make distinctions among these three locations is because grading units in the "contaminated" locations may produce different false call probabilities than those residing in completely blank specimens. Whether or not contaminated locations produce reliable false call statistics is an important design question. Although, it costs money to produce and inspect blank specimens in a round-robin test, flawed specimens cost significantly more to produce than blank material. However, from a test design perspective it would be fortunate if contaminated grading units produced unbi-

Table 9.4. Comparison of Actual and Calculated False Recording Probability

Material	Access		Grading Unit Size			
			1 in. (2.54 cm)	2 in. (5.1 cm)	3 in. (7.6 cm)	8 in. (20.3 cm)
Cast SS	Near	Actual	0.042	0.056	0.085	0.127
		Calc.	0.045	0.063	0.081	0.163
	Far	Actual	0.155	0.169	0.183	0.225
		Calc.	0.147	0.185	0.222	0.382
Clad Ferritic	Near	Actual	0.042	0.042	0.042	0.069
		Calc.	0.041	0.054	0.067	0.129
	Far	Actual	0.042	0.042	0.042	0.069
		Calc.	0.058	0.069	0.079	0.132
Wrought SS	Near	Actual	0.085	0.128	0.159	0.226
		Calc.	0.179	0.241	0.298	0.526
	Far	Actual	0.284	0.372	0.427	0.573
		Calc.	0.444	0.537	0.615	0.846

ased estimates of FCP, so that fewer completely blank specimens would be required in such a test.

Table 9.5 shows that a blank grading unit near a crack does not necessarily produce the same results as a blank grading unit that is on a blank specimen. As expected, blank grading units that are located opposite a crack show a higher false-call rate than isolated grading units. This effect is most pronounced for clad ferritic material, where the false call rate for contaminated blanks is 5 to 10 times that exhibited for isolated blanks. This is because, if several cracks are close to the weld centerline, it is easy in this material for the UT procedures to see through the weld to the other side and detect the crack. Any error in location that places the crack on the wrong side of the weld centerline would produce a false call.

Generally, the contaminated grading units exhibit a higher false call rate. However, the wrought stainless steel material results contain some important features. Wrought stainless steel is the only material in which we employed more than two grading units per specimen and consequently the only material which contained "same-side" contaminated grading units. Most wrought ss specimens contained four grading units, to allow us to include all cracked areas within a grading unit. Because it is important that adjacent grading unit results be independent of each other, all grading units were

separated by at least 2 in. (5 cm) of uncracked material. The fact that the false call probabilities for the same-side grading units are all very close to those for the isolated grading units is strong evidence that the grading units were laid out properly.

Another interesting aspect of the wrought SS results is the fact that the FCP for opposite-side blank grading units is about half the FCP for isolated blank grading units. Apparently teams were very reluctant to mark anything on the far-side, once they had found something on the near-side.

9.3 False Call Hot Spots

The false call data affords some insight into the decision-making process, and may provide some guidance for inspectors. False calls are caused by signals resulting from certain metallurgical and geometrical conditions within the grading unit. Signal effects related to the weld counterbore and the weld root can present a challenge to the inspectors. One of the major skills that an effective inspector must develop is the ability to recognize and classify such signals properly. For example, the welds were manufactured to contain defects such as weld roots exhibiting drop through, suck back, or lack of fusion, and counterbores were randomly located with some nearly vertical and near to the weld-

Table 9.5. Effect of Grading Unit Location on FCP and RP

Material	Access	Probability	Isolated Grading Unit	Contaminated	
				Same-Side	Opposite-Side
Cast SS	Near	FCP	0.04	--	0.09
		RP	0.08	--	0.23
		# Obs	71	0	53
	Far	FCP	0.13	--	0.19
		RP	0.18	--	0.26
		# Obs	71	0	183
Clad Ferritic	Near	FCP	0.01	--	0.16
		RP	0.04	--	0.20
		# Obs	72	0	64
	Far	FCP	0.0	--	0.32
		RP	0.04	--	0.35
		# Obs	72	0	203
Wrought SS	Near	FCP	0.03	0.04	0.12
		RP	0.16	0.13	0.24
		# Obs	328	341	265
	Far	FCP	0.13	0.12	0.08
		RP	0.43	0.38	0.25
		# Obs	328	339	373

* 3-in. (7.62 cm) grading unit size employed

Obs = number of observations

ment. However, we did ensure that the grading units used to estimate FCP did not exhibit suck back or lack of fusion.

In the case of the clad ferritic samples, the cracks were very close to the weld centerline, and any location errors would probably place the defect on the wrong side of the weld. This would yield both an error in detection and a false call. Also, for the clad ferritic samples, the grading unit for cracks spanned the weld centerline and the blank grading units had no cracks on the opposite side of the weld. To illustrate the impact of this effect, the POD for single-side grading units was 56%, while for grading units that spanned both sides of the weld POD was 97%.

For the wrought stainless steel specimens, it appeared that the inspectors could sometimes see through the

weld and detect cracks on its far-side. As an example, inspections were conducted from both sides of the weld on the specimen B213, quadrant 4 with cracks on each side. The results for this quadrant are shown in Figure 9.3, with the A-side results at the top of the figure (see Appendix A for an explanation of hit-o-gram plotting). It is clear from the alignment that the double-peaked response is caused by the defects on each side of the weld and that far-side defects may be plotted incorrectly on the near-side of the weld. The diagrams in Figure 9.3 are extracted from the entire set of hit-o-grams that is reproduced in Appendix A. In Figure 9.4, specimen B214, quadrant 2 shows some false calls in zones where there were no defects. Detection rates for the two cracks in this pipe were very low even though one of the cracks is 55% of the way through the wall. Another example is specimen B215, quadrant 3, for which Figure

9.0 Examination of False Calls

9.5 shows a number of false calls with very low detection rates for the cracks.

In all cases where false calls were made in the blank grading units, these specimens were re-examined by ultrasonic inspection, finger-damping techniques, dye penetrant inspection, radiographic examination, and even some destructive testing to ensure that no unintended defects existed in these blank grading units. In all cases, the defect-free nature of the grading unit was confirmed; and only then were the results included in this analysis.

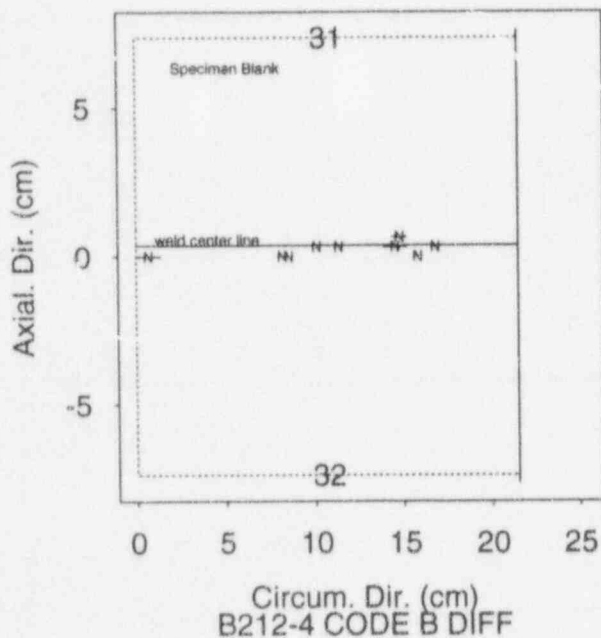


Figure 9.1 Indication Plot by Team #6 on Wrought SS Blank Specimen, Grading Units 31 and 32, Showing Clusters of "N" Decisions

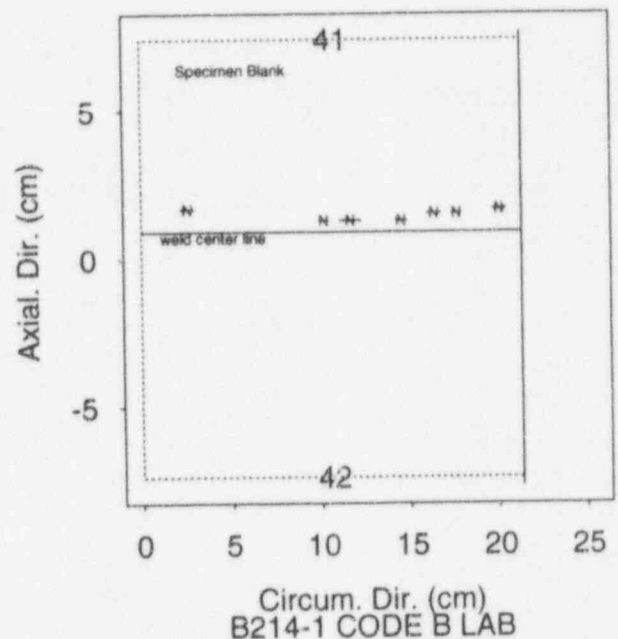
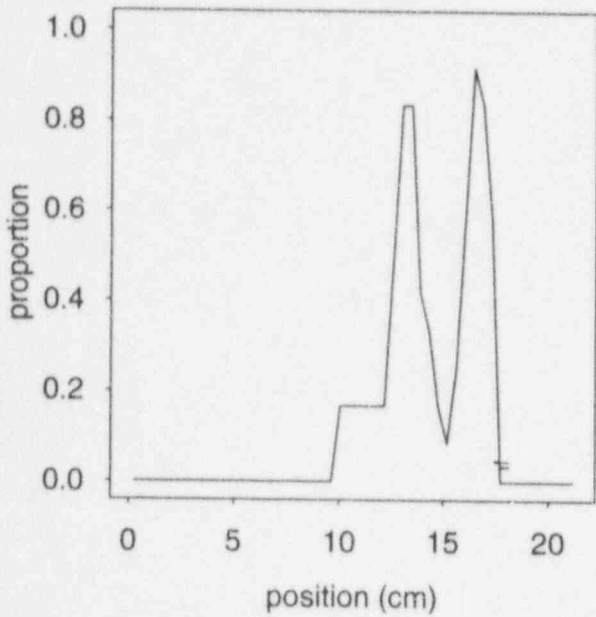
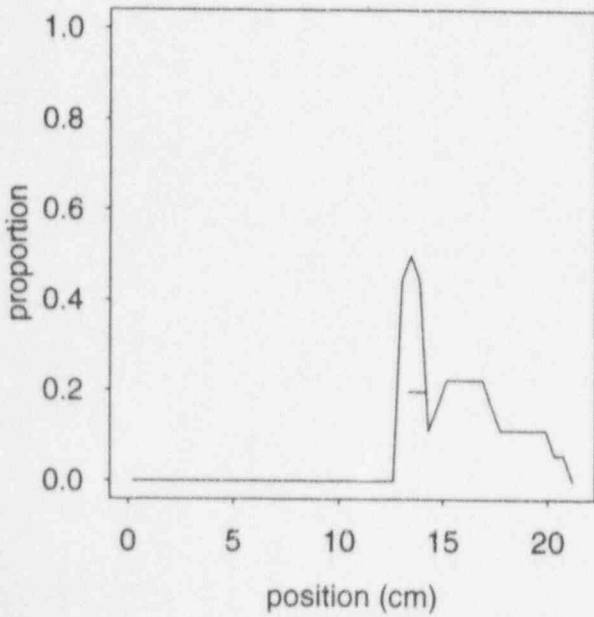


Figure 9.2 Indication Plot by Team #1 on Wrought SS Blank Specimen, Grading Units 41 and 42, Showing Clusters of "N" Decisions



(a)



(b)

Figure 9.3 Hit-O-Gram for wrought SS Specimen B213-4 (number of inspections = 12). "Defect" shows the location of the crack and crack length: Crack depth is the height of this line above the horizontal axis. a) A side Results, b) B side Results

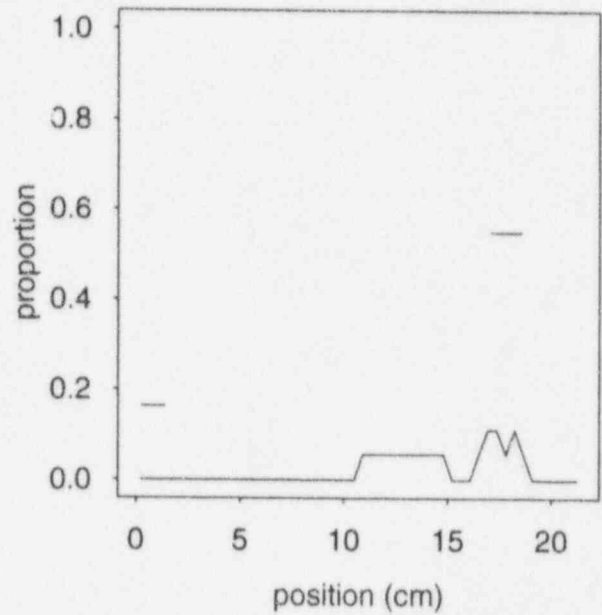


Figure 9.4 Hit-O-Gram for wrought SS Specimen B214-2, (number of inspections = 18).

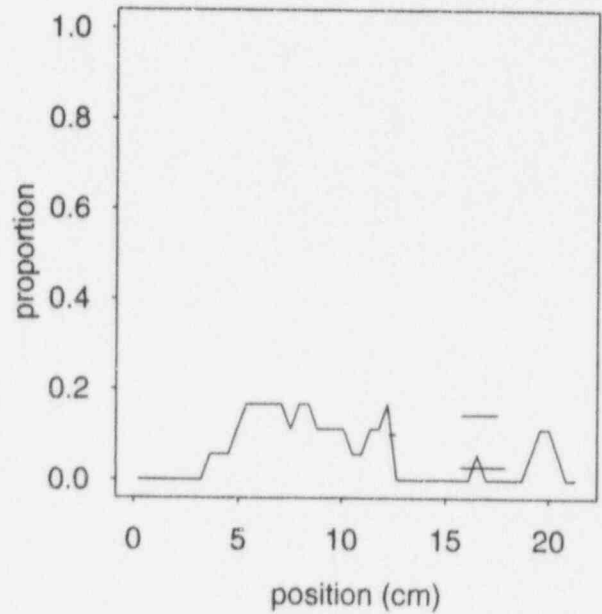


Figure 9.5 Hit-O-Gram for wrought SS Specimen B215-3, (number of inspections = 18).

10.0 Conclusions and Recommendations

Ultrasonic inspection of nuclear piping is meant to serve as one layer of the "defense in depth" philosophy. In order for it to fulfill this task, it must demonstrate high probabilities of crack detection, particularly for cracks that are important to structural integrity. An example illustrating this point would occur when making the assumptions that ISI must ensure a pipe failure rate lower than 10^{-3} per inspection interval and that it is assumed a large crack will grow through-wall during a single inspection interval, then a POD performance above 0.999 is necessary for large cracks.

Good inspection procedures should also be capable of determining the correct size of any detected defect. Without this capability, the decision maker will be confronted with two bad choices: either repair all indicated defects and bear the cost (and risk) of many unnecessary repairs; or ignore the defects indicated as "small", and bear the risk that these defects are sized incorrectly.

10.1 POD Performance

Table 4.10 gives a summary of the variables that affect probability of detection. The experimental results indicate that inspection environment (laboratory vs. simulated difficult field conditions) and inspection procedures (meeting ASME Code requirements (1977 edition through 1978 addenda) vs. a Pacific Northwest National Laboratory improved procedure) have no significant effect on POD. On the other hand, the variables crack type, access, material, and defect size do have a significant effect on POD (with crack type the least significant and defect size the most significant).

Construction of POD curves required us to determine which measure of crack size is most closely related to POD for this data set. For all material types, POD curves and recording probability curves were constructed based upon crack depth and crack length measurements.

We found that for cast stainless steel piping, either crack depth or crack length would be acceptable although crack length is slightly better for this data set. A number of factors were the cause for the results which included very low POD performance, inspectors looking for signals that exceed the high noise levels, and that four of six inspections showed no differences with

the other two showing good results from only a few of the very long cracks.

For clad ferritic piping, crack length is more directly related to POD. However, because crack length and depth were highly correlated in this round robin, the data was not well suited to distinguishing between them.

Depth is a better metric than length for both TFCs and IGSCCs in wrought stainless steel. Length is not a good size metric to use for wrought ss. The results for wrought stainless steel were what had been expected because as a flaw gets larger it generally reflects more acoustic energy providing a larger signal and increasing the detection probability.

The calculated POD curves show that far-side inspections in cast and wrought stainless steel produced results that were little better than guessing (see, for example, Figures 5.5, 5.21, and 5.23). In clad ferritic, near- and far-side detection results were lumped together, because several cracks were too near the weld centerline to permit a realistic test of the access condition. However, in this material the aggregate POD performance was quite good (see Figure 5.14). For cracks 40% through-wall in depth and 3 in. (7.6 cm) in length, the POD was 99.9%. Further analyses showed no practical difference between near- and far-side access results in clad ferritic material.

There was an important difference in the POD performance on thermal fatigue cracks versus IGSC cracks in wrought stainless steel, as illustrated in Figures 5.20 and 5.30. POD for the thermal fatigue cracks was much lower than for the IGSCC. For example, the largest thermal fatigue cracks [0.27 in. (6.9 mm) through-wall and 2.25 in. (5.7 cm) long] yielded a 70% POD, while the largest IGSC cracks (0.15 in. (3.8 mm) through-wall and 2 in. (5.1 cm) long) yielded a 90% POD. This supports the position that thermal fatigue cracks can indeed be used as conservative surrogates for IGSCC in round-robin and qualification testing.

The results obtained from the cast stainless steel inspections were clearly the worst of the three materials examined. Figure 5.2 shows that in this material, the different teams averaged 30% POD on near-side inspection of thermal fatigue cracks 3.5 in. (8.9 cm) long and 20% (12 mm) to 30% (18 mm) through-wall. The best team (Team 6, Figure 5.10) was able to produce a

10.0 Conclusions and Recommendations

42% POD for a thermal fatigue crack 2.5 in. (6.35 cm) long; a result that is much better than the average performance but still unreliable.

For clad ferritic material, the inspection results produced high POD values that are much better than those in cast SS. In this clad ferritic material, the largest cracks produced a POD estimate of 99.9%, while cracks half this size had a 90% chance of being detected, as shown in Figure 5.12.

10.2 False Call Performance

False call probability is primarily of economic concern. If the false call probability is too high, inspections may not be completed in the allotted time; but worse, valuable resources may be expended to repair good welds. Based upon field reports, the perception is that false calls occur infrequently during ISI. It is therefore expected to have low false call rates in round-robin studies. If the false call rates are not low then it may be the result of inspection decision processes changing during the round-robin versus that used during field inspections. The lower the false call rate for each material the better. In this report, false call probability was analyzed in conjunction with POD; false call probability is simply the point on the POD curve associated with a crack size of zero. False call rates were also computed from blank weld specimens and are summarized in Table 9.1. The false call probability exhibited in clad ferritic was the lowest of the three materials. The average false call rate for clad ferritic was approximately 0.024 false calls per foot (0.078 per meter). Therefore, a typical ISI of 25 welds on pipe 26 in. (66 cm) in diameter would yield about four false calls, a number that is quite low in comparison to those rates for the stainless steels.

The observed false call rate in cast stainless steel was much higher than the clad ferritic material and was 0.096 false calls per foot (0.315 per meter), which translates into about 17 false calls in an ISI inspection of 25 welds on pipe 26 in. (66 cm) in diameter. In addition, one must recognize that the POD in this material is low, so these false calls would occur with little offsetting benefit.

The false call probability in wrought stainless steel was highest of the three materials. In wrought SS, a false

call rate of 0.12 false calls per foot (0.394 per meter) was observed, which translates into an average of 20 false calls in an ISI inspection of 25 welds on pipe 26 in. (66 cm) in diameter.

10.3 ROC Performance

Because the PIRR was not designed for ROC analysis and the inspection results were recorded using only two decision thresholds (N=not a crack, C=crack), the complete shapes of the ROC curves could not be estimated accurately. It must be recognized that the ROC curves presented in Section 6 were estimated from only two points in the operating characteristic space, and those two points were frequently not located at an optimum position for good curve estimation. Nevertheless, the ROC curves confirm many of the important conclusions arrived at through the analysis of POD curves.

For example, a comparison of Figures 6.20 and 6.22 shows the strong effect of access condition on detection performance. Furthermore, the shape of the ROC curve in Figure 6.22 indicates that far-side inspection in wrought stainless steel is completely ineffective, no matter what decision threshold is employed.

The ROC-curve analysis also provides some information about the adequacy of the decision threshold embodied in procedures meeting the ASME Code requirements up through the 1978 addenda. For example, Figures 6.9 through 6.12 confirm that the C-decision threshold is set at a value that gives nearly optimum results for clad ferritic material; in each figure, the threshold places the "C" values (FCP,POD) in the upper-left hand corner of the operating characteristics diagram.

On the other hand, the ROC curves for cast stainless steel, Figures 6.2 through 6.4 indicate that the decision threshold might be a bit too stringent here; lowering the threshold would increase POD at the expense of a modest increase in FCP. Of course, these ROC curves also demonstrate that any change in the decision criteria cannot improve the cast SS inspection results dramatically.

10.4 Sizing Performance

Table 7.1 summarizes the depth sizing performance for the various materials. For all teams, the standard deviations ranged from a low of 0.037 in. (0.94 mm) to a high of 0.294 in. (7.5 mm), while the RMSE values ranged from a low of 0.047 in. (1.2 mm) to a high of 0.574 in. (14.6 mm). In addition to a large standard deviation in the depth measurements and a large RMSE, substantial bias also existed in the measurements. Small cracks tended to be oversized and large cracks undersized. Although there were obvious differences in performance between materials, no material displayed accurate depth sizing results, including clad ferritic. All the teams utilized probe motion measurements, and the results support the ineffectiveness of this method for depth sizing of the cracks. Two teams (Team 2 and Team 6) did display positive sizing ability on the clad ferritic and wrought stainless steel. Although their sizing measurements were significantly better than one could do by guessing, their measurements contain significant errors.

Table 7.2 summarizes team length performance. Although one would expect length sizing to be an easier task than depth sizing, the results in Table 7.2 are not markedly better than those in Table 7.1. The standard deviation of the length measurements varies from a low of 0.236 in. (5.8 mm) to a high of 2.152 in. (54.7 mm). The RMSE ranged from a low of 0.4 in. (10.2 mm) to a high of 2.23 in. (56.6 mm). As with depth measurements, we find that the length measurements are biased, with small cracks being oversized and large cracks undersized. Even in the case of clad ferritic, results were disappointing. The average standard deviation was approximately 0.9 in. (23 mm) and the average slope of the regression line was only 0.695. In reviewing this performance analysis, it was concluded that flaw length sizing is a more difficult task than what had been expected. For short flaws the length sizing is difficult because the insonification field is larger than the flaw. This causes the short flaws to be oversized. For the longer flaws the difficulty occurs because near the flaw extremes the flaw is tight (low acoustic response) and is tapering away (the crack cross section is wedge shaped) thus, making it difficult to determine the end of the flaw and generally, leading to an undersizing. It is recommended that in order to improve the performance of length sizing, the sizing must be performed to

the loss of signal for long cracks, inspectors need to have training in making these measurements and care must be taken when making these measurements.

10.5 Team Error Statistics

The objective for this part of the study was to identify the important inspection errors made by the teams and to tabulate their frequency. Therefore during the inspections, extensive monitoring was performed to permit the identification of the errors made by each team. All data was manually reviewed to identify the errors made. These errors were then categorized and tabulated. The major errors of interest were those that lead to a missed crack, those that created a large sizing error and those that created a large location error.

Table 8.3 provides a summary of the inspection errors as classified by consequence. The first thing to note is that there is on average a 0.019 rate for a team to miss a defect because of a inspection error. This statistic indicates that the probability of detection may be no higher than 0.98, even for large cracks. This indicates that more attention needs to be made to human factors in order to lower this important error rate and increase the probability of crack detection.

Table 8.3 also indicates that the chance of some large sizing or location error occurring during an inspection is 0.048. This makes it more difficult to use inservice inspection results to monitor crack growth over time, because there is almost a 5% chance of a large sizing or location error on any individual measurement.

11.0 References

- Agresti, A. 1990. *Categorical Data Analysis*, John Wiley, New York, NY.
- Bates, D. J., S. R. Doctor, P. G. Heasler, and E. Burck. 1987. *Stainless Steel Round Robin Test - Centrifugally Cast Stainless Steel Screening Phase*, NUREG/CR-4970, PNL-6266, PISC III Report No. 3. Pacific Northwest Laboratory, Richland, Washington.
- Berens, A. P., and P. W. Hovey. 1981. *Evaluation of NDE Reliability Characterization*, AFWAL-TR-81-4160, University of Dayton Research Institute.
- Berens, A. P., and P. W. Hovey. 1984. *Flaw Detection Reliability Criteria*, AFWAL-TR-84-4022, University of Dayton Research Institute.
- Bishop, Y. M., S. E. Fienberg, and P. W. Holland. 1976. *Discrete Multivariate Analysis*, MIT Press.
- 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.
- Bush, S. H. 1985. "Statistics of Pressure Vessel and Piping Failures," *Decade of Progress in Pressure Vessel Technology*. American Society of Mechanical Engineers, p. 875.
- 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.
- Heasler, P. G., T. T. Taylor, J. C. Spanner, S. R. Doctor, and J. D. Deffenbaugh. 1990. *Ultrasonic Inspection Reliability for Intergranular Stress Corrosion Cracks: A Round Robin Study of the Effects of Personnel, Procedures, Equipment and Crack Characteristics*. NUREG/CR-4908. 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.
- Nichols, R. W., and S. Crutzen (Eds.). 1988. *Ultrasonic Inspection of Heavy Section Steel Components, The PISC II Final Report*. Elsevier Applied Science Publishers, London, England.
- Swets, J. A. 1983. "Assessment of NDT Systems - Part 1: The Relationship of True and False Detection," *Materials Evaluation*, Vol. 41.
- Swets, J. A., and R. M. Pickett. 1982. *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*, Academic Press.
- 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.

Appendix A

Hit-O-Grams for PIRR Specimens

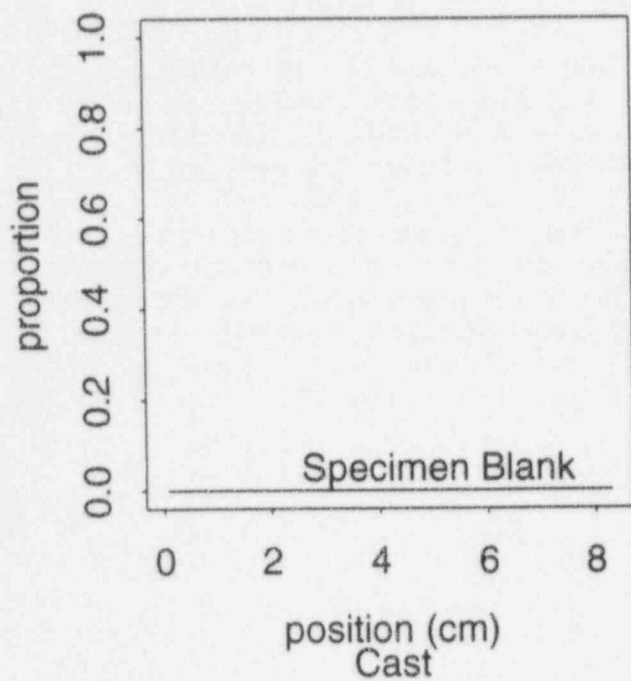
Appendix A

Hit-O-Grams for PIRR Specimens

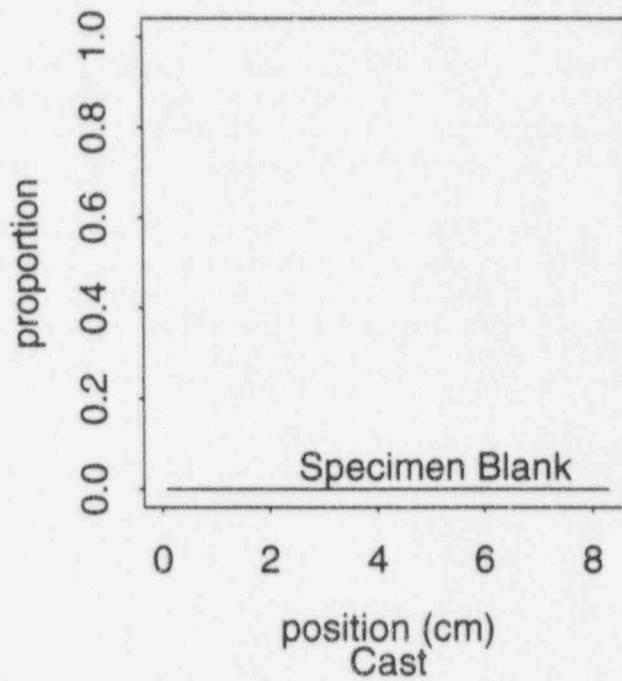
This appendix presents the "hit-o-grams" for each specimen of the PIRR. Hit-o-grams provide a sort of instantaneous probability of detection for each unit of weld inspected. At each point along the weld, the number of inspections that recorded a crack indication is divided by the total number of inspections performed on the weld. This results in a number between 0 and 1 that is quite analogous to a probability of detection calculated with a grading unit of length zero.

Each hit-o-gram is identified with the specimen under consideration, the side being inspected (A or B), and the type of inspection access (near or far). The hit-o-gram title also lists the total number of inspections carried out under the specified conditions. In addition to the hit-o-gram curve, the main plot will also contain horizontal lines that indicate the position and depth of any existing cracks in the specimen. The depth is expressed in terms of proportion of through wall.

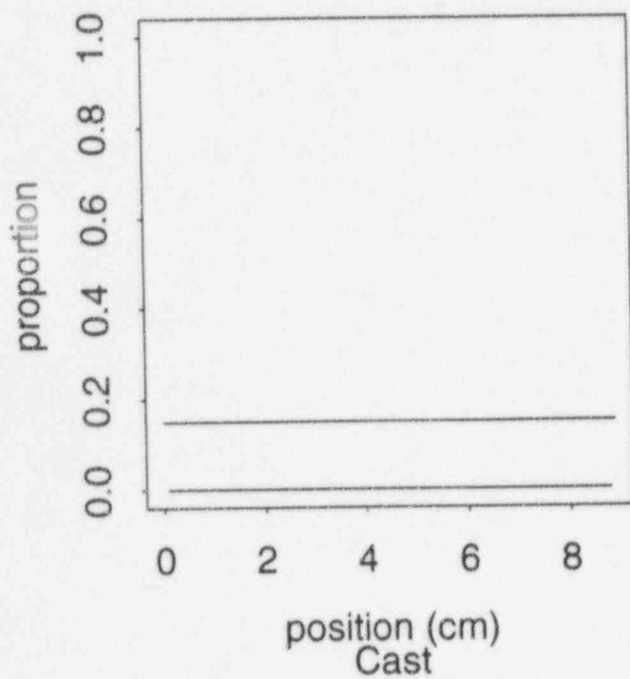
1 A near # insp= 2



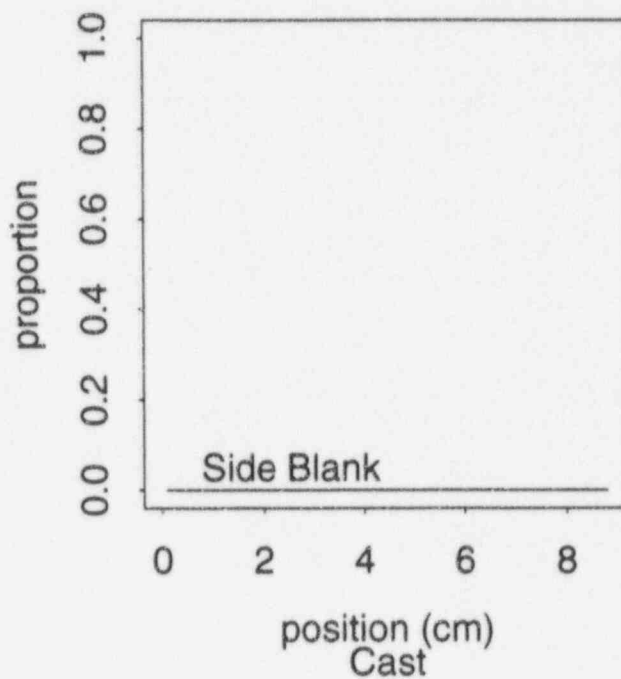
1 B near # insp= 1



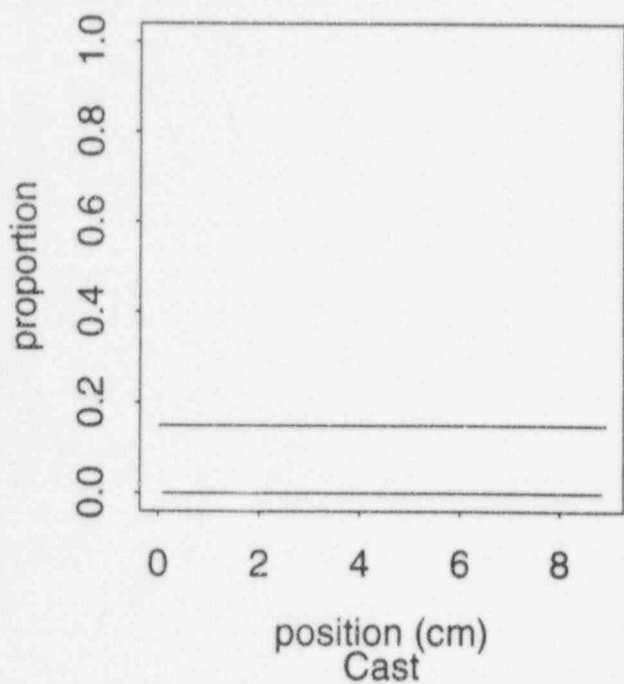
12 A near # insp= 2



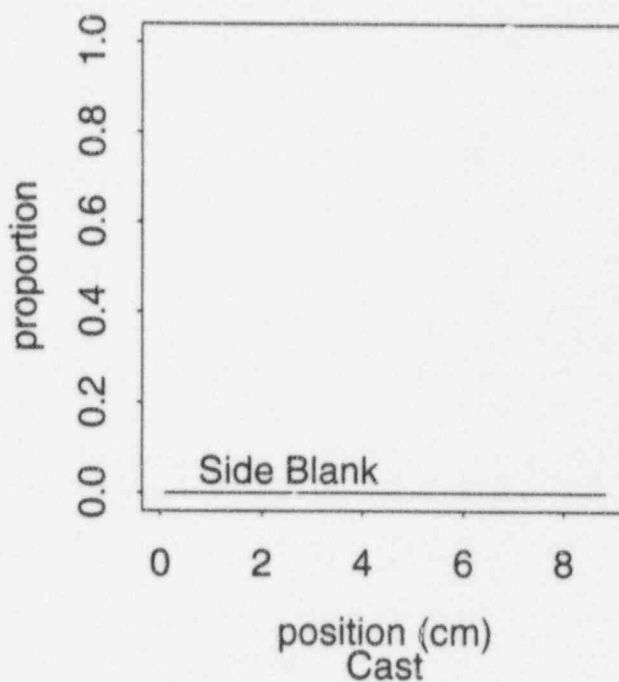
12 B near # insp= 2



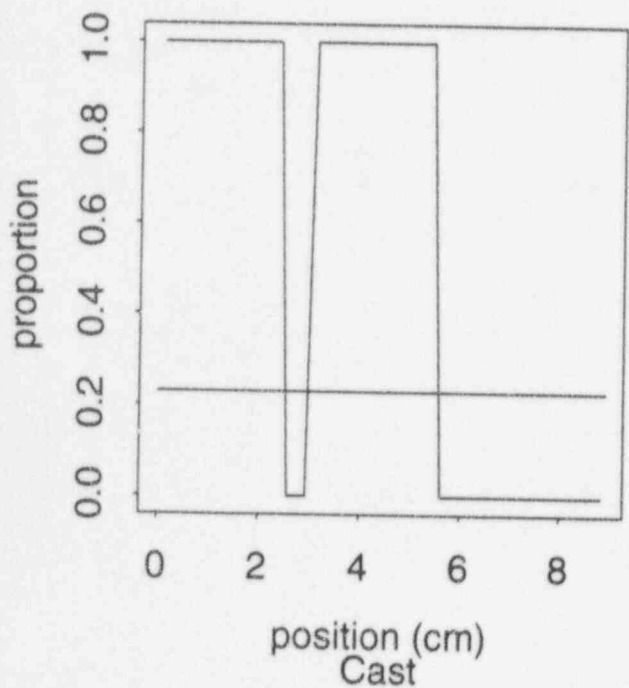
13 A near # insp= 1



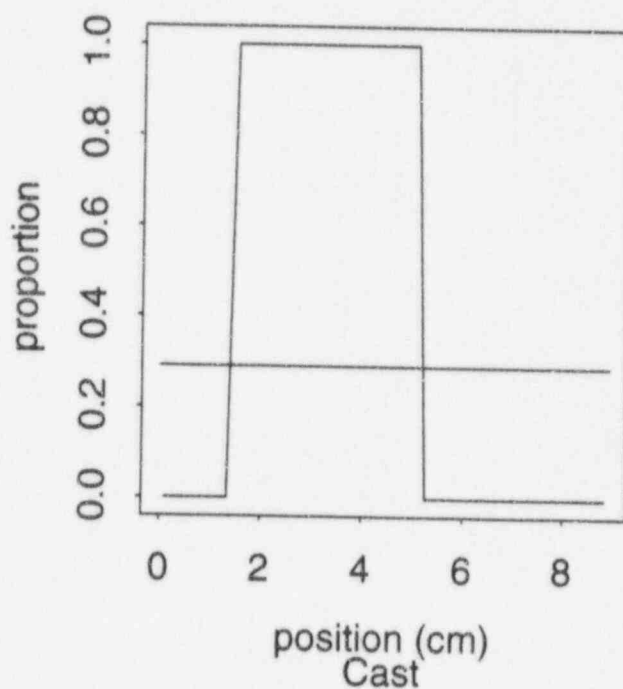
13 B near # insp= 1



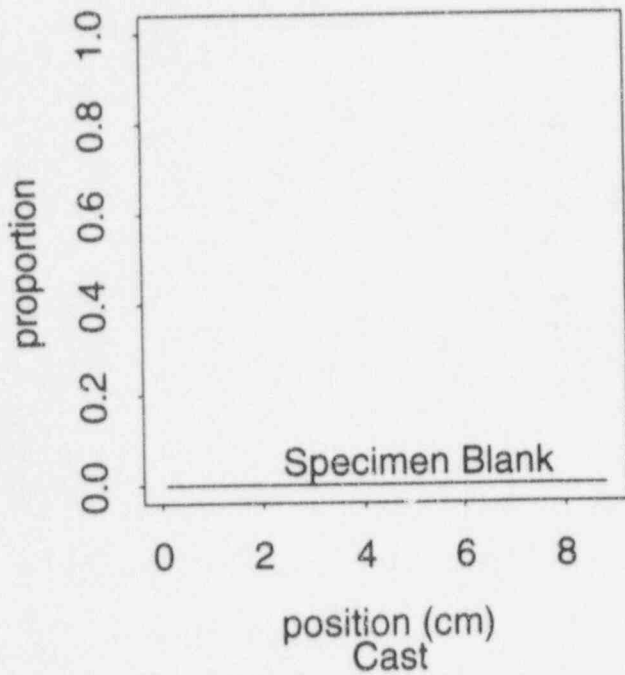
16 A near # insp= 1



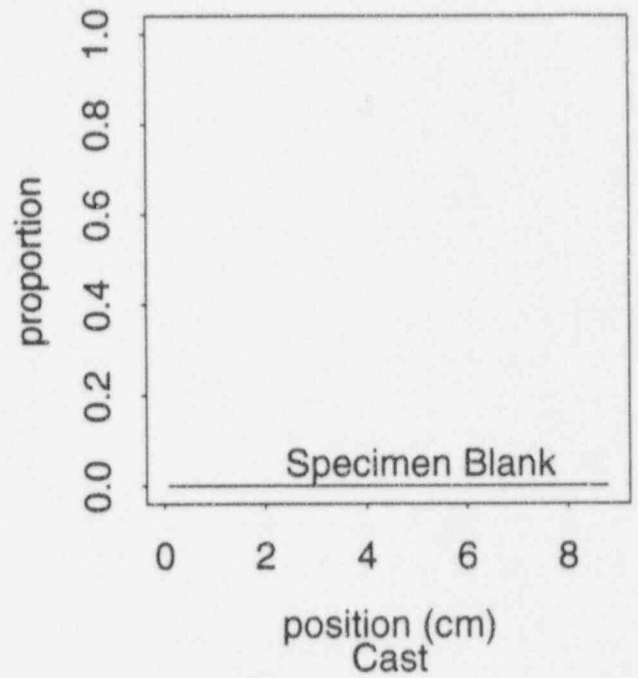
19 A near # insp= 1



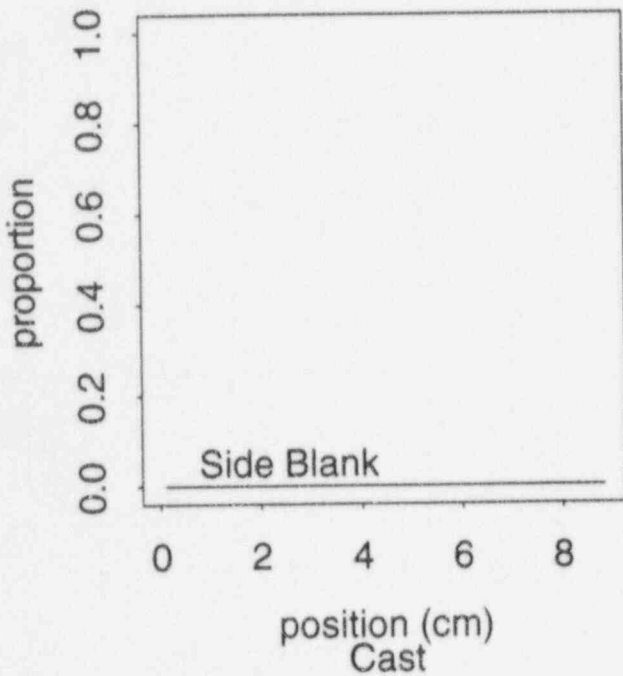
2 A near # insp= 1



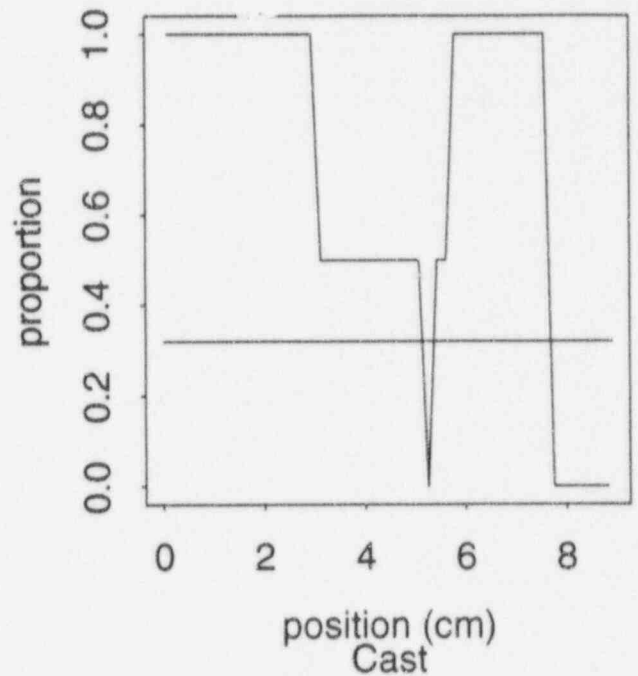
2 B near # insp= 1



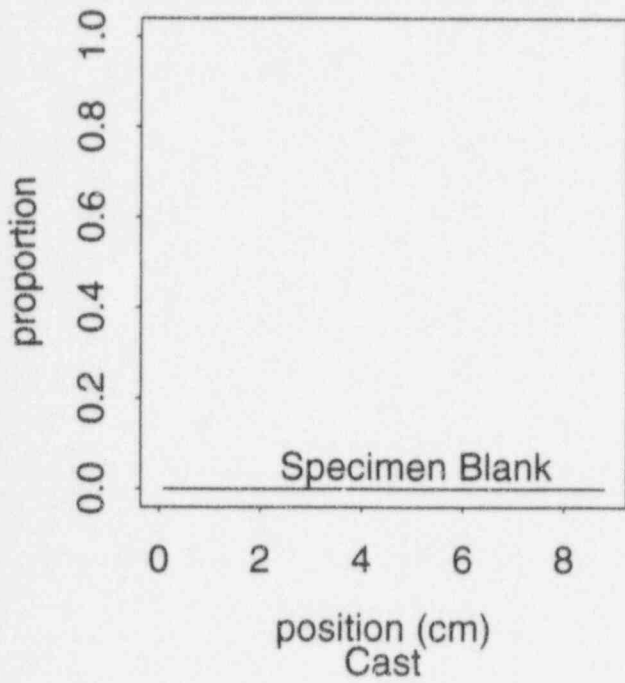
19 B near # insp= 1



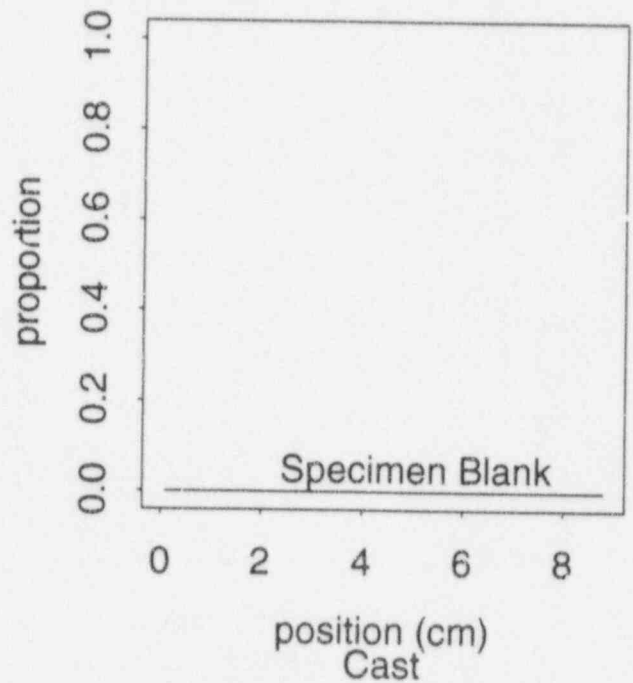
20 A near # insp= 2



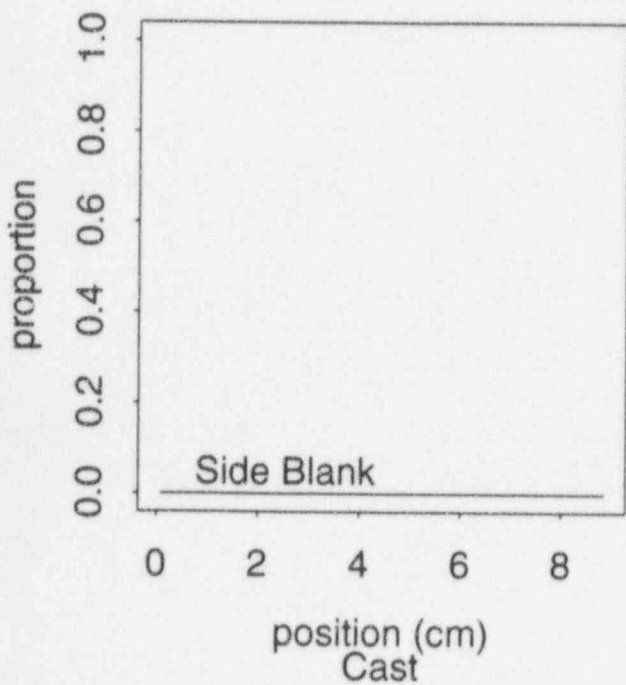
3 A near # insp= 1



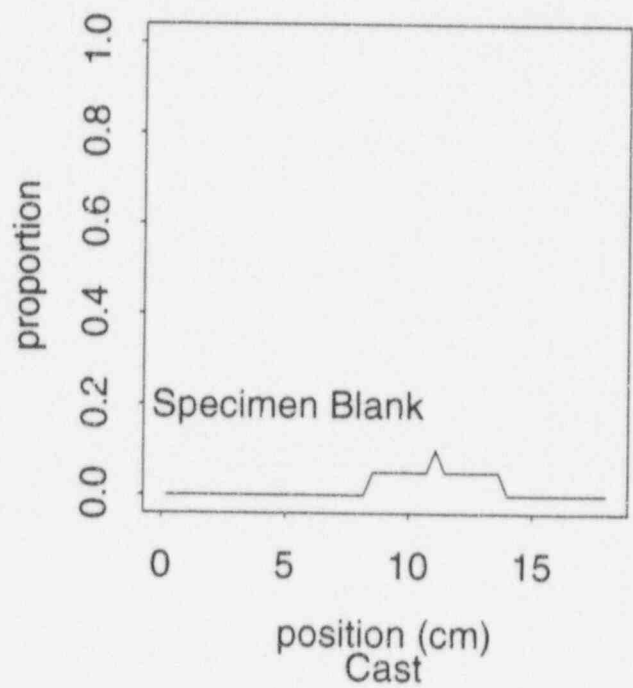
3 B near # insp= 2



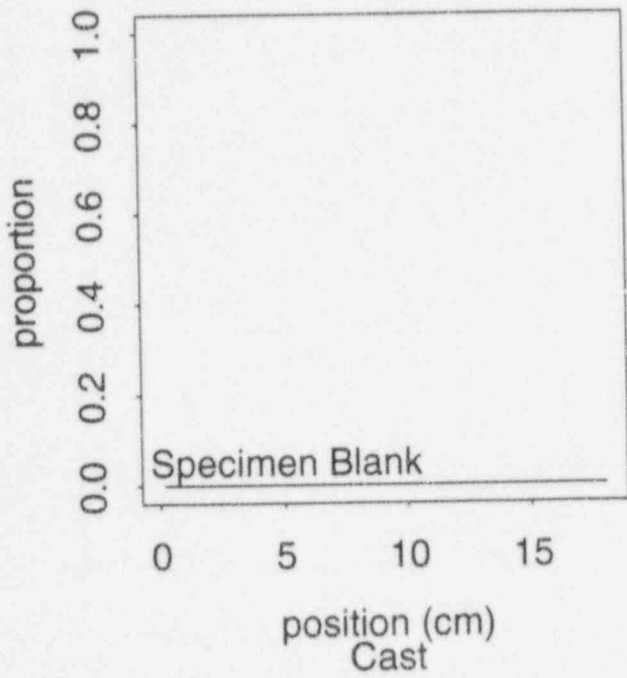
20 B near # insp= 2



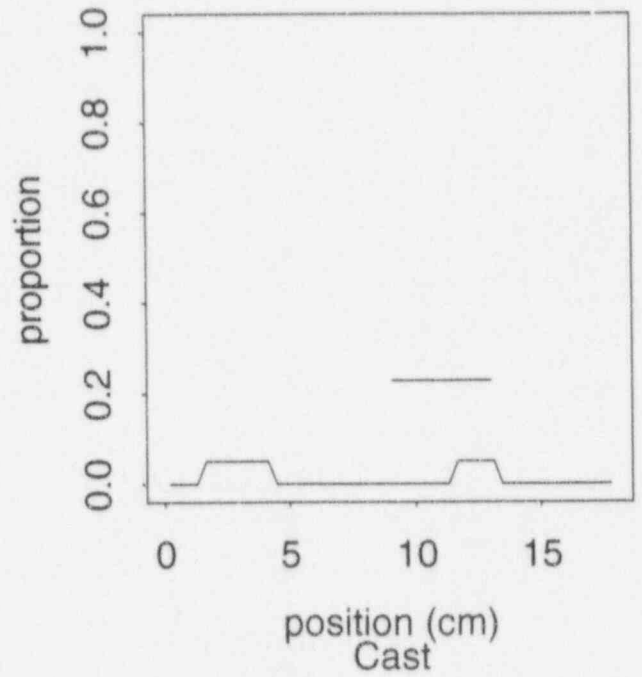
B501 A near # insp= 20



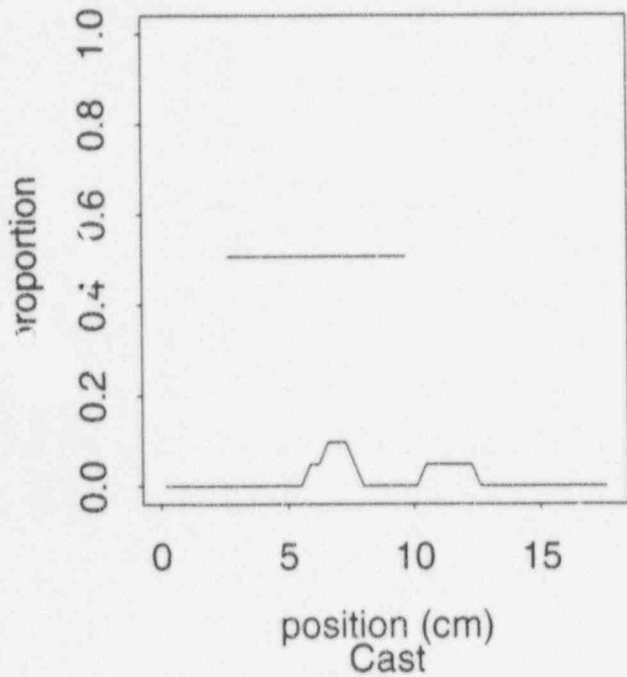
B501 B near # insp= 5



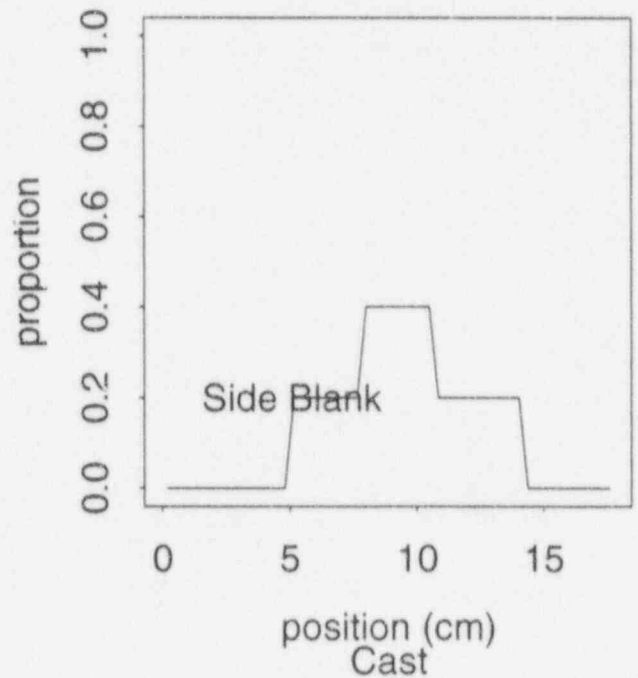
B502 A near # insp= 20



B504 A near # insp= 21

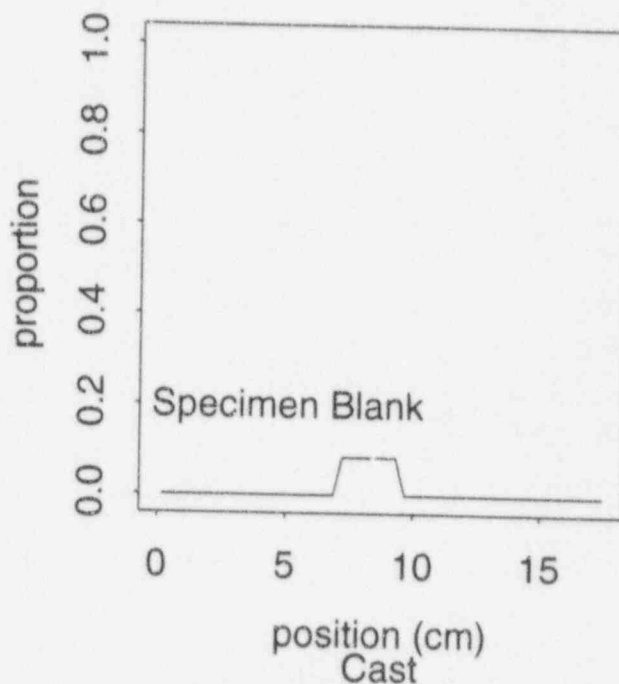
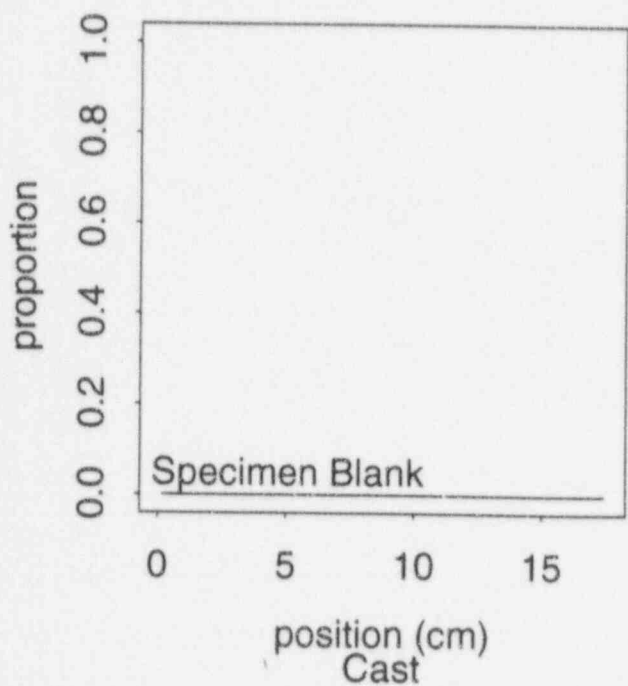


B504 B near # insp= 5



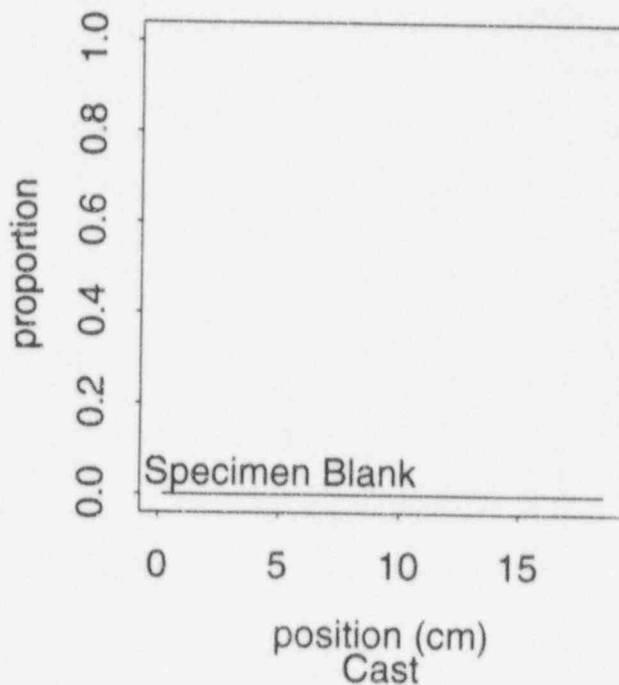
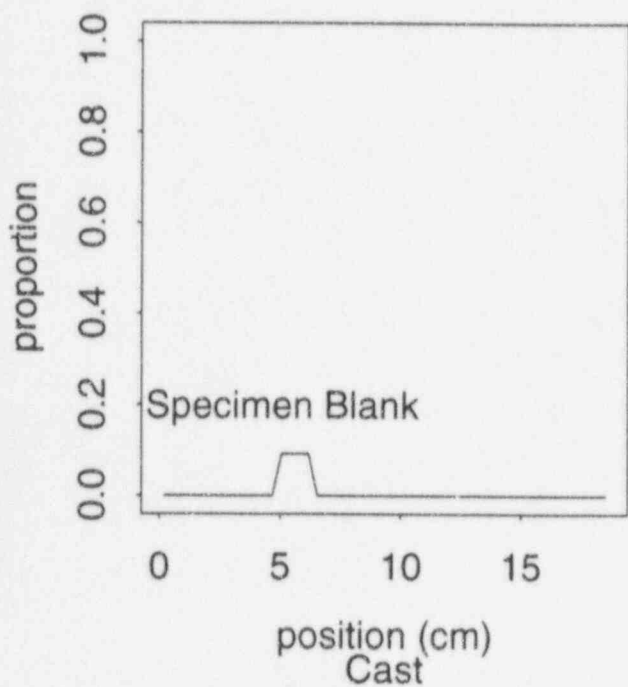
B506 A near # insp= 5

B506 B near # insp= 12

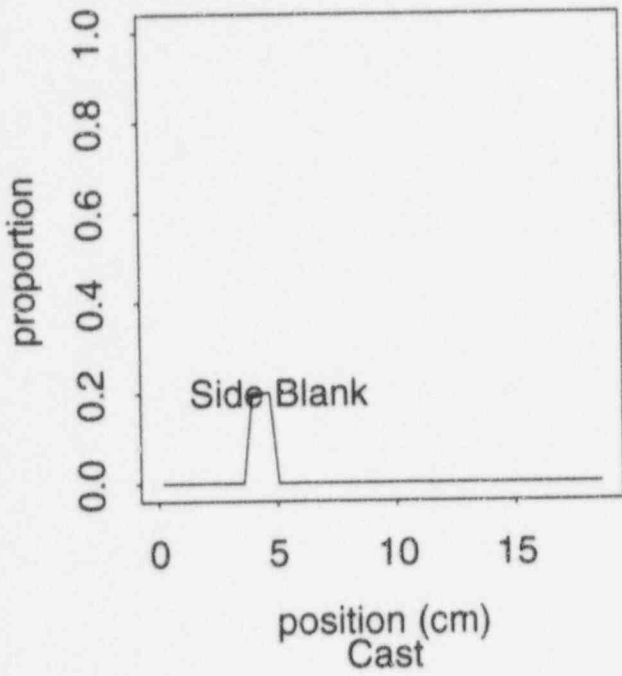


B507 A near # insp= 11

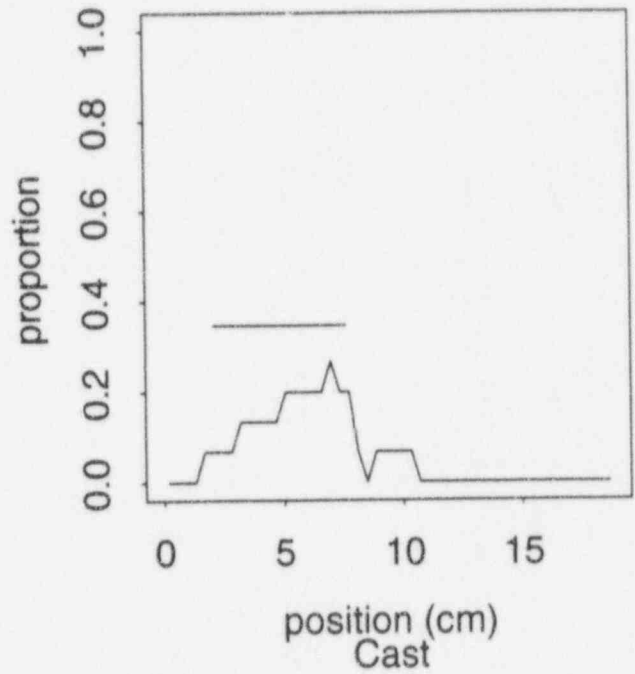
B507 B near # insp= 5



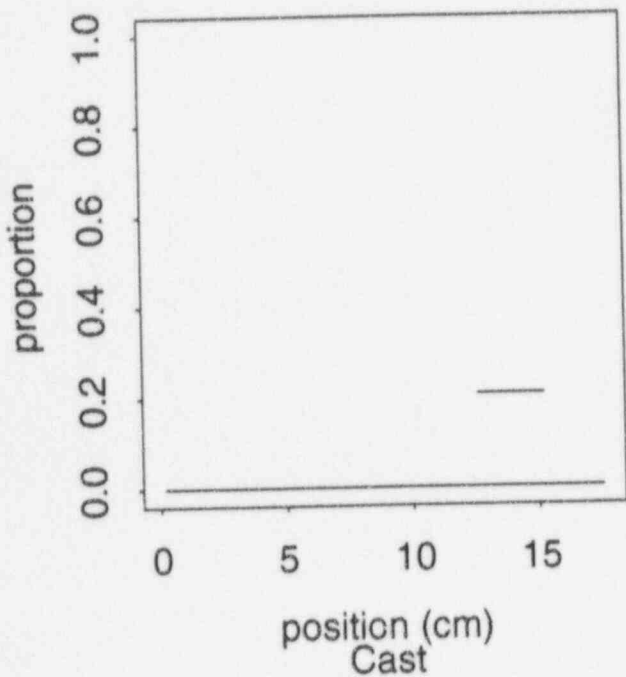
B508 A near # insp= 5



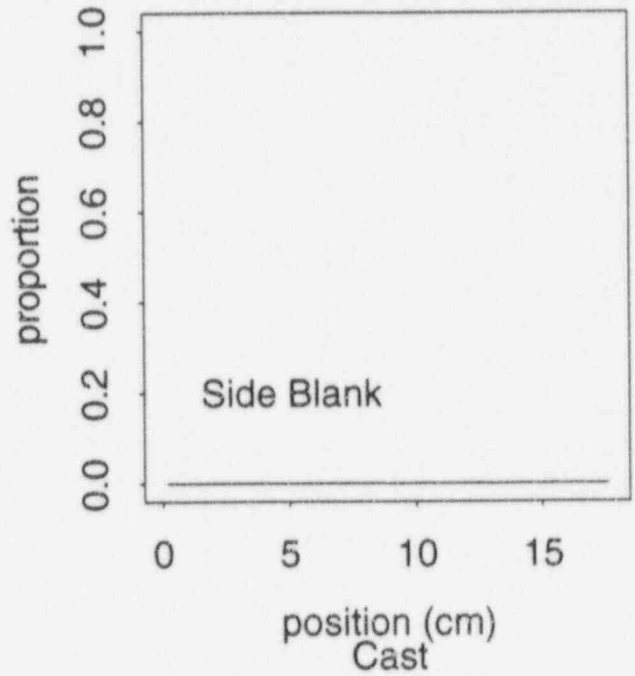
B508 B near # insp= 15



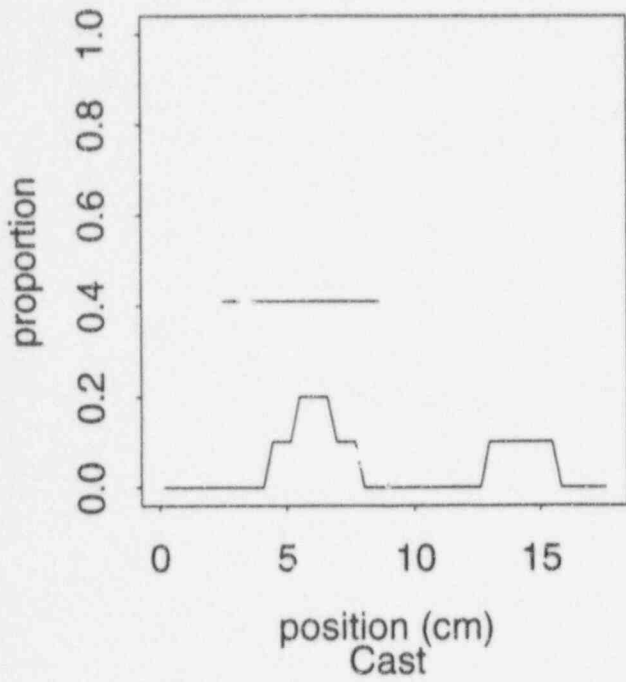
B509 B near # insp= 10



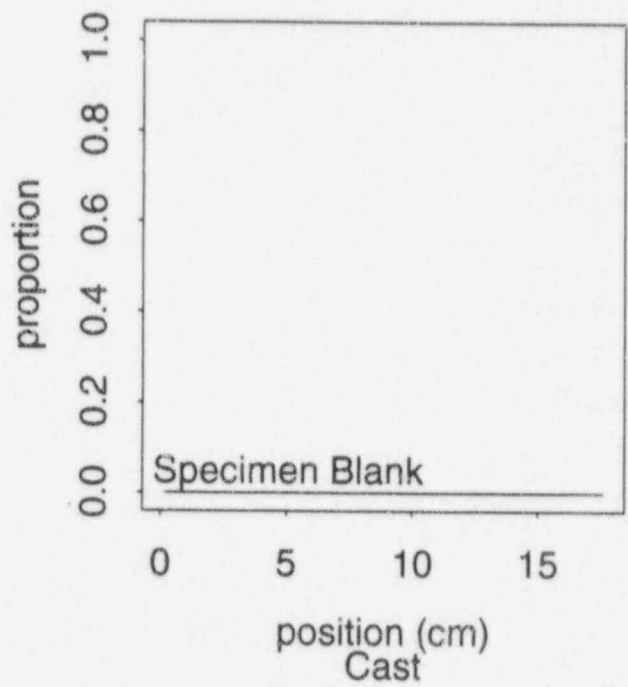
B510 A near # insp= 5



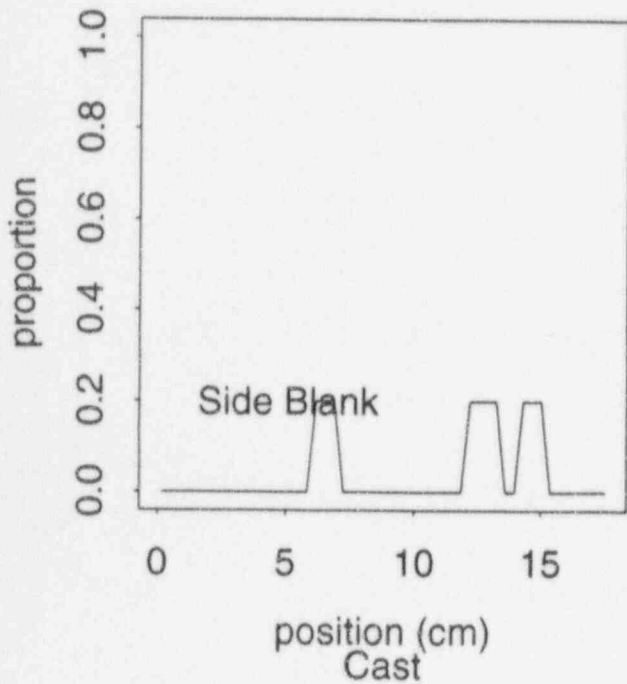
B510 B near # insp= 10



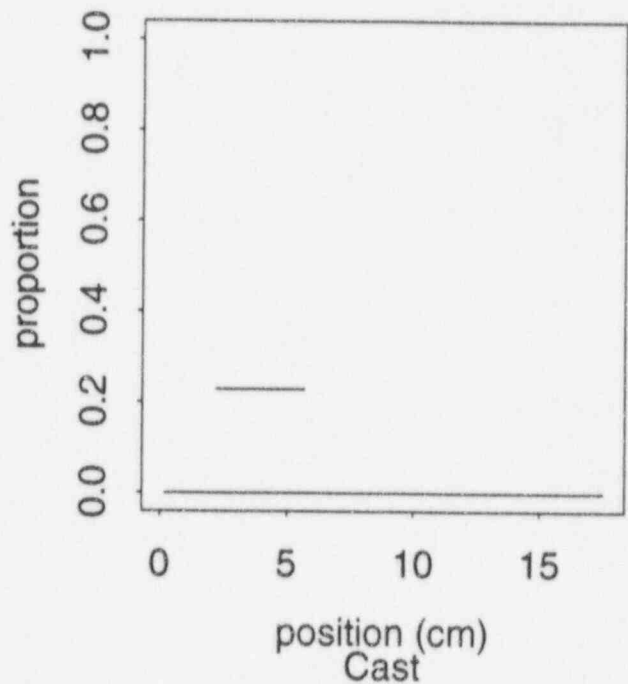
B511 B near # insp= 5



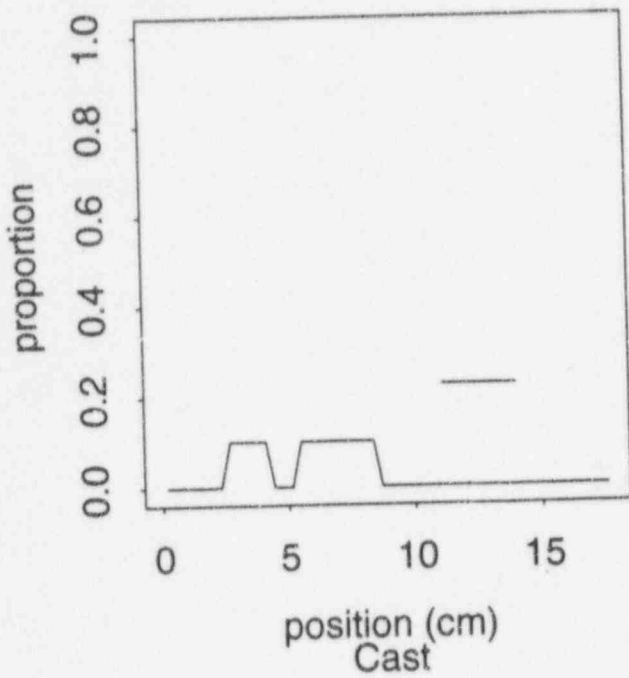
B512 A near # insp= 5



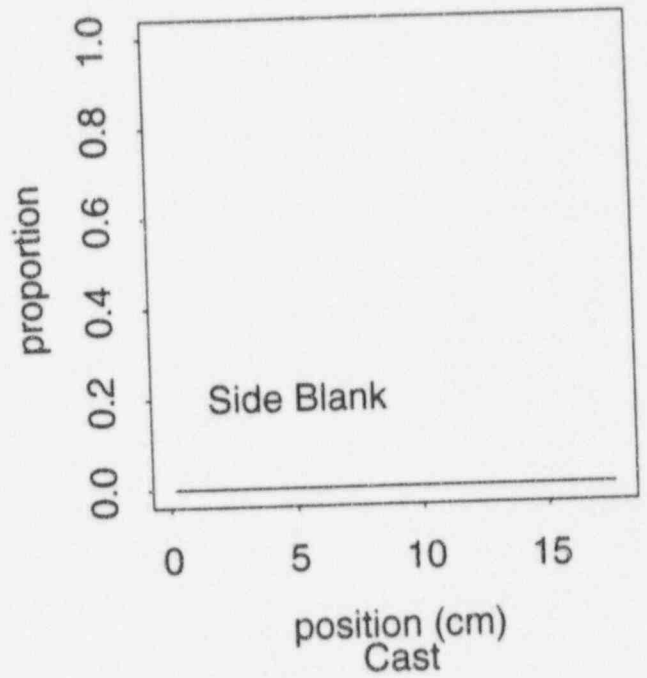
B512 B near # insp= 21



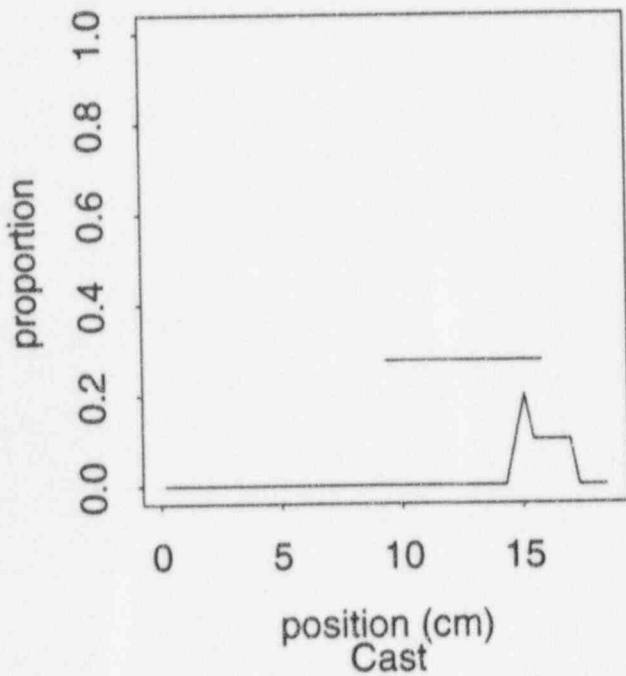
B513 A near # insp= 10



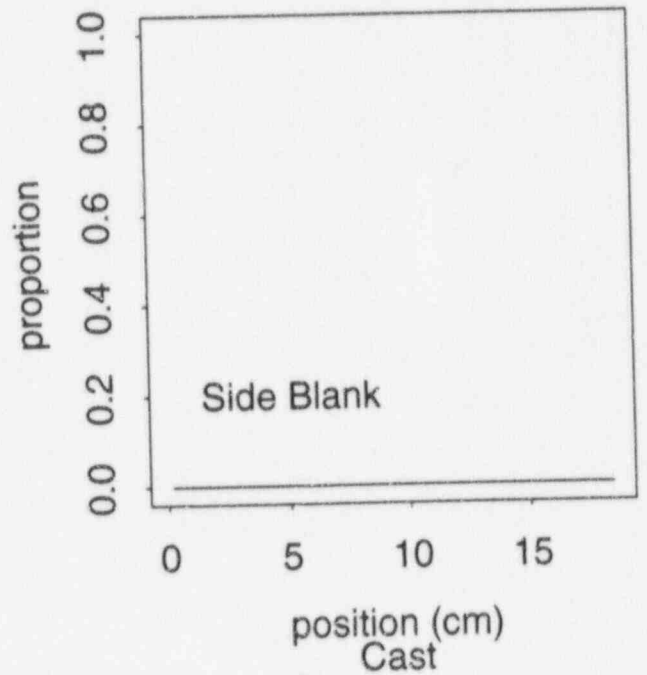
B513 B near # insp= 5



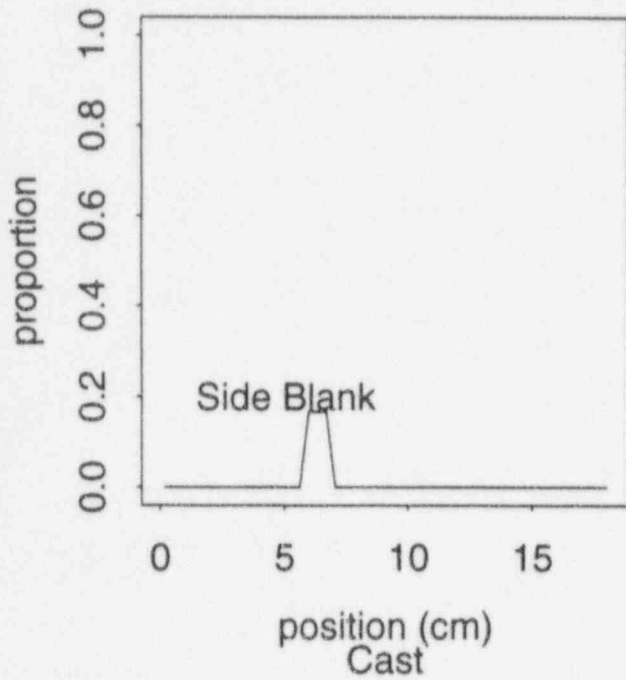
B514 A near # insp= 10



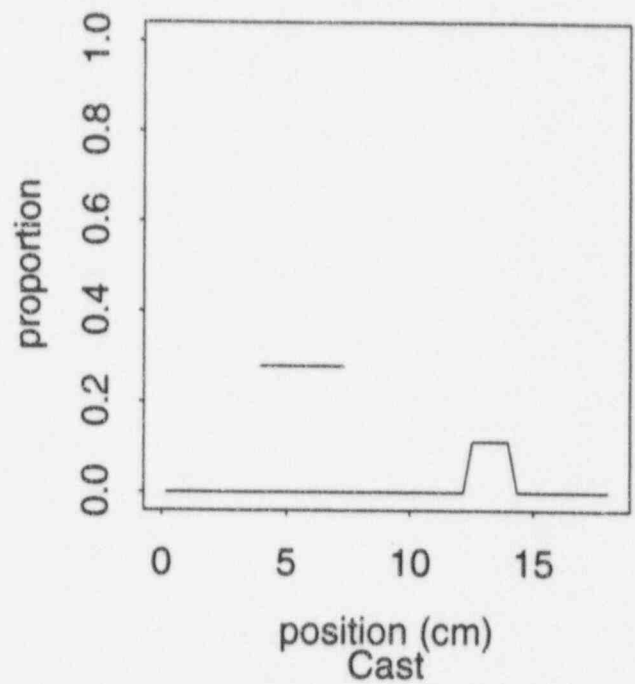
B514 B near # insp= 6



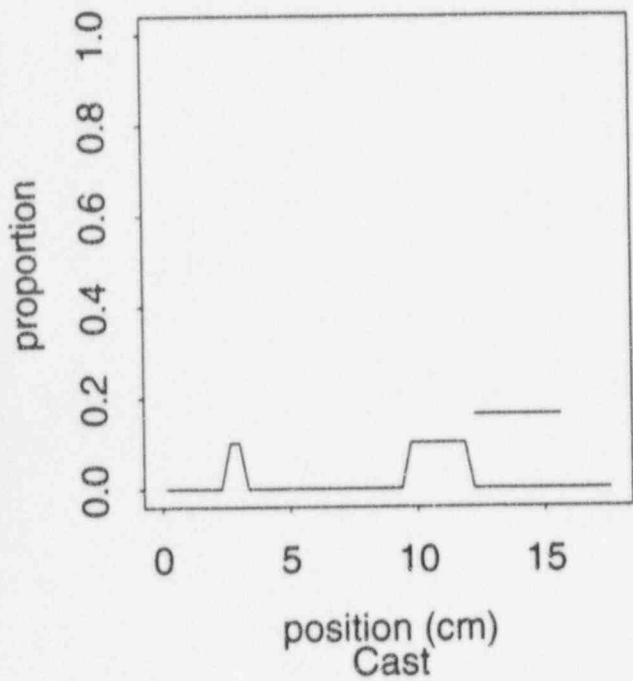
B515 A near # insp= 6



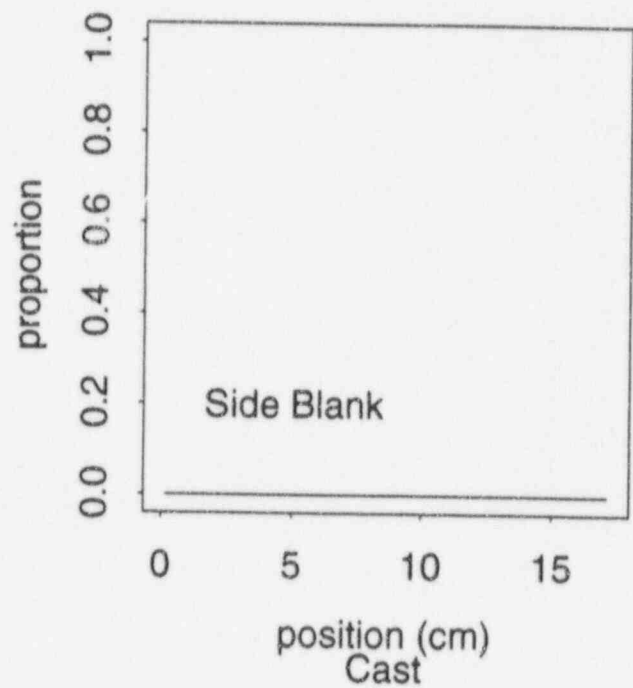
B515 B near # insp= 9



B517 B near # insp= 10

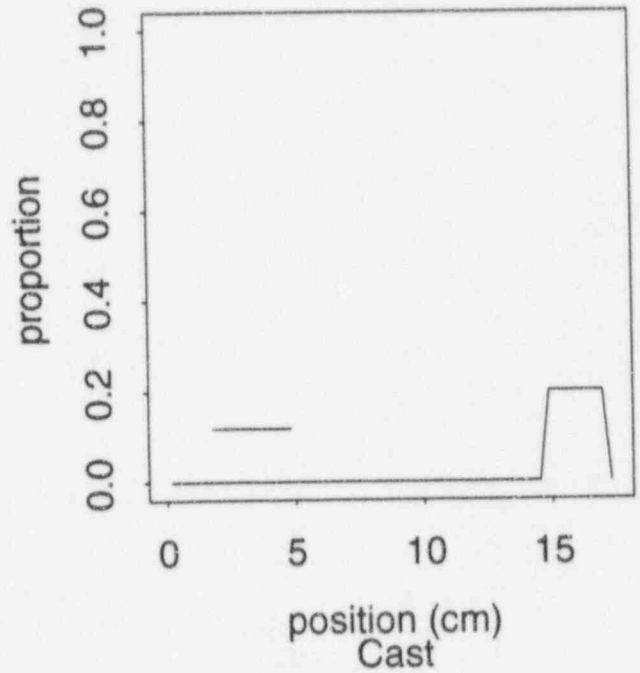
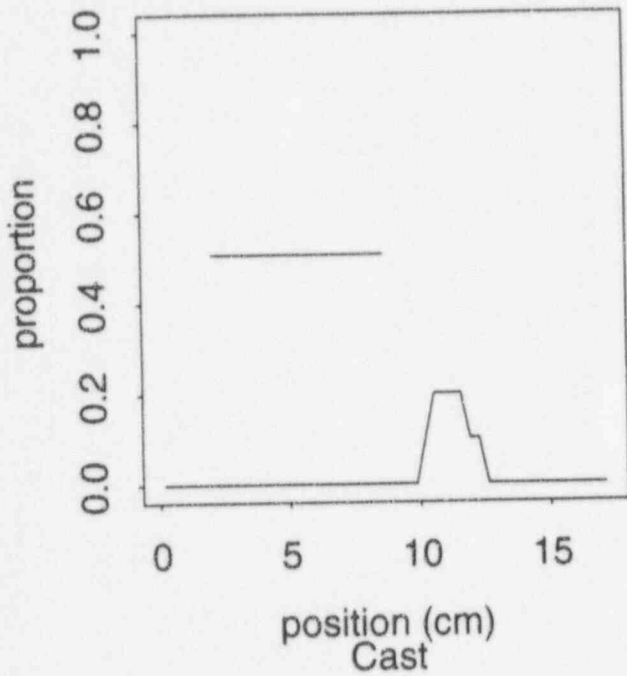


B519 A near # insp= 5



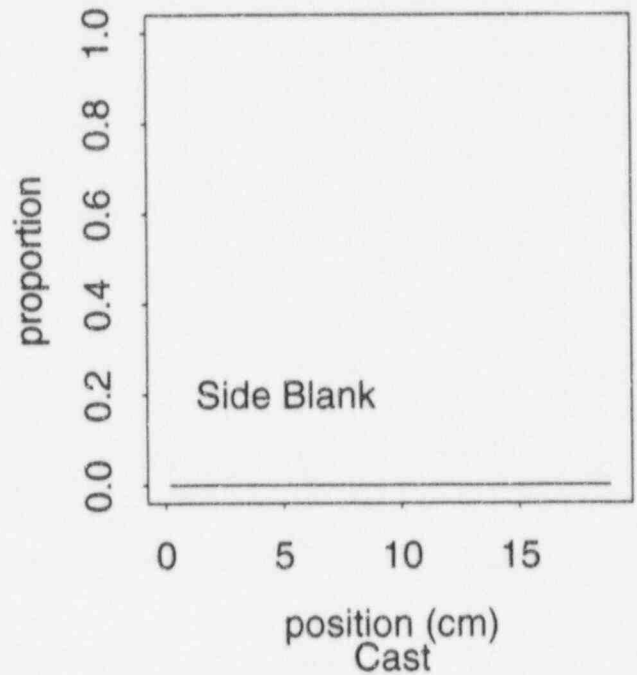
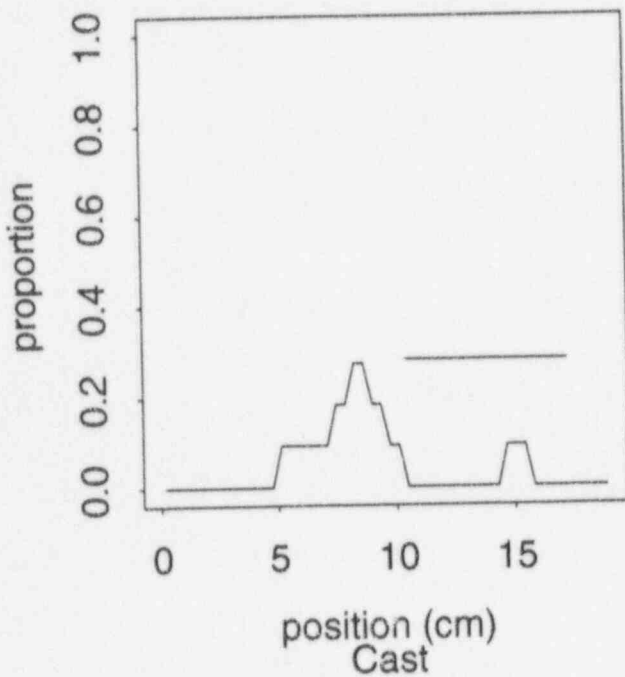
B519 B near # insp= 10

B520 A near # insp= 5

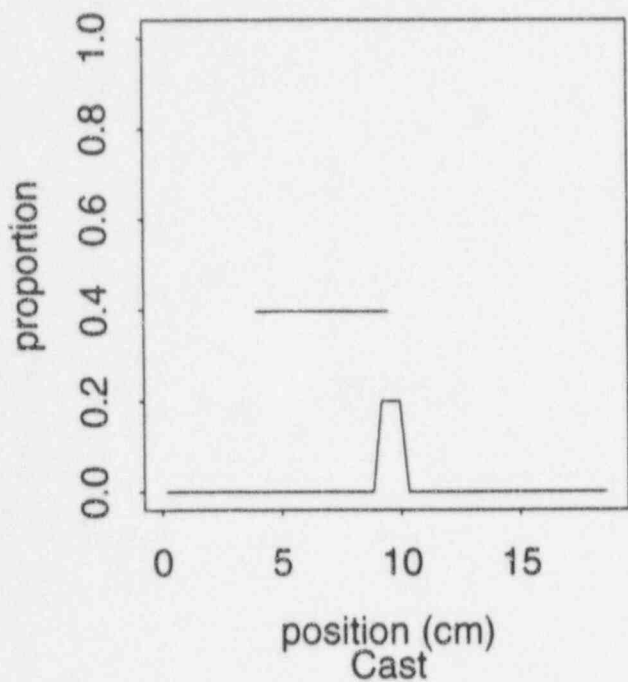


B521 A near # insp= 11

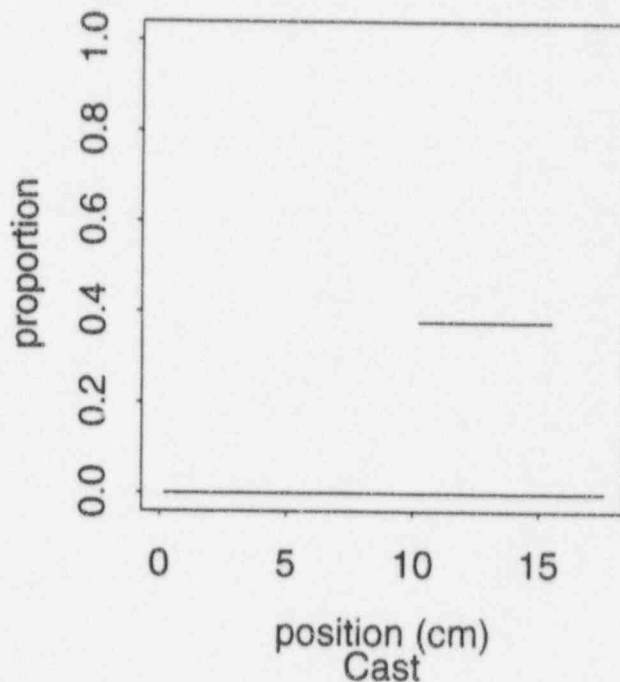
B521 B near # insp= 5



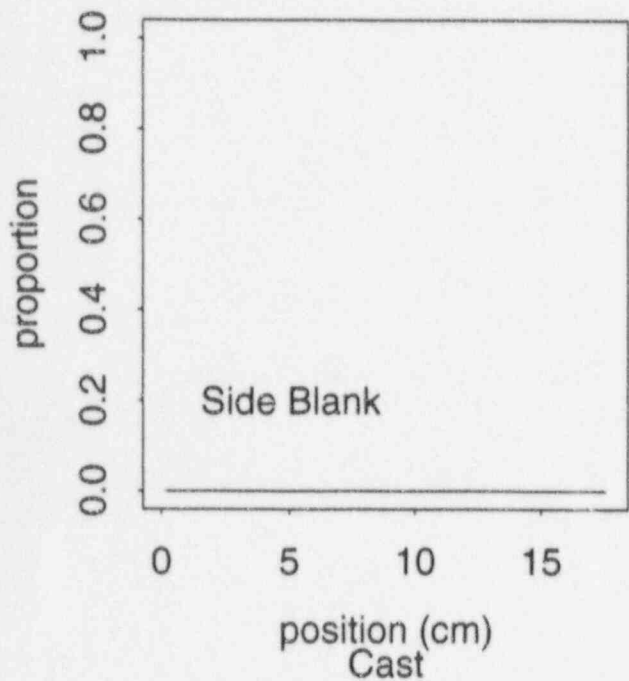
B522 A near # insp= 5



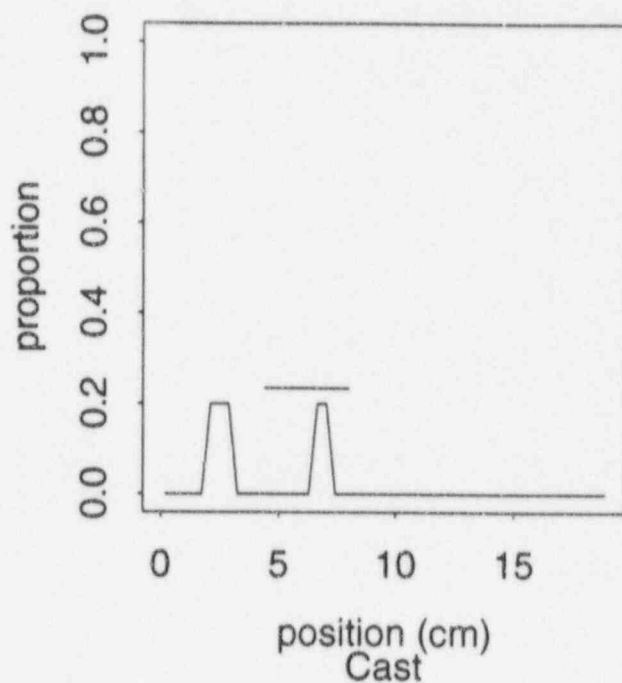
B523 A near # insp= 5



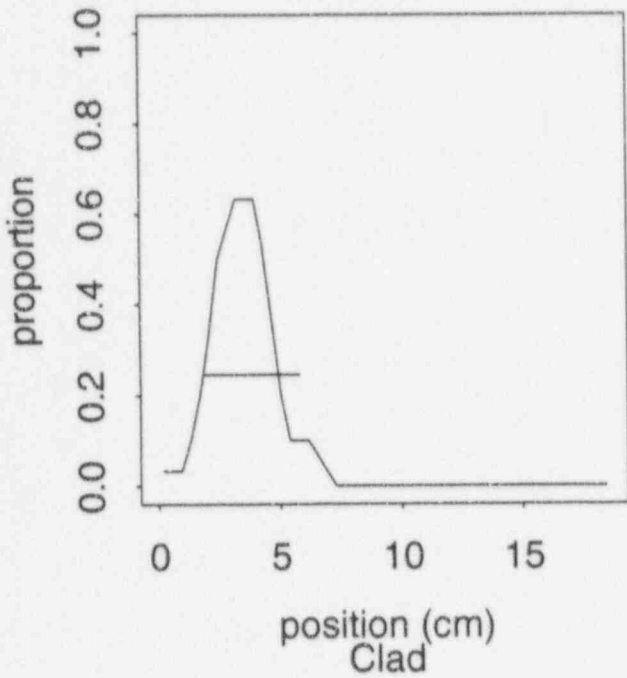
B523 B near # insp= 1



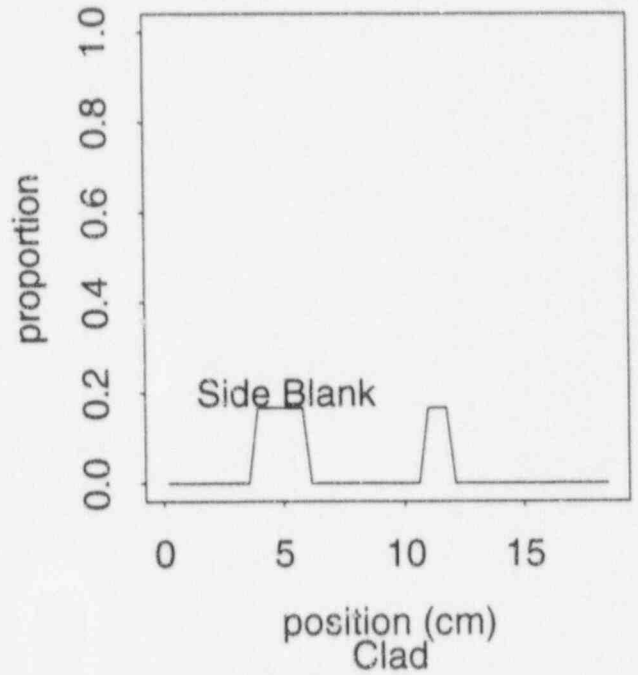
B528 B near # insp= 5



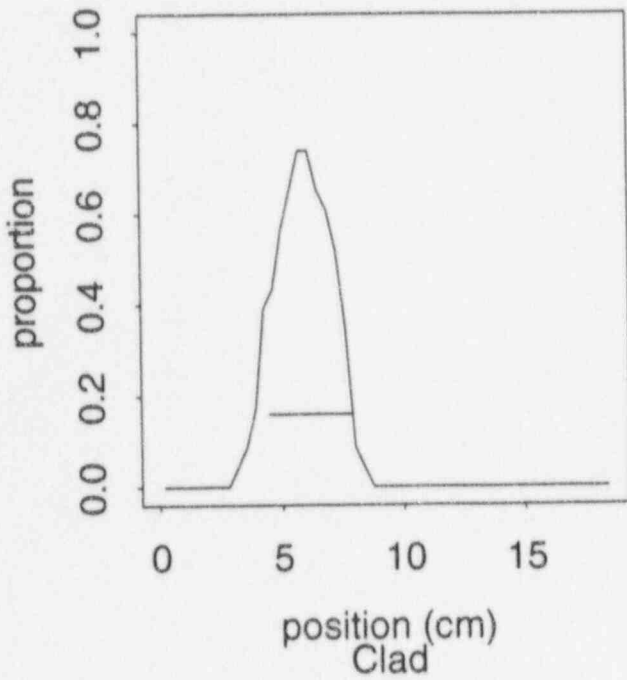
B601 A near # insp= 30



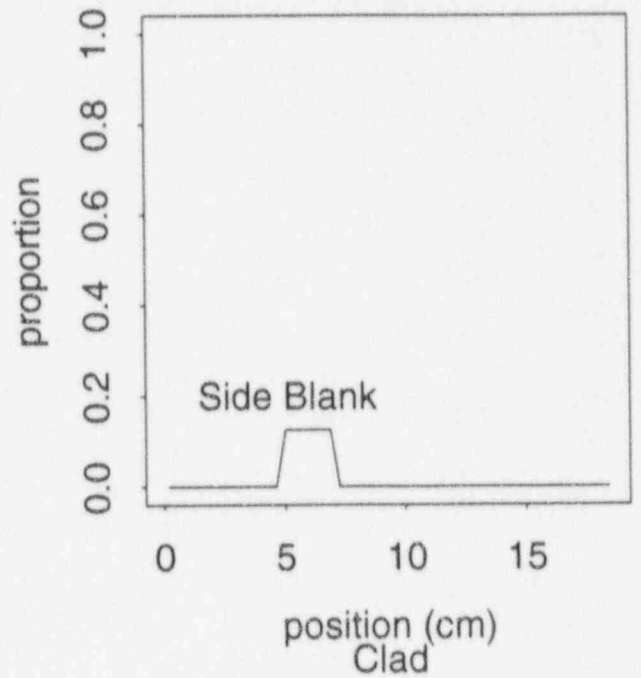
B601 B near # insp= 6



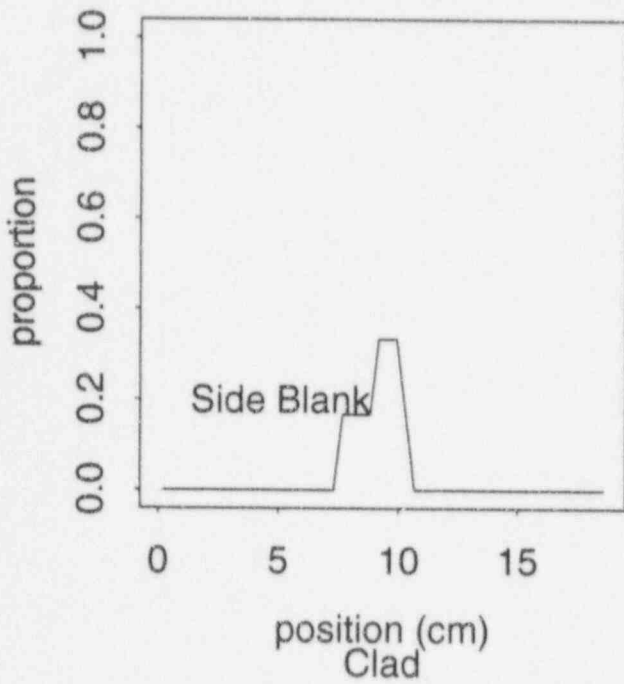
B602 A near # insp= 23



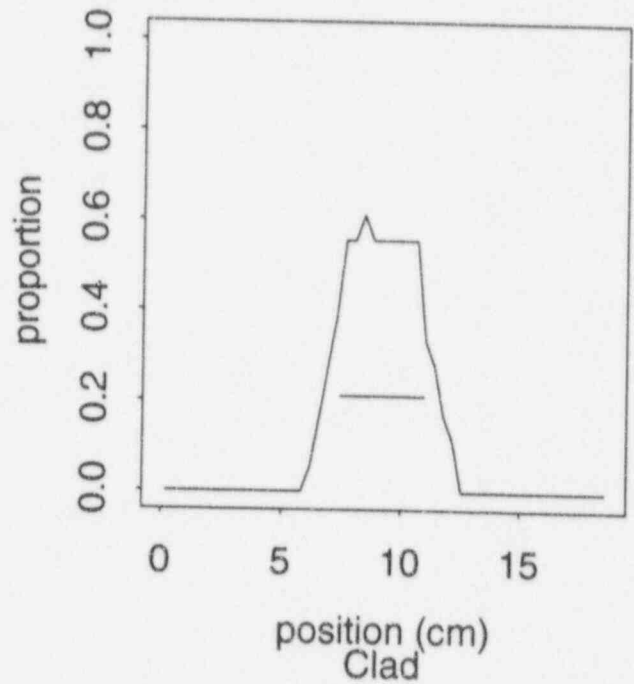
B602 B near # insp= 8



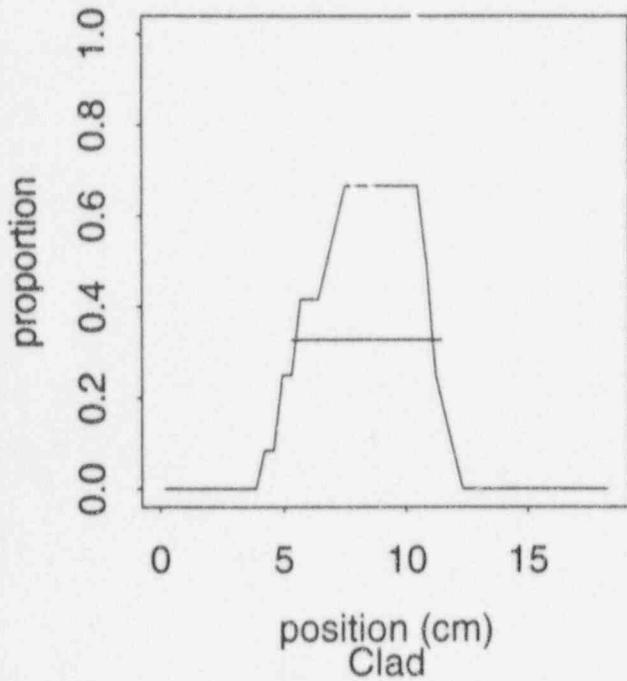
B603 A near # insp= 6



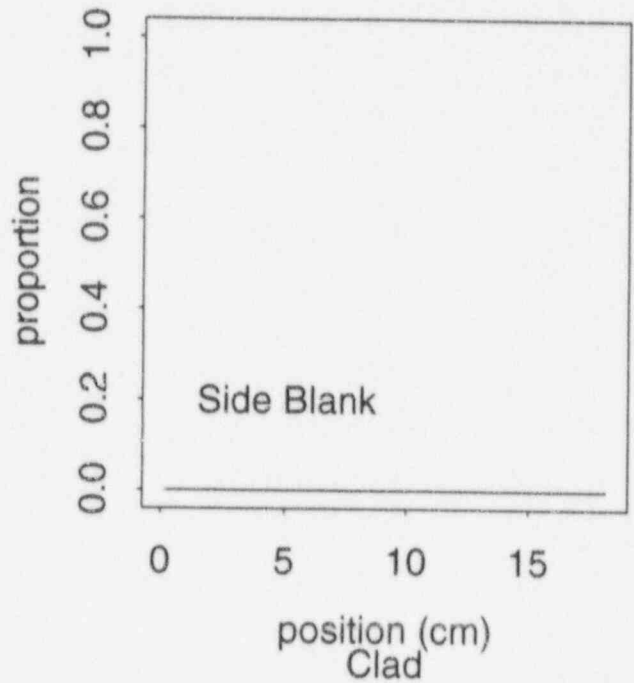
B603 B near # insp= 18



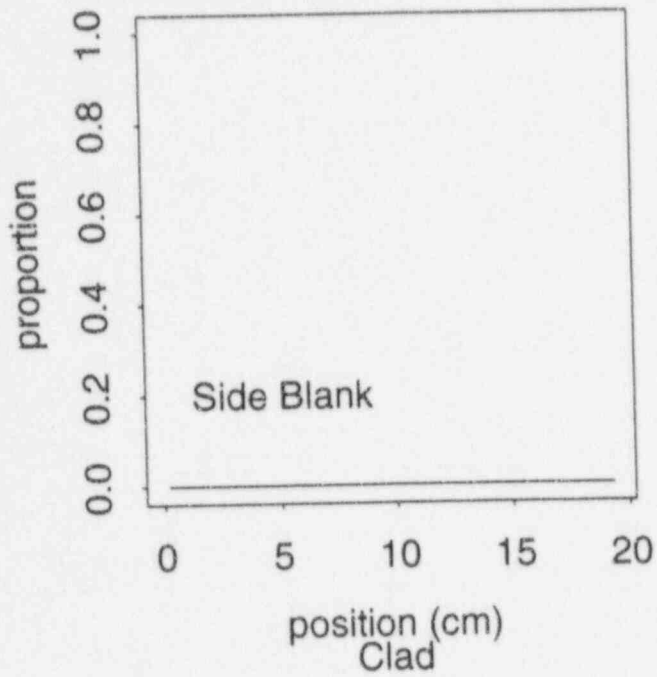
B604 A near # insp= 12



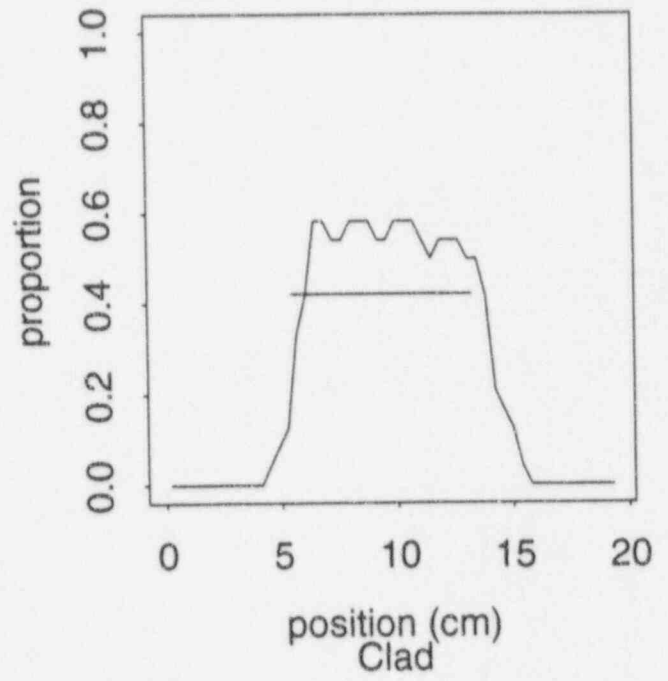
B604 B near # insp= 1



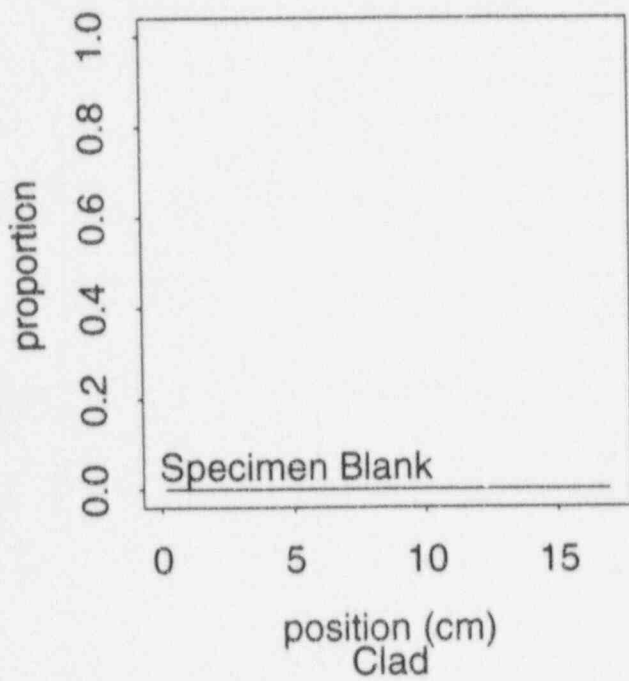
B605 A near # insp= 5



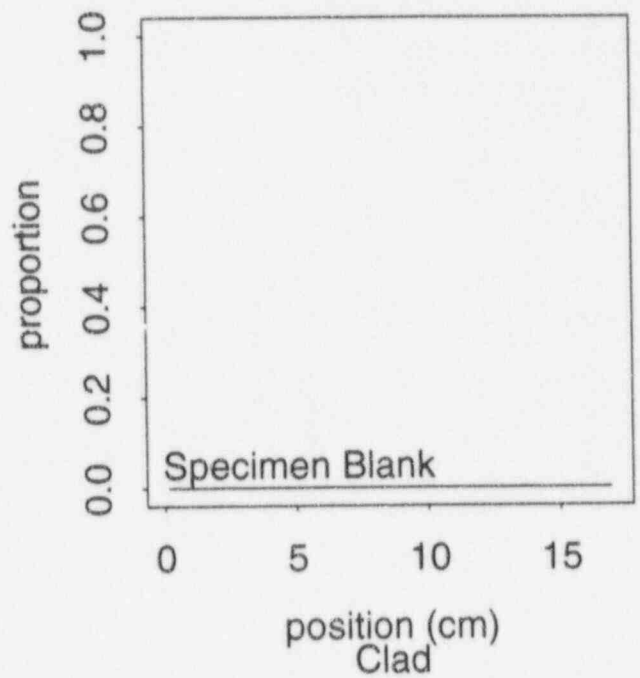
B605 B near # insp= 24



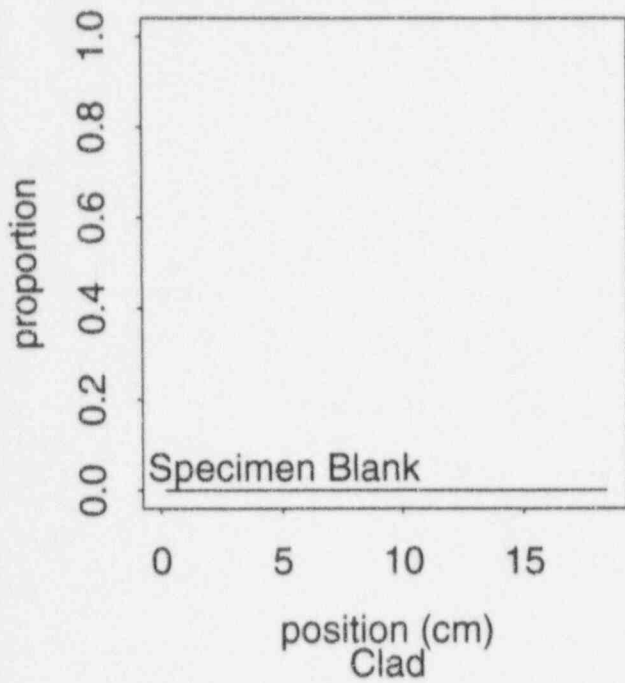
B607 A near # insp= 23



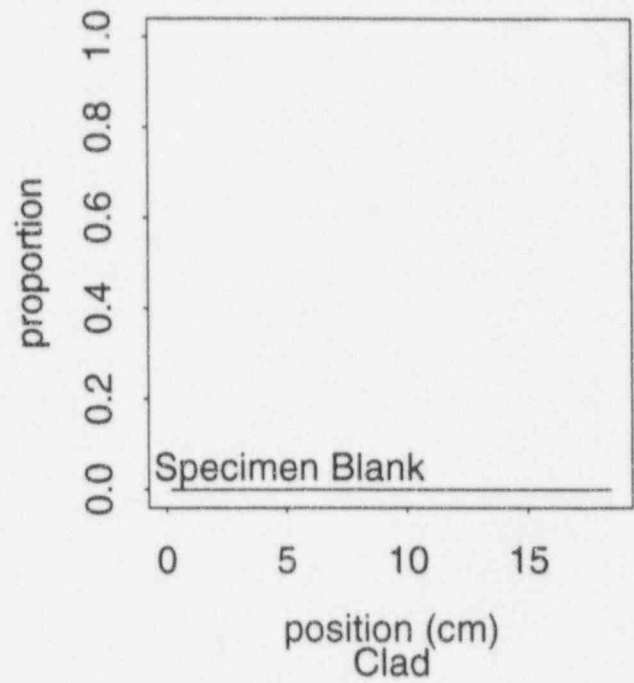
B607 B near # insp= 6



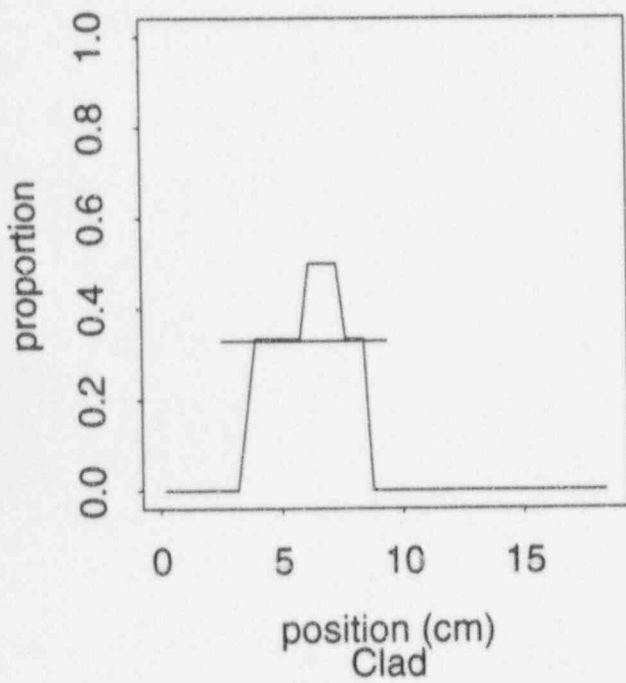
B608 A near # insp= 6



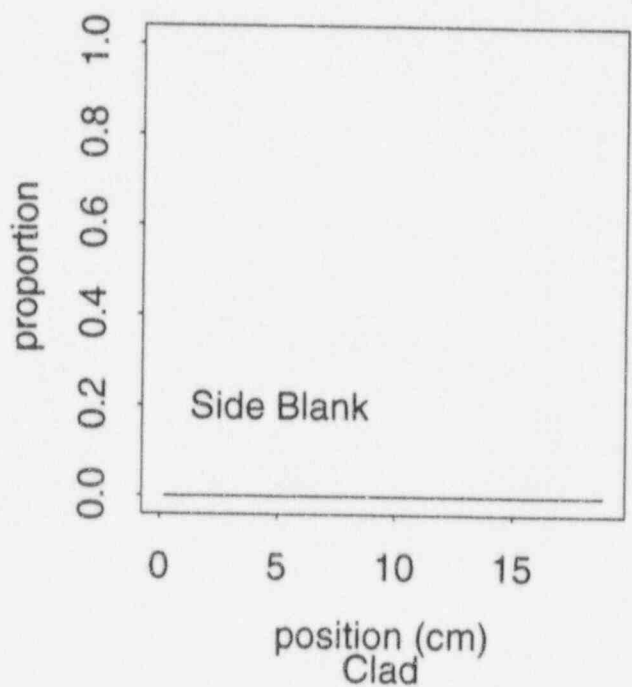
B608 B near # insp= 12



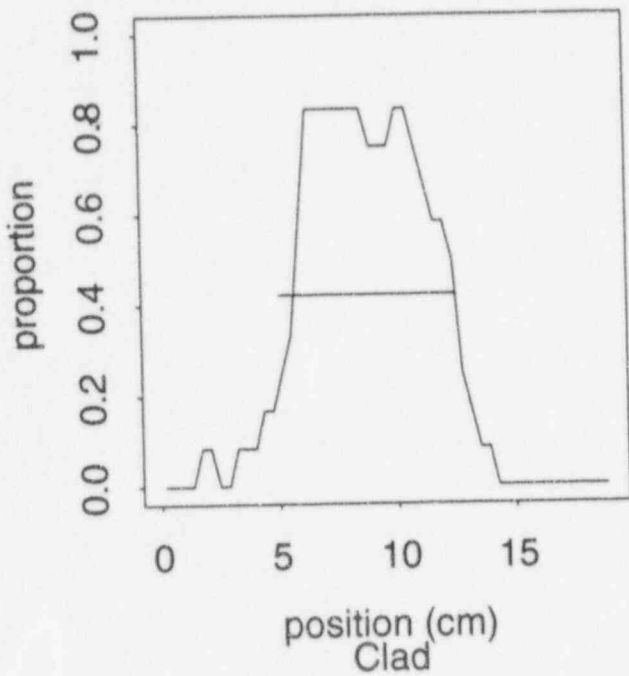
B609 B near # insp= 6



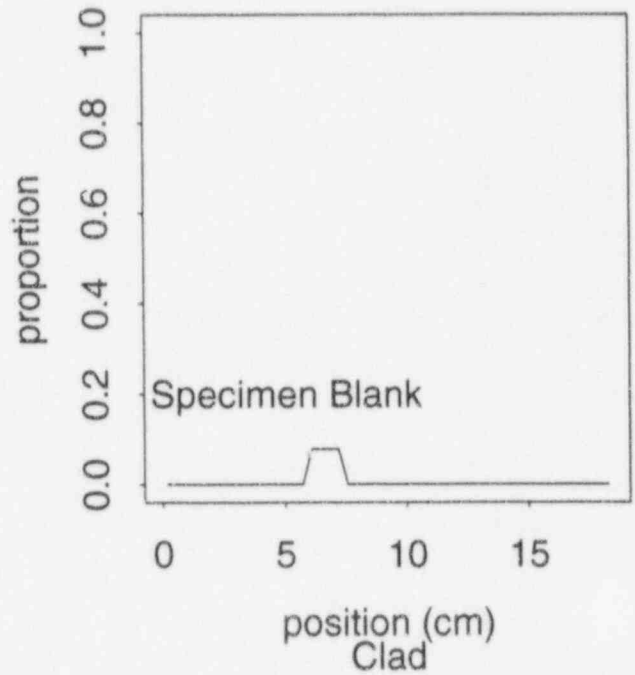
B610 A near # insp= 7



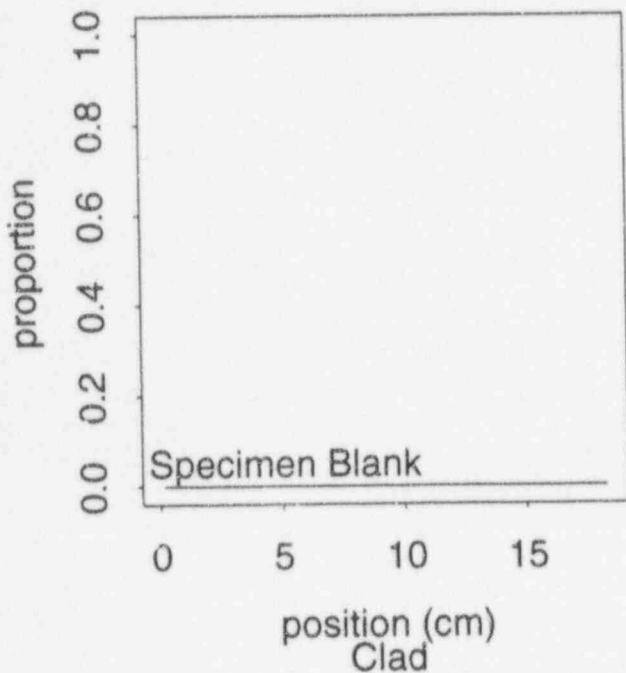
B610 B near # insp= 12



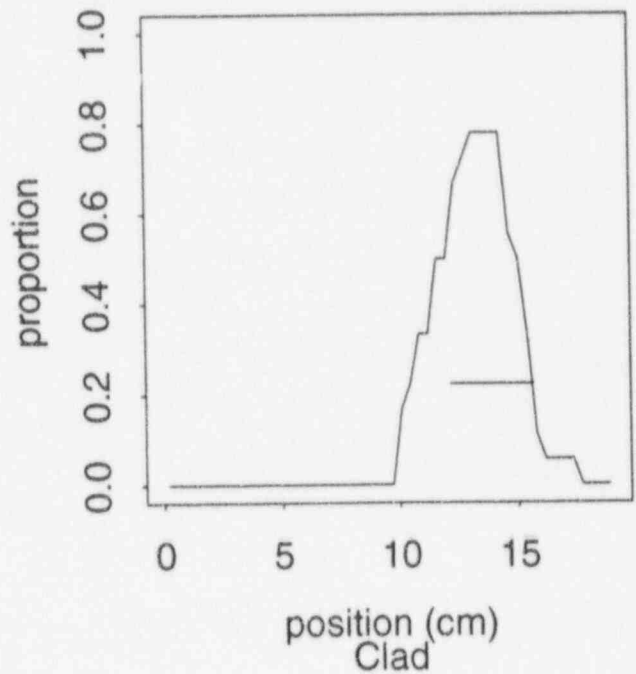
B611 A near # insp= 13



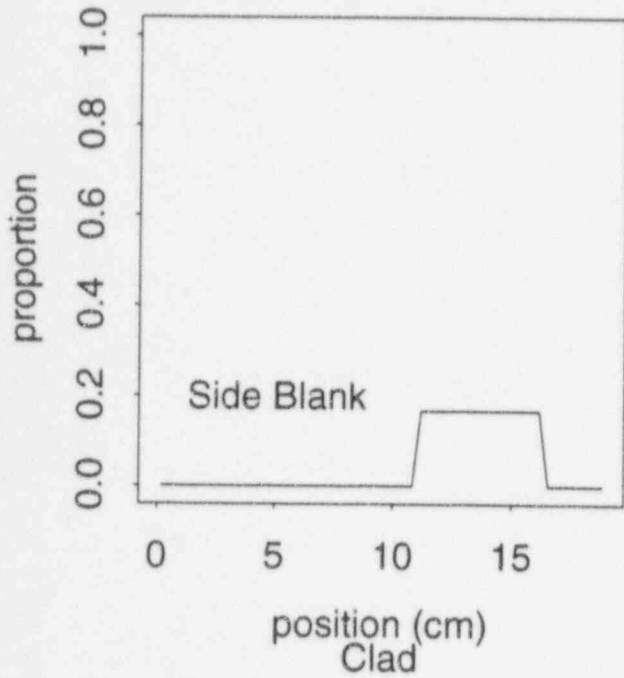
B611 B near # insp= 6



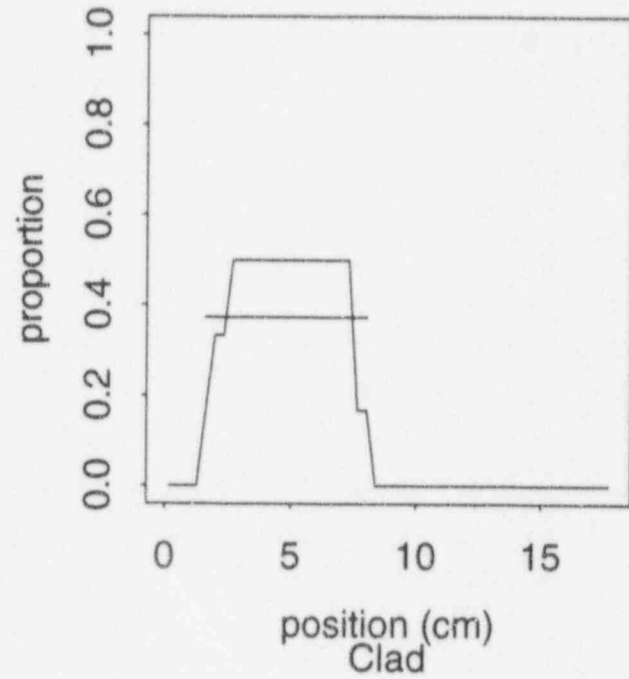
B612 A near # insp= 18



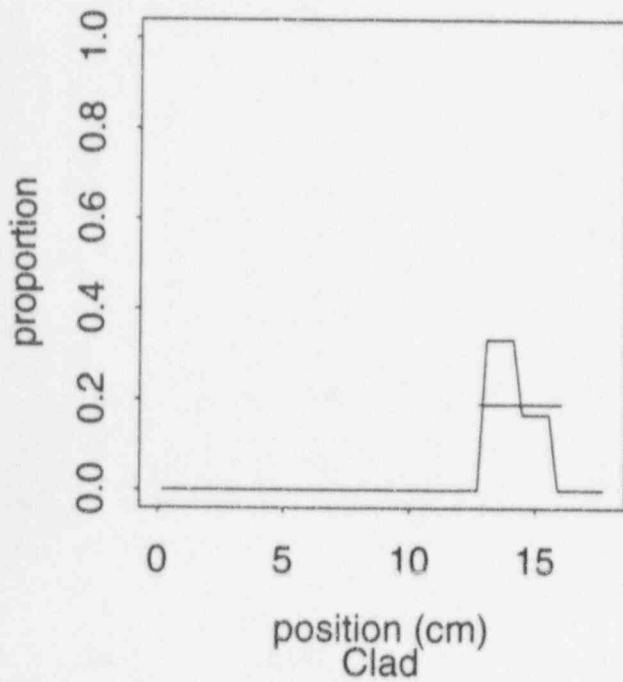
B612 B near # insp= 6



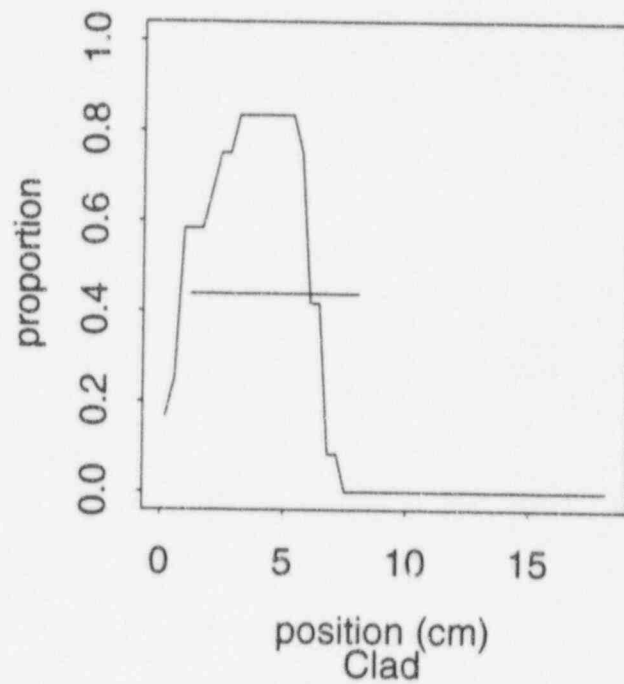
B613 B near # insp= 6



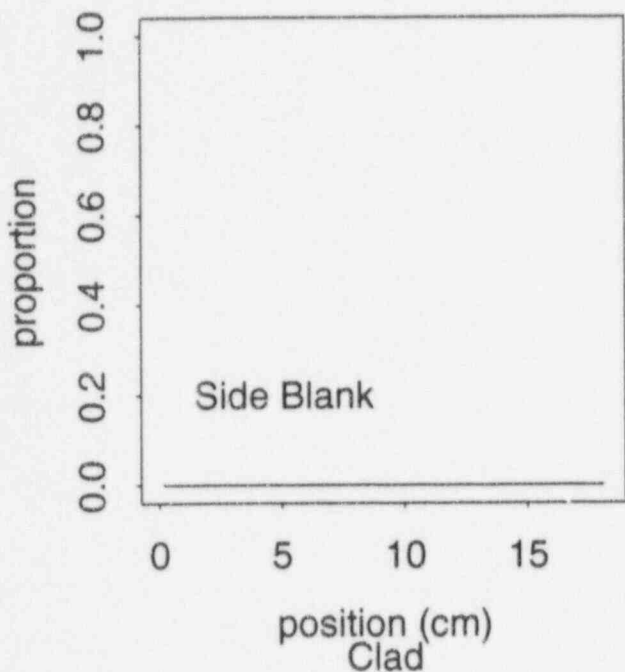
B614 B near # insp= 6



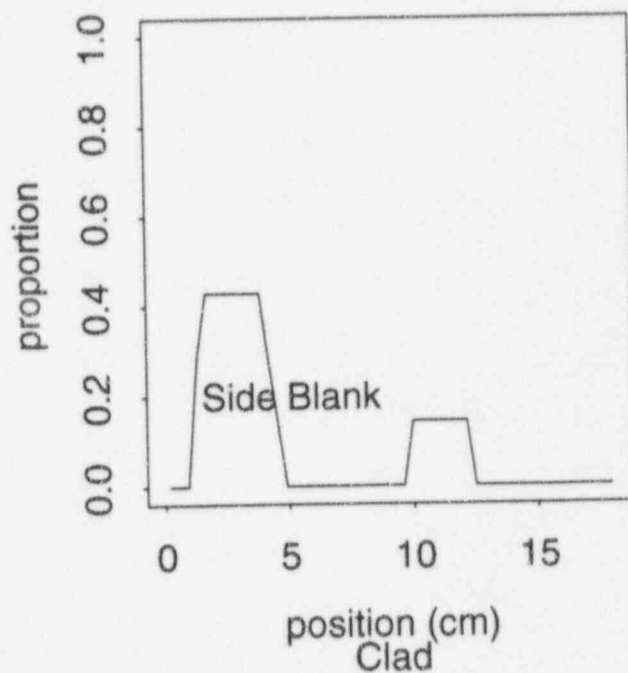
B615 A near # insp= 12



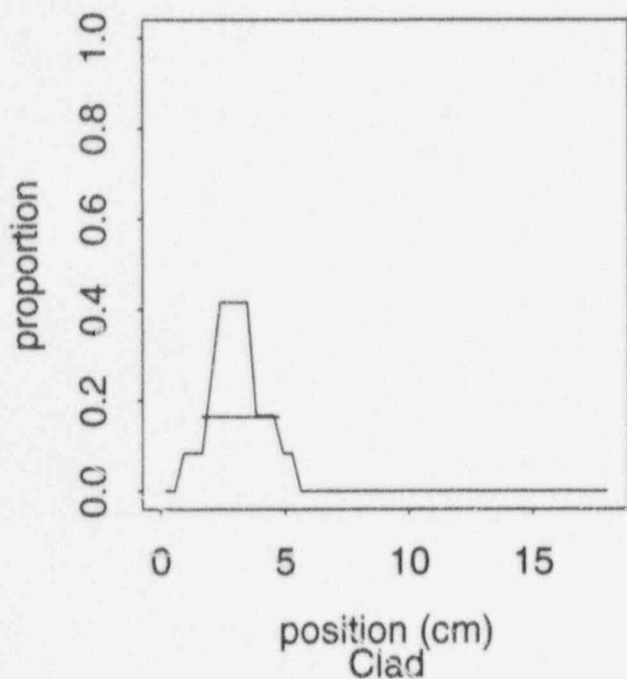
B615 B near # insp= 6



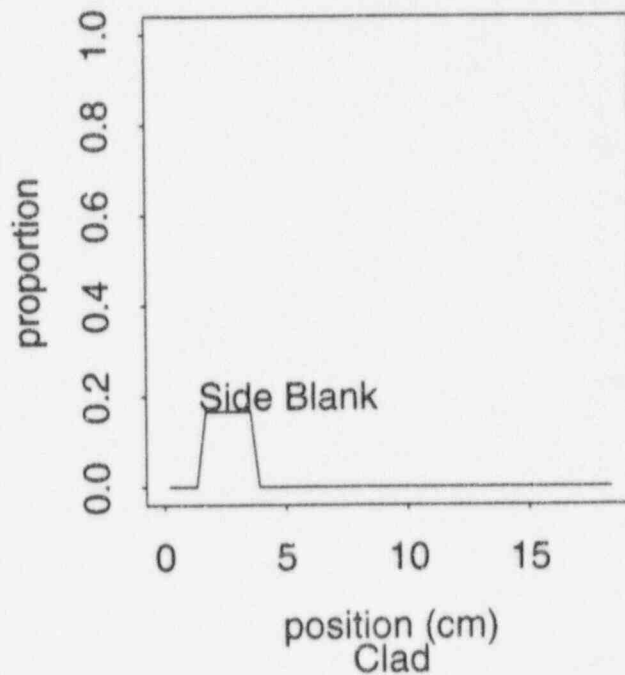
B616 A near # insp= 7



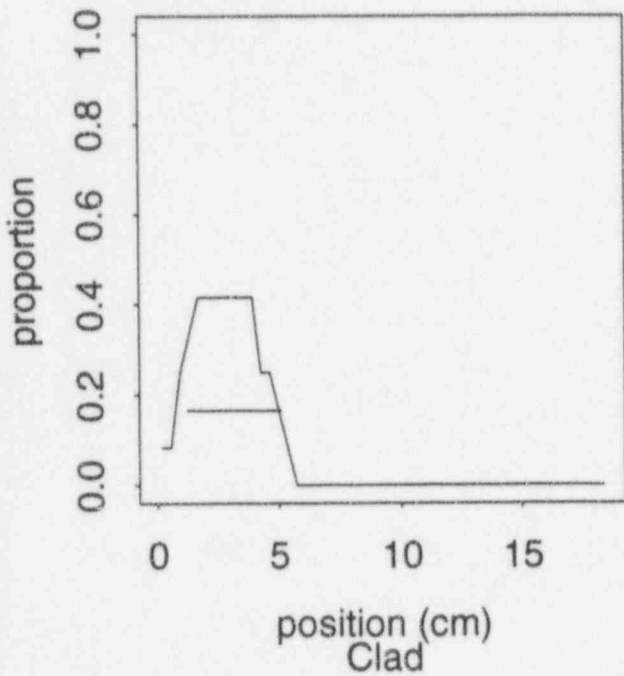
B616 B near # insp= 12



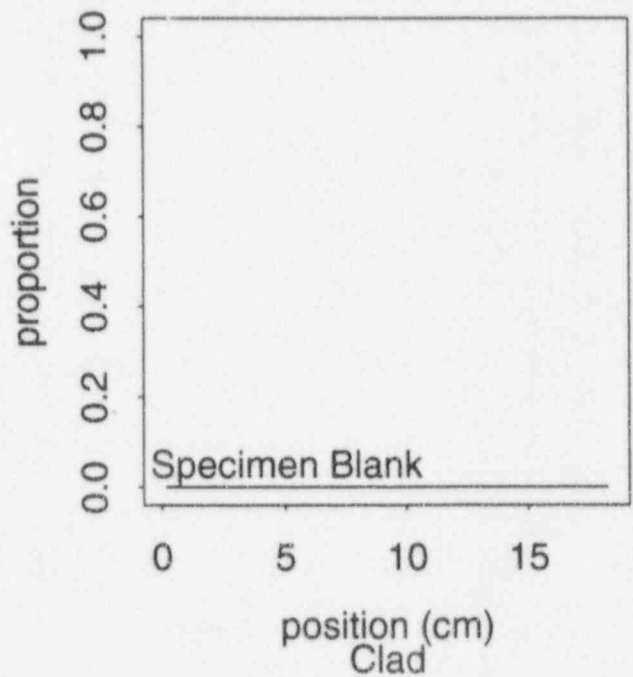
B617 A near # insp= 6



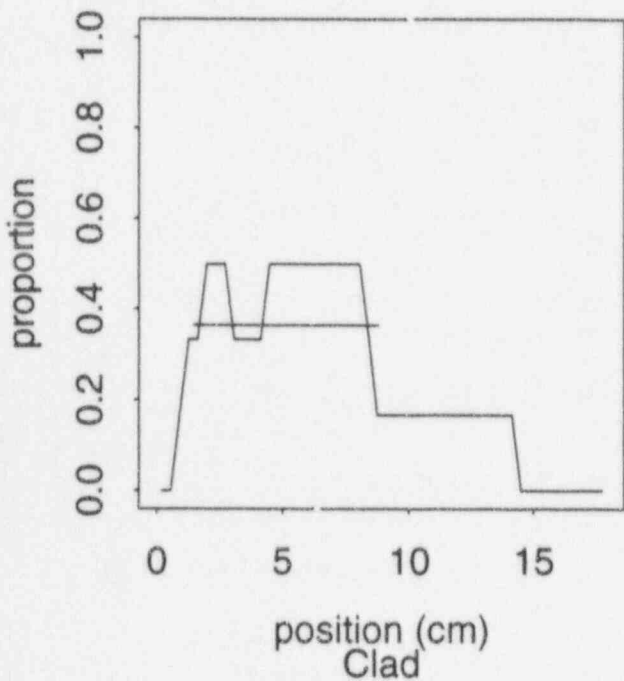
B617 B near # insp= 12



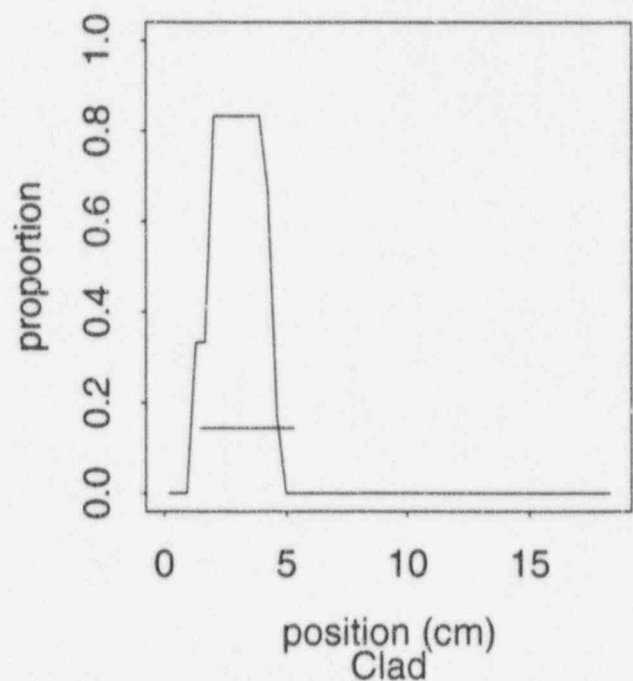
B618 B near # insp= 6



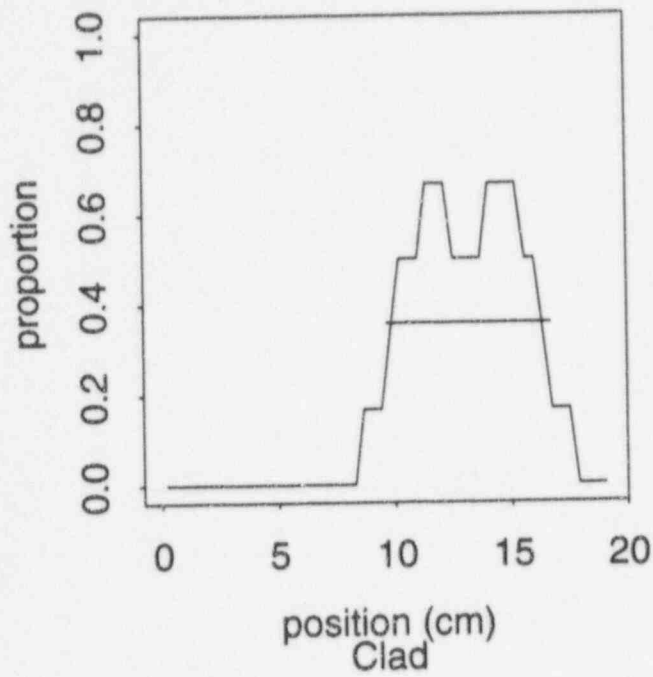
B619 B near # insp= 6



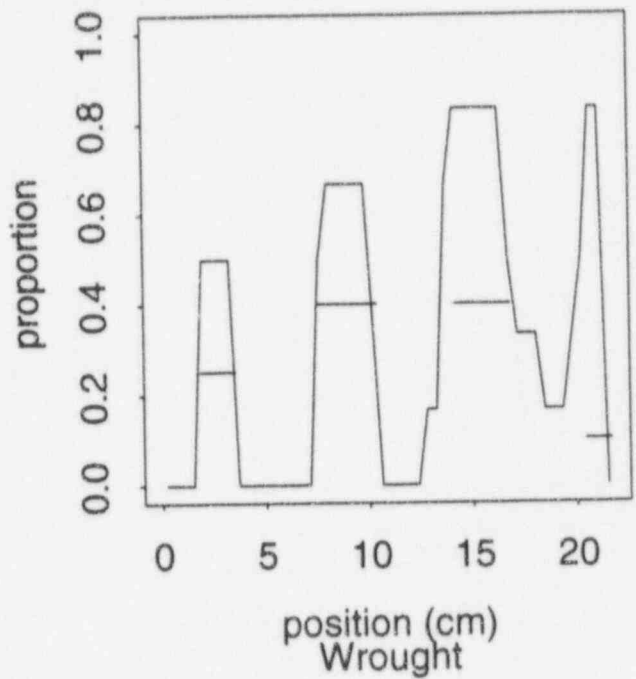
B620 B near # insp= 6



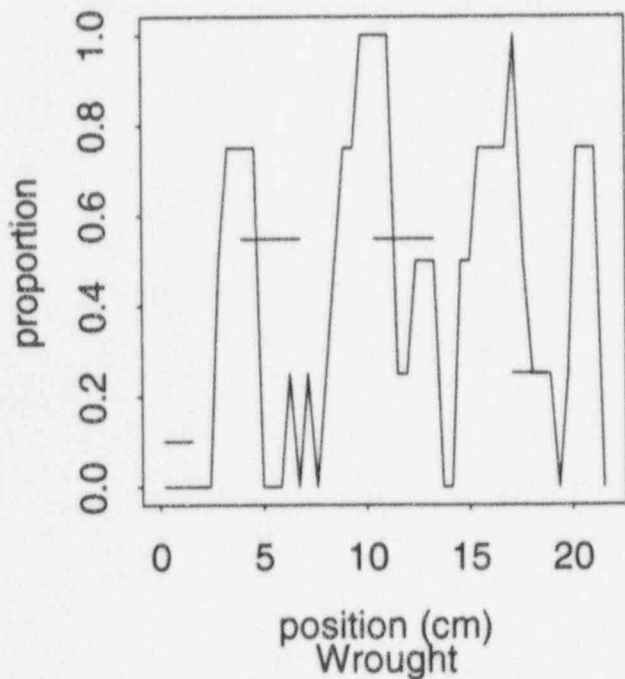
B621 A near # insp= 6



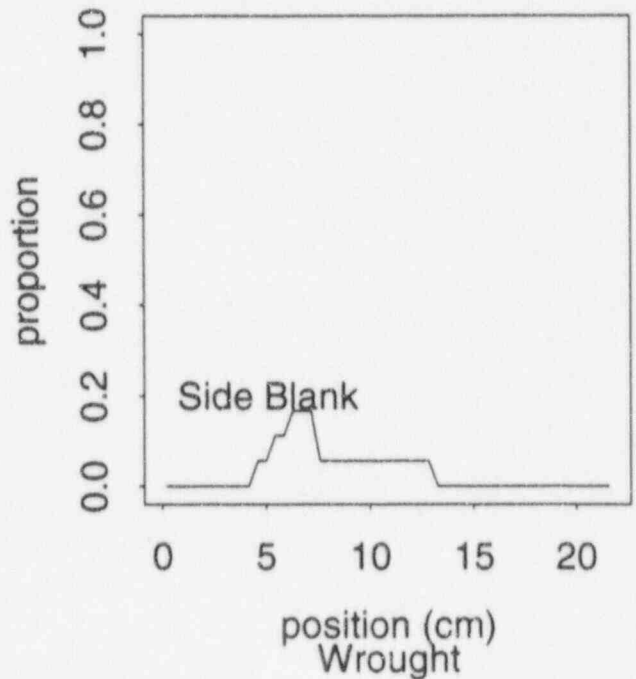
B211-1 A near # insp= 6



B211-1 B near # insp= 4

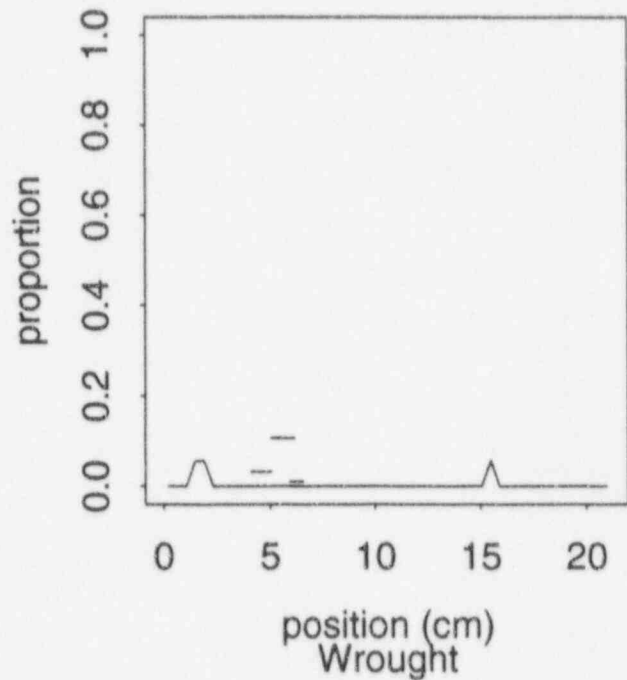
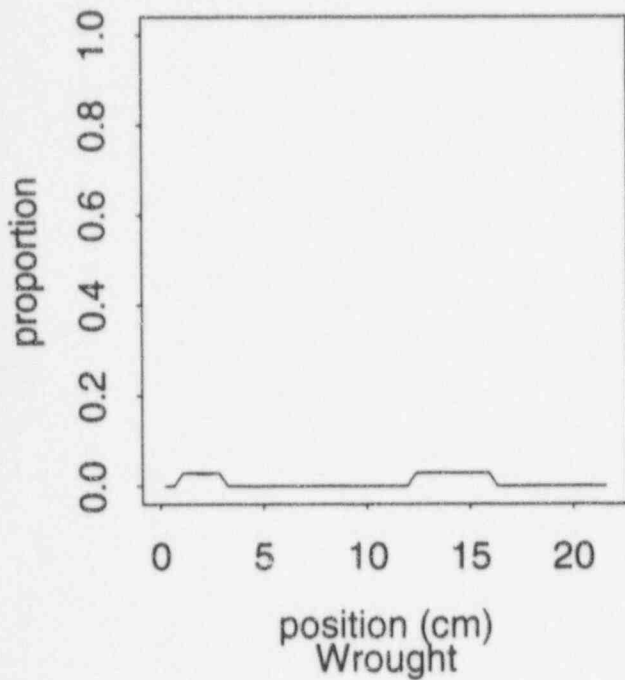


B211-2 A near # insp= 18



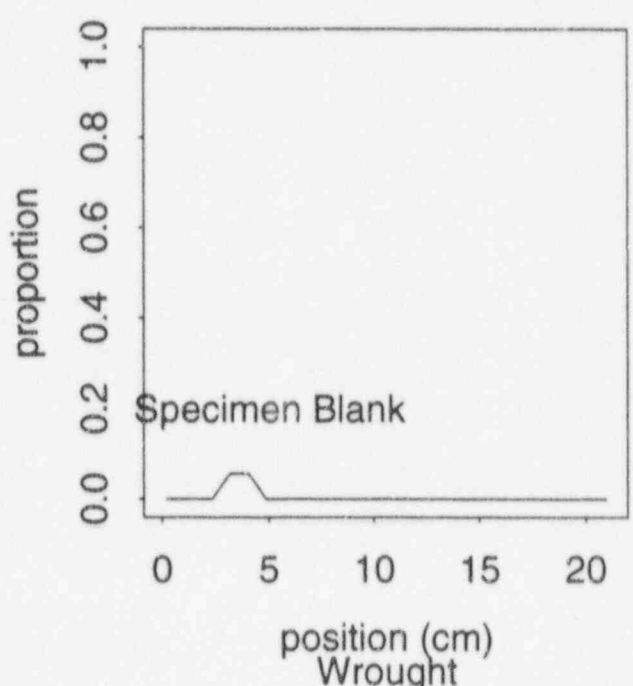
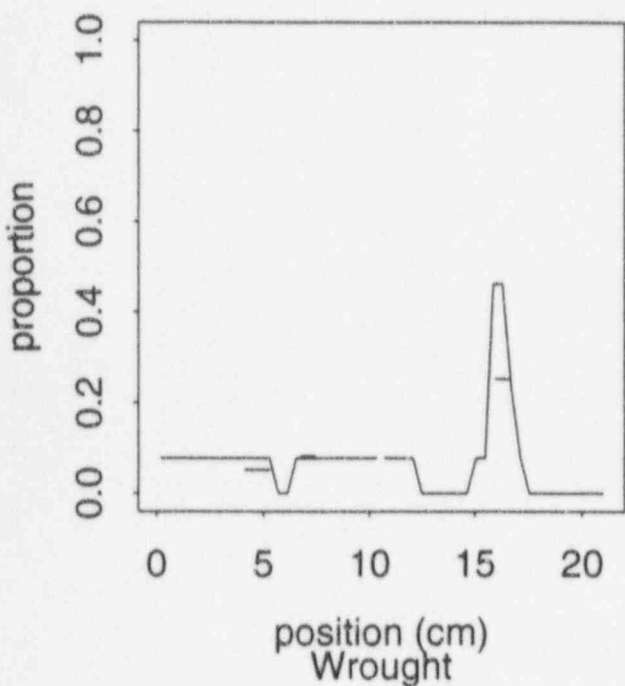
B211-2 B near # insp= 36

B211-3 A near # insp= 18

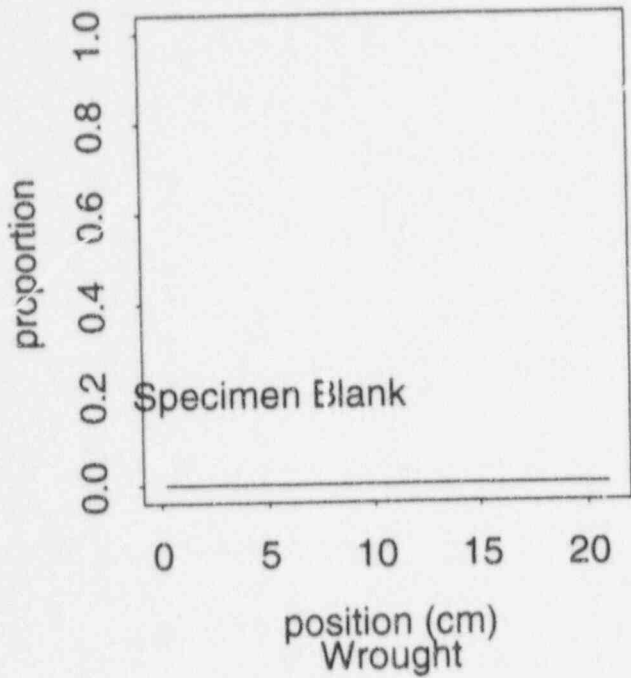


B211-3 B near # insp= 13

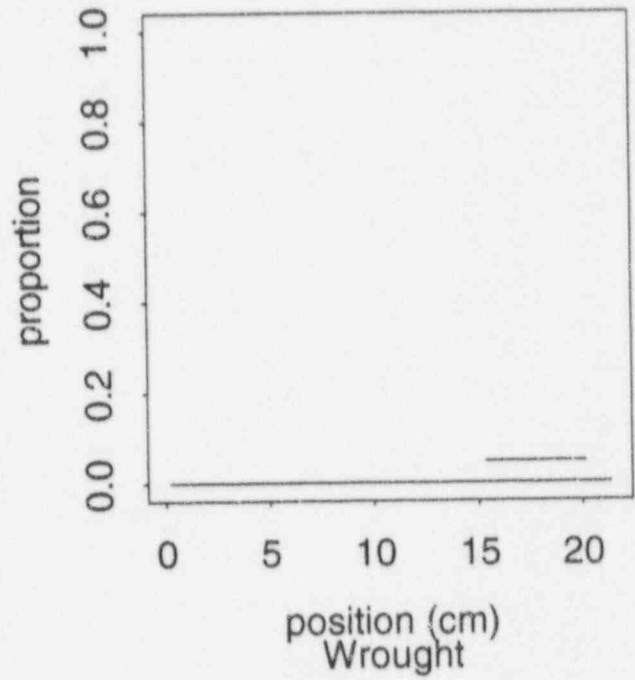
B211-4 A near # insp= 36



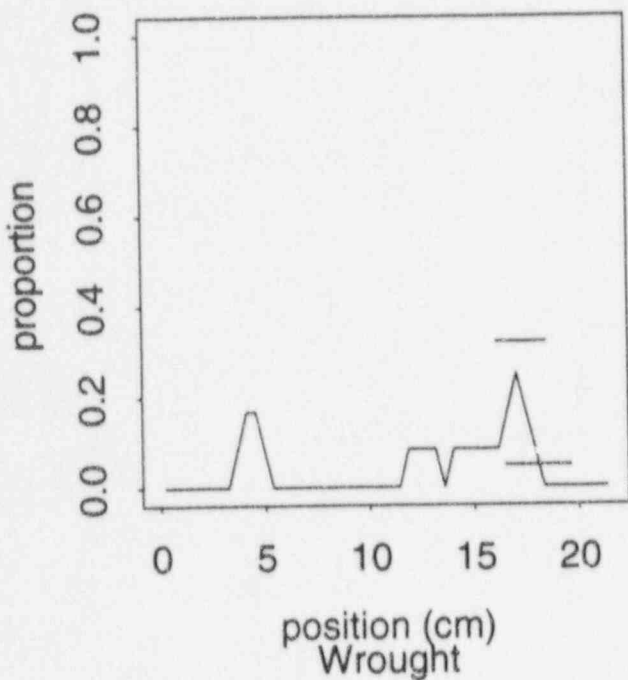
B211-4 B near # insp= 17



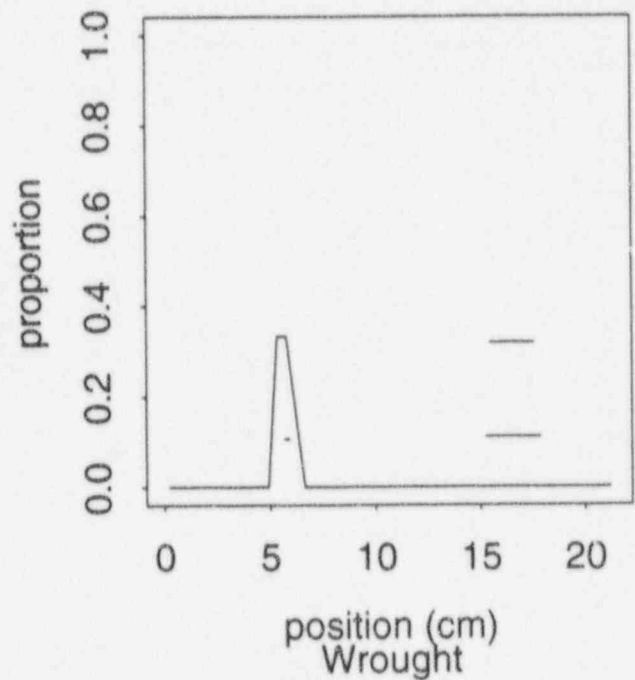
B212-1 A near # insp= 1



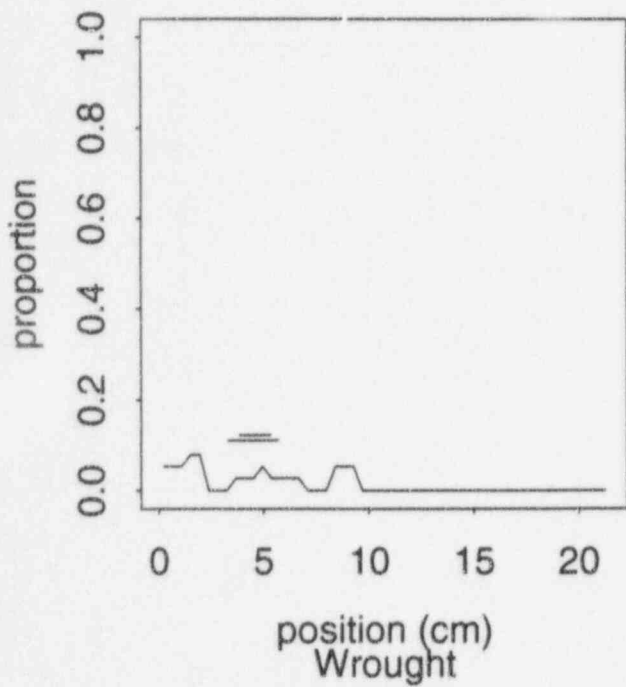
B212-1 B near # insp= 12



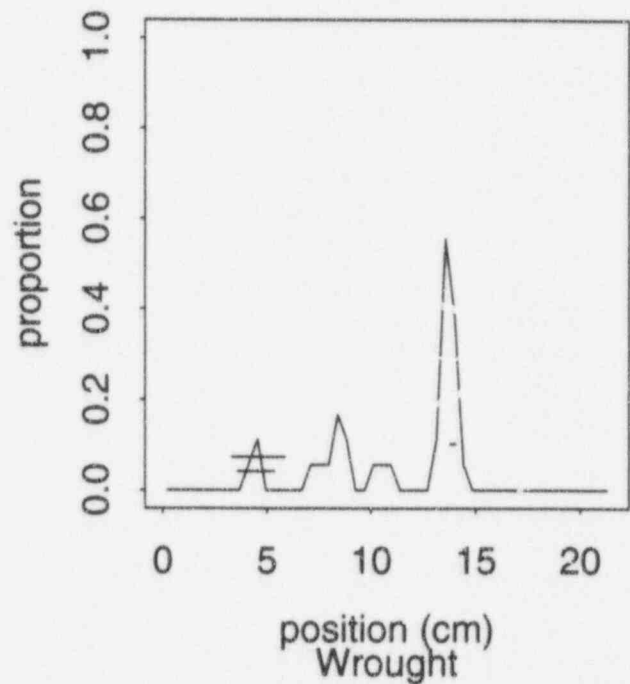
B212-2 A near # insp= 6



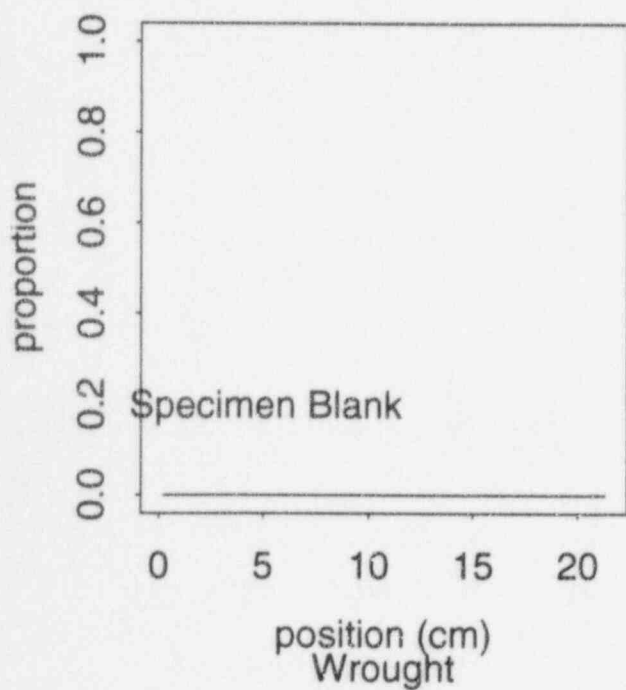
E212-3 A near # insp= 38



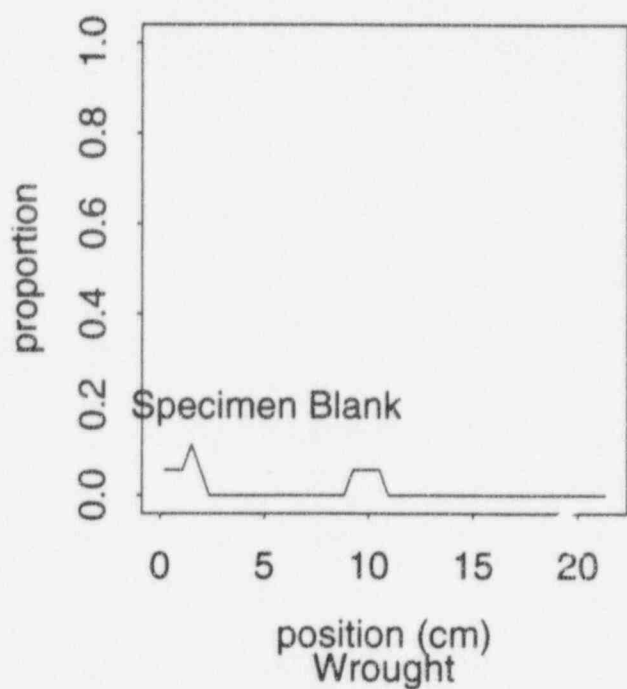
B212-3 B near # insp= 18



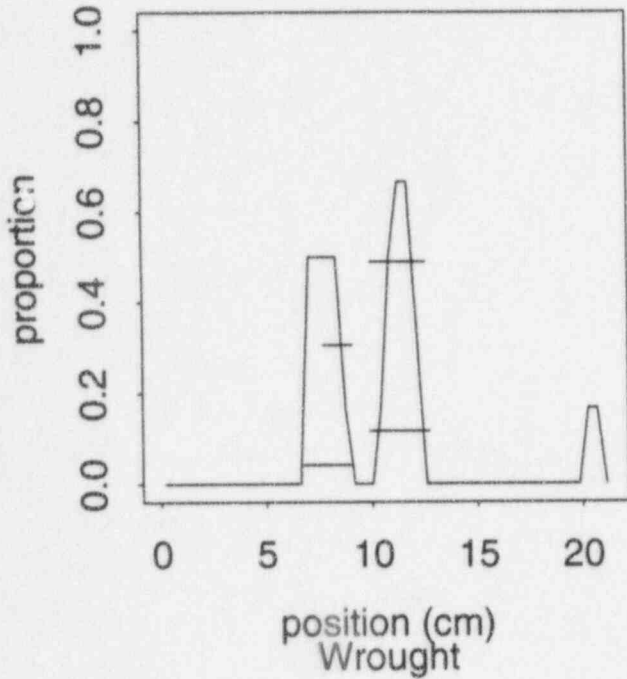
B212-4 A near # insp= 13



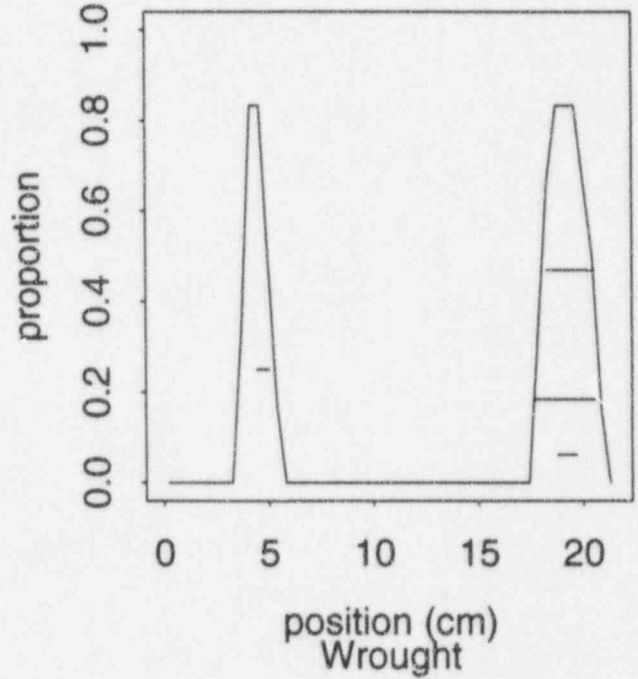
B212-4 B near # insp= 18



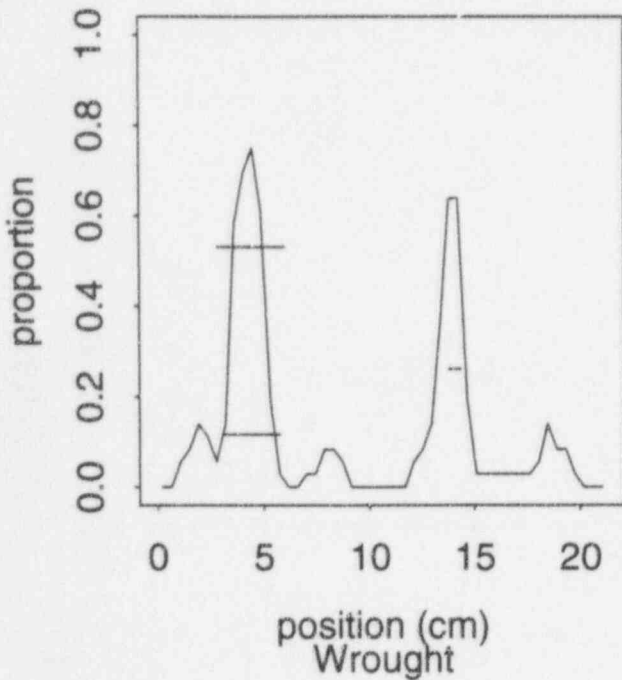
B213-1 A near # insp= 6



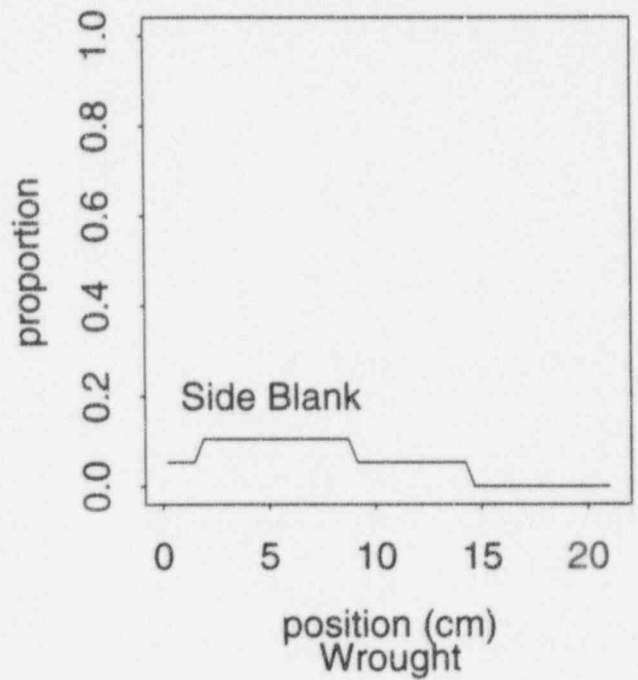
B213-2 A near # insp= 6



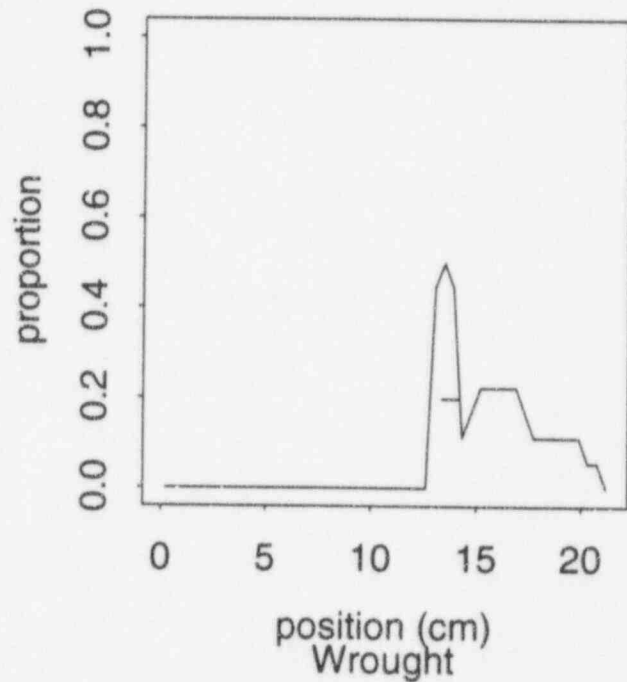
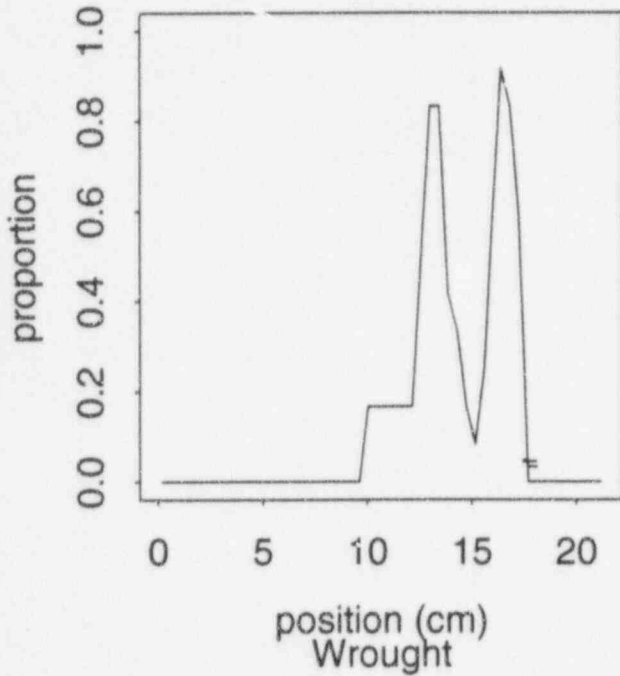
B213-3 A near # insp= 36



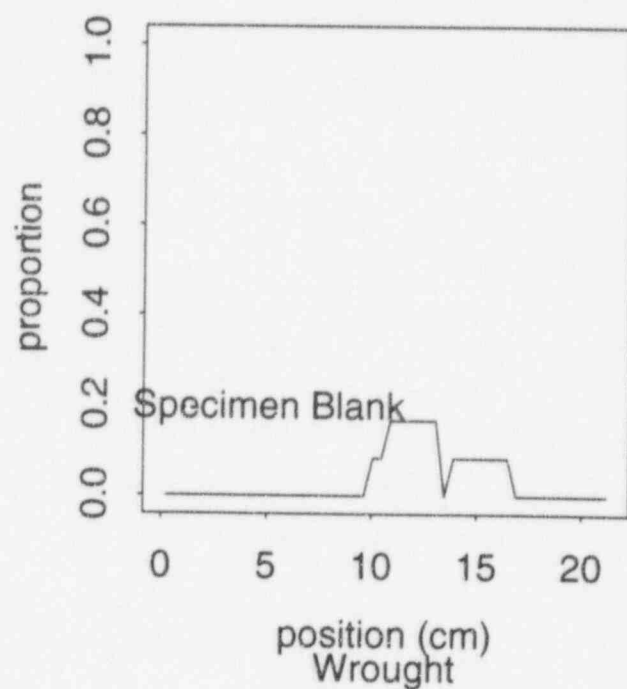
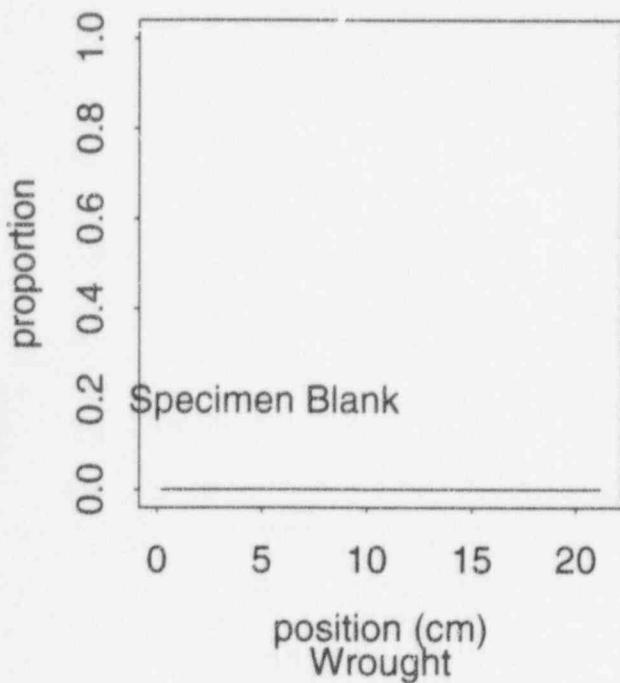
B213-3 B near # insp= 19



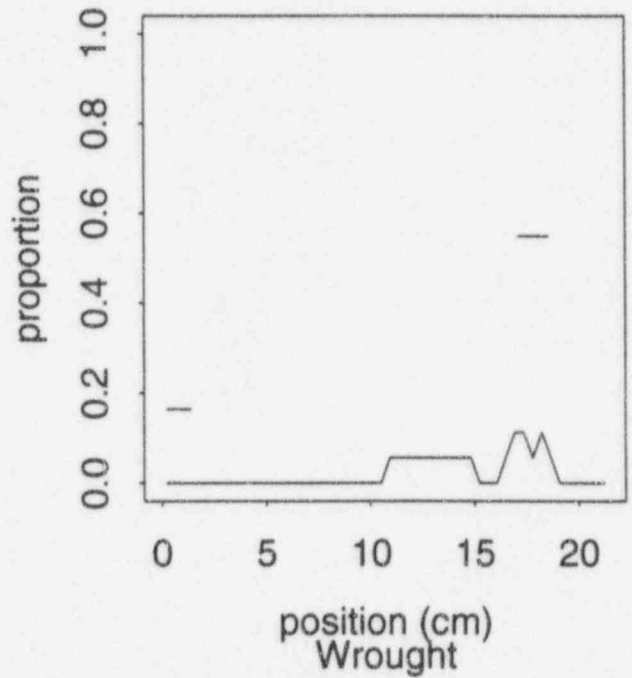
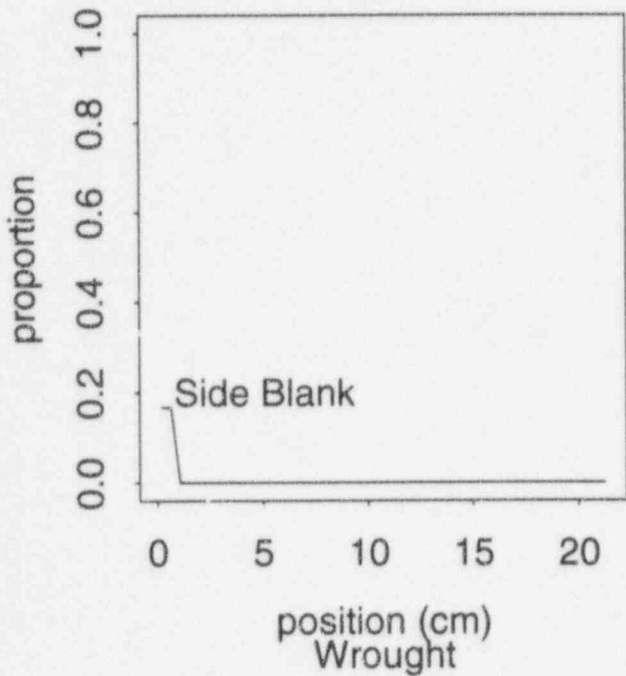
B213-4 A near # insp= 12 B213-4 B near # insp= 18



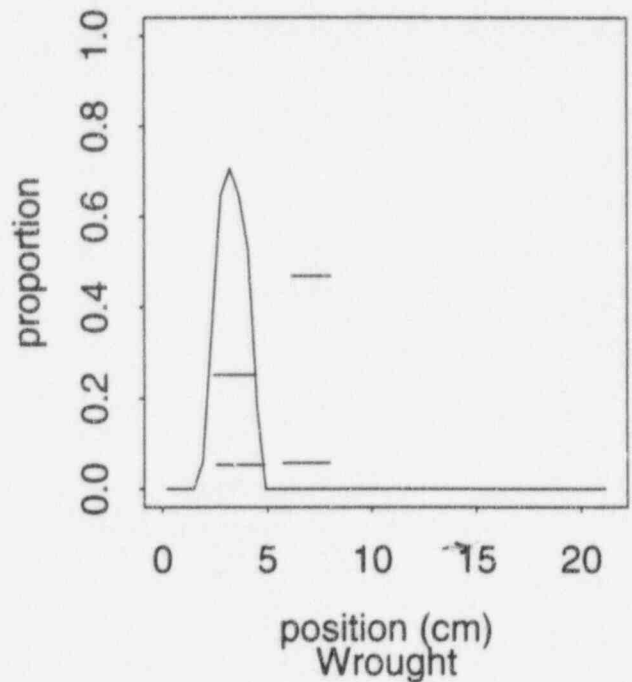
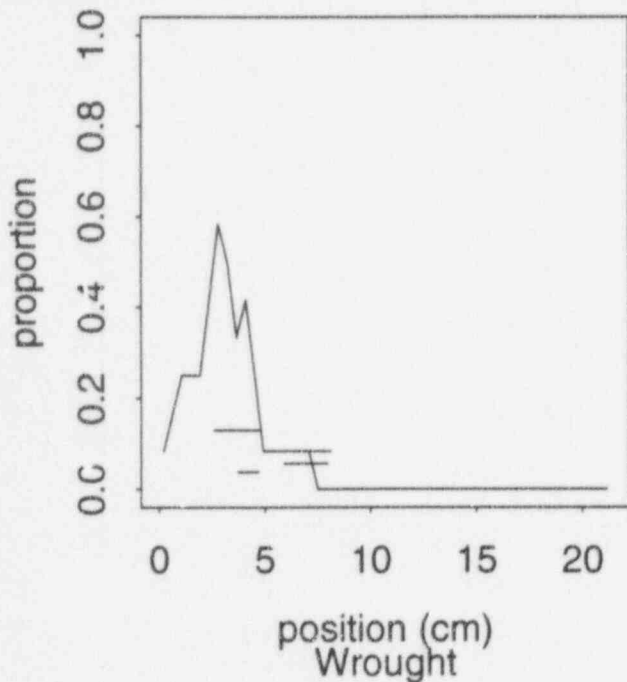
B214-1 A near # insp= 18 B214-1 B near # insp= 12



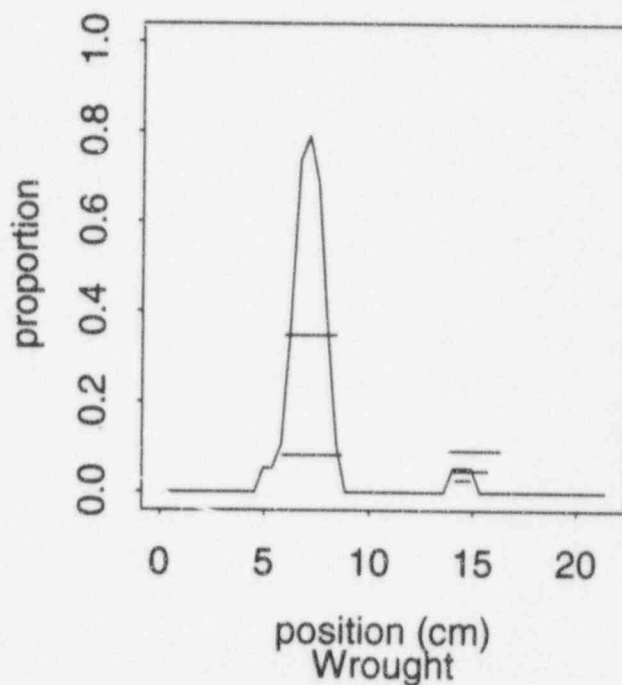
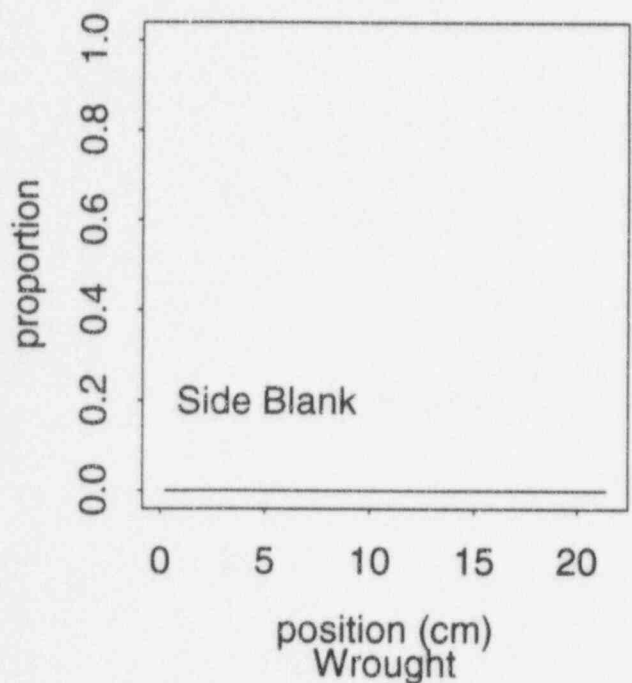
B214-2 A near # insp= 12 B214-2 B near # insp= 18



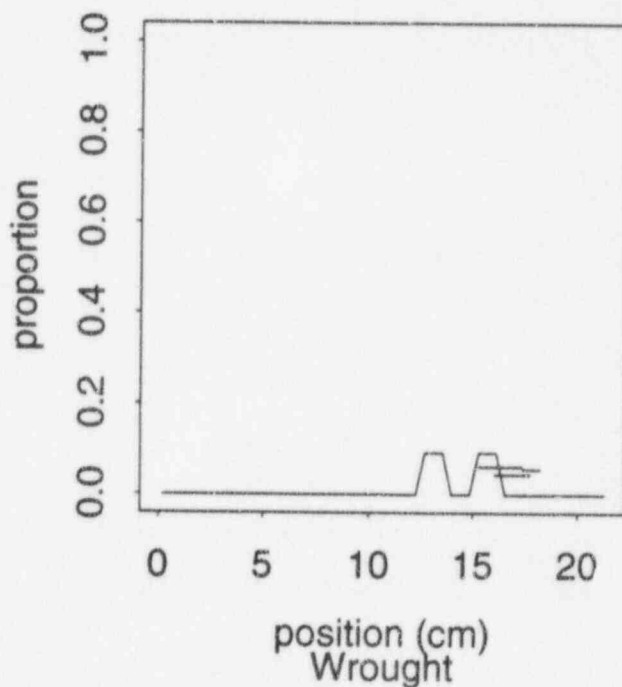
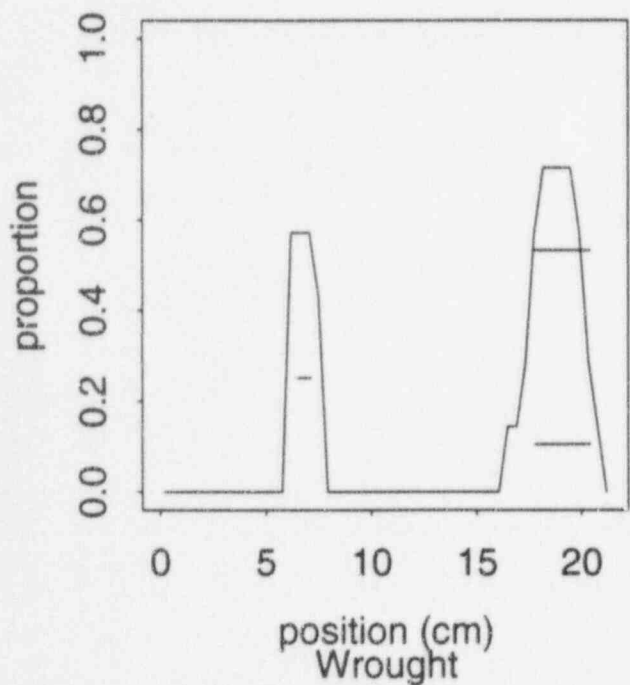
B214-3 A near # insp= 12 B214-3 B near # insp= 17



B214-4 A near # insp= 13 B214-4 B near # insp= 19

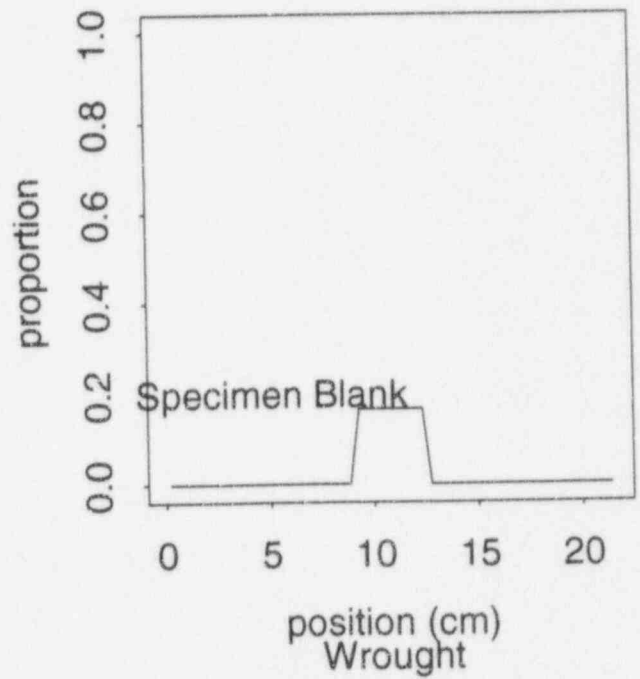
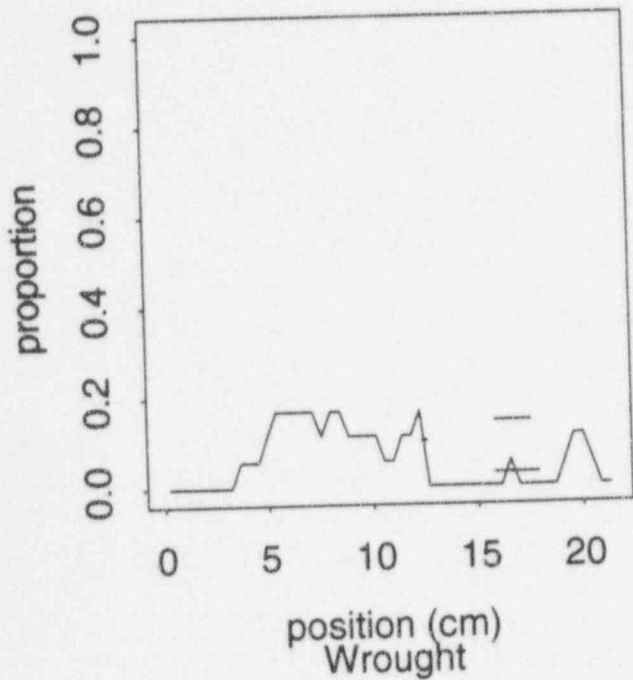


B215-1 A near # insp= 7 B215-3 A near # insp= 11



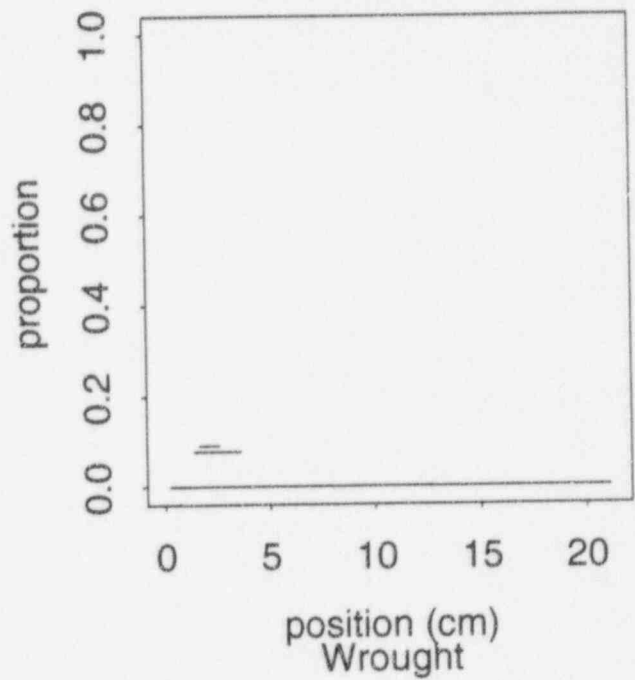
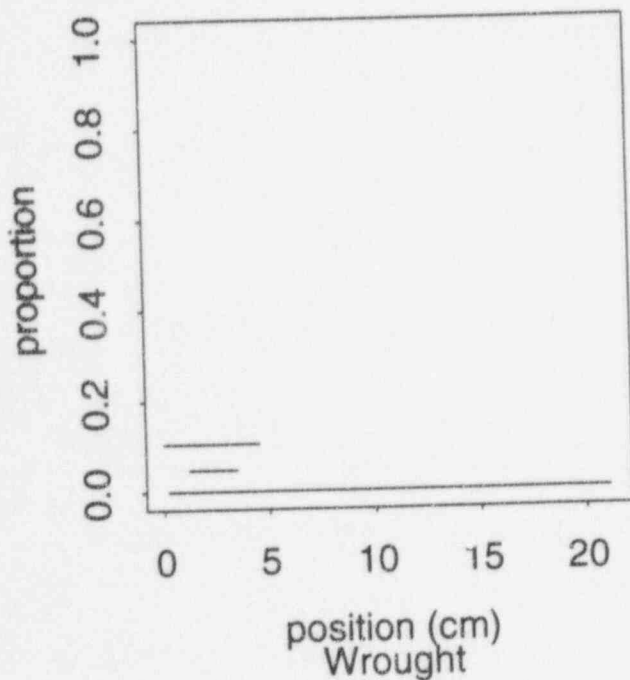
B215-3 B near # insp= 18

B215-4 B near # insp= 6

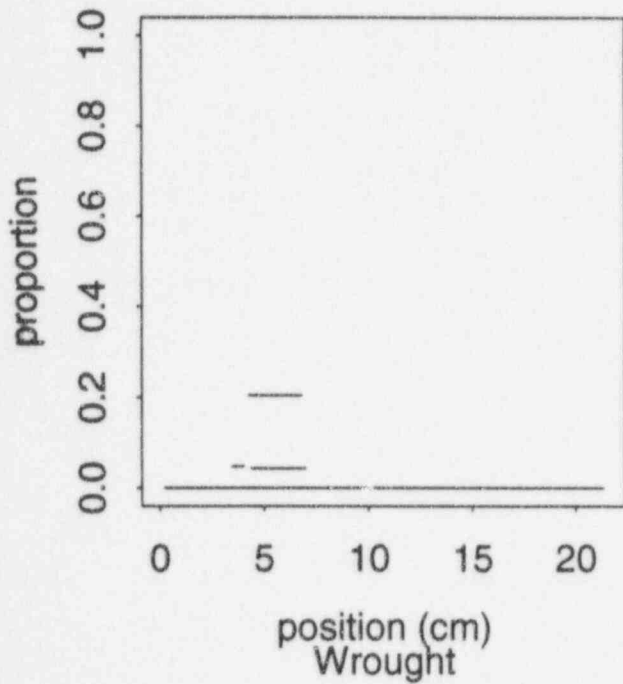


B216-1 A near # insp= 1

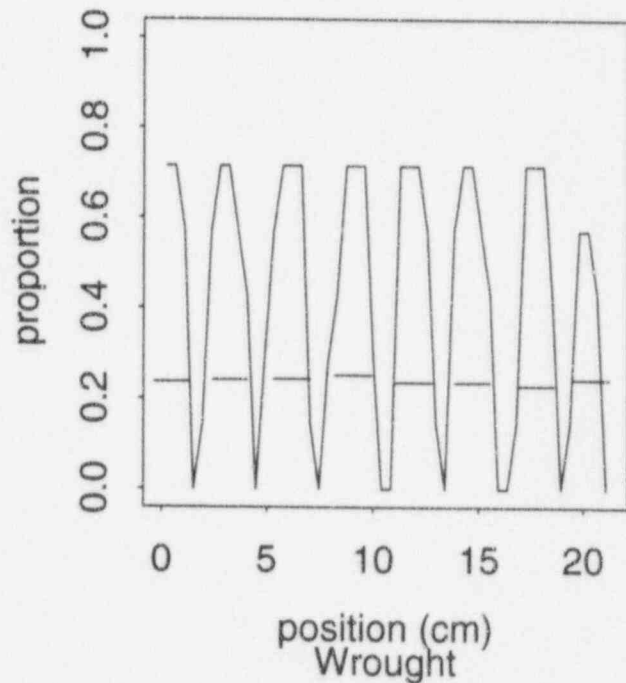
B216-1 B near # insp= 6



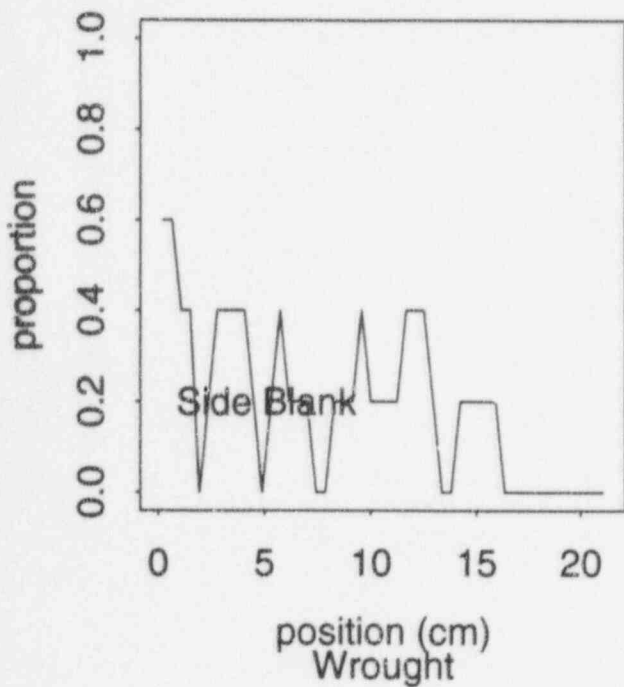
B216-2 A near # insp= 6



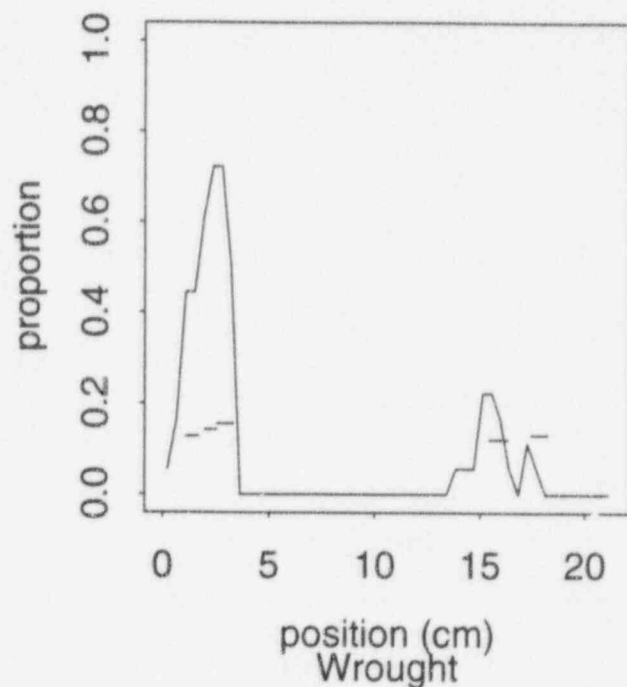
B216-4 A near # insp= 7



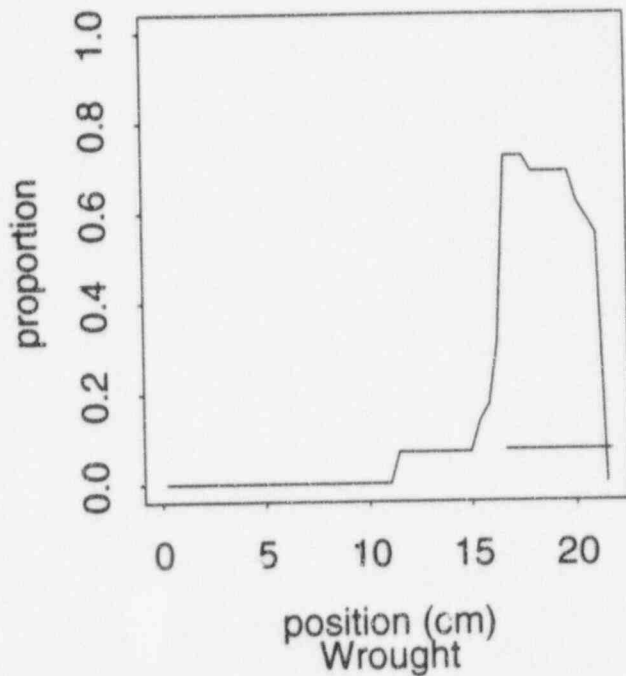
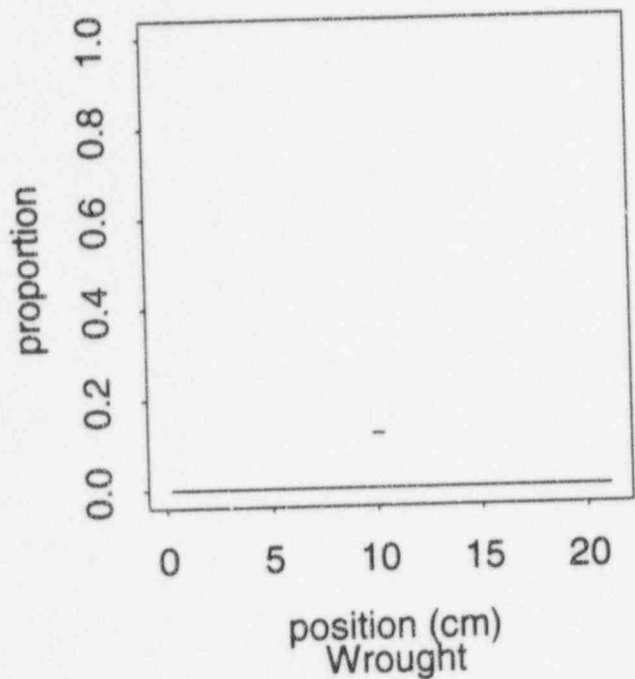
B216-4 B near # insp= 5



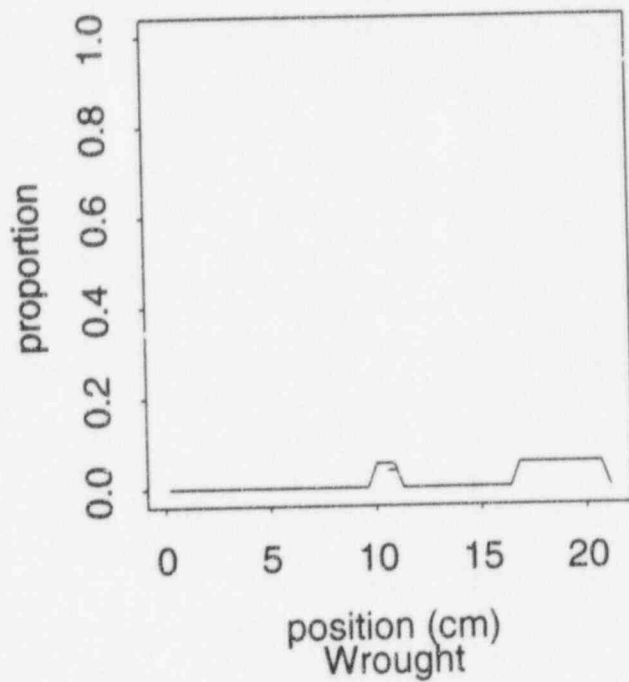
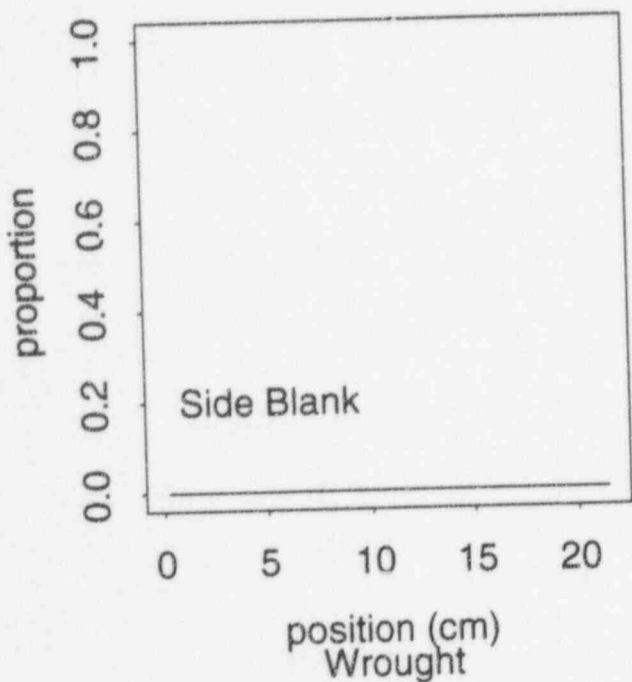
B217-2 A near # insp= 18



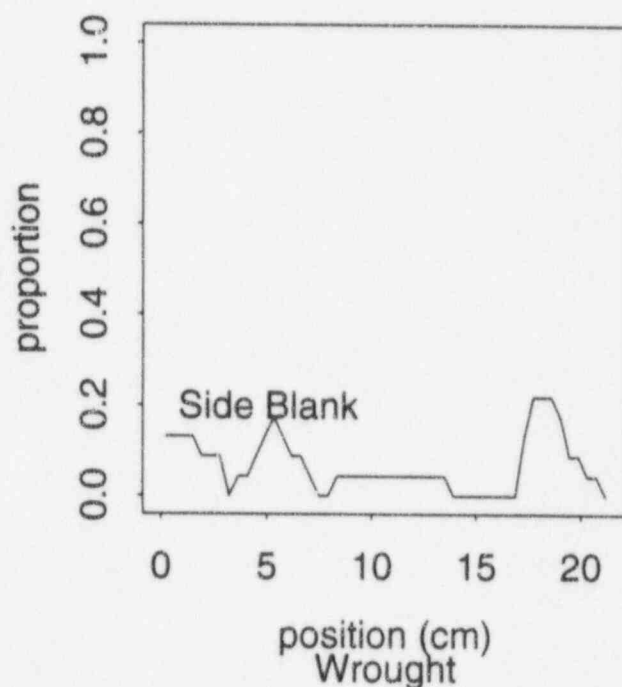
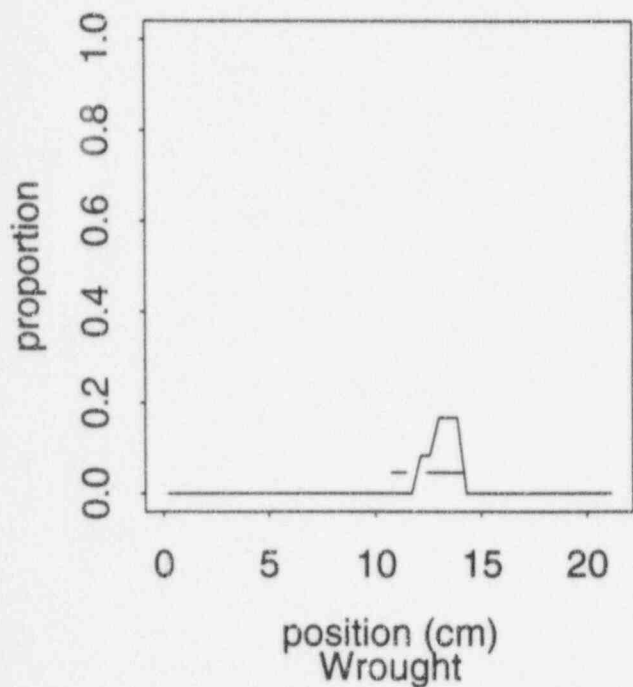
B217-2 B near # insp= 6 B217-3 A near # insp= 29



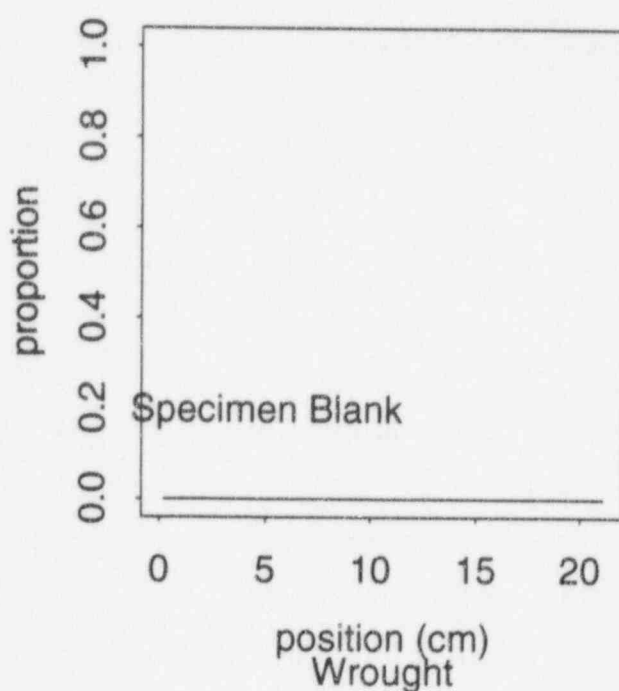
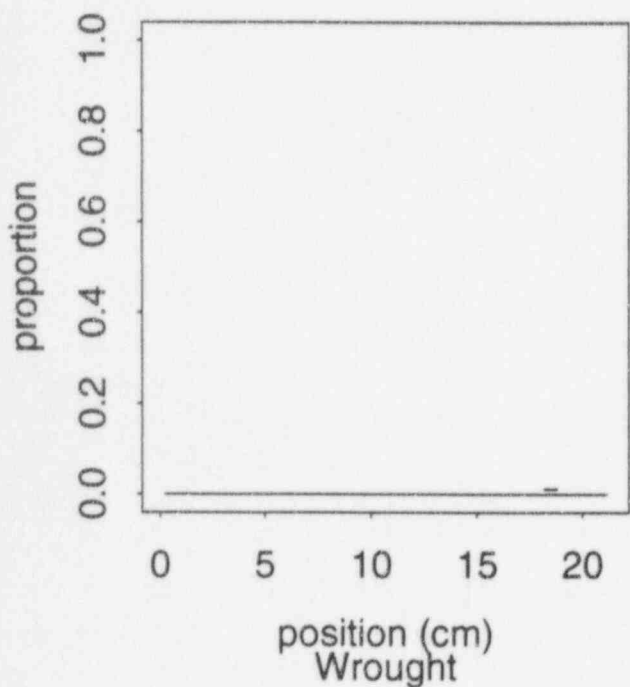
B217-3 B near # insp= 4 B217-4 A near # insp= 19



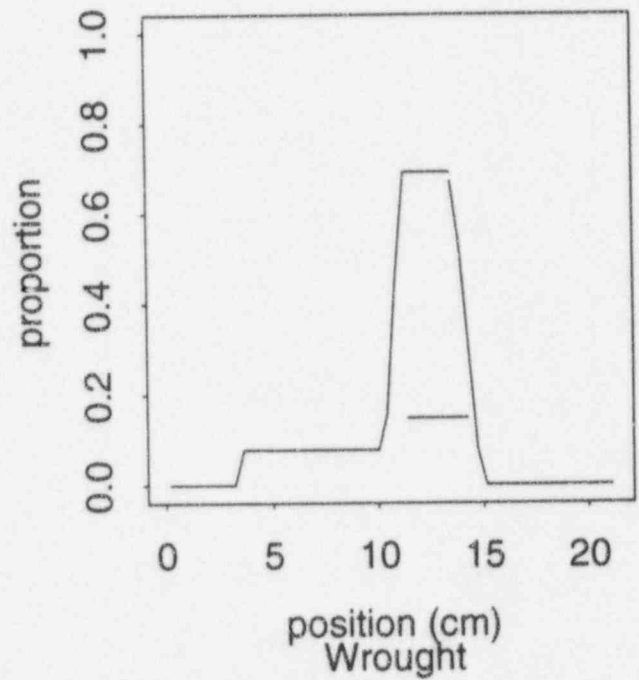
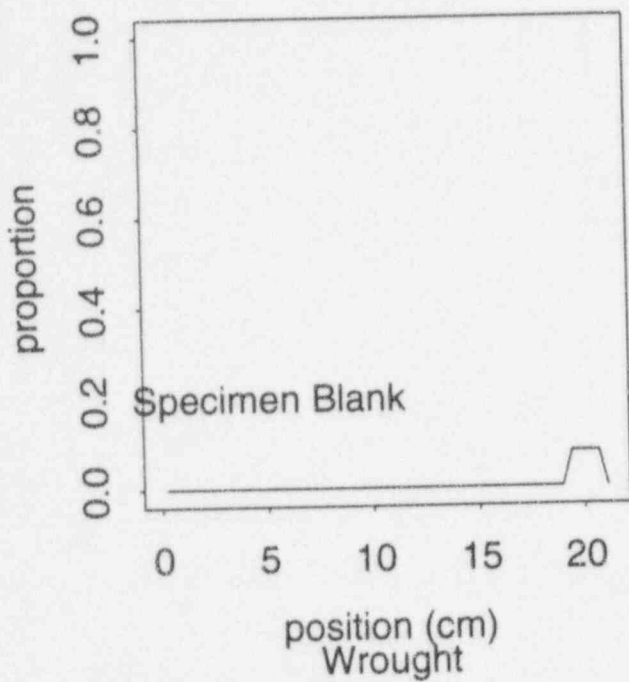
B217-4 B near # insp= 12 D101-1 A near # insp= 23



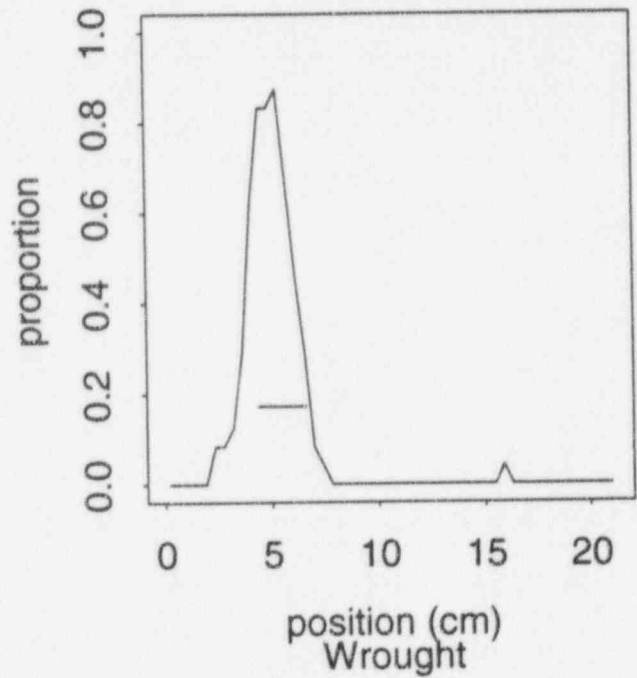
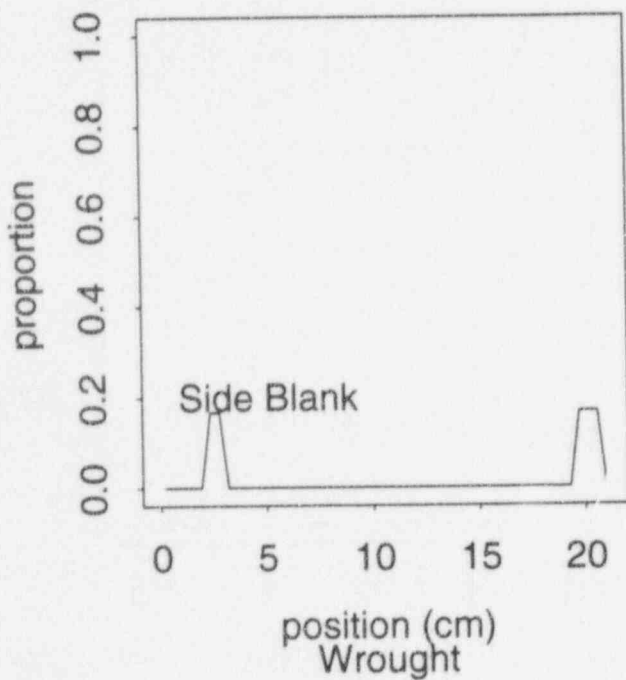
D101-1 B near # insp= 6 D101-2 A near # insp= 5



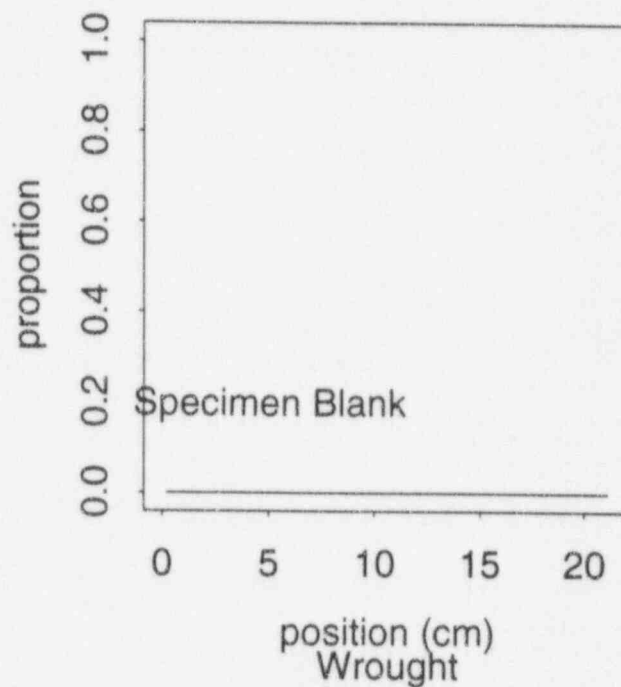
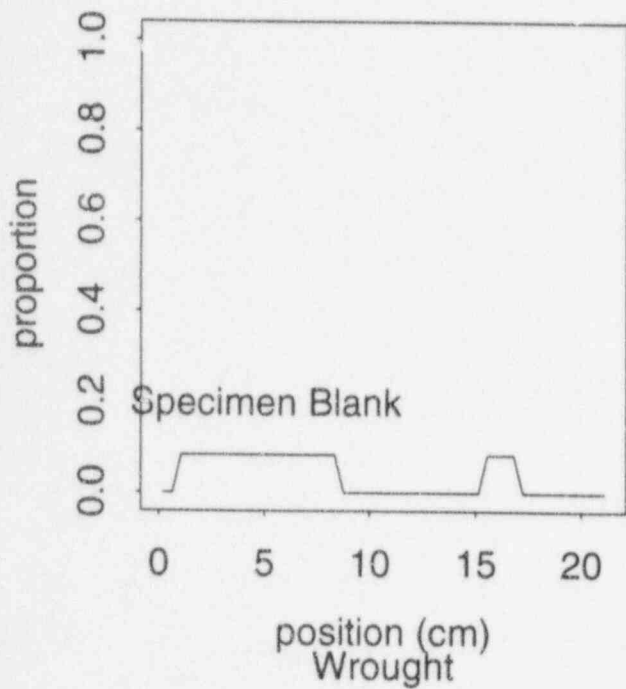
D101-2 B near # insp= 13 D101-3 B near # insp= 13



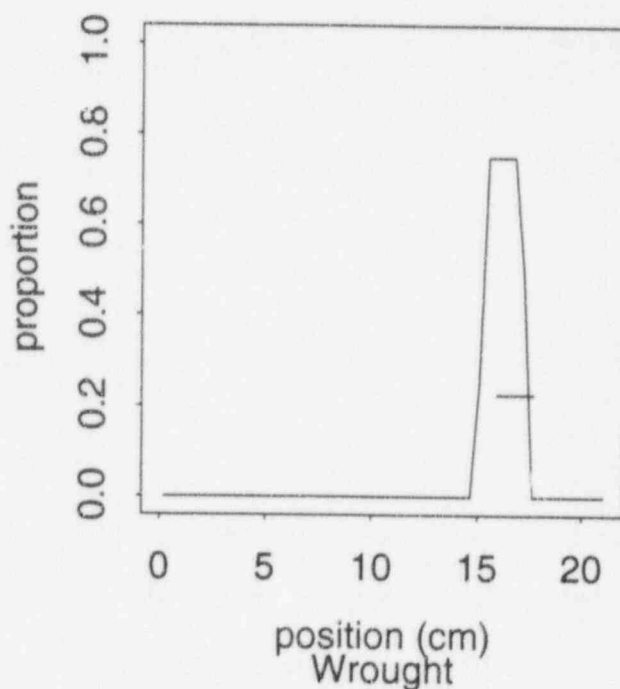
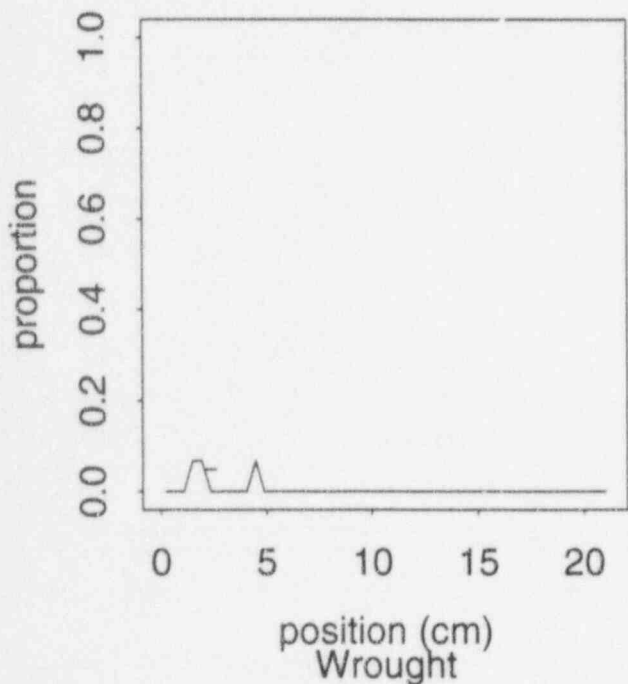
D101-4 A near # insp= 6 D101-4 B near # insp= 24



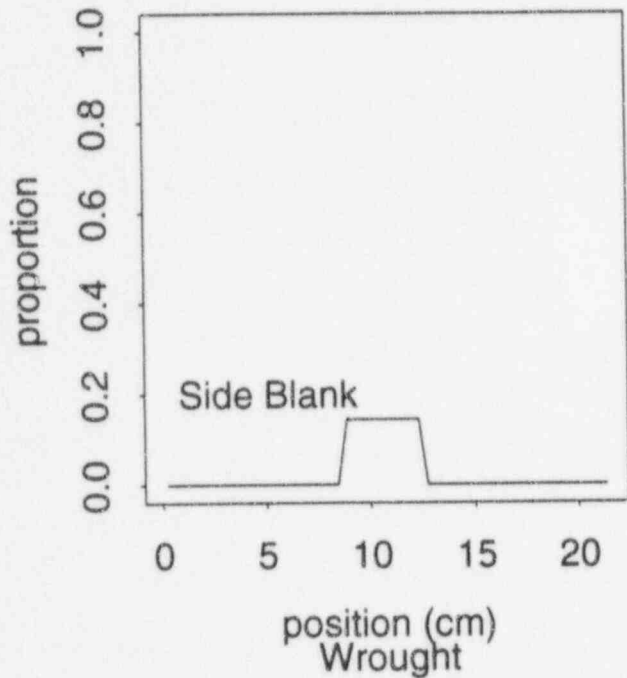
D102-1 A near # insp= 12 D102-1 B near # insp= 6



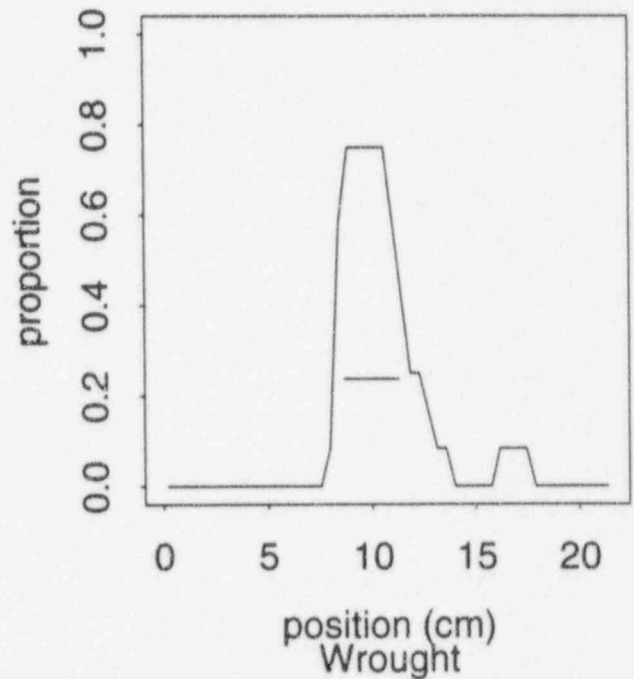
D102-2 A near # insp= 15 D102-2 B near # insp= 12



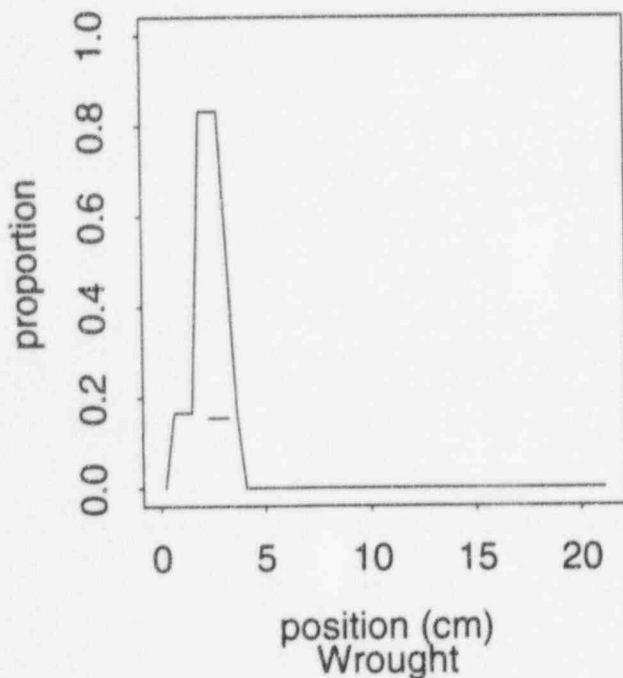
D102-3 A near # insp= 7



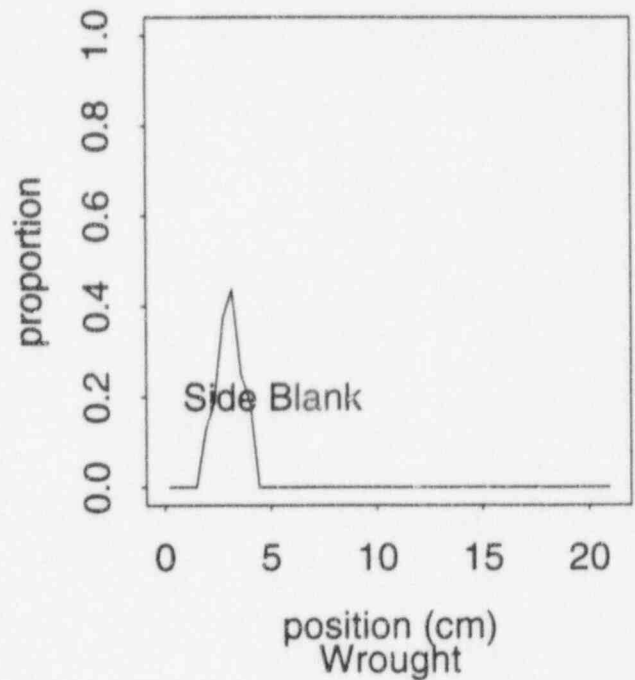
D102-3 B near # insp= 12



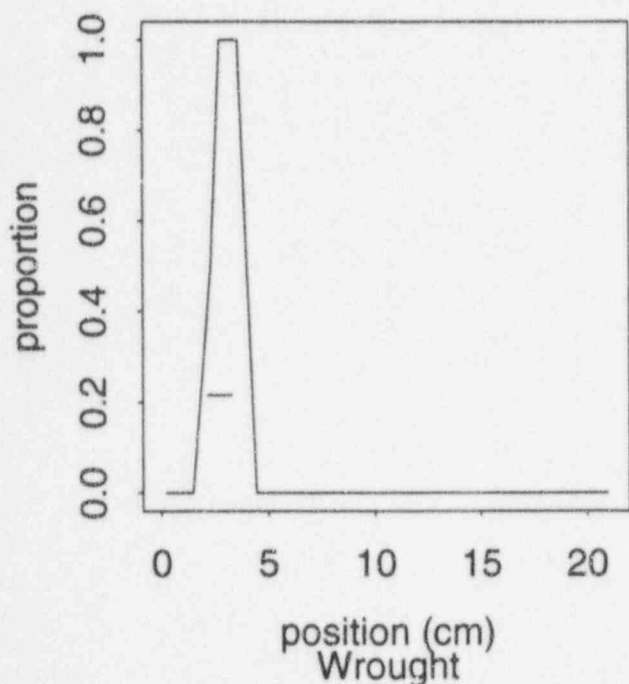
D102-4 B near # insp= 6



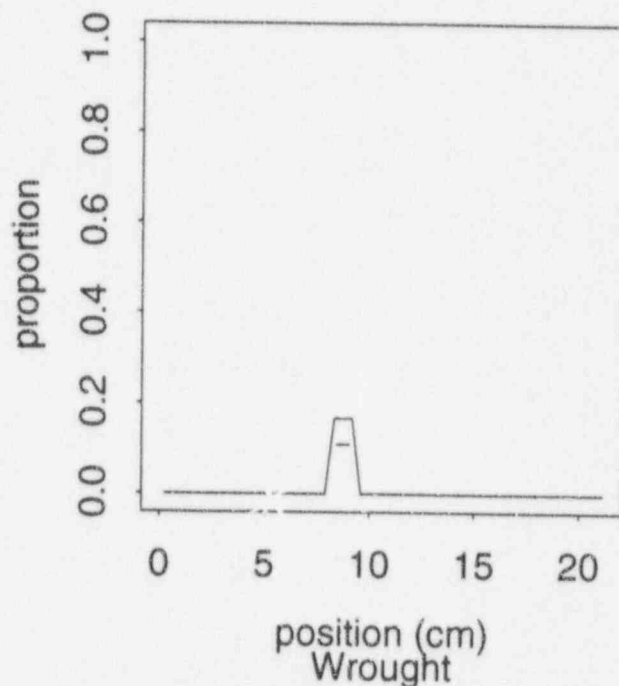
D105-1 A near # insp= 16



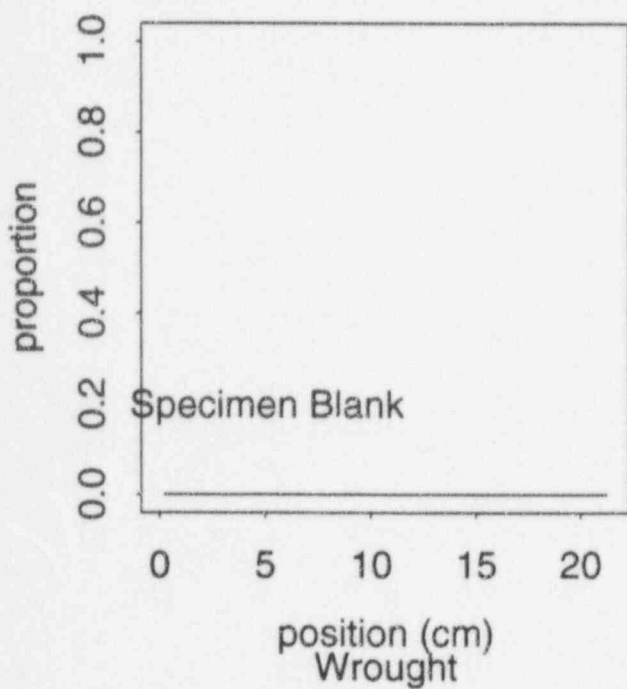
D105-1 B near # insp= 4



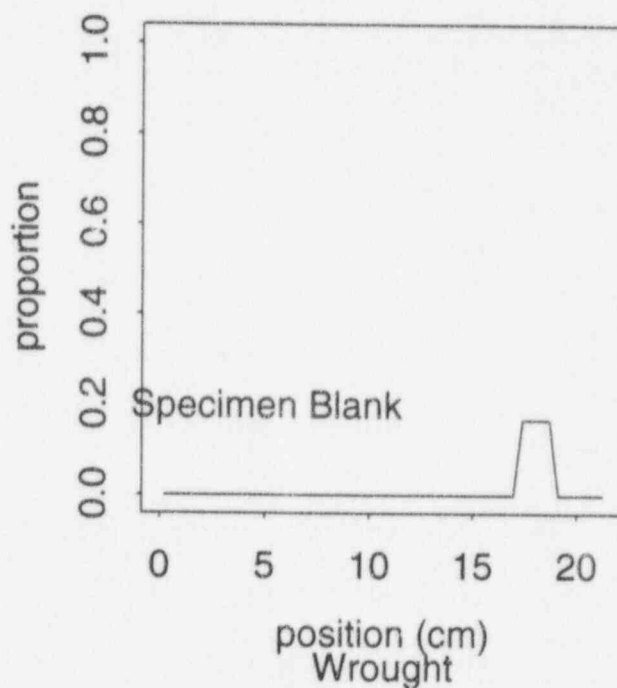
D105-4 A near # insp= 6



D107-1 A near # insp= 1

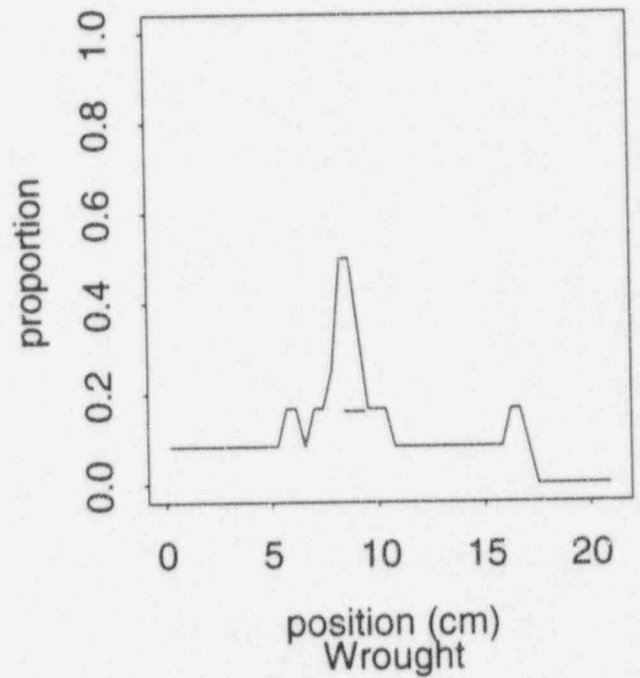
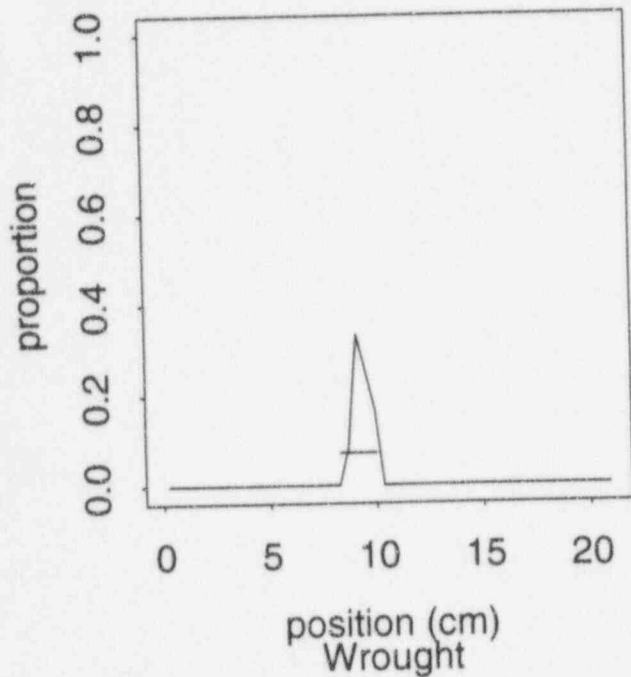


D107-1 B near # insp= 6



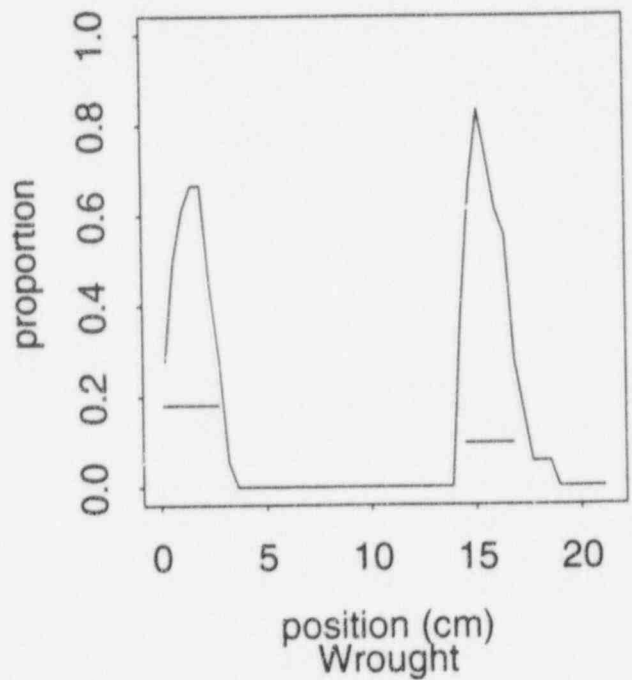
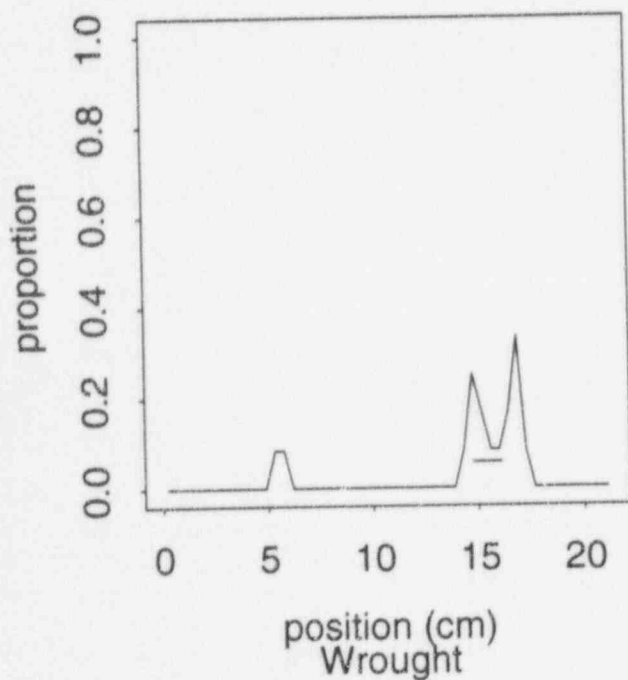
D107-2 A near # insp= 12

D107-2 B near # insp= 12

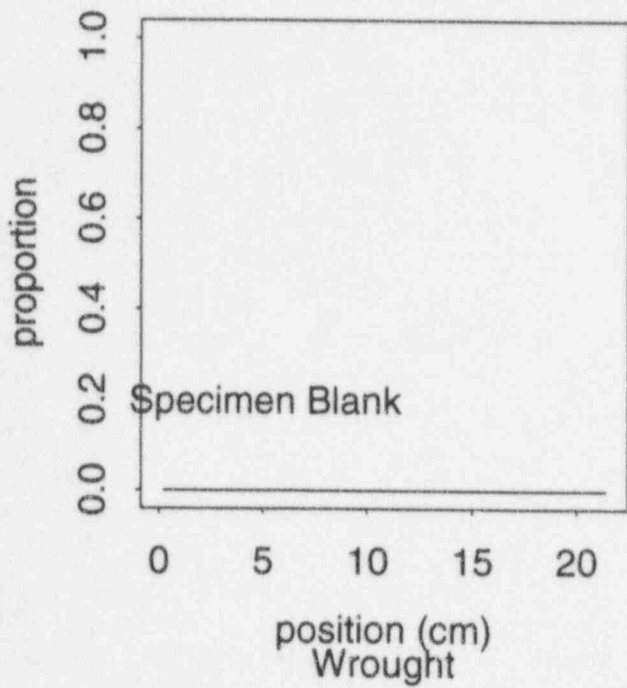


D107-3 A near # insp= 12

D107-3 B near # insp= 18



D107-4 B near # insp= 1



Appendix B

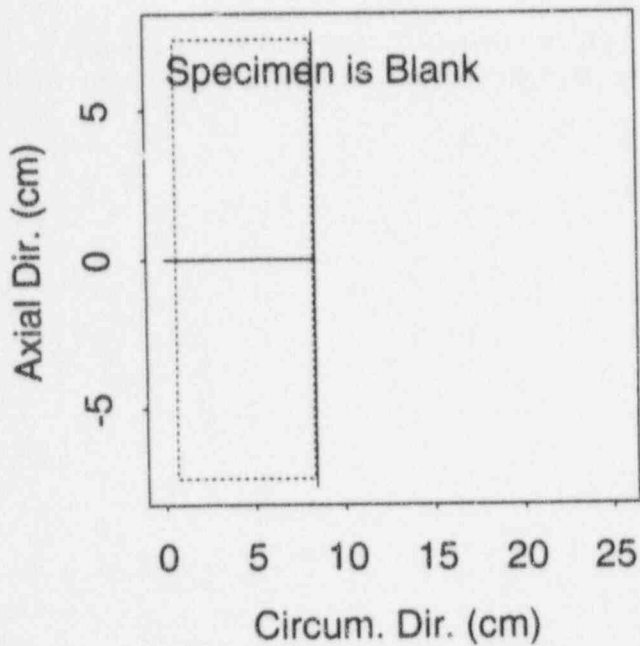
True-State Plots for PIRR Specimens

Appendix B

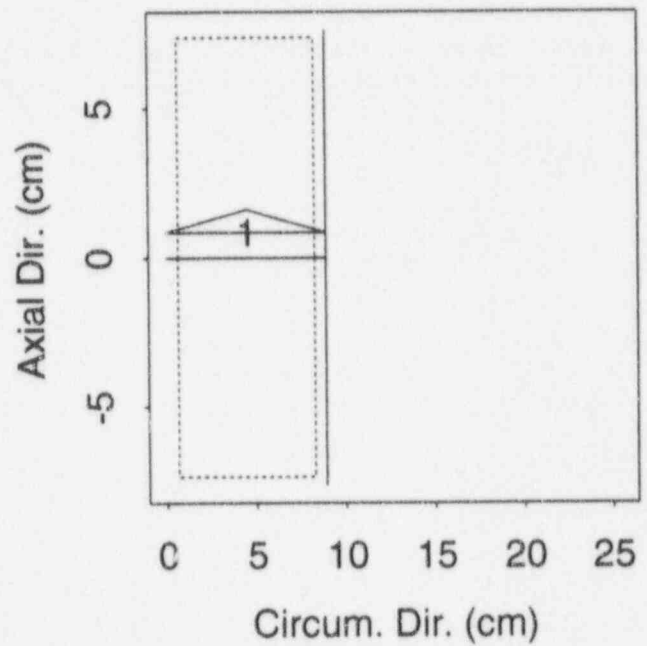
True-State Plots for PIRR Specimens

This appendix presents the true-state plots of all specimens used in the PIRR exercise. Grading units are marked as dotted lines and the depth of cracks are displayed as a triangle; the height of the triangle represents the depth of the crack.

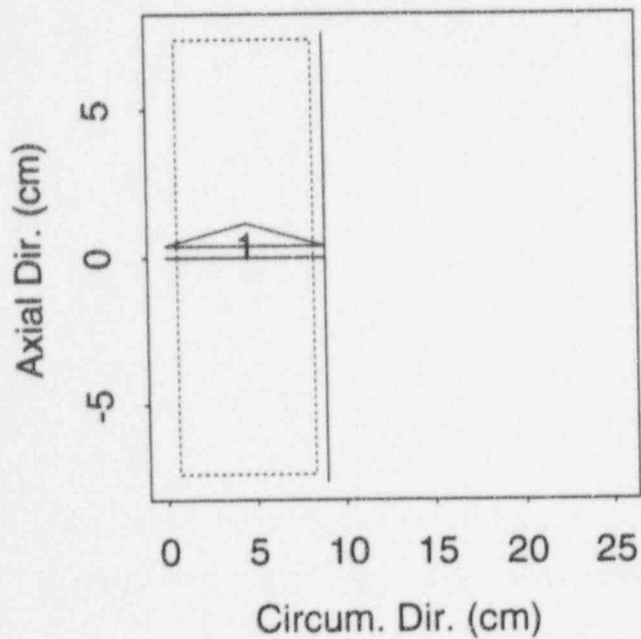
True State of 1 Cast



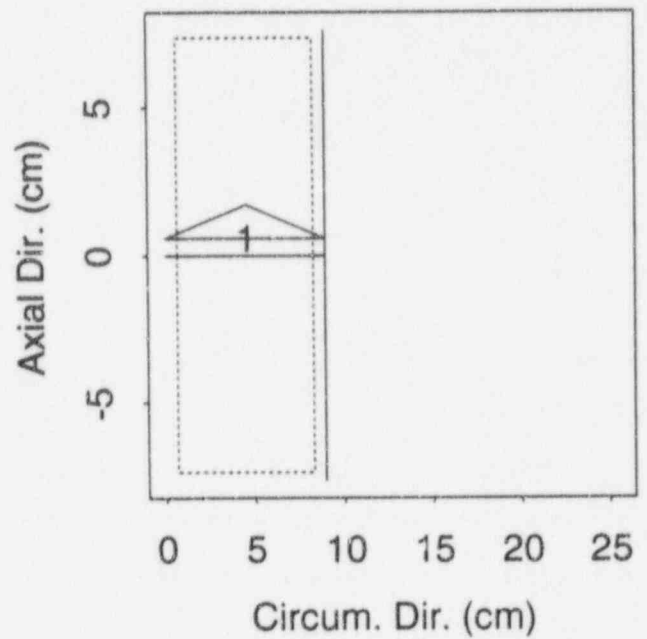
True State of 12 Cast



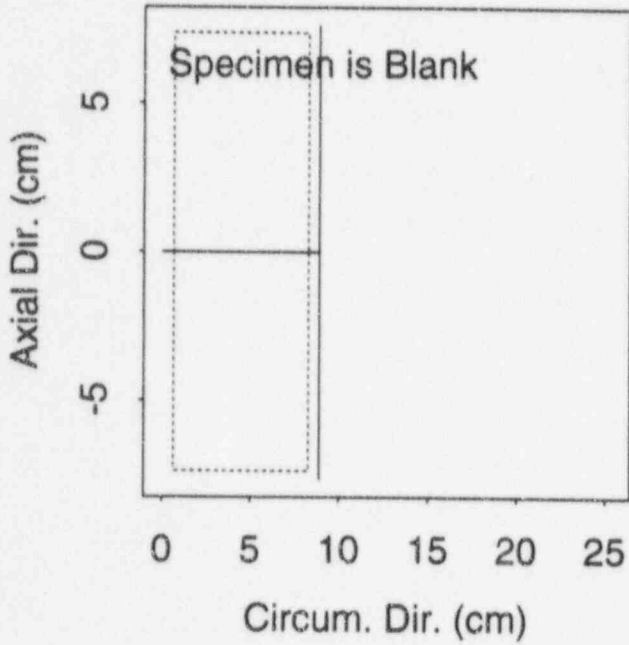
True State of 13 Cast



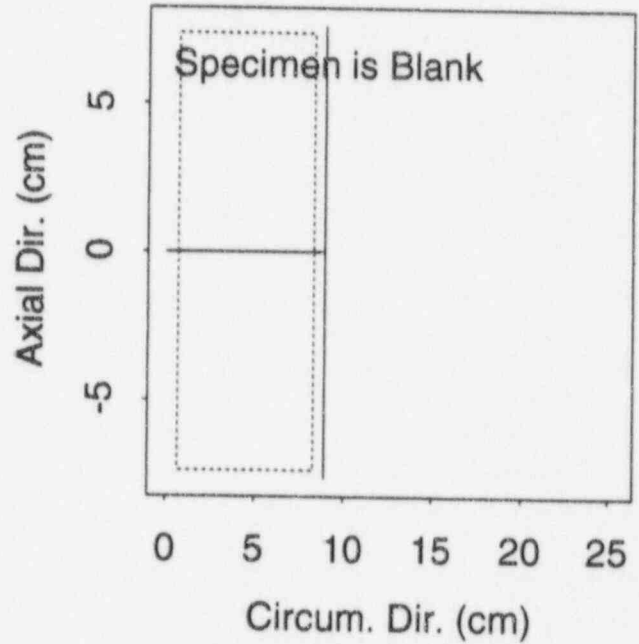
True State of 16 Cast



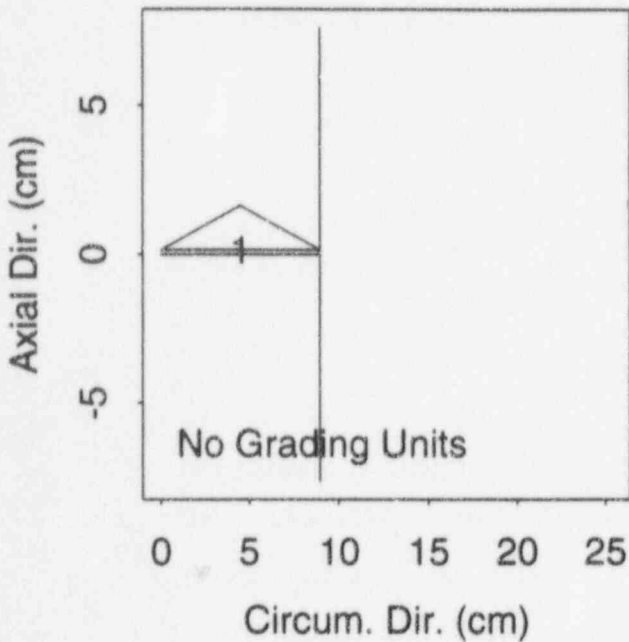
True State of 2 Cast



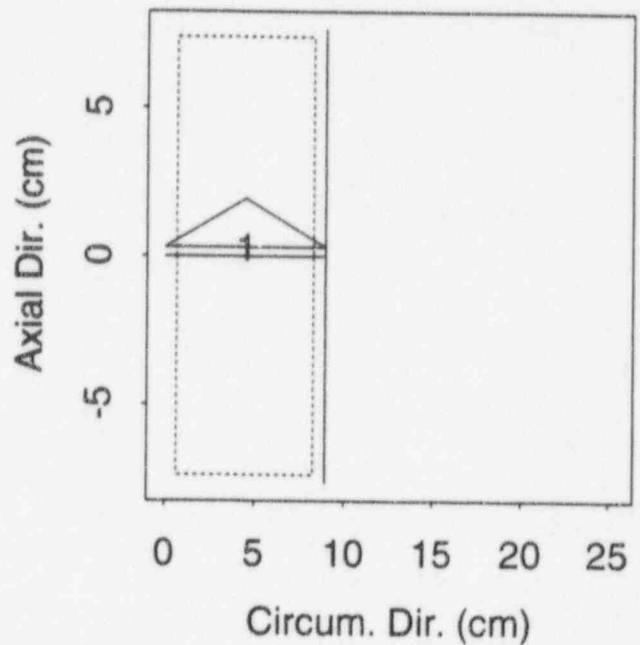
True State of 3 Cast



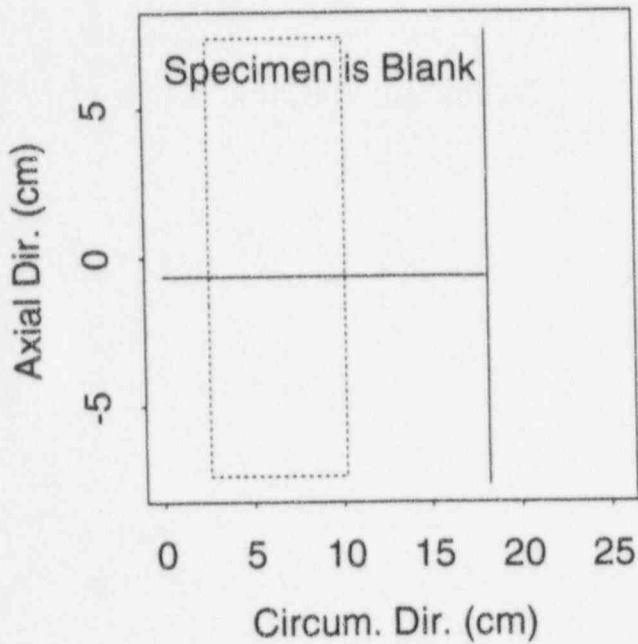
True State of 19 Cast



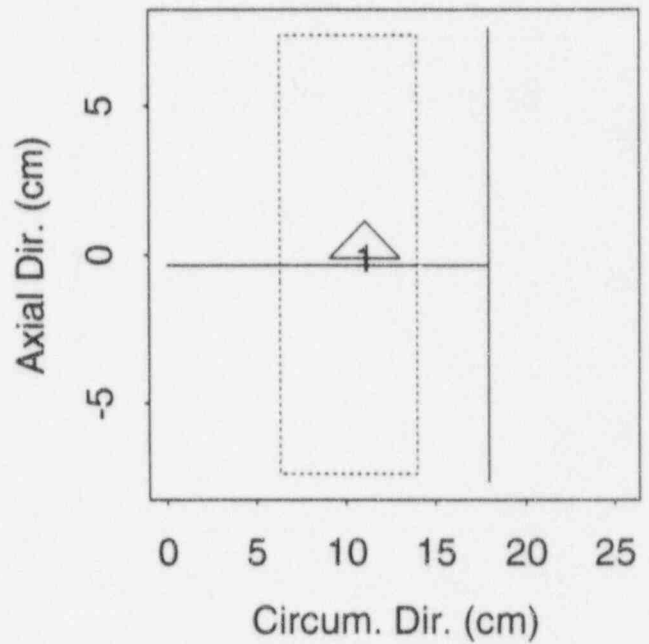
True State of 20 Cast



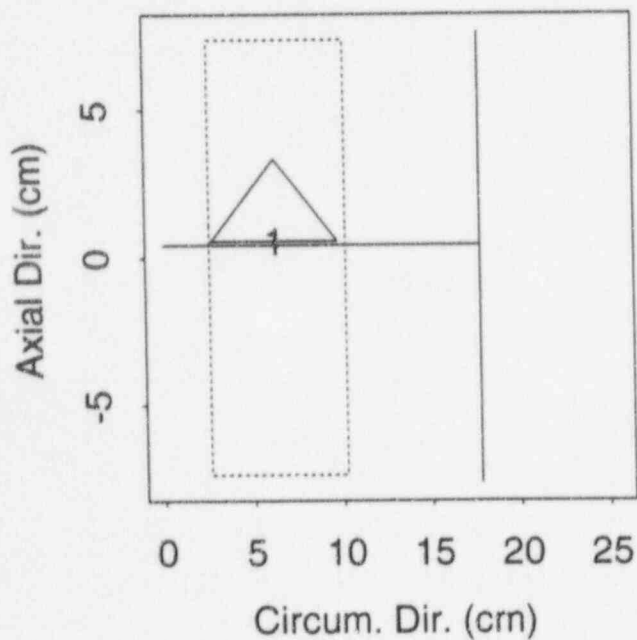
True State of B501 Cast



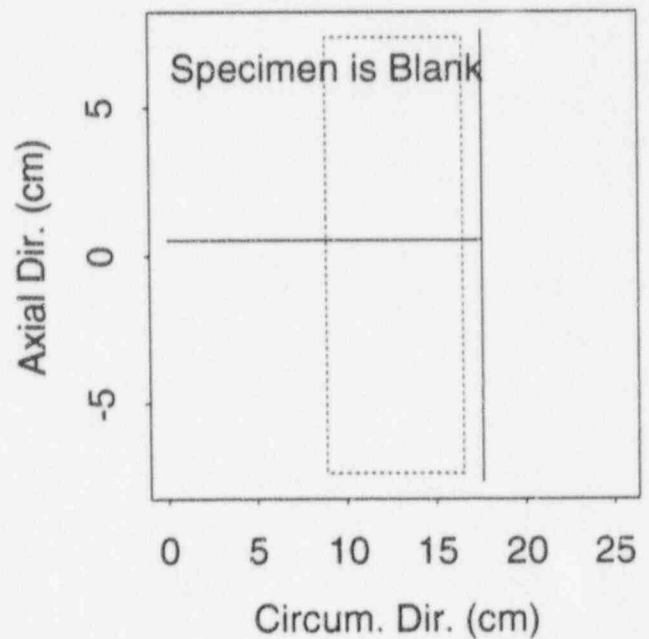
True State of B502 Cast



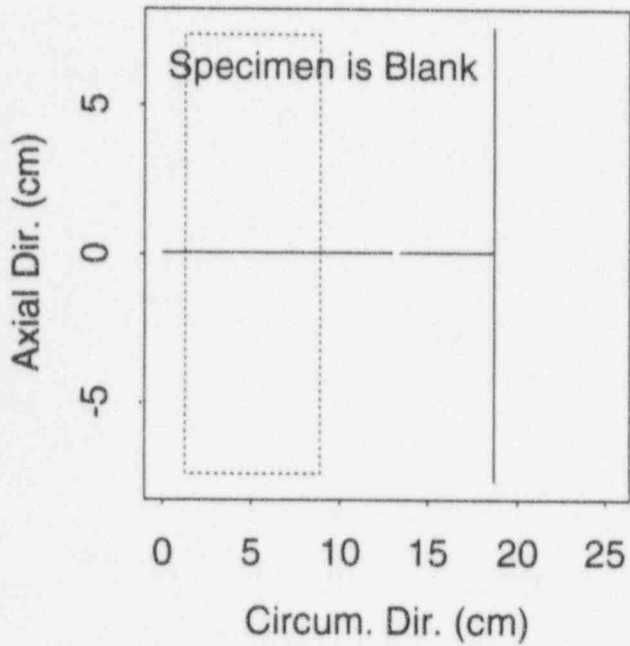
True State of B504 Cast



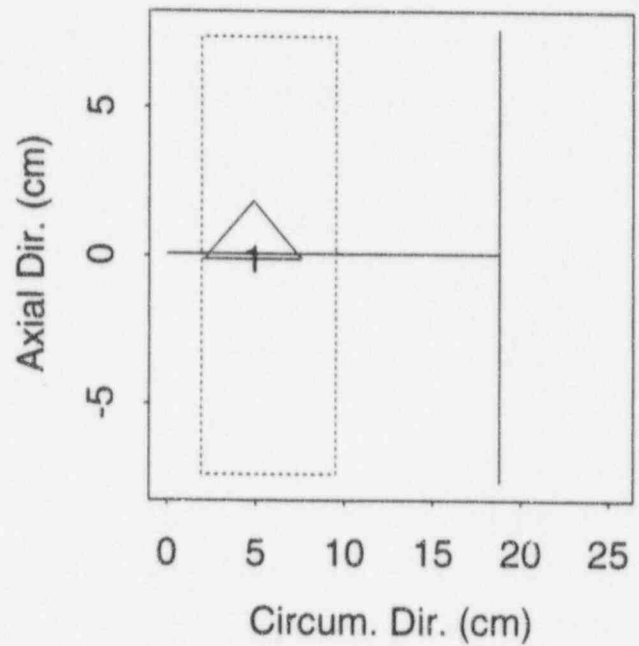
True State of B506 Cast



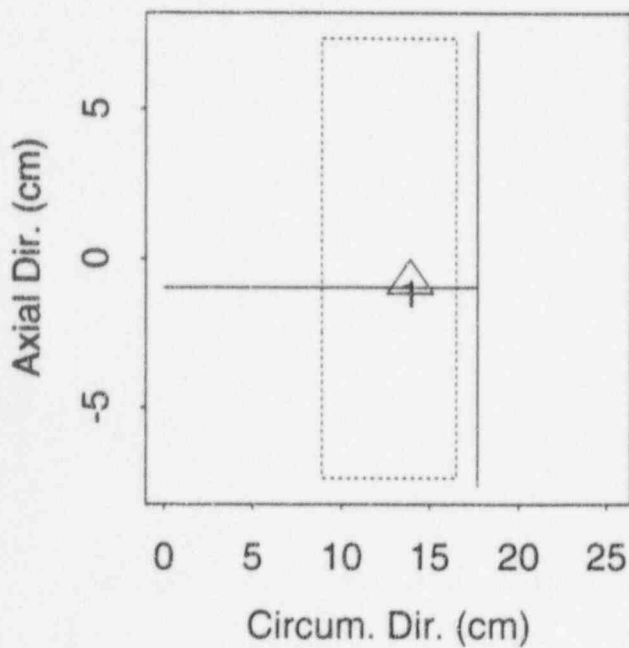
True State of B507 Cast



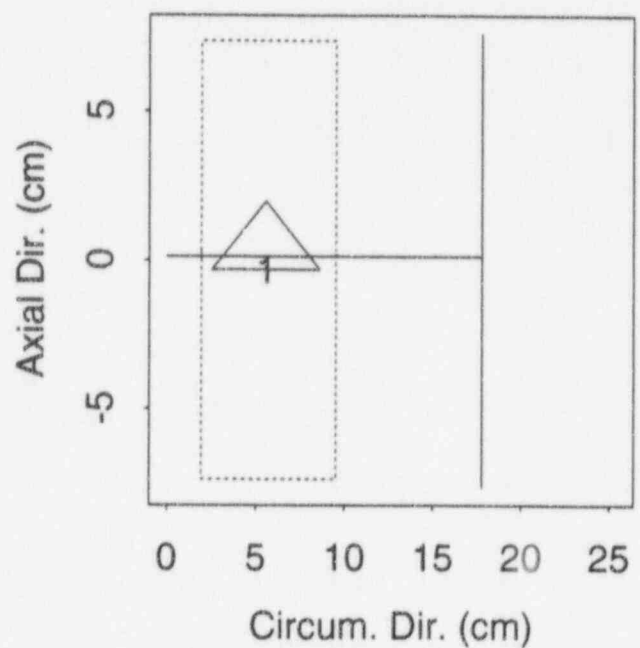
True State of B508 Cast



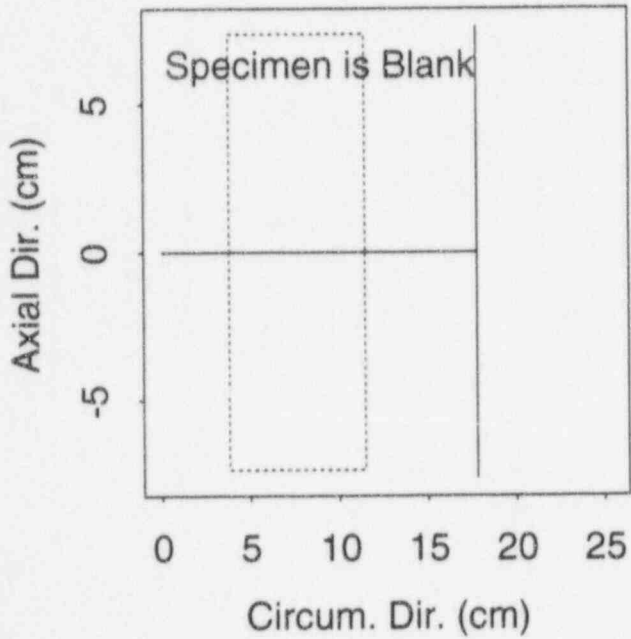
True State of B509 Cast



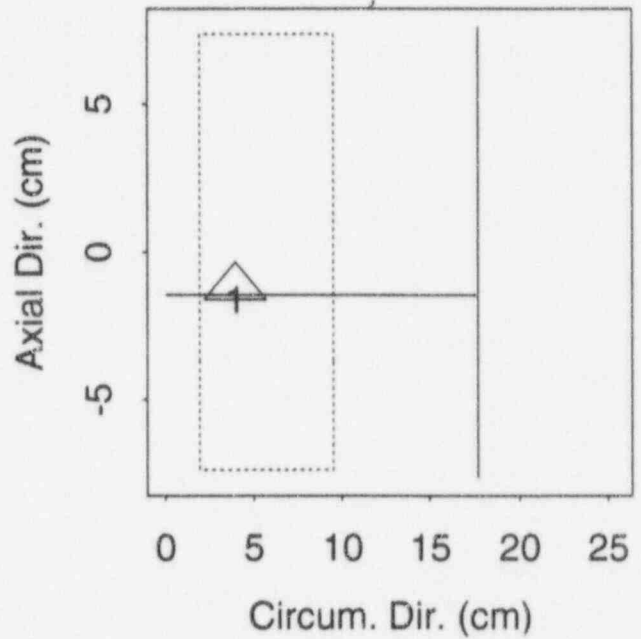
True State of B510 Cast



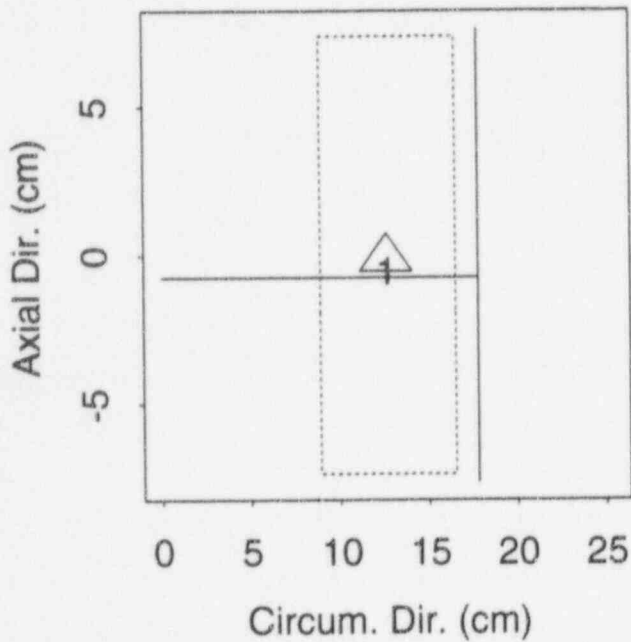
True State of B511 Cast



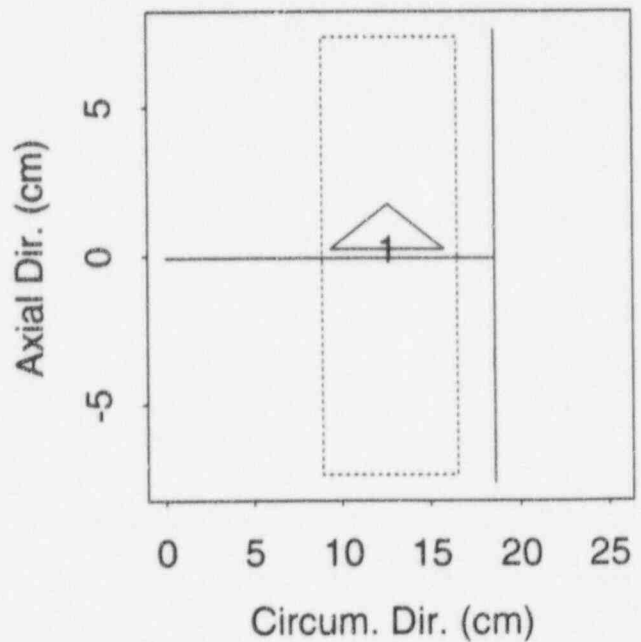
True State of B512 Cast



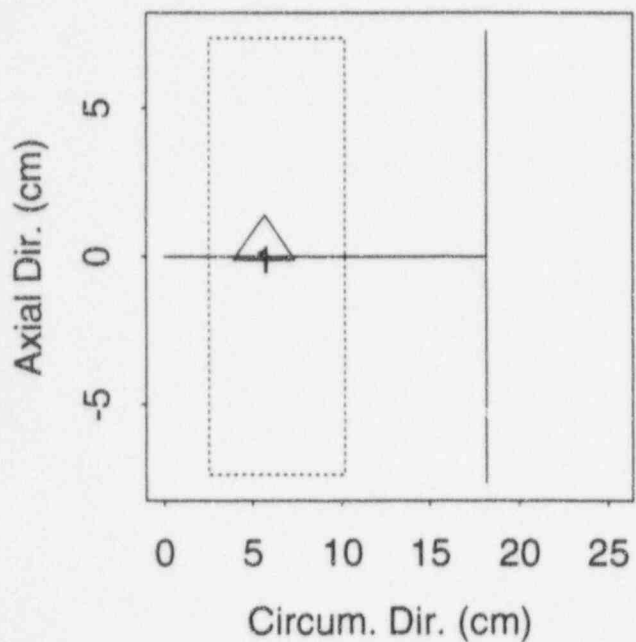
True State of B513 Cast



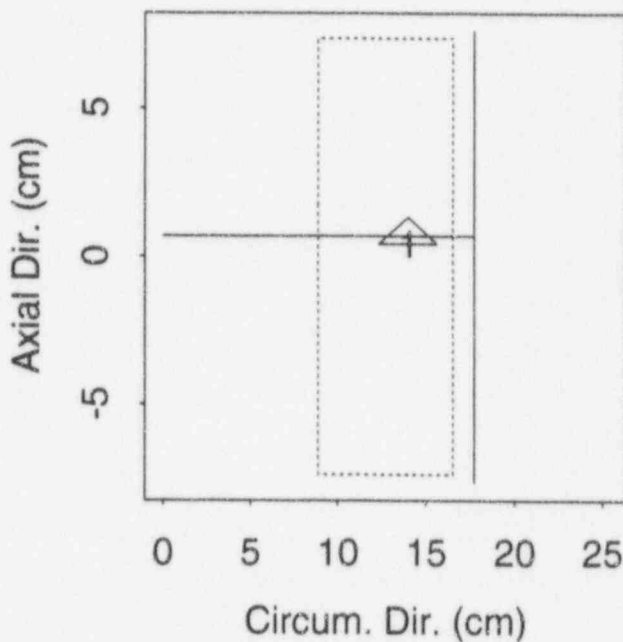
True State of B514 Cast



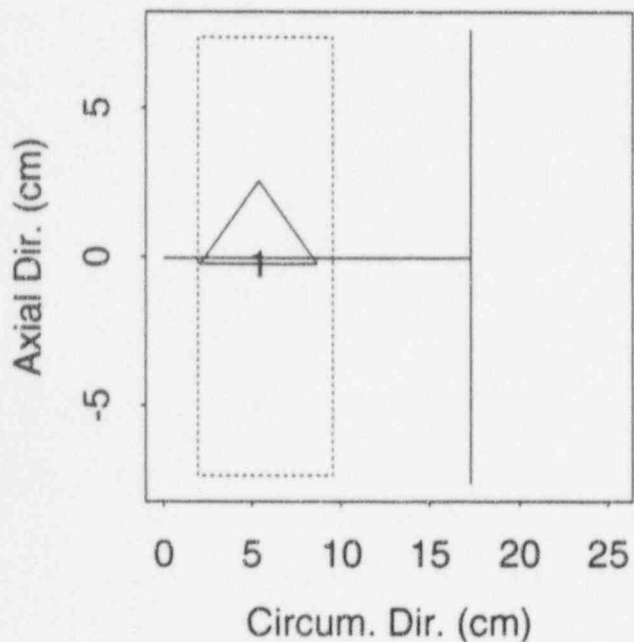
True State of B515 Cast



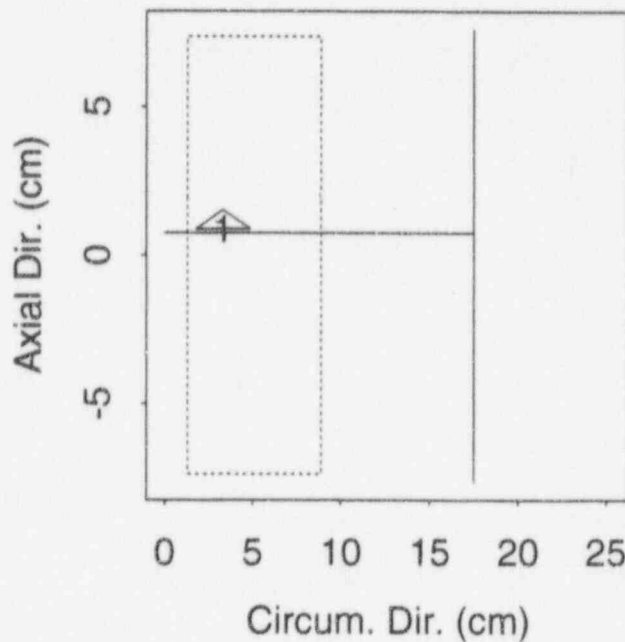
True State of B517 Cast



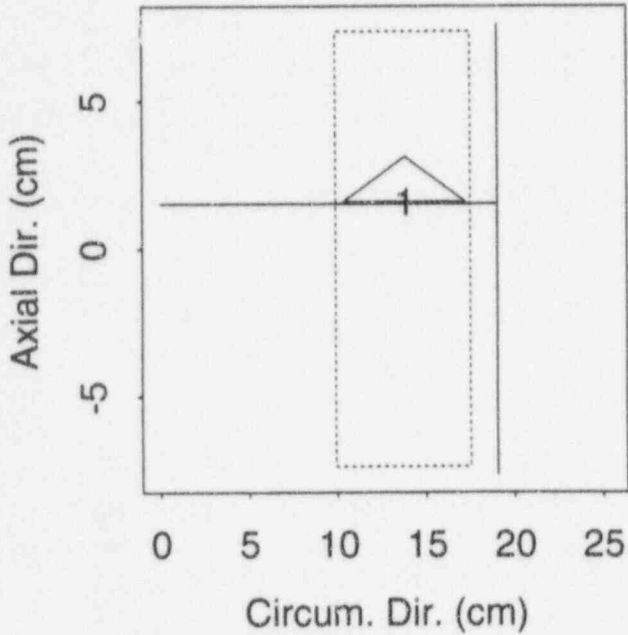
True State of B519 Cast



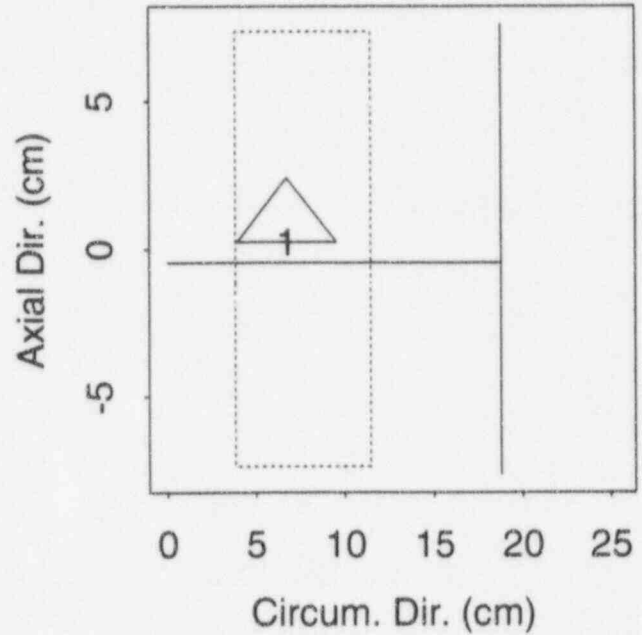
True State of B520 Cast



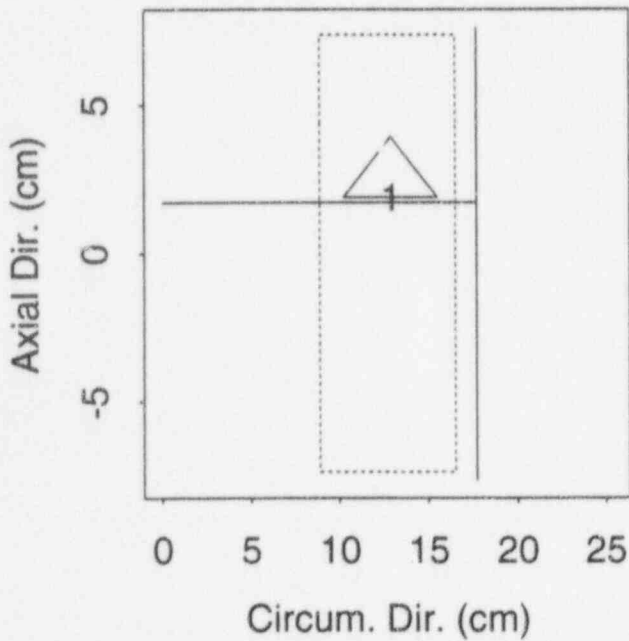
True State of B521 Cast



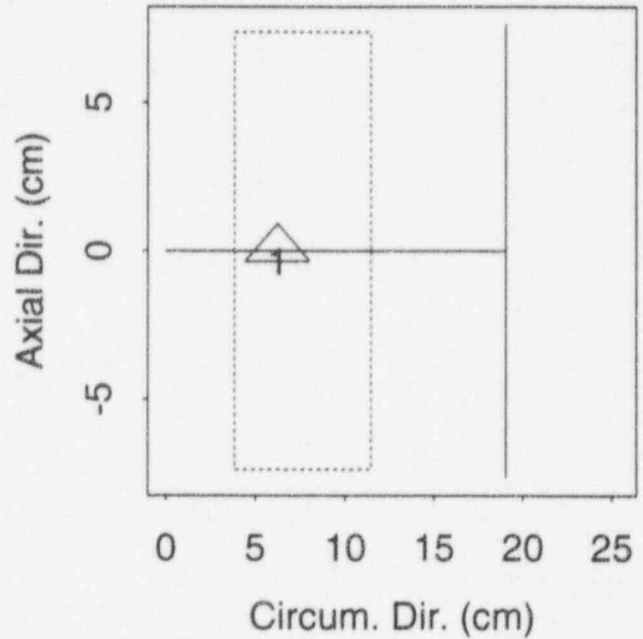
True State of B522 Cast



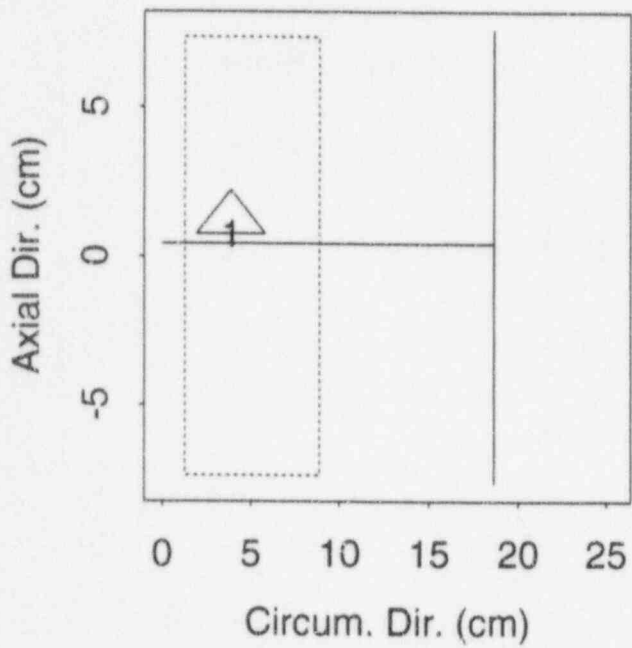
True State of B523 Cast



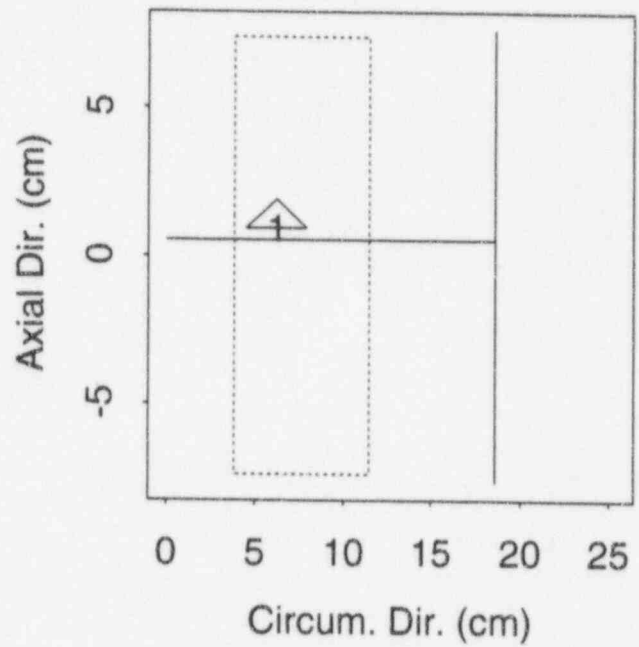
True State of B528 Cast



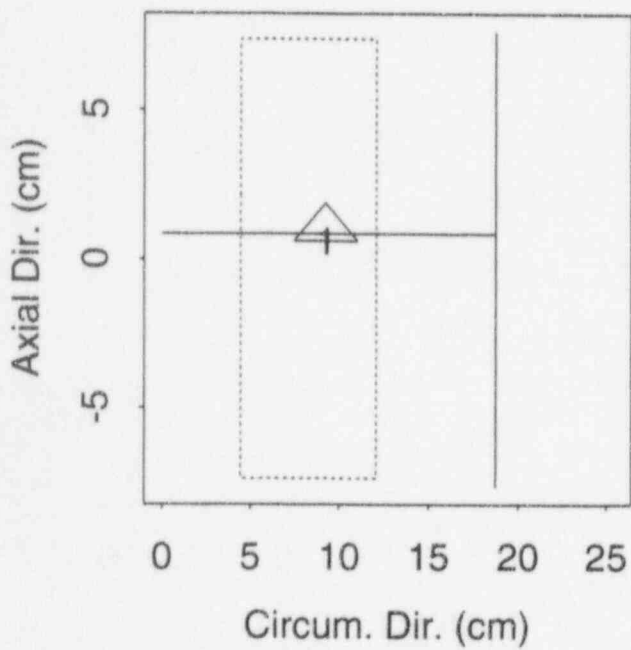
True State of B601 Clad



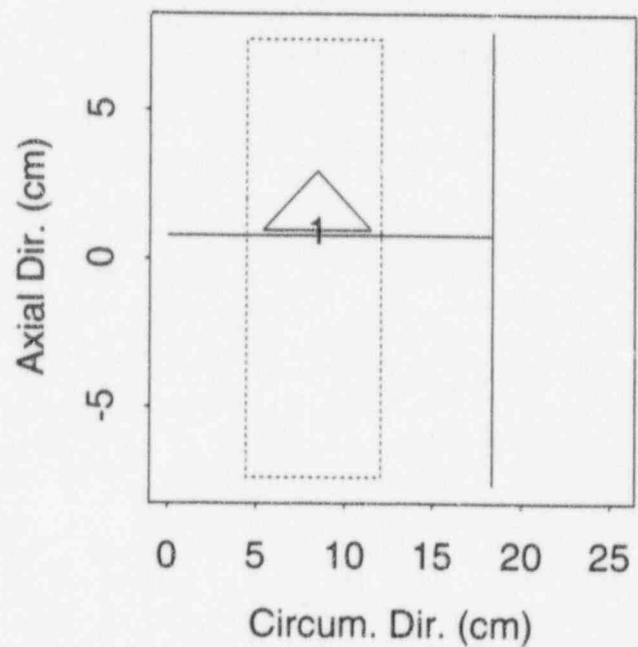
True State of B602 Clad



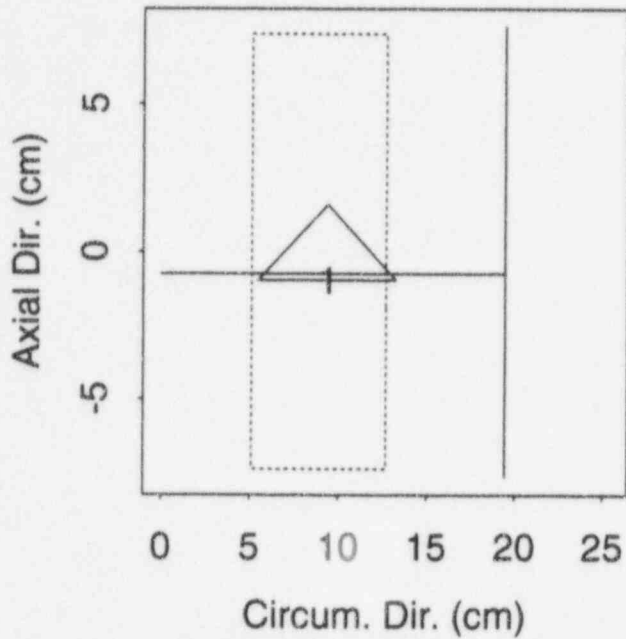
True State of B603 Clad



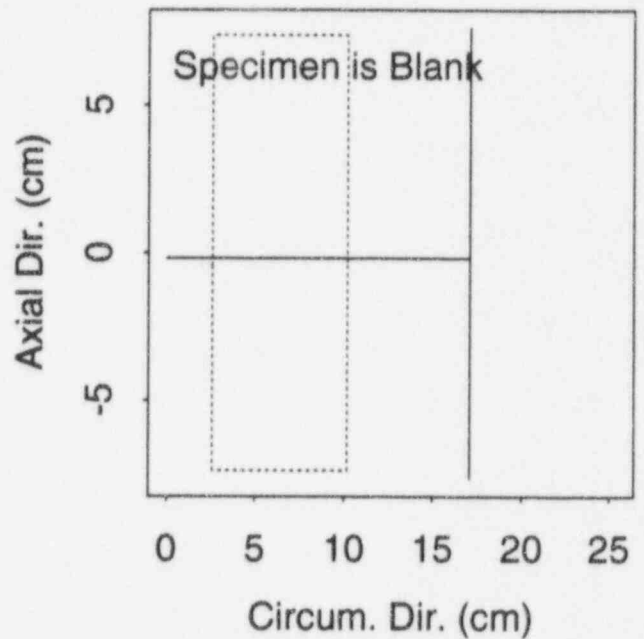
True State of B604 Clad



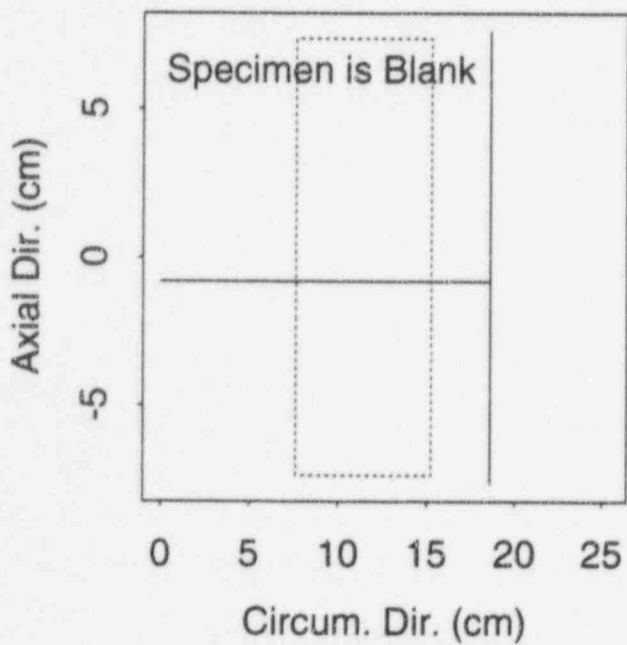
True State of B605 Clad



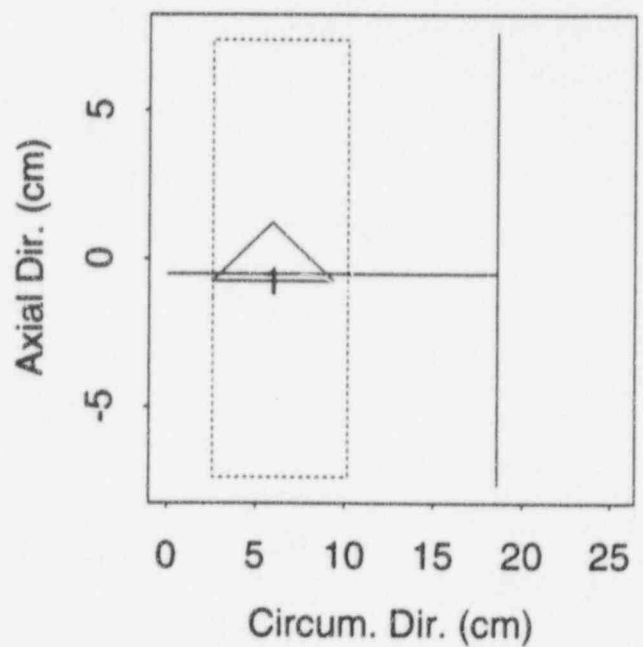
True State of B607 Clad



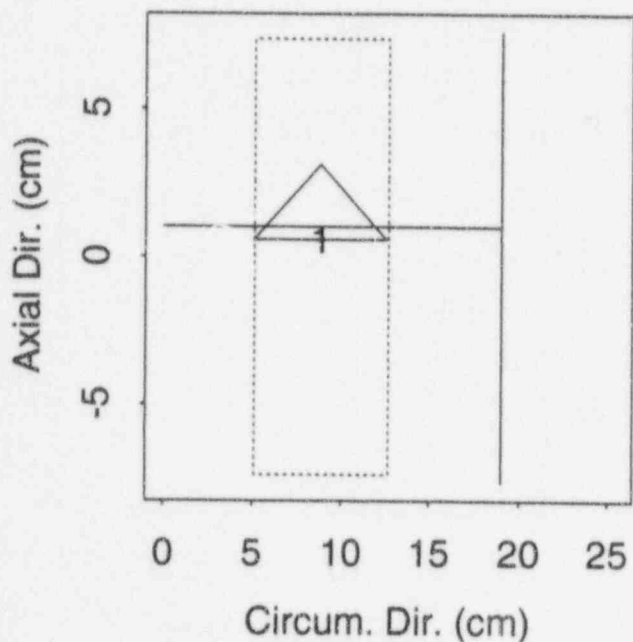
True State of B608 Clad



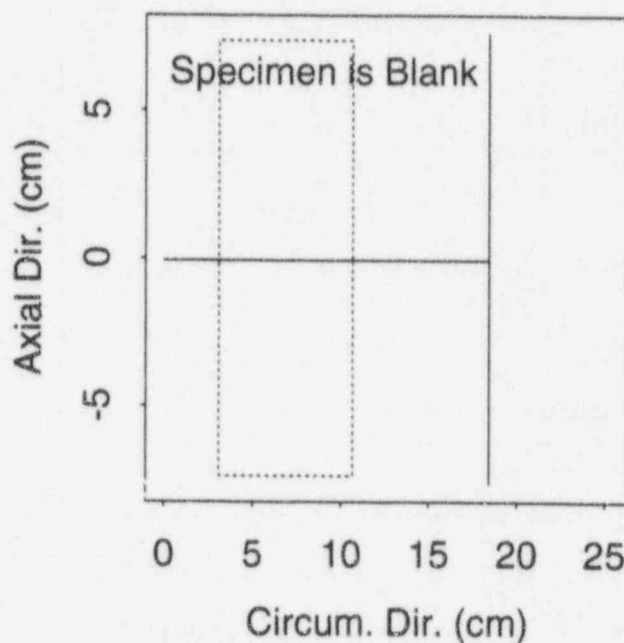
True State of B609 Clad



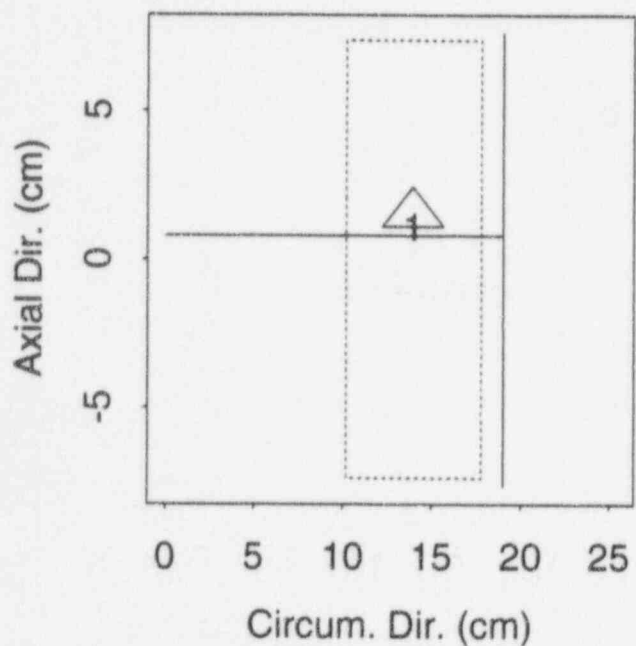
True State of B610 Clad



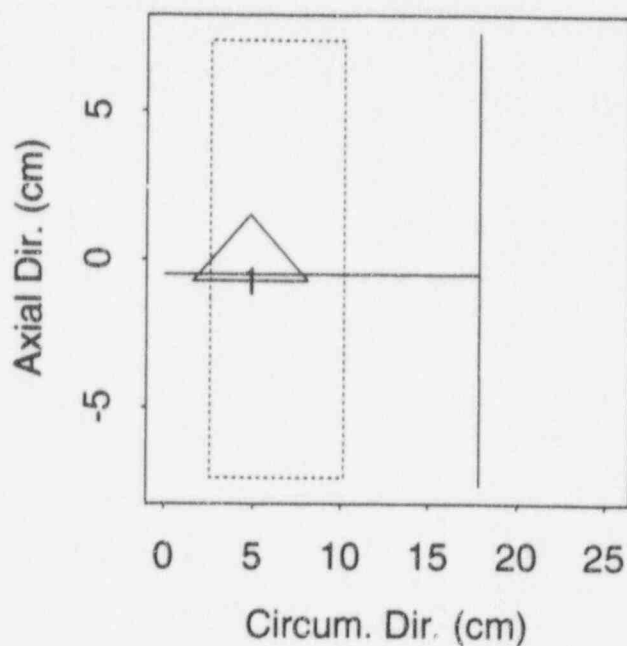
True State of B611 Clad



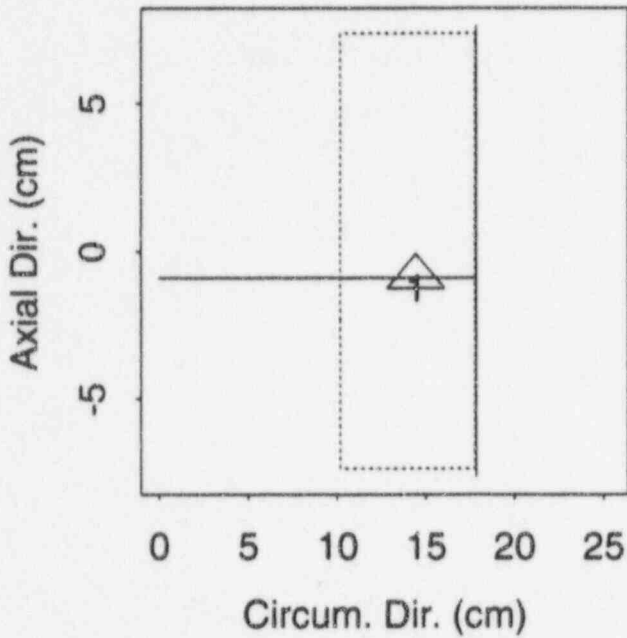
True State of B612 Clad



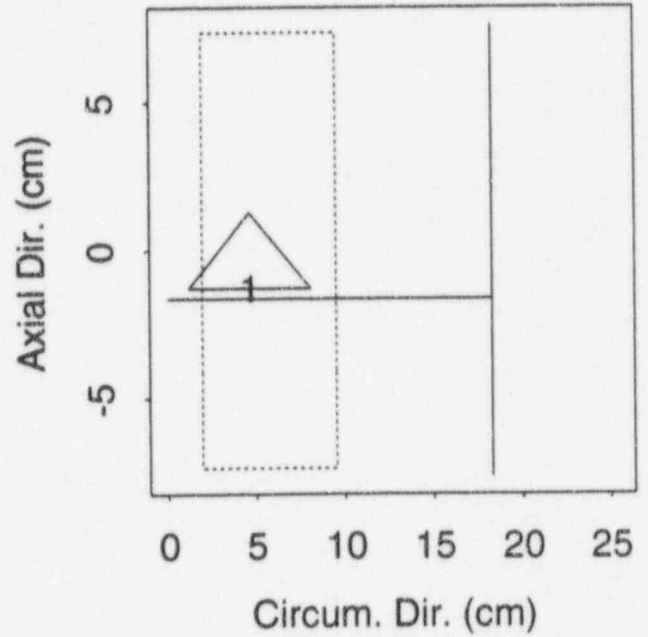
True State of B613 Clad



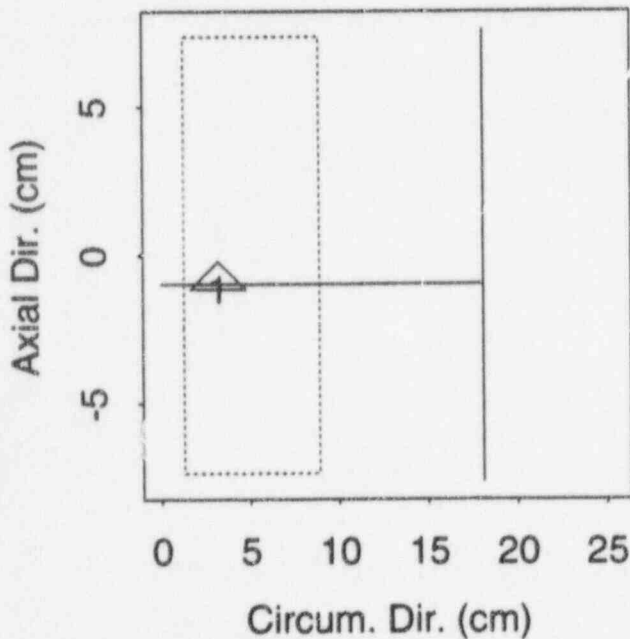
True State of B614 Clad



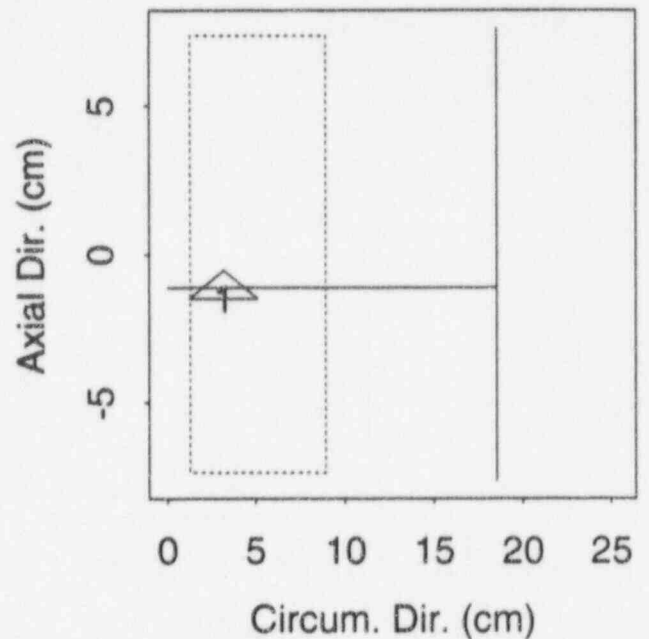
True State of B615 Clad



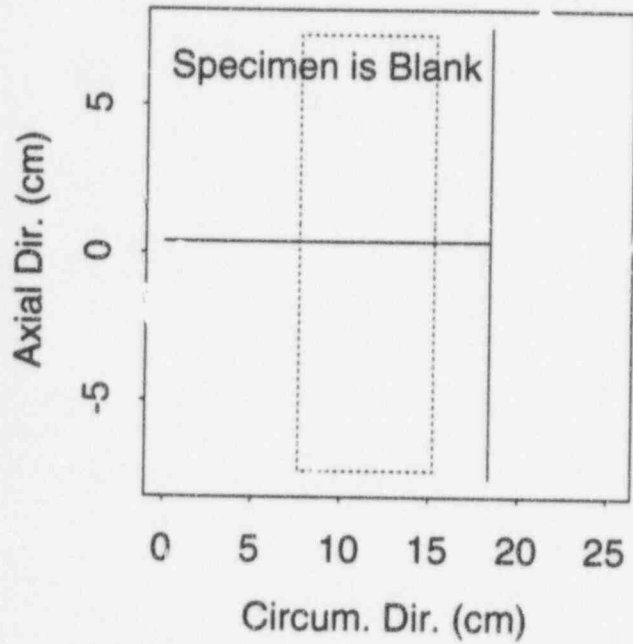
True State of B616 Clad



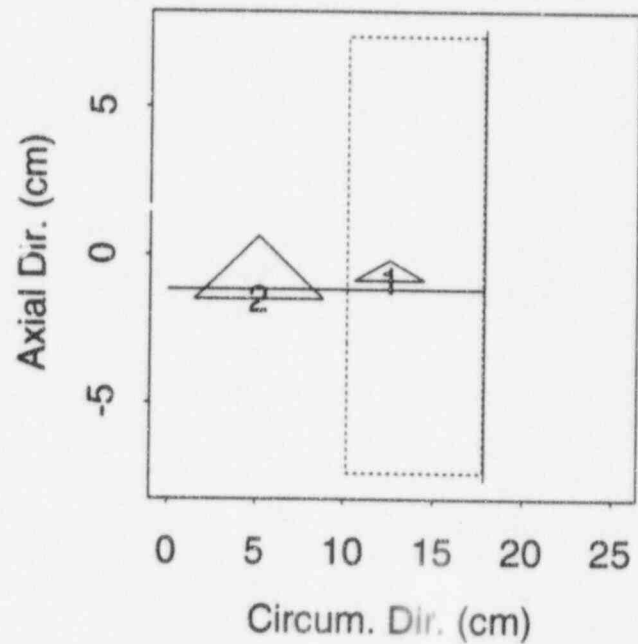
True State of B617 Clad



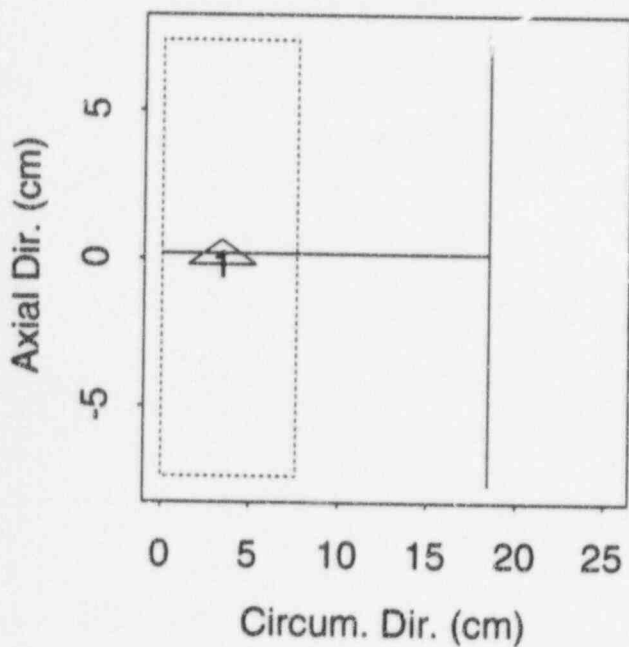
True State of B618 Clad



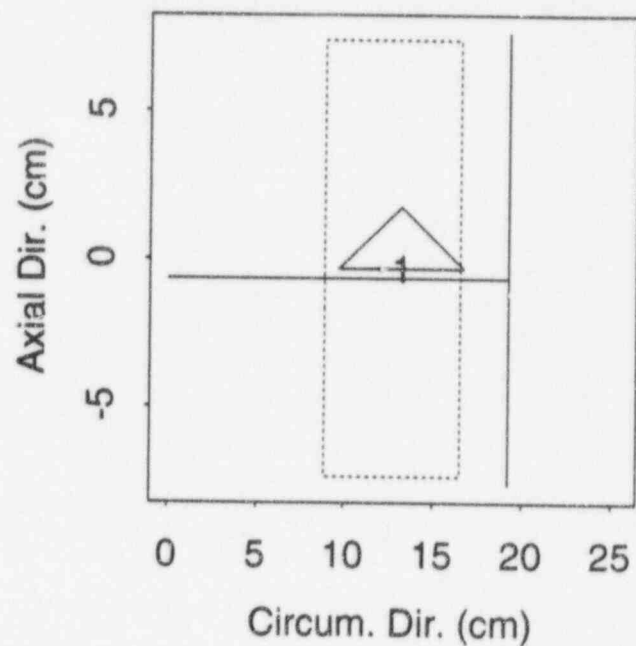
True State of B619 Clad



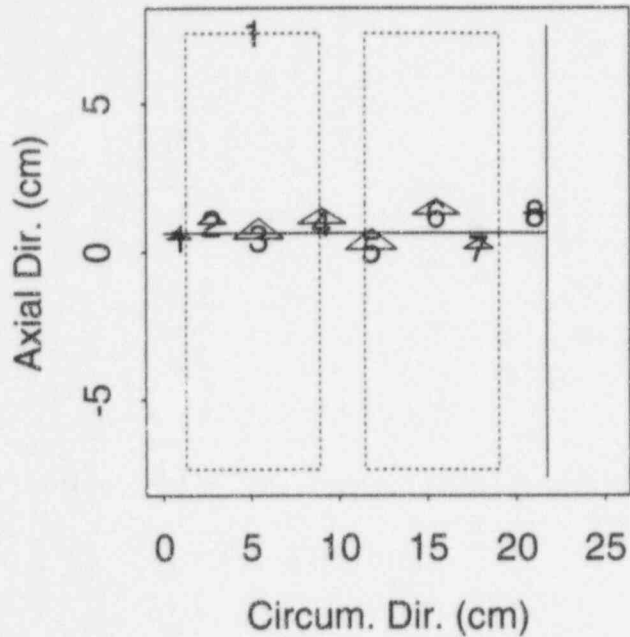
True State of B620 Clad



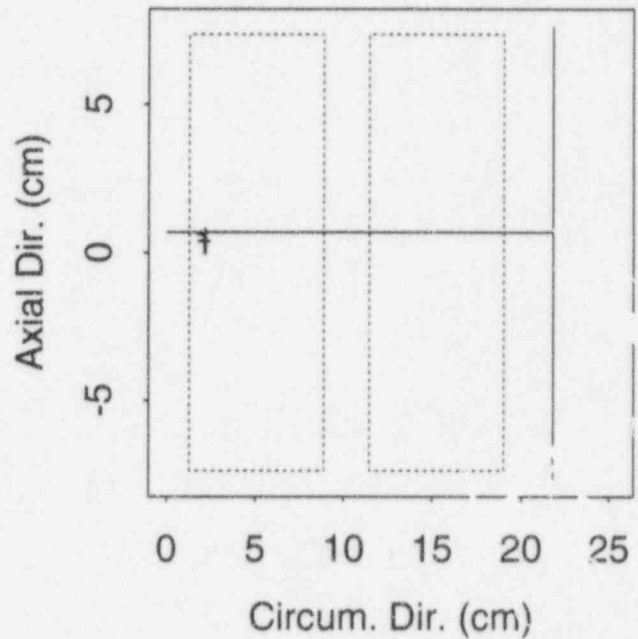
True State of B621 Clad



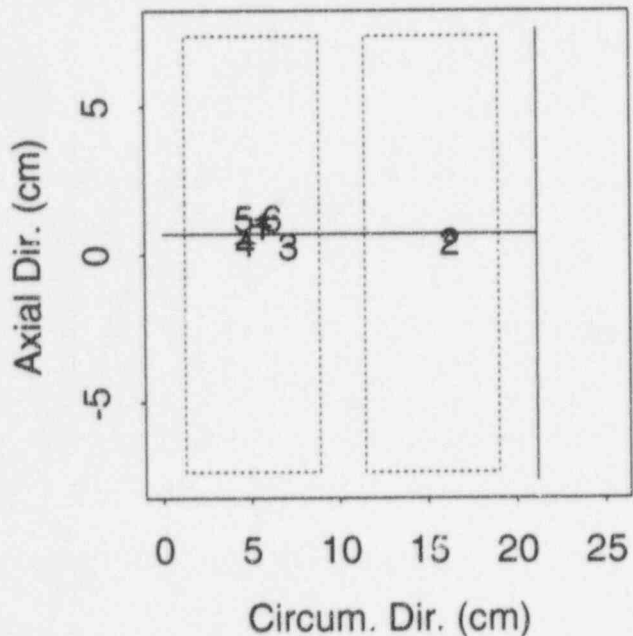
True State of
B211-1 Wrought



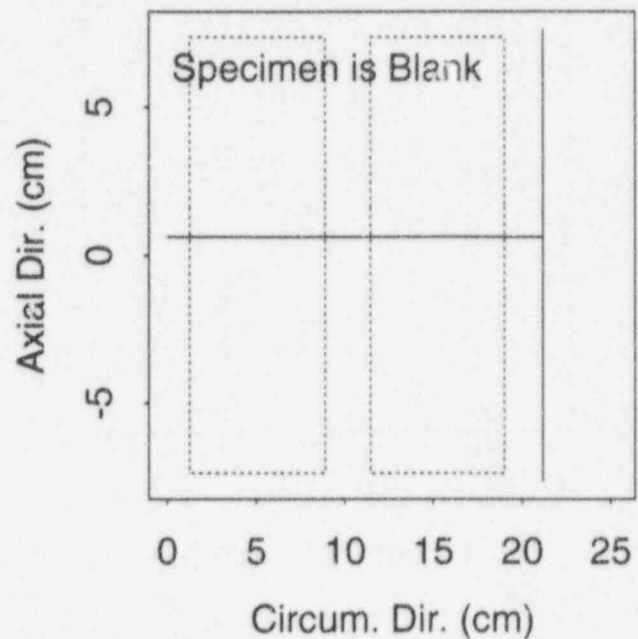
True State of
B211-2 Wrought



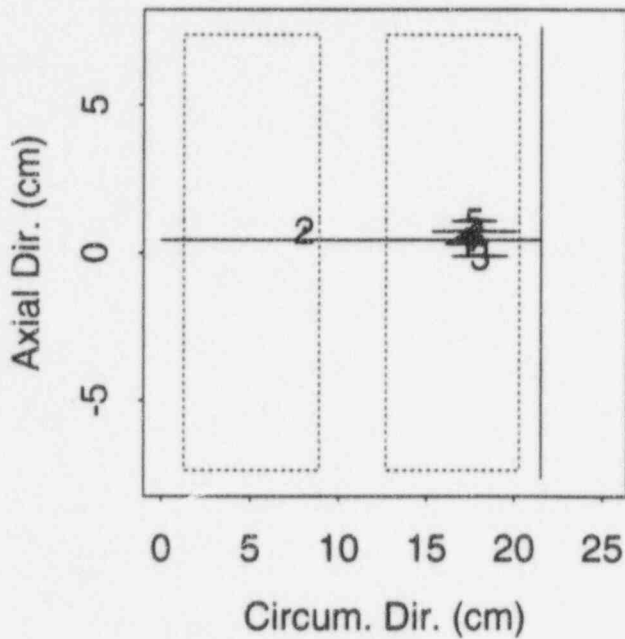
True State of
B211-3 Wrought



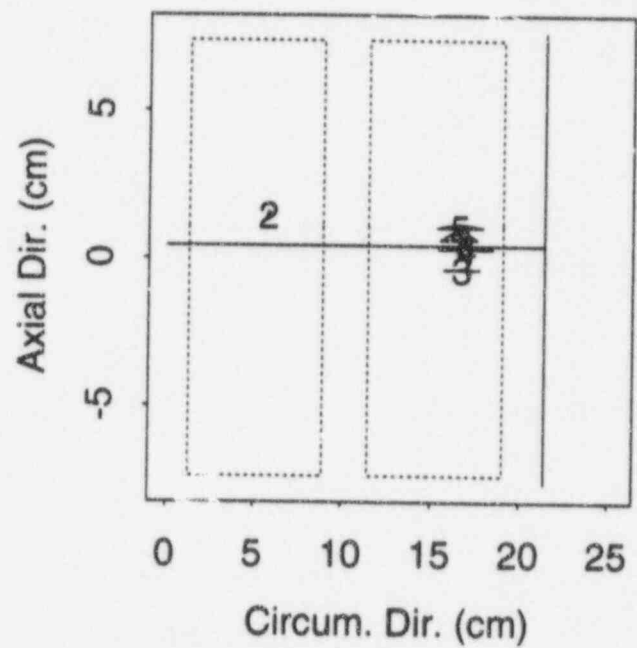
True State of
B211-4 Wrought



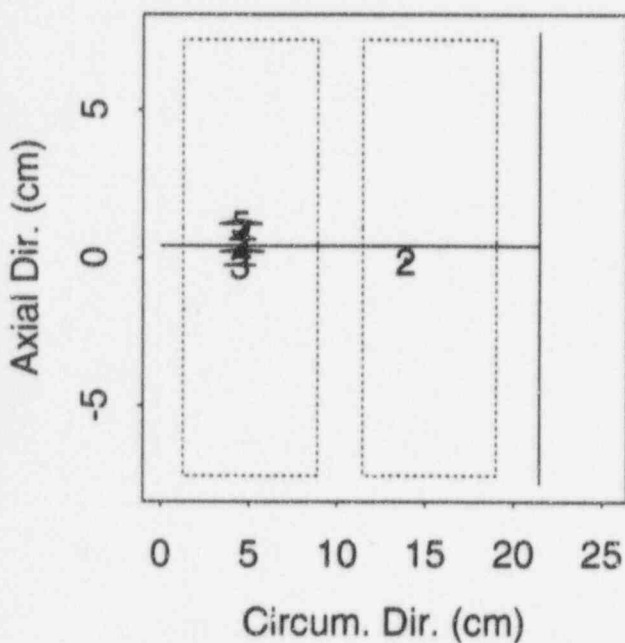
True State of
B212-1 Wrought



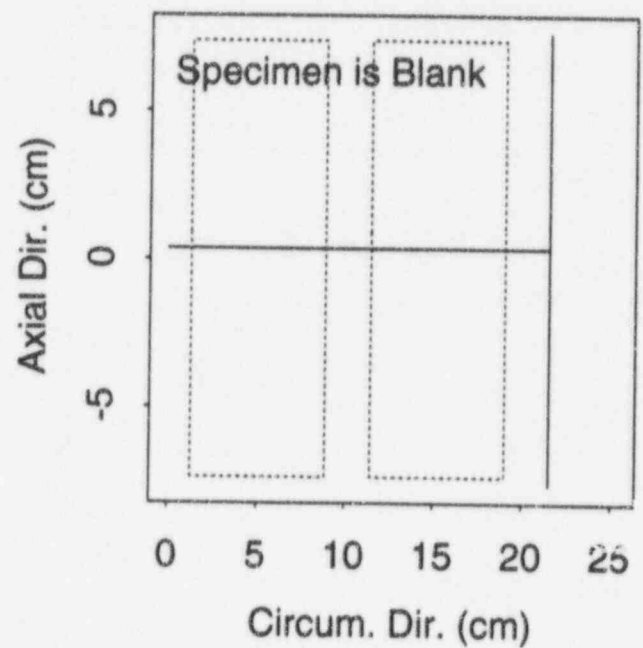
True State of
B212-2 Wrought



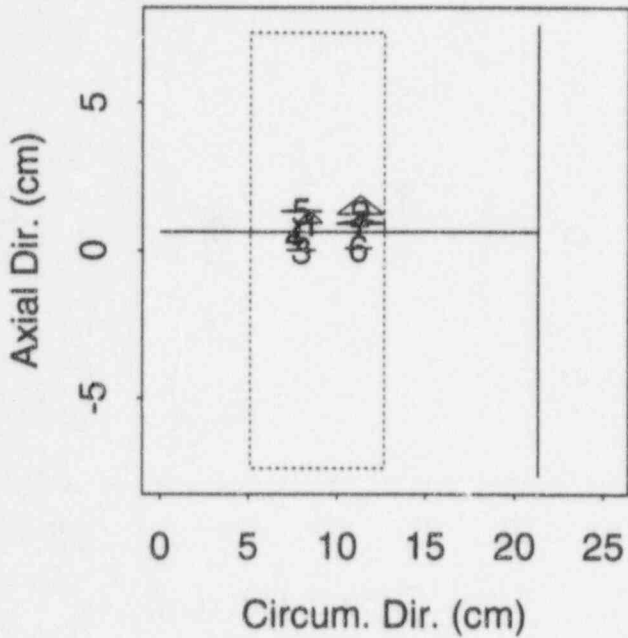
True State of
B212-3 Wrought



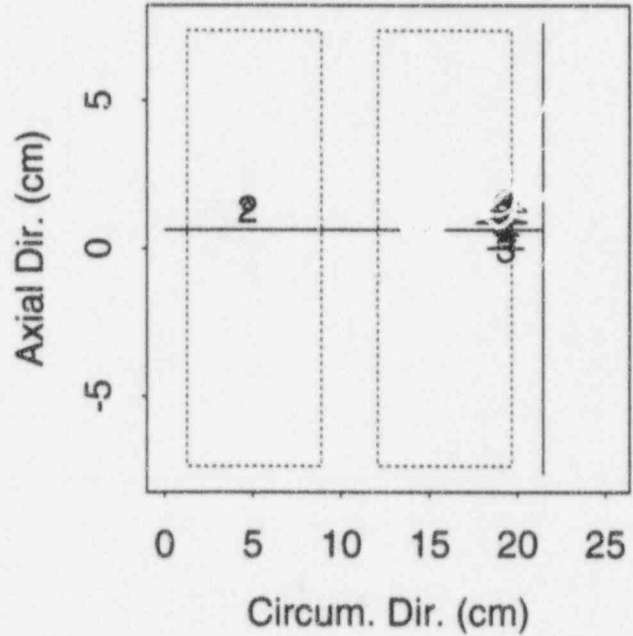
True State of
B212-4 Wrought



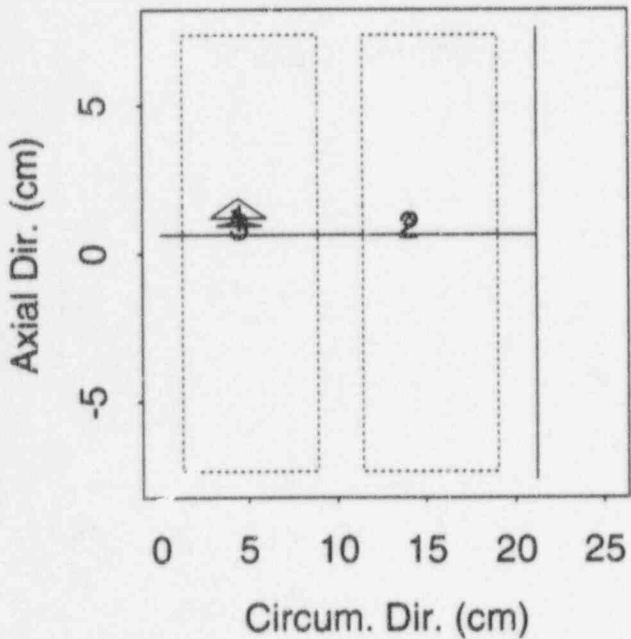
True State of B213-1 Wrought



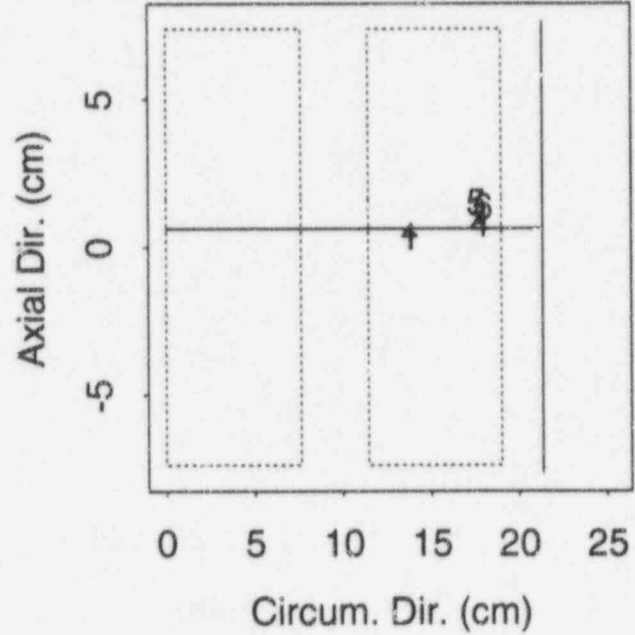
True State of B213-2 Wrought



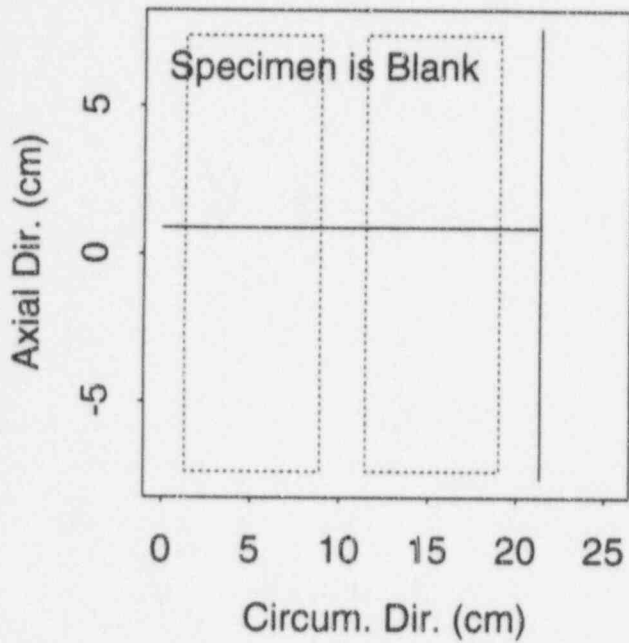
True State of B213-3 Wrought



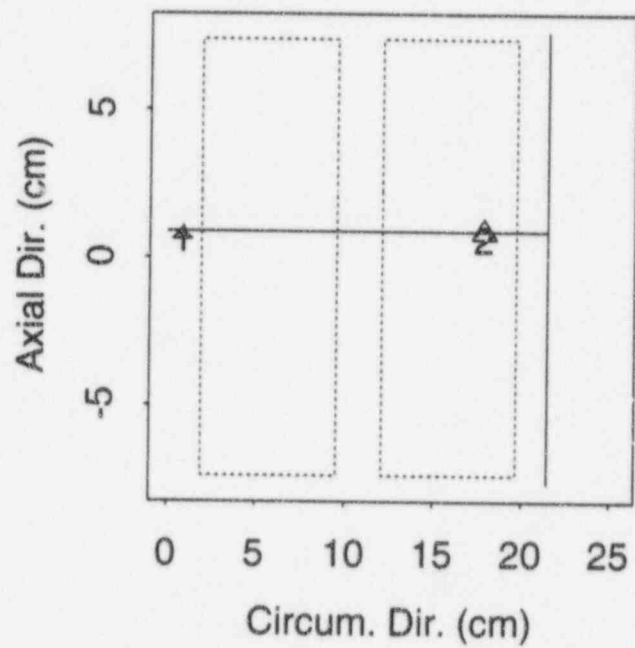
True State of B213-4 Wrought



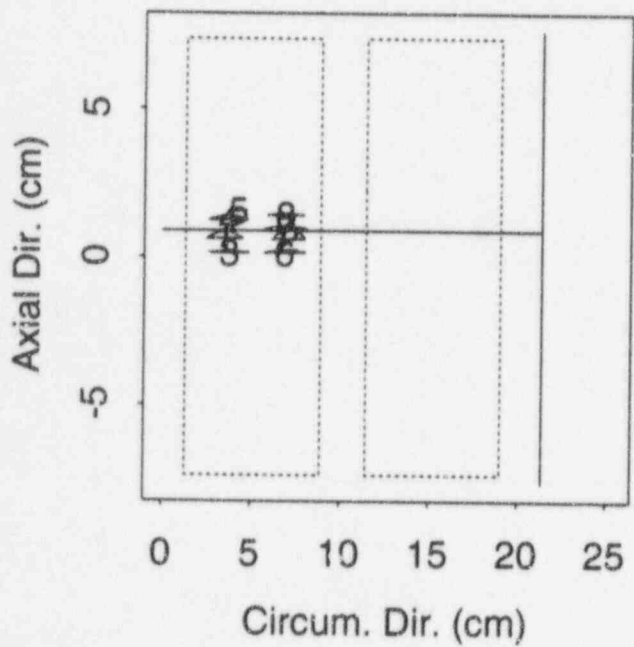
True State of B214-1 Wrought



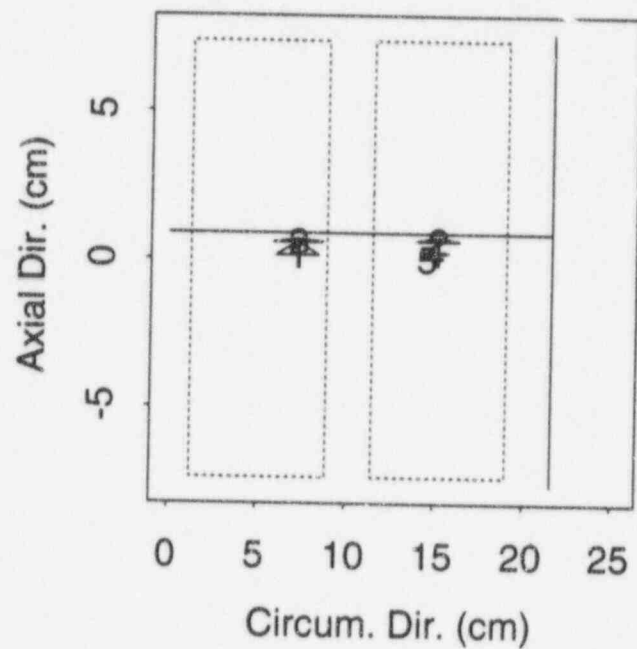
True State of B214-2 Wrought



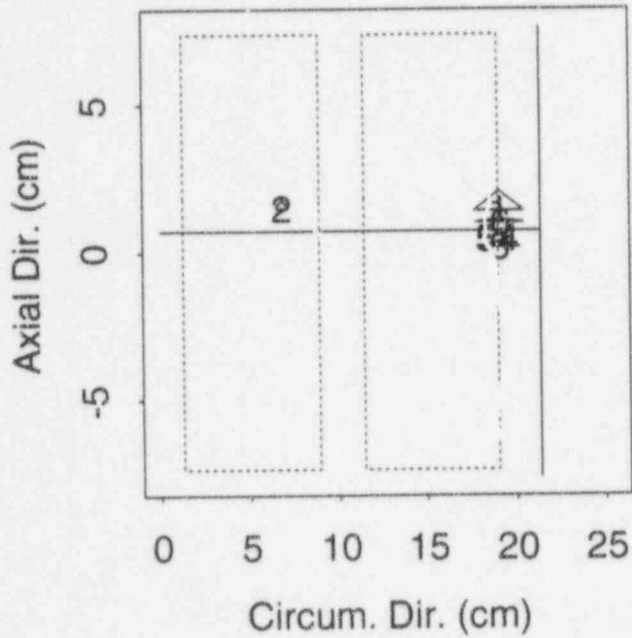
True State of B214-3 Wrought



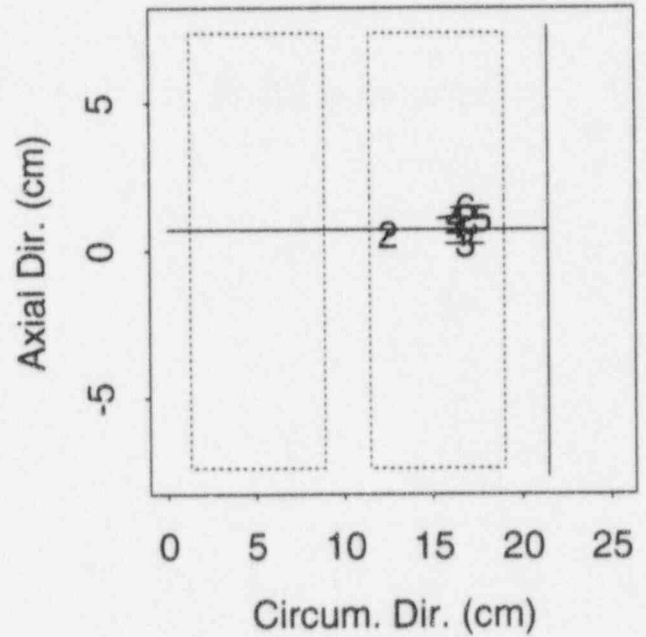
True State of B214-4 Wrought



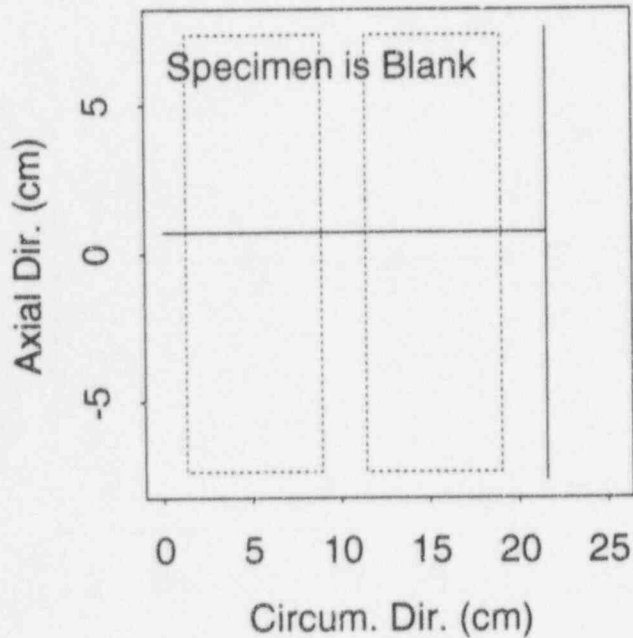
True State of
B215-1 Wrought



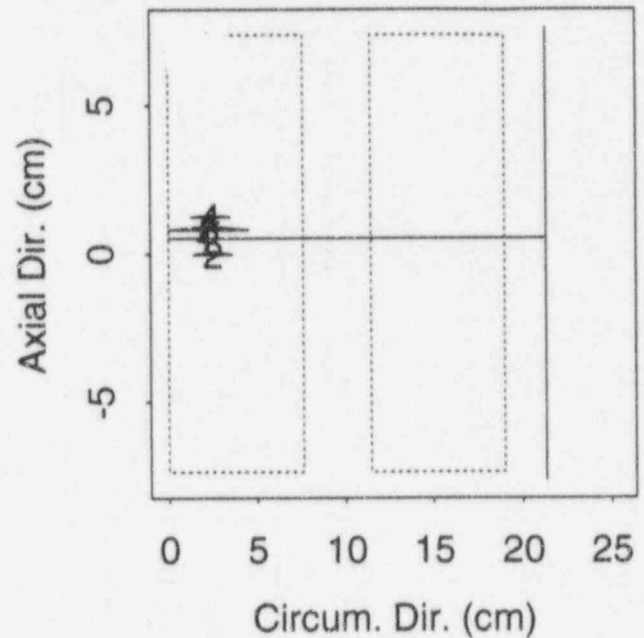
True State of
B215-3 Wrought



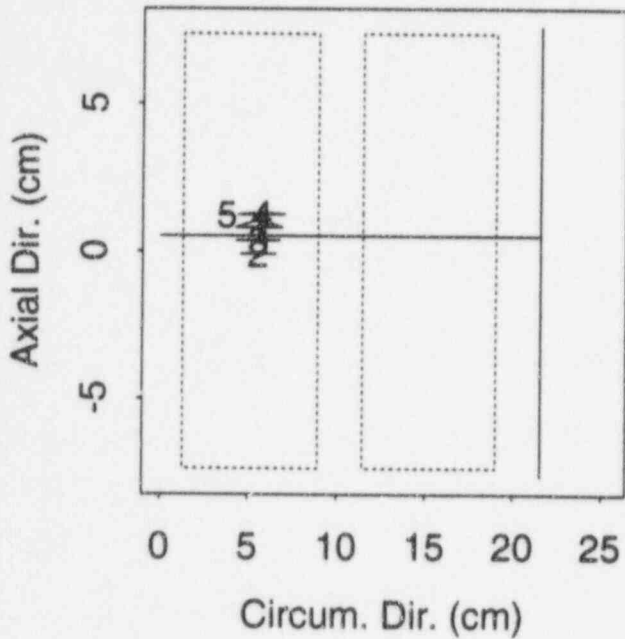
True State of
B215-4 Wrought



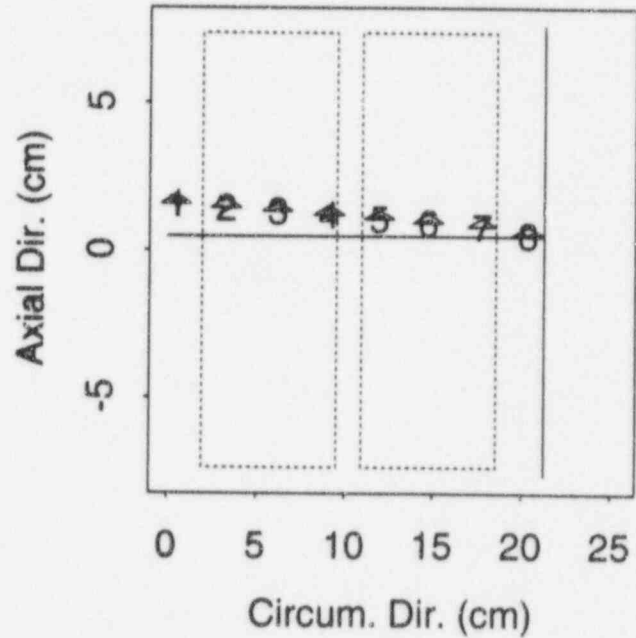
True State of
B216-1 Wrought



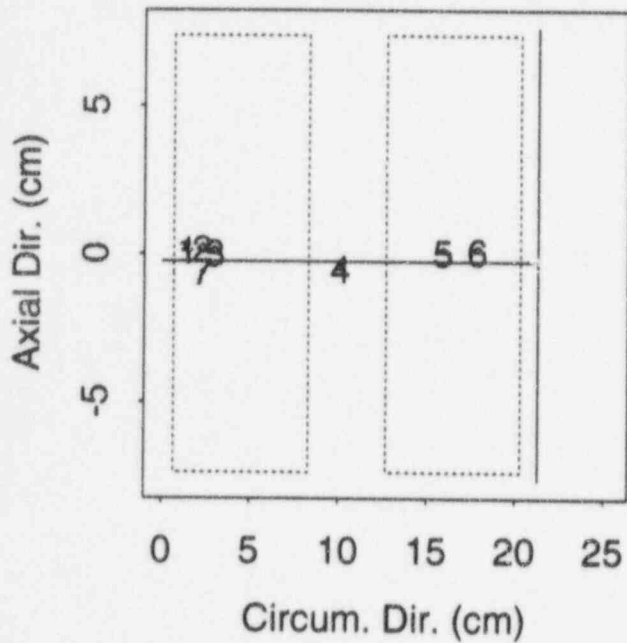
True State of
B216-2 Wrought



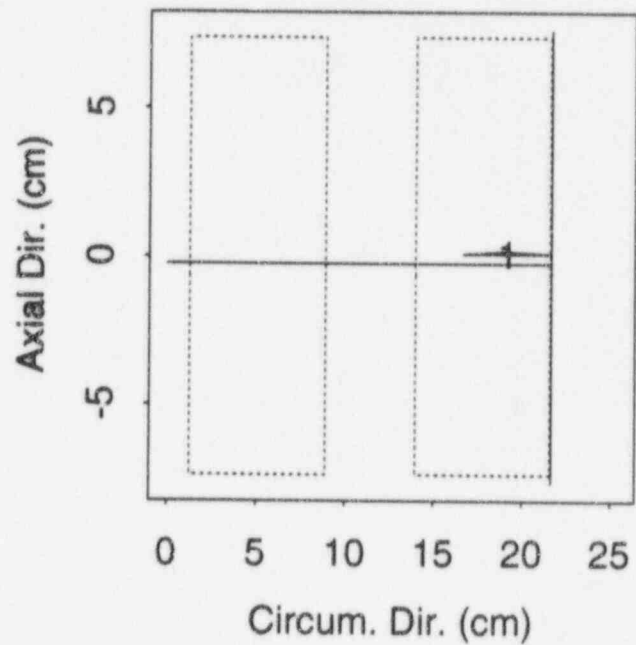
True State of
B216-4 Wrought



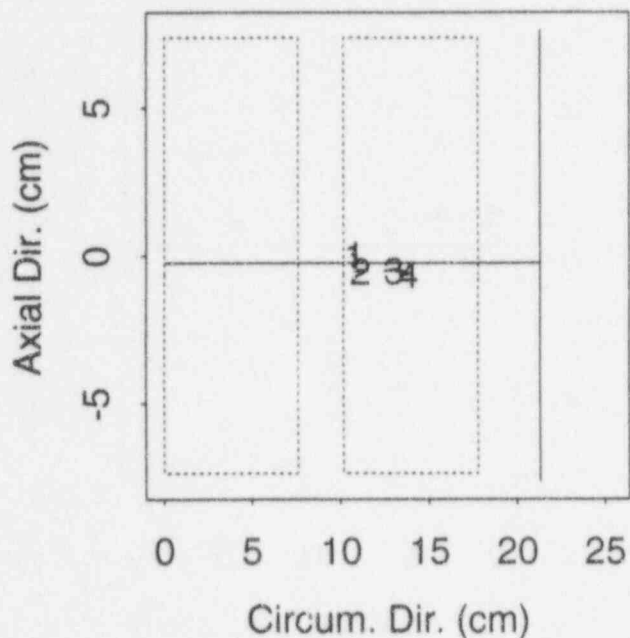
True State of
B217-2 Wrought



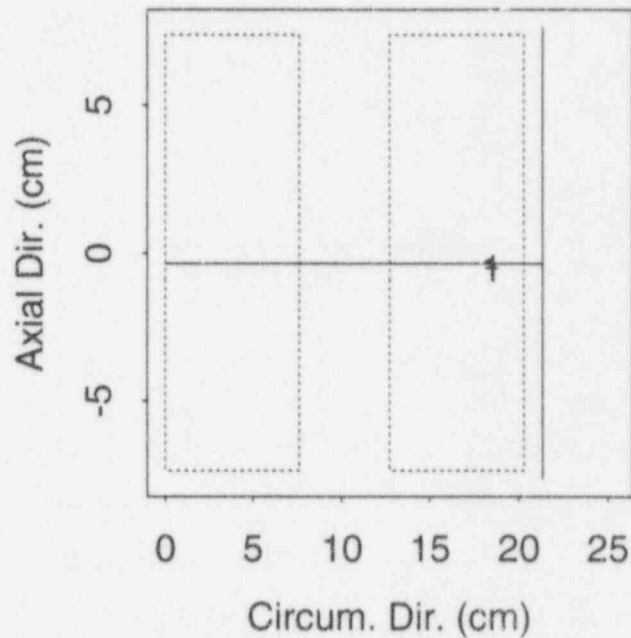
True State of
B217-3 Wrought



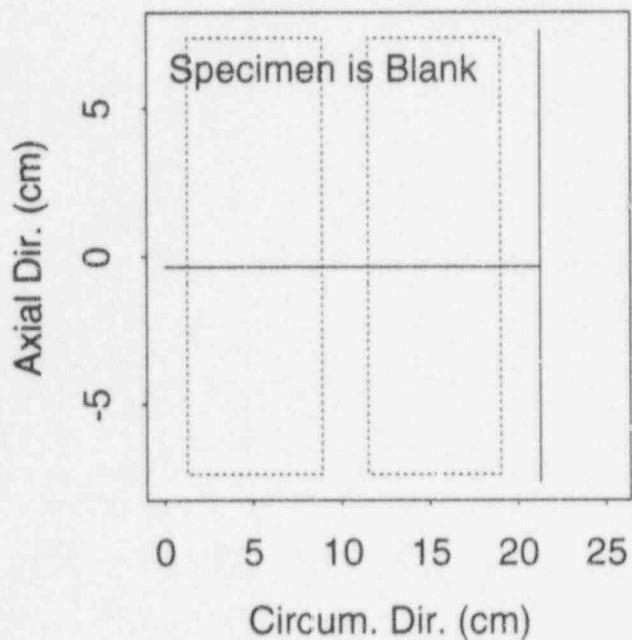
True State of B217-4 Wrought



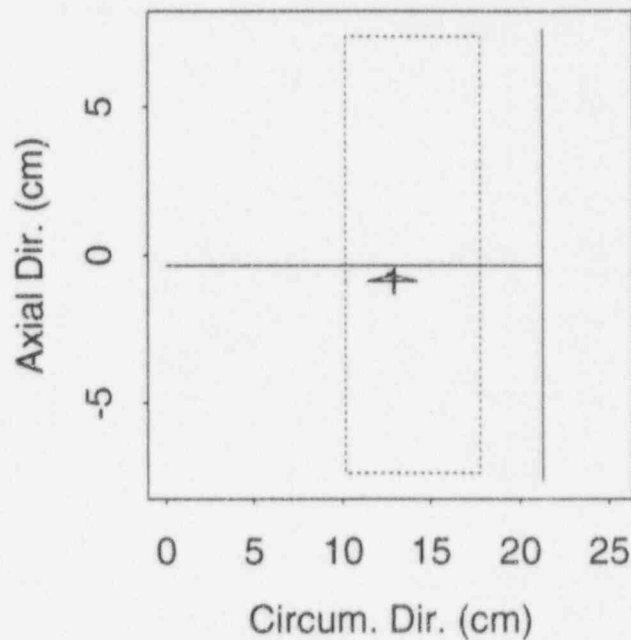
True State of D101-1 Wrought



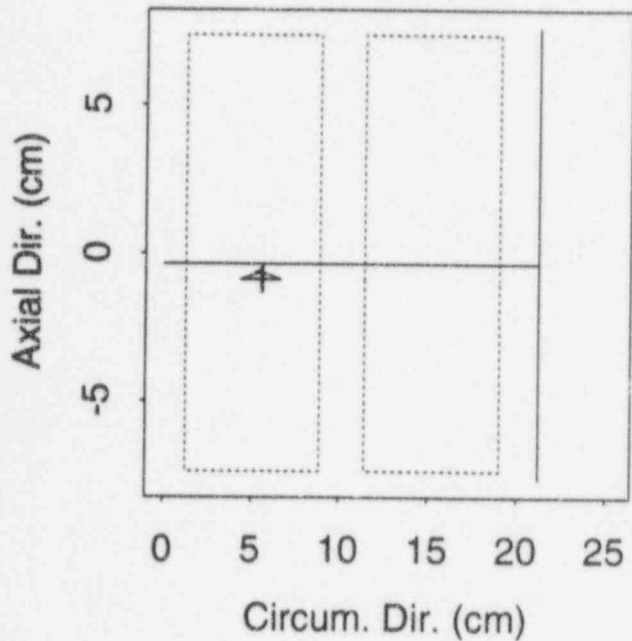
True State of D101-2 Wrought



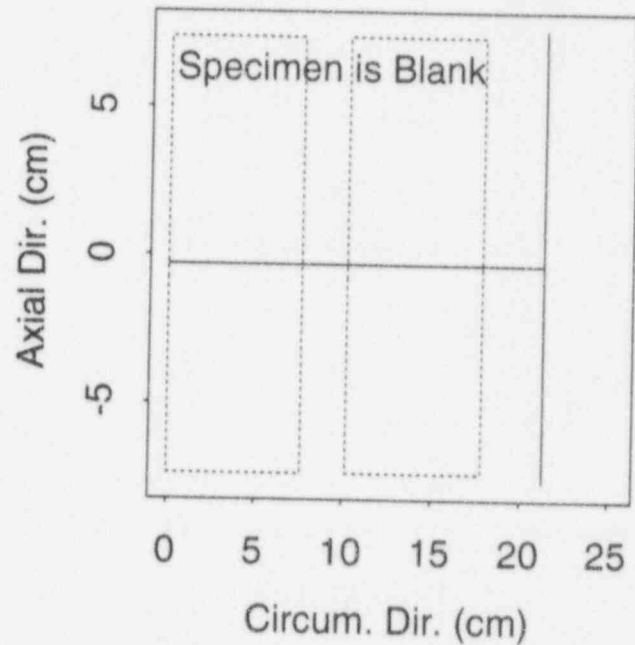
True State of D101-3 Wrought



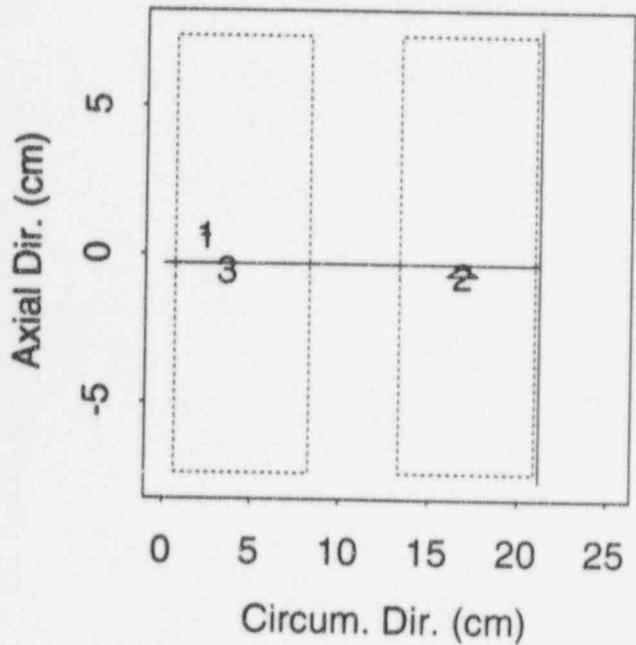
True State of
D101-4 Wrought



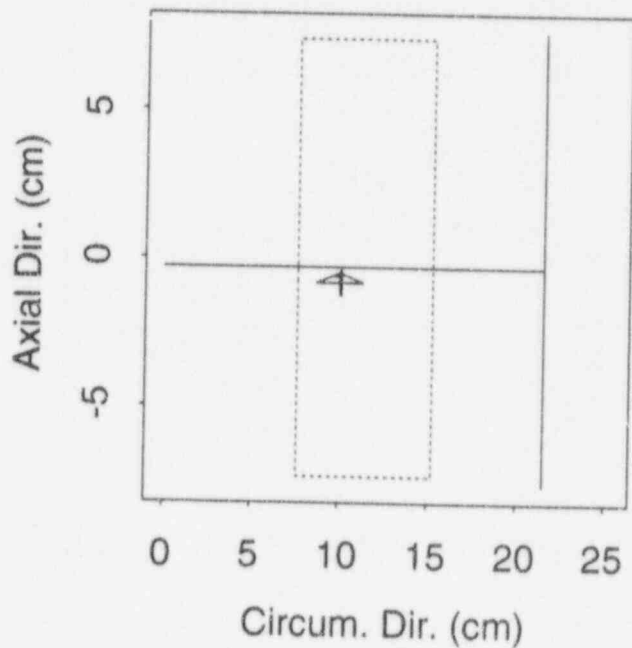
True State of
D102-1 Wrought



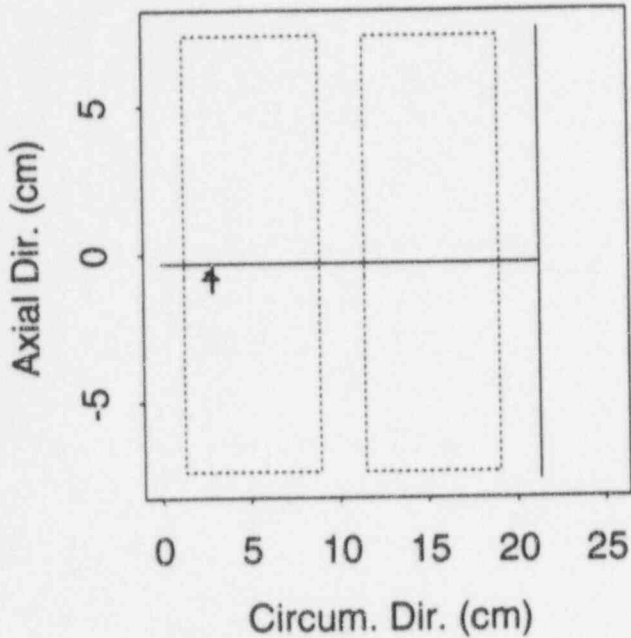
True State of
D102-2 Wrought



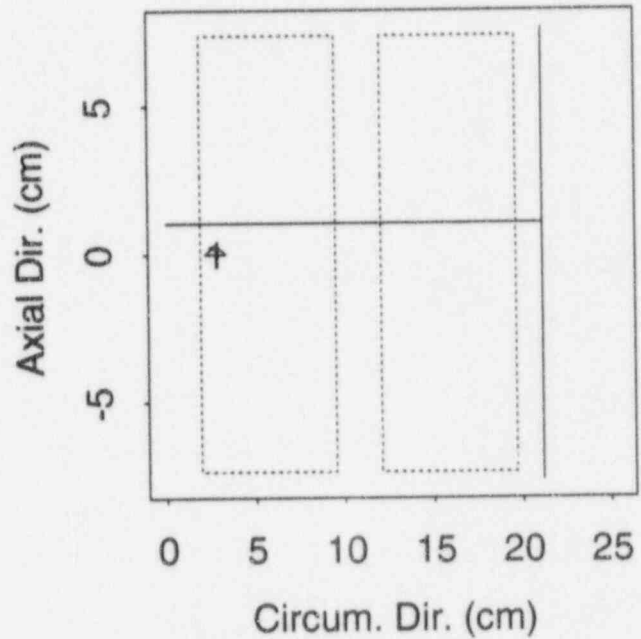
True State of
D102-3 Wrought



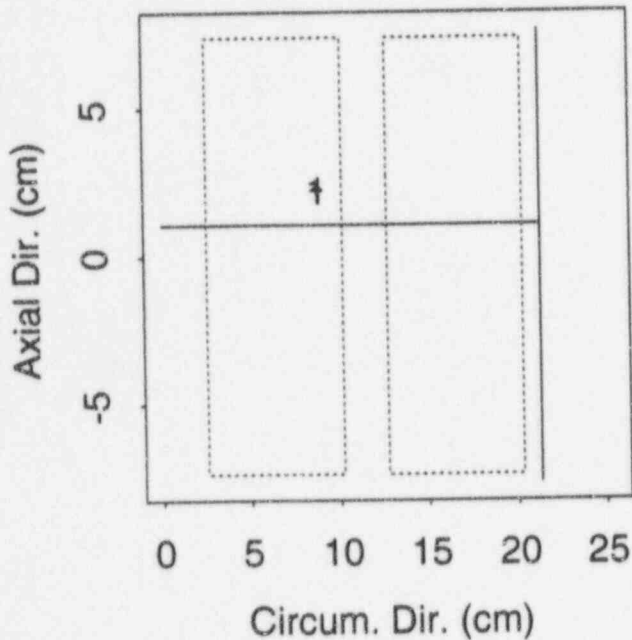
True State of D102-4 Wrought



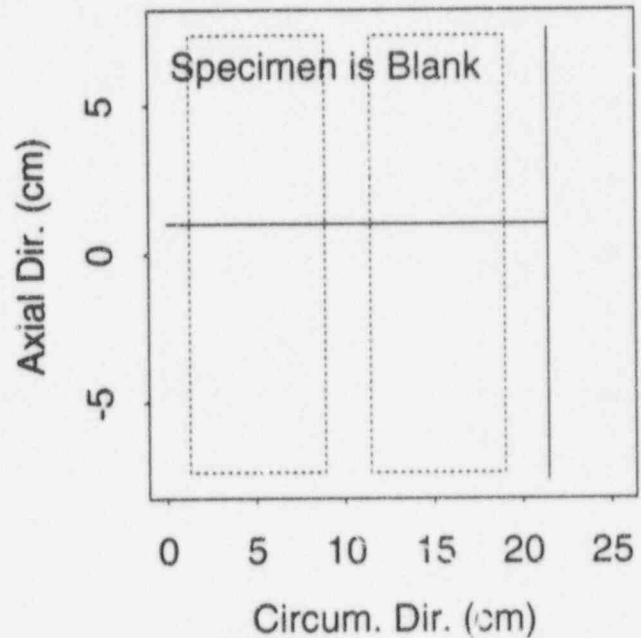
True State of D105-1 Wrought



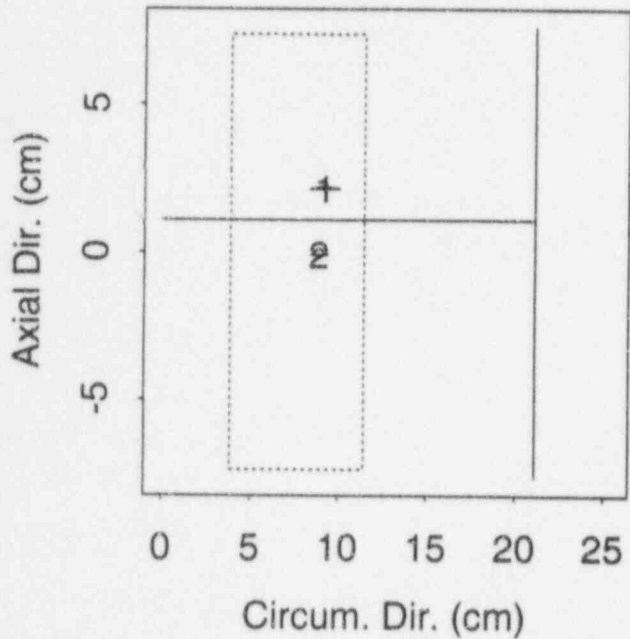
True State of D105-4 Wrought



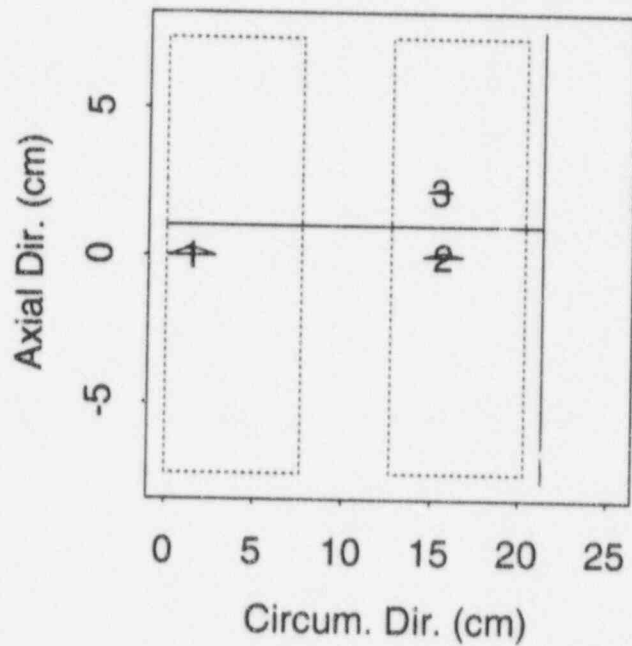
True State of D107-1 Wrought



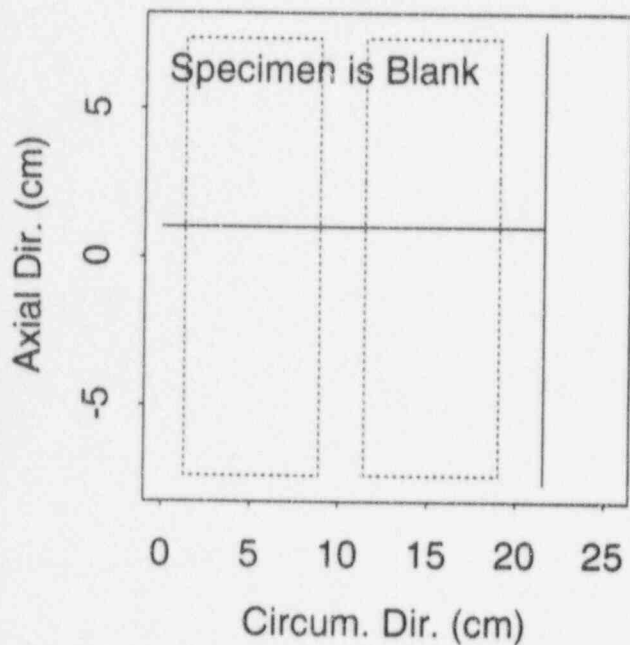
True State of D107-2 Wrought



True State of D107-3 Wrought



True State of D107-4 Wrought



DISTRIBUTION

No. of
Copies

OFFSITE

J. Muscara (2)
NRC/RES
Mail Stop 10 E10

M. E. Mayfield
NRC/RES
Mail Stop 10 E10

C. A. Hrabal
NRC/RES
Mail Stop 10 C9

F. L. Becker
J. A. Jones Applied Research Co.
P.O. Box 217097
1300 Harris Blvd.
Charlotte, NC 28221

G. Selby
J. A. Jones Applied Research Co.
P.O. Box 217097
1300 Harris Blvd.
Charlotte, NC 28221

S. Walker
J. A. Jones Applied Research Co.
P.O. Box 217097
1300 Harris Blvd.
Charlotte, NC 28221

D. Hotthaus
J. A. Jones Applied Research Co.
P.O. Box 217097
1300 Harris Blvd.
Charlotte, NC 28221

T. Taylor
EPRI
1300 Harris Blvd.
Charlotte, NC 28221

T. Lambert
Lambert McGill Thomas, Inc.
515 Aldo Ave.
Santa Clara, CA 95050

A. E. Holt
Southwest Research Institute
6220 Culebra Road
San Antonio, TX 78228-0510

M. G. Hacker
B&W Nuclear Service Co.
P.O. Box 10935
Lynchburg, VA 24506-0935

D. Adamonis
Westinghouse Electric Corp.
P.O. Box 2728
Pittsburgh, PA 15230-2728

M. Sherwin
MQS Inspection, Inc.
2301 Arthur Ave.
Elk Grove, IL 60007

Nuclear Energy Services, Inc.
44 Shelter Rock Road
Danbury, CT 06810

J. Self
General Electric
12200 Herbert Wayne Court
Suite 100
Huntersville, NC 28078

FOREIGN

No. of
Copies

S. Crutzen
IAM
EC Joint Research Centre
P.O. Box No. 2
NL 1755-ZG PETTEN
NETHERLANDS

P. Lemaitre
IAM
EC Joint Research Centre
P.O. Box No. 2
NL 1755-ZG PETTEN
NETHERLANDS

L. Skanberg
SKI
Box 27106
S-10252 STOCKHOLM
SWEDEN

O. Forli
Det Norske Veritas
Veritasveien 1
N-1322 HOVIK
NORWAY

PACIFIC NORTHWEST NATIONAL LABORATORY
(50)

S. H. Bush
S. R. Doctor (31)
P. G. Heasler
F. A. Simonen
Technical Report Files (5)
Publishing Coordination

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-5068
PNNL-10475

2. TITLE AND SUBTITLE

Piping Inspection Round Robin

3. DATE REPORT PUBLISHED

MONTH YEAR

April 1996

4. FIN OR GRANT NUMBER

B2289

5. AUTHOR(S)

P. G. Heasler
S. R. Doctor

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 National Laboratory
Richland, WA 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 Technology
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

10. SUPPLEMENTARY NOTES

J. Muscara, NRC Project Manager

11. ABSTRACT (200 words or less)

The piping inspection round robin was conducted in 1981 at the Pacific Northwest National Laboratory (PNNL) to quantify the capability of ultrasonics for inservice inspection and to address some aspects of reliability for this type of nondestructive evaluation (NDE). The research was sponsored by the U.S. Nuclear Regulatory Commission, Office of Research, under a program entitled, "Evaluation and Improvement of NDE Reliability for Inservice Inspection of Light Water Reactors."

The round robin measured the crack detection capabilities of seven field inspection teams who employed procedures that met or exceeded the 1977 edition through the 1978 addenda of the American Society of Mechanical Engineers (ASME) Section XI Code requirements. Three different types of material were employed in the study (cast stainless steel, clad ferritic, and wrought stainless steel), and two different types of flaws were implanted into the specimens (intergranular stress corrosion cracks (IGSCCs) and thermal fatigue cracks (TFCs)). When considering near-side inspection, far-side inspection, and false call rate, the overall performance was found to be best in clad ferritic, less effective in wrought stainless steel and the worst in cast stainless steel. Depth sizing performance showed little correlation with the true crack depths.

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

Ultrasonic Testing, Inservice Inspection, Nondestructive Testing,
Nondestructive Examination, Nondestructive Evaluation, Round Robin
Testing of Piping

13. AVAILABILITY STATEMENT

Unlimited

14. SECURITY CLASSIFICATION

(This Page)

Unclassified

(This Report)

Unclassified

15. NUMBER OF PAGES

16. PRICE



Federal Recycling Program

UNITED STATES
NUCLEAR REGULATORY COMMISSION
WASHINGTON, DC 20555-0001

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, \$300

SPECIAL FOURTH-CLASS MAIL
POSTAGE AND FEES PAID
USNRC
PERMIT NO. G-67

120555139531 1 1AN1R5
US NRC-OADM
DIV FOIA & PUBLICATIONS SVCS
TPS-PDR-NUREG
2WFN-6E7
WASHINGTON DC 20555