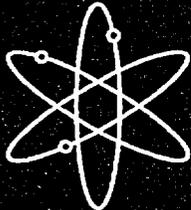




# **A Comparison of Three Round Robin Studies on ISI Reliability of Wrought Stainless Steel Piping**



**Pacific Northwest National Laboratory**



**U.S. Nuclear Regulatory Commission  
Office of Nuclear Regulatory Research  
Washington, DC 20555-0001**



## AVAILABILITY OF REFERENCE MATERIALS IN NRC PUBLICATIONS

### NRC Reference Material

As of November 1999, you may electronically access NUREG-series publications and other NRC records at NRC's Public Electronic Reading Room at <http://www.nrc.gov/reading-rm.html>. Publicly released records include, to name a few, NUREG-series publications; *Federal Register* notices; applicant, licensee, and vendor documents and correspondence; NRC correspondence and internal memoranda; bulletins and information notices; inspection and investigative reports; licensee event reports; and Commission papers and their attachments.

NRC publications in the NUREG series, NRC regulations, and *Title 10, Energy*, in the Code of *Federal Regulations* may also be purchased from one of these two sources.

1. The Superintendent of Documents  
U.S. Government Printing Office  
Mail Stop SSOP  
Washington, DC 20402-0001  
Internet: [bookstore.gpo.gov](http://bookstore.gpo.gov)  
Telephone: 202-512-1800  
Fax: 202-512-2250
2. The National Technical Information Service  
Springfield, VA 22161-0002  
[www.ntis.gov](http://www.ntis.gov)  
1-800-553-6847 or, locally, 703-605-6000

A single copy of each NRC draft report for comment is available free, to the extent of supply, upon written request as follows:

Address: Office of the Chief Information Officer,  
Reproduction and Distribution  
Services Section  
U.S. Nuclear Regulatory Commission  
Washington, DC 20555-0001

E-mail: [DISTRIBUTION@nrc.gov](mailto:DISTRIBUTION@nrc.gov)  
Facsimile: 301-415-2289

Some publications in the NUREG series that are posted at NRC's Web site address <http://www.nrc.gov/reading-rm/doc-collections/nuregs> are updated periodically and may differ from the last printed version. Although references to material found on a Web site bear the date the material was accessed, the material available on the date cited may subsequently be removed from the site.

### Non-NRC Reference Material

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. Such documents as theses, dissertations, foreign reports and translations, and non-NRC conference proceedings may be purchased from their sponsoring organization.

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at—

The NRC Technical Library  
Two White Flint North  
11545 Rockville Pike  
Rockville, MD 20852-2738

These standards are available in the library for reference use by the public. Codes and standards are usually copyrighted and may be purchased from the originating organization or, if they are American National Standards, from—

American National Standards Institute  
11 West 42<sup>nd</sup> Street  
New York, NY 10036-8002  
[www.ansi.org](http://www.ansi.org)  
212-642-4900

Legally binding regulatory requirements are stated only in laws; NRC regulations; licenses, including technical specifications; or orders, not in NUREG-series publications. The views expressed in contractor-prepared publications in this series are not necessarily those of the NRC.

The NUREG series comprises (1) technical and administrative reports and books prepared by the staff (NUREG-XXXX) or agency contractors (NUREG/CR-XXXX), (2) proceedings of conferences (NUREG/CP-XXXX), (3) reports resulting from international agreements (NUREG/IA-XXXX), (4) brochures (NUREG/BR-XXXX), and (5) compilations of legal decisions and orders of the Commission and Atomic and Safety Licensing Boards and of Directors' decisions under Section 2.206 of NRC's regulations (NUREG-0750).

**DISCLAIMER:** This report was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any employee, 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 publication, or represents that its use by such third party would not infringe privately owned rights.

NUREG/CR-6795  
PNNL-13873

---

---

# **A Comparison of Three Round Robin Studies on ISI Reliability of Wrought Stainless Steel Piping**

---

---

Manuscript Completed: September 2002  
Date Published: February 2003

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

Pacific Northwest National Laboratory  
Richland, WA 99352

D. A. Jackson, 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 Y6604



## ABSTRACT

The Pacific Northwest National Laboratory (PNNL) is conducting a multi-year program sponsored by the Nuclear Regulatory Commission (NRC) to address issues related to the reliability of ultrasonic testing (UT) and the development of improved programs for inservice inspection (ISI). This includes establishing the accuracy and reliability of UT methods for ISI of light water reactor components. From 1981 through 1990, three major round robin studies were conducted to quantify the capability of ISI inspectors to detect cracks in wrought stainless steel, and their accuracy in crack sizing. This report concentrates on analysis techniques to estimate comparable ISI detection and sizing statistics from the three round robins. This analysis provides a tool for evaluating the effect of technological advances in UT capability during the 1980s, and to assess combining the data from the three studies to form a better overview of inspection performance.

# CONTENTS

|  |      |
|--|------|
| Abstract .....   | iii  |
| Executive Summary .....                                  | x    |
| Acknowledgments .....                                    | xii  |
| Acronyms .....   | xiii |
| 1 Introduction .....                                     | 1.1  |
| 2 Overview of the Round Robin Exercises .....            | 2.1  |
| 2.1 Piping Inspection Round Robin .....                  | 2.1  |
| 2.2 MRR Background .....                                 | 2.2  |
| 2.3 PISC-AST Background .....                            | 2.2  |
| 2.4 Comparison of Data from the Round Robins .....       | 2.3  |
| 3 Statistical Evaluation of Round Robin Data .....       | 3.1  |
| 3.1 Statistical Analysis of Round Robin Data .....       | 3.2  |
| 3.2 Flaw Sizing Regression Model .....                   | 3.2  |
| 3.3 POD Regression Model .....                           | 3.5  |
| 4 Depth Sizing Evaluations .....                         | 4.1  |
| 4.1 PIRR Depth Sizing Results .....                      | 4.3  |
| 4.2 MRR Depth Sizing Results .....                       | 4.5  |
| 4.3 PISC-AST Depth Sizing Results .....                  | 4.9  |
| 5 Length Sizing Capability .....                         | 5.1  |
| 5.1 PIRR Length Sizing Results .....                     | 5.2  |
| 5.2 MRR Length Sizing Results .....                      | 5.3  |
| 5.3 PISC-AST Length Sizing Results .....                 | 5.7  |
| 6 Detection Capability .....                             | 6.1  |
| 6.1 PIRR POD Results .....                               | 6.3  |
| 6.2 MRR POD Results .....                                | 6.3  |
| 6.3 PISC-AST POD Results .....                           | 6.5  |
| 7 Combined Results .....                                 | 7.1  |
| 8 Comparison of ANOVA Results to Published Results ..... | 8.1  |
| 8.1 PIRR Evaluations .....                               | 8.1  |
| 8.2 PISC-ASt Evaluations .....                           | 8.1  |
| 9 Conclusions .....                                      | 9.1  |

|   |      |
|---|------|
| 10 Bibliography .....                       | 10.1 |
| A Round Robin Data Used in This Study ..... | A.1  |
| A.1 PIRR Data .....                         | A.1  |
| A.2 MRR Data .....                          | A.4  |
| A.2.1 MRR POD and Length Data .....         | A.4  |
| A.2.2 MRR Depth Data .....                  | A.6  |
| A.3 PISC-AST Data .....                     | A.6  |

## FIGURES

|      |   |      |
|------|---|------|
| 4.1  | Comparison of Depth Sizing in Round Robin Showing Regressions Fits and 95% Confidence Bounds .....              | 4.2  |
| 4.2  | Depth Sizing Regression for PIRR Round Robin Including Teams, the Average, and 95% Confidence Bounds .....      | 4.4  |
| 4.3  | PIRR RMSE Depth Sizing Errors .....   | 4.4  |
| 4.4  | PIRR Depth Sizing Regressions for Individual Teams .....  | 4.6  |
| 4.5  | Depth Sizing Regression for MRR Round Robin Including Teams, Average, and 95% Confidence Bounds .....           | 4.7  |
| 4.6  | MRR RMSE Sizing Errors .....  | 4.7  |
| 4.7  | MRR Depth Sizing Regressions for Individual Teams .....   | 4.8  |
| 4.8  | Depth Sizing Regression for PISC-AST Round Robin Including Teams, Average, and 95% Confidence Bounds .....      | 4.10 |
| 4.9  | Best PISC-AST Teams and Average Performance for Depth Sizing Regression for PISC-AST .....                      | 4.11 |
| 4.10 | PISC-AST RMSE Sizing Errors .....   | 4.11 |
| 4.11 | PISC-AST Depth Sizing Regressions for Individual Teams .....  | 4.12 |
| 5.1  | Comparison of Round Robins for Length Sizing Showing the Average and 95% Confidence Bounds .....                | 5.2  |
| 5.2  | Length Sizing Regression for PIRR Round Robin Including all Teams, the Average, and 95% Confidence Bounds ..... | 5.3  |
| 5.3  | PIRR RMSE Length Sizing Errors .....  | 5.4  |
| 5.4  | PIRR Length Sizing Regressions for Individual Teams .....   | 5.5  |
| 5.5  | Length Sizing Regression for MRR Round Robin Including all Teams, the Average, and 95% Confidence Bounds .....  | 5.6  |
| 5.6  | MRR RMSE Length Sizing Errors .....   | 5.6  |

|      |  |      |
|------|--|------|
| 5.7  | MRR Length Sizing Regressions for Individual Teams .....   | 5.8  |
| 5.8  | Length Sizing Regression for PISC-AST Round Robin Including all Teams, the Average,<br>and 95% Confidence Bounds ..... | 5.10 |
| 5.9  | PISC-AST RMSE Length Sizing Errors .....   | 5.11 |
| 5.10 | PISC-AST Length Sizing Regressions for Individual Teams .....  | 5.12 |
| 6.1  | Comparison of Round Robins .....   | 6.2  |
| 6.2  | POD Regression for PIRR Round Robin Showing the Average Performance and<br>95% Confidence Bounds in Dashed Lines ..... | 6.3  |
| 6.3  | PIRR POD Regressions for Individual Teams with 95% Confidence Bounds .....   | 6.4  |
| 6.4  | POD Regression for MRR Round Robin Showing the Average Performance and<br>95% Confidence Bounds in Dashed Lines .....  | 6.5  |
| 6.5  | POD Regressions for Individual MRR Teams with 95% Confidence Bounds .....  | 6.6  |
| 6.6  | POD for PISC-AST Round Robin Showing Average Team Results (dashed line) Along<br>with 95% Confidence Bounds .....      | 6.8  |
| 6.7  | PISC-AST POD Regressions for Individual Teams with 95% Bounds .....  | 6.9  |
| 7.1  | Combined MRR and PISC-AST Measurements .....   | 7.3  |
| 7.2  | Combined MRR and PISC-AST POD with 95% Bounds .....  | 7.4  |

## TABLES

|     |   |     |
|-----|---|-----|
| 2.1 | Summary of Round Robin Data .....   | 2.4 |
| 3.1 | Typical Structure of Round Robin Data .....                                     | 3.1 |
| 3.2 | Illustration of Flaw by Team Relationships That are Obscured by Averaging ..... | 3.2 |
| 3.3 | Illustration of Team/Team and Flaw/Flaw Variability .....                       | 3.4 |
| 4.1 | Summary of Depth Sizing Fits .....  | 4.1 |

|     |   |     |
|-----|---|-----|
| 7.1 | Summary of Combined Fits .....                                  | 7.2 |
| 8.1 | Comparison of Present Results with PIRR Report .....            | 8.2 |
| 8.2 | Comparison of Present Results with PISC-AST Report No. 33 ..... | 8.2 |

## EXECUTIVE SUMMARY

PNNL is conducting a multi-year program for the Nuclear Regulatory Commission (NRC) to assess the reliability of ultrasonic testing (UT) in detecting flaws in piping and pressure vessels, and to develop improved programs for inservice inspection (ISI). This involves establishing the accuracy and reliability of UT for ISI. To accomplish this objective, PNNL has been involved in a number of national and international round robin inspection exercises. This report contains an integrated analysis of these round robin studies on wrought stainless steel piping.

PNNL conducted its first round robin study in 1981-82 to study the capability of the U.S. nuclear industry to detect and size cracks in primary coolant piping materials used in nuclear power plants in the U.S (NUREG/CR-5068). This study was called the Piping Inspection Round Robin (PIRR). The PIRR was conducted prior to the development of the NRC Inspection and Enforcement Bulletin (IEB) 82-03. IEB 82-03, issued in 1982, requires that the personnel, equipment and procedures used to perform ISI on wrought stainless steels that may contain intergranular stress corrosion cracks (IGSCC), must demonstrate that they can detect this degradation mode. Thus, the PIRR represents the effectiveness of ISI before industry efforts evolved to create special training courses and performance demonstration methods.

PNNL conducted a second study in 1986 to assess the effects of training and testing programs being used to qualify inspectors for detecting and sizing IGSCC in wrought stainless steel piping (NUREG/CR-4908). This second study was called the Mini-Round Robin (MRR), to identify its limited scope of capability assessment.

PNNL also participated in the international Programme for the Inspection of Steel Components (PISC-III), austenitic steel testing (AST) study on wrought stainless steel. Only the PISC-III AST is included here since it was the only study that involved wrought stainless steel. The PISC-AST study was conducted in 1989-1990.

This report contains an overview of all three round robin studies. These three round robin studies cover nearly 10 years of capability assessment for the detection and sizing of cracks in wrought stainless steel piping. From the results of these studies, one can evaluate the effect of technological advances on UT capability. The analysis presented in this report concentrates on estimating comparable UT detection and sizing statistics from the three round robins.

This report describes the statistical models and processes that are applied to the three round robin studies. The statistical models include four terms to produce a comprehensive description of inspection error. The first term describes the effect of measurement bias while the other three represent the contributions of team/team, flaw/flaw and measurement variability. The bias or systematic error represents an off set that occurs in the regression model fit relative to ideal performance. The other terms represent the variability of the data about the regression model fit. For the sizing analysis, a linear regression model is used; and for detection, a logistic regression model is used. The selection of the models for detection and sizing is based in part on the physics of UT and these models are believed to be a reasonably accurate fit to the data.

The three round robin data bases were edited to represent the inspection of wrought stainless steel piping with good access. This means that the results represent an upper bound on performance. Depth sizing analysis of the three round robin data bases showed that a gross depth sizing error (a sizing error greater than three sigma in the regression) occurred about three in a thousand depth measurements. The probability of a gross depth sizing error seems to be about the same in all three round robin studies.

In ranking the depth sizing performance in the three studies, the MRR was best, the PISC-AST was next and the PIRR the worst. However, the PISC-AST and the PIRR required the teams to inspect a specimen and detect the crack and then determine the deepest portion of the crack. In contrast, the MRR teams conducted separate detection and depth sizing tests. In the MRR depth sizing test, 25 mm wide areas of weld were presented to the inspectors for depth sizing. Thus, the protocols were quite different for the three studies, which complicates combining the studies to compare depth sizing performance.

In length sizing, much larger gross errors can occur than in depth sizing, because the upper bound on flaw length is the size of the specimen being inspected. It was found that the probability of a gross sizing error was similar in all three studies, and that about one and one half of a percent of the measurements result in a gross length sizing error. In ranking the performance for length sizing, the PISC-AST results are the best, the MRR are next, and the PIRR results are the least accurate.

The detection performance is analyzed using the probability of detection (POD) regression parameter. Ranking detection performance (POD) in the three studies would place the PISC-AST results as the best, the MRR next, and the PIRR as least effective. The sources of error were analyzed and it was found that the flaw/flaw variability is large for both the PISC-AST and PIRR studies. Part of this can be attributed to the different flaws used in the studies (IGSCC, thermal fatigue cracks (TFC), and mechanical fatigue cracks (MFC)). However, this has important implications for multiple inspections, a technique which has been advocated to increase POD on important components. The model that has been advocated is correct as long as the inspections are independent and the flaw/flaw variability is small. These round robin studies show that the flaw/flaw variability is not small, so multiple inspections will have only limited usefulness.

In summary, the three round robins show that, during the 1980s when these studies were conducted, there was a clear and significant improvement in the capability of procedures, equipment, and personnel to more reliably detect and to more accurately size cracks in wrought stainless steel.

## ACKNOWLEDGMENTS

The authors would like to thank all of the teams that performed inspections and supporters of the programs that created the round robin data bases used in this report. The NRC and PISC had the foresight to see a need for developing these important data bases.

The authors would like to thank Feng Gao (now at the Pfizer Research Center) for conducting some of the analysis in this report, Nell Cliff for getting the report into a word processor so it could be easily manipulated and placed in NRC format, and finally, Rose Urbina, from Visual Communications who assisted in getting the report finalized for publication.

## ACRONYMS

|              |  |
|--------------|--|
| $\Gamma_T$   | Team-to-team covariance matrix                                     |
| $\sigma_E^2$ | Random Measurement variability                                     |
| $\sigma_F^2$ | Flaw-to-flaw variability   |
| ANOVA        | Analysis of Variance   |
| AST          | Austenitic Steel Tests   |
| DAC          | Distance Amplitude Correction                                      |
| EDM          | Electro-Discharge Machine  |
| EN           | Embedded Notch   |
| EPRI         | Electric Power Research Institute                                  |
| IGSCC        | Intergranular Stress Corrosion Crack                               |
| IEB          | NRC Inspection and Enforcement Bulletin                            |
| ISI          | Inservice Inspection   |
| Logit        | $\text{Logit}(z) = (1 + \exp(-z))^{-1}$                            |
| MFC          | Mechanical Fatigue Crack   |
| MRR          | Mini-Round Robin   |
| NDE          | Non-Destructive Evaluation   |
| NRC          | U.S. Nuclear Regulatory Commission                                 |
| PIRR         | Piping Inspection Round Robin                                      |
| PISC         | Programme for the Inspection of Steel Components                   |
| PND          | Probability of Non-Detection (1-POD)                               |
| POD          | Probability of Detection   |
| PVRC         | Pressure Vessel Research Committee of the Welding Research Council |

|      |   |
|------|---|
| RMSE | Root mean squared error                 |
| ROC  | Receiver Operating Characteristic Curve |
| TFC  | Thermal Fatigue Crack                   |
| UT   | Ultrasonic Testing                      |

# 1 INTRODUCTION

The Nuclear Regulatory Commission (NRC) is conducting a multi-year program (JCN W6275) at the Pacific Northwest National Laboratory (PNNL) to address a number of issues relating to the reliability of ultrasonic testing (UT) and the development of improved programs for inservice inspection (ISI). The specific objectives of the project are to:

- establish the accuracy and reliability of non-destructive evaluation (NDE) methods for ISI,
- provide technical bases and improved ISI programs for important reactor systems and components,
- evaluate the impact of ISI reliability on reactor system integrity, and
- provide recommended changes to codes and standards to improve the effectiveness and adequacy of ISI methods and programs.

Ultrasonic inspection is employed as one of the layers in the defense-in-depth approach to ensuring the integrity of pressurized components in commercial nuclear reactors. One important portion of the reactor primary pressure boundary consists of the reactor pressure vessel and associated large diameter piping. A flaw in these components could challenge the structural integrity of the reactor, and consequently it is important to determine the effectiveness of UT inspections. Over the past twenty years, several significant studies have quantified the capabilities of the ultrasonic inspection techniques being used in nuclear power plants.

The capabilities of inservice inspection have typically been measured through round robin exercises, which are tests that present a group of inspection teams with mockups containing known flaws. This report describes an analysis of data from three round robins, the PIRR (Piping

Inspection Round Robin [4]), the MRR (Mini Round Robin, [3]), and the PISC-AST<sup>a</sup> (Programme for the Inspection of Steel Components-Austenitic Piping [1]). The PIRR was conducted in 1981-1982, the MRR in 1986, and the PISC-AST in 1989-1990. These three round robins are well spaced to allow an evaluation of the effect of technological advances in NDE capability.

This analysis concentrates on estimating comparable ISI detection and sizing statistics from the three round robins. Also, consideration is given to combining the data from the studies to form a better overview of inspection performance. The analysis is limited to one type of piping material, austenitic stainless steel.

This report provides a systematic review and analysis of the three data bases, examines the data for trends, and tries to quantify improvements in NDE that have occurred over the decade during which these studies were conducted. Section 2 of this report contains a brief overview of each of the three round robin studies that are compared here. Section 3 describes the statistical analysis models and processes that are applied to the three round robin studies. Section 4 provides a detailed analysis and comparison of the depth sizing performance in each study. Section 5 contains an analysis of the length sizing capability. Section 6 analyzes the probability of detection (POD) of flaws in wrought stainless steel. Section 7 presents a best estimate of inspection capability by combining information from all three studies. Section 8 compares the results obtained in this study to other published results. Finally, Section 9 contains the conclusions that can be drawn from this integrated analysis of these data bases. References are provided for the work, and an appendix contains the data used in this evaluation.

---

<sup>a</sup>In this report, the results referred to as PISC-AST are only those from PISC III AST wrought stainless steel round robin.

## 2 OVERVIEW OF THE ROUND ROBIN STUDIES

### 2.1 Piping Inspection Round Robin Background

The Piping Inspection Round Robin was conducted by Pacific Northwest National Laboratory in 1981-82 as part of a multi-year program for the NRC. The program was concerned with quantifying the effectiveness of ISI, identifying improvements to NDE that could be achieved, and using fracture mechanics to assess the impact of NDE reliability on the structural integrity of nuclear plant components.

The round robin was conducted to determine the detection and sizing capabilities of ultrasonic inspection teams that were inspecting nuclear plant piping at that time. These teams had to employ procedures that met or exceeded the 1977 ASME Section XI Code, including the 1978 addendum requirements. Seven teams participated in the round robin. Each team was selected from a major ISI vendor, and the set of seven included most of the companies then performing inspections at nuclear power plants.

An individual team consisted of three people (Level I, II, and III inspectors). During the round robin, a team conducted approximately 250 inspections on clad ferritic, cast, and wrought stainless steel weldments. Five teams inspected all material, one team inspected everything but the cast stainless steel specimens, and the last inspected only the cast stainless steel. The inspection data shown in the Appendix is for the six teams that inspected the wrought stainless steel specimens.

A total of approximately 100 flaws were present in the round robin weldment set, and they were distributed over about 100 weldments. When the round robin was completed, a total of approximately 1500 inspections were recorded.

The inspections were organized so that the effects of several variables on inspection performance could be measured. These variables included:

**Procedure:** As practiced versus improved. The improved procedure required the team to report indications at 20% DAC which is a more sensitive examination than working at 50% DAC as required by ASME Code at that time.

**Environment:** Laboratory versus difficult field conditions. Difficult field conditions consisted of upside-down access to the weld in a confined space.

**Access:** Near versus far-side access to defects. Inspection was limited to one side of the weld, so far-side access meant that the inspector had to detect the defect by insonification through the weld.

**Material:** Cast stainless steel, clad ferritic, and wrought stainless steel.

**Crack Type:** TFC, IGSCC, and electro-discharge machine (EDM) notches.

**Flaw Size:** As measured by depth and by length. Flaws ranged from 0 (blank) to approximately 7 mm in depth and up to 90 mm in length for the wrought stainless steel.

The round robin test design called for each team to perform inspections under the different conditions defined by the above variables. The effects of the variables were assessed by calculating detection and sizing statistics under the different conditions.

For the analysis in this report, only inspections on wrought stainless steel specimens were used. These specimens were 254 mm (10 inches) in

diameter, and contained thermal fatigue and laboratory grown IGSCC flaws. Although near versus far-side access was found to be important to inspection performance, the far-side inspections were eliminated to make the data more comparable to the PISC-AST round robin. Both improved and field inspections were used, because the analysis showed there was no significant performance difference between these two conditions.

## 2.2 MRR Background

The Mini-Round Robin was conducted at PNNL in 1985 as part of the NRC sponsored program to quantify ISI capability and to examine ways to improve NDE reliability for ISI of light water reactors. The MRR [3] was conducted to evaluate the effect of new standards (IEB 83-02) and training (such as the EPRI administered sizing course) on inspection performance for IGSCC in wrought stainless piping. These requirements and training were put into effect after the PIRR was completed, so the MRR was conducted to determine the improvement in performance these changes might have caused.

Nine inspection teams participated in the testing. All inspectors except for one team using advanced technology had successfully passed the IEB 83-02 based detection test administered at the EPRI NDE Center. Eight of the inspectors had successfully passed the EPRI NDE Center administered IGSCC crack depth sizing test. The teams performed about 300 inspections on approximately 70 flaws. All flaws were in wrought stainless steel assemblies, with the flaws ranging up to 130 mm in length and 12 mm in depth. All flaws were IGSCC.

The teams participating in this exercise included two that used automated procedures. The rest of the teams used improved procedure variants based on minimum ASME requirements. The minimum ASME Section XI Code requirements had a 50% DAC indication recording level and a 100% DAC level for investigation of an

indication. The Code recognized the challenge of inspecting austenitic materials but gave little guidance in Appendix III Supplement 7 other than to suggest that angles in addition to 45 degrees might be needed.

As its name implies, the mini-round robin was not designed to measure inspection performance as finely as the PIRR. The primary objective was to validate if there has been a relatively large improvement in inspection performance (over that exhibited in PIRR).

It should also be noted that the depth-sizing measurements were conducted separately from the flaw detection round robin inspections. To evaluate the team's depth-sizing capabilities, each team was directed to size flaws at specific locations (the teams did not have to find the flaws before sizing them). This presented the teams with a sizing problem similar to the testing of inspectors at the EPRI NDE Center. Because they did not have to worry about the flaw's location, the teams faced an easier task than they would be presented with in the field.

## 2.3 PISC-AST Background

PISC, coordinated by the Commission of the European Communities Joint Research Center, was a 15-year international effort to evaluate the capabilities of inservice inspection procedures. The effort has been chronologically divided into three phases (PISC I 1975-80, PISC II 1981-85, and PISC III 1986-95), and round robin exercises have been a central feature of each phase. In PISC I, pressure vessel material provided by PVRC [10] was examined. In PISC II, a more realistic set of pressure vessel plates and flaws were examined. Finally in PISC III several round robin exercises were run which investigated ISI on the following materials: wrought and cast stainless steel (called the AST - Austenitic Steel Test), dissimilar metal weldments, pressure vessels and nozzles, and steam generator tubing.

The PISC round robins have a very broad selection of inspection teams, including teams from the U.S., Japan, and most western European countries. Because of this international participation, the PISC round robins present a view of worldwide ISI capability. Twenty-three teams participated in the PISC-AST, including six U.S. teams.

The round robin performed on wrought stainless steel (identified as the PISC-AST round robin in this report) was conducted in 1989-90 on six assemblies. Five of the assemblies were provided by the Japanese Atomic Power Engineering and Inspection Corporation (JAPEIC) to the PISC program. The material for the sixth assembly was provided by the NRC, with the welding and flaws introduced by the Joint Research Centre. Four assemblies consisted of pipe-to-pipe welds, while the other two consisted of elbow-to-pipe welds. The assemblies consisted of 320 mm OD diameter piping sections approximately 1000 mm long with wall thickness from 11 – 25 mm.

Twenty six flaws were introduced in the assemblies. The flaws included thermal fatigue cracks, IGSCC, mechanical fatigue cracks, and EDM notches. Most of the flaws were oriented circumferentially or parallel to the weld, but three were axial or perpendicular to the weld. The flaws ranged in size to 14 mm in depth and up to 100 mm in length.

Most flaws were in the heat affected zone on the inside diameter of the pipe. A few flaws were also placed on the edge of the counterbore. Since the counterbores for these welds were much wider than usual, some teams did not even scan the counterbore edges because they were farther than one wall thickness from the weld.

The PISC-AST data contains approximately 110 inspections by the 23 teams. The teams participating in this study utilized a variety of inspection techniques, including procedures based on ASME requirements, as well as special and automated procedures (SAFT, TOFD, twin crystals, creeping waves, etc). Since this round

robin was conducted approximately 10 years after the PIRR, the number of automatic techniques represented in the PISC-AST is greater than in the PIRR.

## 2.4 Comparison of Data from the Round Robins

In order to examine comparable results from the three round robins, some data was excluded. Obviously, data not from wrought stainless steel components was excluded, and the following decisions were also made:

- (1) For the PIRR and MRR studies, the far-side inspections were excluded so that inspection access in these studies was comparable to PISC-AST (i.e., only near-side inspections were used).
- (2) The axial flaws in PISC-AST were excluded from the data, but flaws in the counterbore were retained.
- (3) All teams in PISC-AST, all in the MRR, and all of those in PIRR that inspected wrought material were used.
- (4) All depth and length measurements were converted to mm, and the results were analyzed in terms of these units.
- (5) No false call information was included from any of the analyses, because it is not comparable. The PIRR and MRR analyses used a different set of false call statistics than the PISC-AST round robin.

Table 2.1 summarizes the data used from the three round robins in this study. It should be noted that the PIRR and MRR used a pipe quadrant (approximately 203 mm (8 in.) of weld) as an inspection unit, so the "no. of assemblies" listed in the table refers to the number of these quadrants. However, the PISC-AST used the entire weld as an inspection unit.

| <b>Table 2.1 Summary of Round Robin Data</b>  |             |                          |                 |
|---|-------------|--------------------------|-----------------|
|   | <b>PIRR</b> | <b>MRR for Detection</b> | <b>PISC-AST</b> |
| No. of Inspections  | 553         | 309                      | 133             |
| No. of Teams  | 7           | 15 <sup>a</sup>          | 23              |
| No. of Assemblies   | 86          | 20                       | 6               |
| Avg. Wall Thickness, mm   | 14          | 14                       | 21              |
| Flaw Depth, mm:   |             |                          |                 |
| Min   | 0.33        | 0.83                     | 0.40            |
| Median  | 2.41        | 4.78                     | 4.50            |
| Max   | 6.83        | 11.44                    | 14.10           |
| Flaw Length, mm   |             |                          |                 |
| Min   | 3.05        | 3.30                     | 0.52            |
| Median  | 26.42       | 21.59                    | 46.39           |
| Max   | 59.19       | 130.80                   | 108.20          |
| No. of EDM  |             |                          | 9               |
| No. of IGSCC  | 21          | 15                       | 12              |
| No. of MFC  |             |                          | 3 <sup>b</sup>  |
| No. of TFC  | 24          |                          | 2               |
| Total flaws   | 45          | 15                       | 26              |
| <sup>a</sup> 8 inspectors performed sizing, 15 inspectors performed detection<br><sup>b</sup> 2 mechanical fatigue cracks and 1 lack of weld root penetration |             |                          |                 |

It should also be noted that one MRR flaw extended around the entire weld circumference (1005 mm), but because the inspectors evaluated only a quadrant of any weld, the effective length of this flaw was less than the quadrant length.

Finally, in the analysis presented for these three round robins in this report, there is no correlation between teams for these three round robin studies. Team numbers/codes were assigned randomly.

### 3 STATISTICAL EVALUATION OF ROUND ROBIN DATA

In a round robin study, a fixed set of assemblies is inspected by the participating inspection teams. When the round robin is completed, each flaw has been inspected by each team. This results in inspection data which can be organized into a rectangular array, as illustrated in Table 3.1, with the rows in the array representing flaws and the columns representing teams. A cell in this array describes a particular team's results for a particular flaw.

Table 3.1 presents a portion of the results from PISC-AST for the purposes of illustration (10 teams on 15 flaws). The complete detection table would contain 24 rows and 23 columns. A "1" in this array signifies a detection, a 0 indicates a non-detection (miss), and an "NA" identifies a flaw that was not inspected (some teams did not inspect all assemblies and some teams' inspection zones did not include all flaw locations). By averaging over columns and rows, one can

calculate POD statistics for teams and flaws. This averaging allows one to make a crude comparison between teams, and between flaws. For example, the row averages shows that the POD for team KM, at 47%, is one of the highest (for the 15 flaws in the matrix), while that for team NR, 21%, is the lowest. Averaging across columns shows us that there are some flaws with a high POD (flaw 3 at 100%) and other flaws that can't be detected at all (flaw 4 at 0%).

Table 3.1 also illustrates a deviation from the ideal round robin structure that occurs frequently. Sometimes all flaws can't be inspected by all teams because of some limitation (time, scan access, etc.) and some of the cells in the matrix are therefore blank (such as flaw 12, for team EI). Blank cells cause the matrix to be unbalanced, and a severely unbalanced data set can make it much more difficult to compare teams and flaws to each other. It should be noted that the sizing

**Table 3.1 Typical Structure of Round Robin Data**

| Flaw ID      | Team IDs |    |    |    |    |    |    |    |    |    | Flaw POD % |     |
|--------------|----------|----|----|----|----|----|----|----|----|----|------------|-----|
|              | EI       | FJ | GK | HL | JN | KM | LP | MK | NJ | NR |            |     |
| 1            | 1        | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 0          | 90  |
| 2            | 0        | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 1  | 0          | 20  |
| 3            | 1        | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1          | 100 |
| 4            | 0        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0          | 0   |
| 5            | 0        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0          | 0   |
| 6            | 0        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0          | 0   |
| 7            | 0        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0          | 0   |
| 8            | 0        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0          | 0   |
| 9            | 1        | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1          | 100 |
| 10           | 1        | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1          | 100 |
| 11           | NA       | 1  | 1  | 1  | 1  | 1  | NA | 1  | 1  | 1  | 0          | 88  |
| 12           | NA       | 1  | 1  | 1  | 1  | 1  | NA | 1  | 1  | NA | 100        |     |
| 13           | 0        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0          | 0   |
| 14           | 0        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0          | 0   |
| 15           | 0        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0          | 0   |
| Team POD (%) | 31       | 40 | 40 | 40 | 40 | 47 | 31 | 40 | 47 | 21 |            |     |

1 represents a detection, 0 a non-detection.

data originating from most round robins is unbalanced, because teams size only the flaws that they find, and generally, there are always some undetected flaws.

The data from each of the three round robins can be organized into matrices as described above. Although none of the matrices are perfectly balanced, all are largely complete. Each round robin actually produces three matrices, one for detection (as illustrated in Table 3.1), one for depth sizing, and one for length sizing.

### 3.1 Statistical Analysis of Round Robin Data

The flaw-by-team data matrices discussed in the last section determine the type of statistical analysis that can be applied to the data. The simplest analyses involve calculating column (flaw) and row (team) averages. However, it is difficult to evaluate the uncertainty in these estimates without the use of a more complete statistical model.

Also, averages tend to obscure the flaw by team relationships that may be available in the data. Table 3.2 illustrates this problem in its most extreme form. In this table, POD is 50% for each flaw and each team, suggesting that there is no difference between flaws or teams. But one can see that teams 1, 2 and 3 detect exactly the same

set of flaws, while teams 4, 5, and 6 detect an entirely different set. To deal with data like this, one must build a statistical model capable of describing the important flaw by team interactions.

The values in a round robin data matrix can be described by the variable  $Y_{ij}$ , where  $i$  represents flaw  $i$  and  $j$  represents team  $j$ . This variable may represent one of three inspection results: a detection statistic (as illustrated in the previous tables), a flaw depth, or a flaw length. In each case, we expect  $Y_{ij}$  to be related to the true flaw size, and this relationship to change from team to team. To evaluate this relationship, we construct a regression model that relates flaw size to the data  $Y_{ij}$ .

This regression model is very similar to models used in previous analyses (see [1], [3], [4]), except that the previous regression models described a single team. The regression model for detection data differs from the sizing models, because detection data is binary. However, all three models have the same general form.

### 3.2 Flaw Sizing Regression Model

An inspection flaw depth,  $Y_i$ , is assumed to be linearly related to the true flaw depth  $X_i$  through the formula:

**Table 3.2 Illustration of Flaw by Team Relationships That are Obscured by Averaging**

| Flaw ID      | Team IDs              |    |    |    |    |    | Flaw POD (%) |
|--------------|-----------------------|----|----|----|----|----|--------------|
|              | 1                     | 2  | 3  | 4  | 5  | 6  |              |
|              | Flaw/Flaw Variability |    |    |    |    |    |              |
| 1            | 1                     | 1  | 1  | 0  | 0  | 0  | 50           |
| 2            | 1                     | 1  | 1  | 0  | 0  | 0  | 50           |
| 3            | 0                     | 0  | 0  | 1  | 1  | 1  | 50           |
| 4            | 0                     | 0  | 0  | 1  | 1  | 1  | 50           |
| 5            | 1                     | 1  | 1  | 0  | 0  | 0  | 50           |
| 6            | 0                     | 0  | 0  | 1  | 1  | 1  | 50           |
| Team POD (%) | 50                    | 50 | 50 | 50 | 50 | 50 |              |

1 represents a detection, 0 a non-detection

$$\hat{Y}_{ij} = B_{1j} + B_{2j}X_i + F_i + E_{ij} \quad (3.1)$$

The coefficients  $B_{1j}$ ,  $B_{2j}$  describe the linear relationship between measured and true flaw size for team  $j$ . This linear relationship is not perfect, and the actual measurement  $Y_{ij}$  may deviate from the linear equation by random amounts. The random deviations present are represented by terms  $F_i$  and  $E_{ij}$ .

The term  $F_i$  describes the variation a particular flaw may exhibit from its true size. A particular flaw in the round robin may be systematically undersized or oversized;  $F_i$  represents this effect. The population of flaws is described by the flaw to flaw variability, which is defined as

$\text{Var}(F_i) = \sigma_F^2$ . The model also contains the error term  $E_{ij}$  which describes the standard inspection-to-inspection sizing errors. This variability is described by  $\text{Var}(E_{ij}) = \sigma_E^2$ , and is typically due to instrument or ultrasonic noise, so we will refer to this as instrument variability.

Since the objective of this study is to describe the population of inspection teams, the slope and intercept terms  $B_{1j}$ ,  $B_{2j}$  in the regression model are assumed to be random variables with:

$$E(B_{1j}, B_{2j}) = (\beta_1, \beta_2) \quad (3.2)$$

and

$$\text{Cov}(B_{1j}, B_{2j}) = \Gamma_T = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \quad (3.3)$$

This means that the regression lines are “distributed around” the line defined by  $Y = \beta_1 + \beta_2 X$  with a “variance” of  $\Gamma_T$ . This distribution allows one to describe the team-to-team variability in the population. If the teams selected for the round robin are a representative sample of commercial inspectors, then  $\Gamma_T$  and  $(\beta_1, \beta_2)$  describe the total population of commercial inspectors.

The regression algorithm (see [8] and [9]) that fits the model described in Equation 3.1 therefore produces the following results:

- Estimates of the average regression line  $Y = \beta_1 + \beta_2 X$  along with its uncertainty.
- Estimates of the team-to-team variability,  $\Gamma_T$ .
- Estimates of the flaw-to-flaw variability,  $\sigma_F^2$ .
- Estimates of the random inspection (instrument) measurement variability,  $\sigma_E^2$ .
- Individual team regression lines  $(B_{1j}, B_{2j})$ .

Once this model has been fit to the data, one can calculate the depth sizing error for a flaw of any size. The sizing error will be a mixture of random errors and fixed bias. The population root mean square error for a flaw of depth  $X$  is given by:

$$\text{RMSE}(x)^2 = (\beta_1 + (\beta_2 - 1)x)^2 + \begin{bmatrix} 1 \\ x \end{bmatrix}^T \Gamma_T \begin{bmatrix} 1 \\ x \end{bmatrix} + \sigma_F^2 + \sigma_E^2 \quad (3.4)$$

which is the sum of four error terms. The first error term describes the effect of measurement bias (systematic offset), and is called RMSE due to bias. The other terms represent the contributions of team/team, flaw/flaw, and instrument variability to RMSE. This root mean square error describes how far a depth measurement will deviate from truth for any arbitrary team chosen from the population. A specific inspection team may have a RMSE that is larger or smaller than this.

Equation 3.1 was developed for depth sizing analysis. An analogous model is used for length sizing;  $Y_{ij}$  represents a measured length and  $X_i$  the true length.

As one can see, the regression model has been constructed to give a detailed description of the variabilities in the inspection data. The principal variabilities are flaw to flaw variability as described by  $\sigma_F^2$ , team to team variability as described by  $\Gamma_T$ , and measurement (or random) variability, as described by  $\sigma_E^2$ . Previous regression analyses of round robin data have not attempted to determine all three of these sources of variability. Determination of these three sources of variability is important because each variability source has a different effect on inspection capability.

Table 3.3 illustrates the different consequences of team/team versus flaw/flaw variability. Large flaw/flaw variability is an indication that the inspection technique can't make consistent measurements on flaws. Particular properties of the flaw (i.e., orientation, tightness) affects the measurement and causes this variability. Large team/team variability is an indication of

inconsistent training, or an inspection procedure that is very difficult to master. Finally, large instrument variability may be an indication of equipment problems.

In the first part of Table 3.3, the inspections exhibit extremely high flaw/flaw variability and no team/team variability, as one can see from the row and column averages. When the variability is concentrated in one of the components, simple summary statistics computed from such data may be inadequate. For example, if one is interested in the mean POD exhibited by flaws of this type, the simple average of the data (which is 50%) would produce an unbiased estimate of the desired value. But if one computed the standard error associated with this estimate (which might be employed to build a confidence bound), the result is too small. The regular formula for standard error is  $\sqrt{POD(1-POD)/N}$  or  $\sqrt{.25/36}$ , but when only flaw/flaw variability is present, the correct value is  $\sqrt{.25/6}$ .

| Table 3.3 Illustration of Team/Team and Flaw/Flaw Variability |          |     |    |     |    |     |              |
|---|----------|-----|----|-----|----|-----|--------------|
| Flaw ID   | Team IDs |     |    |     |    |     | Flaw POD (%) |
|   | 1        | 2   | 3  | 4   | 5  | 6   |              |
| Flaw/Flaw Variability   |          |     |    |     |    |     |              |
| 1   | 1        | 1   | 1  | 1   | 1  | 1   | 100          |
| 2   | 0        | 0   | 0  | 0   | 0  | 0   | 0            |
| 3   | 1        | 1   | 1  | 1   | 1  | 1   | 100          |
| 4   | 1        | 1   | 1  | 1   | 1  | 1   | 100          |
| 5   | 0        | 0   | 0  | 0   | 0  | 0   | 0            |
| 6   | 0        | 0   | 0  | 0   | 0  | 0   | 0            |
| Team POD %  | 50       | 50  | 50 | 50  | 50 | 50  |              |
| Team/Team Variability   |          |     |    |     |    |     |              |
| 1   | 1        | 1   | 0  | 1   | 0  | 1   | 67           |
| 2   | 1        | 1   | 0  | 1   | 0  | 1   | 67           |
| 3   | 1        | 1   | 0  | 1   | 0  | 1   | 67           |
| 4   | 1        | 1   | 0  | 1   | 0  | 1   | 67           |
| 5   | 1        | 1   | 0  | 1   | 0  | 1   | 67           |
| 6   | 1        | 1   | 0  | 1   | 0  | 1   | 67           |
| Team Pod (%)  | 100      | 100 | 0  | 100 | 0  | 100 |              |

1 represents a detection, 0 a non-detection.

The lower part of the table represents the case when team/team variability is dominant, and the same problems occur with the analysis of this data. If one were presented with data containing only one form of variability (team/team, flaw/flaw, or instrument), it would be easy to construct simple summary statistics that would be adequate in describing the inspection data. However, when the data contains all three sources of variability, mixed together in unknown amounts, an ANOVA model as presented in Equation 3.1 is the simplest methodology of accounting for this.

### 3.3 POD Regression Model

To evaluate probability of detection, a logistic regression model (see [6] and [7]) with the following form was utilized:

$$Y_{ij} = \text{Logit} (B_{1j} + B_{2j}X_i + F_i) + E_{ij} \quad (3.5)$$

where  $Y_{ij}$  represents the detection/non-detection of flaw  $I$  by team  $j$  ( $1 = \text{detection}$ ),  $\text{Logit}(z)$  represents the logistic function defined by:

$$\text{Logit}(z) = (1 + \exp(-z))^{-1} \quad (3.6)$$

and  $X_i$  represents the depth of the flaw. Since the expected value of the variable  $Y_{ij}$  is the Probability of Detection under conditions  $ij$ , this regression produces a prediction of POD. The terms inside the logistic function are exactly equivalent to those in Equation 3.1, except that the measurement error term  $E_{ij}$  is now written outside the logistic function and represents binomial error. Aside from the error term, all other terms are interpreted as they were in the sizing model.

A logistic regression fit produces estimates for POD as a function of flaw depth:

$$\text{POD}(x) = \text{Logit} (\beta_1 + \beta_2 X) \quad (3.7)$$

as well as the team to team variance, and flaw-to-flaw variance. This allows one to place a bound on the estimate that incorporates team-to-team and flaw-to-flaw variability, as well as the binomial variability incorporated by standard logistic regression (see [6] and [7] for the statistical details).

## 4 DEPTH SIZING EVALUATIONS

This section describes the fitting of the regression model given by Equation 3.1 to the three round robin data sets. As noted in Section 2.4, the data from each round robin exercise was edited so that the resulting inspections are comparable. In general, the data has been edited so that it describes field inspection of wrought stainless steel piping with good access. This means that this data represents an upper bound on performance.

Table 4.1 presents a summary of the regression fits for depth sizing performance. The parameter estimates for the regression model are presented, as well as RMSE values for selected depths. One can see that the RMSE values are fairly large, particularly for 15 mm flaws. Since the average pipe thickness is 14 mm for the PIRR, 16.6 mm for MRR, and from 11 - 25 mm for the PISC-AST, one can see that RMSE is a large

percentage of the pipe thickness; two-sigma bounds for all three studies are typically larger than the pipe thickness. In relative terms, RMSE is more than 45% of the true flaw depth for all listed cases (the smallest relative RMSE is for the MRR on 15 mm flaws,  $6.69/15=45\%$ ). This means that the depth sizing errors were large for all flaws.

The reported RMSE values account for all sources of error or variability (i.e., teams, flaws, instrument measurement error, and bias). It is instructive to determine which of these sources of error are causing RMSE to be so large. For two round robins, the measurement bias is the dominant source of error. In the PIRR and PISC-AST round robins, inspectors substantially undersized large flaws and oversized small ones, which results in a regression slope (see  $\beta_2$  in Table 4.1) that is substantially less than one.

|   | PIRR             |                    | MRR   |       | PISC-AST |       |
|---|------------------|--------------------|-------|-------|----------|-------|
|   | Est <sup>b</sup> | StDev <sup>c</sup> | Est   | StDev | Est      | StDev |
| No. of teams  | 6                |                    | 8     |       | 23       |       |
| No. of flaws  | 36               |                    | 10    |       | 26       |       |
| No. of tests  | 267              |                    | 80    |       | 374      |       |
| Avg. Flaw Depth, mm   | 2.78             |                    | 4.87  |       | 4.93     |       |
| $\beta_1$ mm  | 2.91             | 0.36               | 1.58  | 1.47  | 2.9      | 0.53  |
| $\beta_2$   | 0.07             | 0.12               | 0.87  | 0.29  | 0.28     | 0.08  |
| $\gamma_{1,1}$ mm <sup>2</sup>  | 0.529            |                    | 0.468 |       | 4.356    |       |
| $\gamma_{1,2}$ mm   | -0.045           |                    | 0     |       | -        |       |
| $\gamma_{2,2}$  | 0.059            |                    | 0.143 |       | 0.371    |       |
| $\sigma_F^2$ mm <sup>2</sup>  | 0.13             |                    | 4.18  |       | 0.075    |       |
| $\sigma_F^2$ mm <sup>2</sup>  | 2.02             |                    | 7.78  |       | 0.45     |       |
| RMSE(1) <sup>a</sup> mm   | 2.56             |                    | 3.83  |       | 3.59     |       |
| RMSE(5) mm  | 2.59             |                    | 4.11  |       | 2.72     |       |
| RMSE(10) mm   | 6.93             |                    | 5.18  |       | 5.20     |       |
| RMSE(15) mm   | 11.63            |                    | 6.69  |       | 8.73     |       |
| <sup>a</sup> RMSE for flaw of 1 mm depth.<br><sup>b</sup> Estimate.<br><sup>c</sup> Standard Deviation. |                  |                    |       |       |          |       |

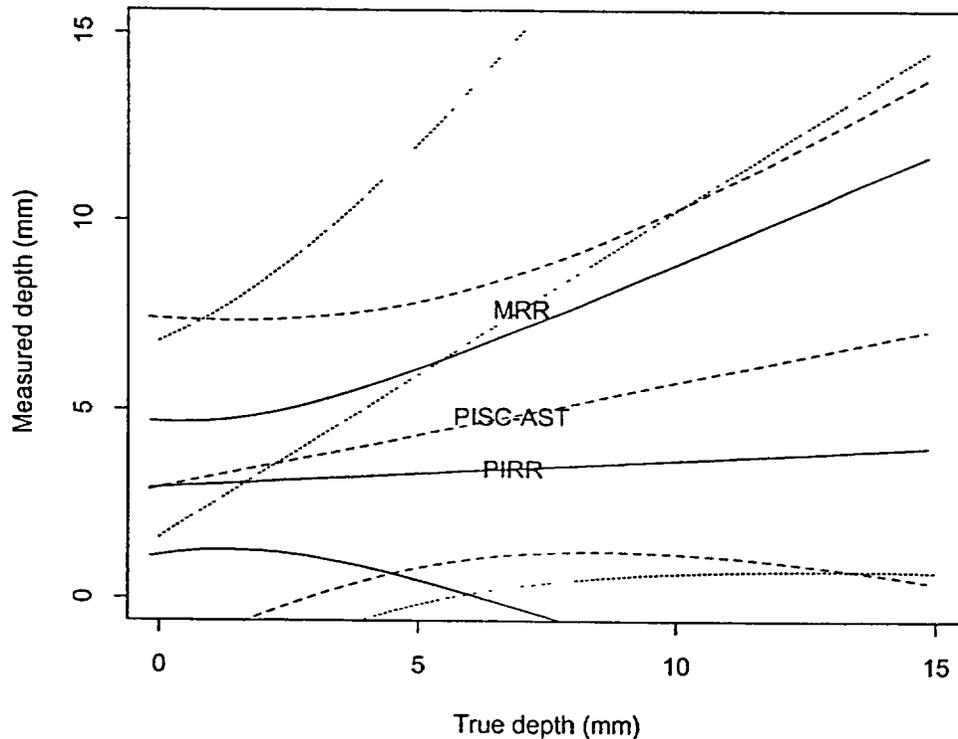
From the table, we see that the PIRR slope is  $\beta_2 = 0.07$ , a value that is not significantly different from zero.

The PISC-AST slope of  $\beta_2 = 0.28$  is significantly different from zero, indicating that the PISC-AST teams sized flaws more accurately, but it is too far from the target value of  $\beta_2$  to represent acceptable depth sizing performance.

If the inspectors in the PISC-AST study were capable of making depth measurements without any bias, the resulting RMSE would be less than 3.0 mm for depth sizing, a much more respectable result.<sup>a</sup> Of course, an effort to decrease the bias by using a regression calibration curve would result in larger measurement variability  $\sigma_E^2$ .

The MRR results reflect the attempt by the U.S. nuclear industry to improve sizing. The inspectors in the MRR all had recently taken and passed the sizing course on IGSCC administered at the EPRI NDE Center. In the MRR, they sized flaws almost exactly as they did during their training. It is clear that the MRR teams produced the best slope of the three studies,  $\beta_2 = 0.87$ . This slope produces the smallest RMSE, and if team/team variability were lower (see  $\gamma_{22} = 0.143$ ), the MRR would have exhibited an RMSE of about 4 mm.

It should be noted that the MRR results are somewhat ambiguous, because the regression results (see Figure 4.1) are strongly influenced by the one large (12 mm) flaw in the study. It is



**Figure 4.1** Comparison of Depth Sizing in Round Robin Showing Regressions Fits and 95% Confidence Bounds

<sup>a</sup>In the ASME Boiler and Pressure Vessel Code, Section XI Appendix VIII Supplement 2 for wrought austenitic piping welds, the depth sizing performance is considered acceptable if the RMSE does not exceed 3.2 mm (0.125")

possible that the MRR depth sizing bias is so much better than the other two studies because this one flaw may have been particularly easy to size.

In both the PISC-AST and the PIRR, the teams performed an inspection and were requested to depth size all flaws at their deepest points that they detected. In the MRR, the detection and sizing were separate activities: First, a detection was made, dispositioned and length sized. In a second test, 25 mm wide areas on welds were presented to inspectors for depth sizing. When performing depth sizing, the MRR inspectors did not have to worry about locating the flaw, as the inspectors in the other studies did.

In both the PISC-AST and PIRR round robins, gross depth sizing errors were identified. We define a gross sizing error as any error greater than 3 sigma in the regression. In both cases, these errors resulted from an over-sizing of a small flaw. In PISC-AST, a flaw 2.9 mm in depth was sized as 20.8 mm where the PISC-AST wall thicknesses went from 11 to 25 mm; while in the PIRR, one inspection reported a size of 13.97 mm (essentially through-wall) on a 2.4 mm flaw. The MRR inspections produced no gross sizing errors.

It is unlikely that such large errors are the result of standard measurement error. Since depth sizing requires the inspector to locate the crack tip, the most plausible explanation is that the inspector completely missed the crack tip and identified some randomly located reflector. There may be more occurrences of this effect than we have detected with our gross error test. It is quite probable that when the inspectors miss the crack tip, they tend to identify a signal that is close to the "typical" flaw size they have seen during inspection, which is about 20% of wall thickness. Such misses could account for the biases we see in the studies, and in particular for the low values of the slope parameter.

From the three round robin studies, it is possible to estimate a "gross depth sizing error probability," which would be:

|          |       |                    |
|----------|-------|--------------------|
| PIRR     | 1/267 | 0.004 <sup>a</sup> |
| MRR      | 0/80  | 0.0                |
| PISC-AST | 1/374 | 0.003              |
| Total    | 2/721 | 0.003              |

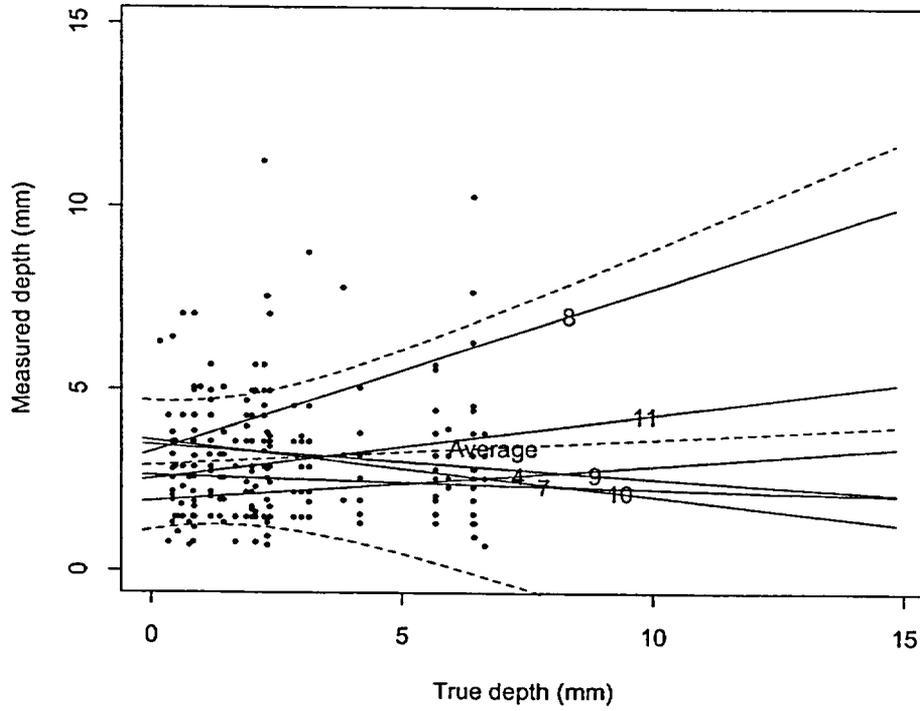
<sup>a</sup>gross errors/measurement

In other words, gross sizing errors seem to occur in about three in a thousand depth measurements. As one can see from the individual round robin statistics, the gross depth sizing error seems to have been about the same in all three studies. Furthermore, gross sizing errors only occurred in over sizing a small flaw and not in under sizing a large flaw.

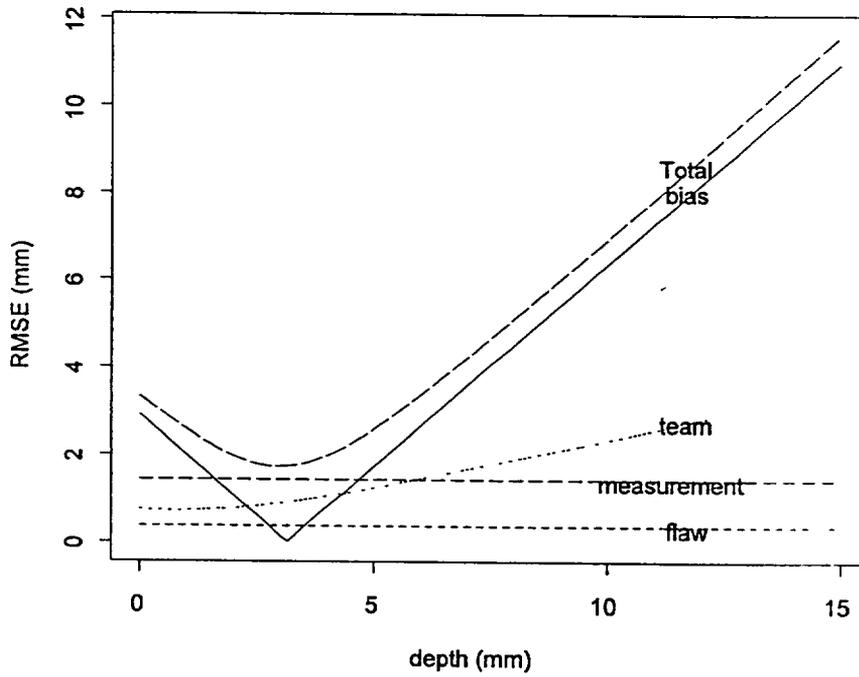
#### 4.1 PIRR Depth Sizing Results

Figure 4.2 presents the depth sizing results for the PIRR in detail. The points in this figure represent individual depth sizing measurements, while the lines describe the regression fit. The dashed line labeled "Average" depicts the size that an average team would report. The numbered lines are each associated with a participating team and each describes the measurement bias of that team. Finally, the curved lines represent 95% bounds on the population of teams; if a randomly selected inspector performs a measurement on a randomly selected flaw, it would fall within the displayed bounds 95% of the time.

From the plot, one can see that the average regression line is nearly flat, indicating a poor correlation between actual and measured depths. A line going from the lower left-hand corner to the upper right-hand corner would represent ideal performance in this figure. For true depth approaching zero, the plotted curves are all positive values, which means that they are over sizing small cracks. Likewise for large cracks approaching 15 mm, the curves are substantially less than this value, indicating that there is under sizing of large cracks. In addition, the data shows that there is extensive scatter in the depth measurements. For example, the largest flaw (with a depth of 7 mm) has measured depths



**Figure 4.2** Depth Sizing Regression for PIRR Round Robin Including Teams, the Average, and 95% Confidence Bounds



**Figure 4.3** PIRR RMSE Depth Sizing Errors

ranging from 1 mm to 10 mm. In other words, the measurements range from 7 to 71% through-wall.

Figure 4.3 plots RMSE for depth measurements, and also subdivides it as to error source by using Equation 3.4. From this plot, it is obvious that the largest contributor to RMSE is measurement bias.

It would be interesting to know if any individual team did substantially better than the average. Figure 4.4 presents regression curves and sizing measurement plots for six individual teams. The plots show that one of the six teams did better than the average (Team 8), but their performance is still far from the ideal regression line of  $0 + 1X$ . There is no evidence that this team used any improved sizing procedure to achieve these results; it just represents the best member within this population. Team 8 is the only team that shows any capability to size flaws, but it also seems to display a larger measurement variability than the other teams. Most of the Team 8 data points are outside the confidence bounds as compared to the other five teams.

## 4.2 MRR Depth Sizing Results

Figure 4.5 presents the regression results for the MRR study. The "Average" regression line from this fit is much closer to the ideal,  $0 + 1X$ , which shows that NDE training for IGSCC did have a significant effect on depth sizing capability; PIRR performance shows a lack of depth sizing capability, while MRR shows a significant relationship between measured and true depth. A line going from the lower left-hand corner to the upper right-hand corner would represent ideal performance in this figure. Four of the teams (14, 2, 10, and 8) had systematic performance like that found in the PIRR where there was over sizing of the small cracks and under sizing of the larger cracks. The other four teams (9, 1, 11, and 5) had a systematic performance that over sized every crack.

As mentioned previously, this improved capability comes at the expense of increased team/team variability, as shown in Figure 4.6. For example, the variance on the team slope is 0.14 for the MRR, about twice the value exhibited in the other two round robins. From Figure 4.5, one sees that the most extreme teams are 8 (low) and 9 (high). While there is no strong statistical (or physical) evidence for excluding these teams from the population, it is worth noting that excluding them would result in a team/team variability that is comparable to that exhibited in both the PIRR and PISC-AST.

It should be noted that the MRR results may be atypical for two reasons: First, this study was conducted soon after these teams had just successfully passed the EPRI NDE Center IGSCC training and demonstration test. Secondly, the test depth sizing protocol closely resembled the EPRI NDE Center test protocol, which gives the inspector an easier task than the typical in-field inspection. In the MRR, a specific location always containing a flaw was presented to the inspector, and he was asked to size the flaw in that specific location. This contrasts with the protocol in the other round robins (and in the field), where an inspector is presented material, required to find (detect) any flaws in the material, and then size them. If a flaw is not detected it will not be sized.

Figure 4.7 presents the sizing results for the eight individual teams, surrounded by 95% bounds on the regression curve. One can see that the measured depth of the largest flaw (12mm) has a large influence on the regression slope. The results obtained in this study are heavily influenced by this flaw. If this flaw were more difficult to size, the results may not have been so positive.

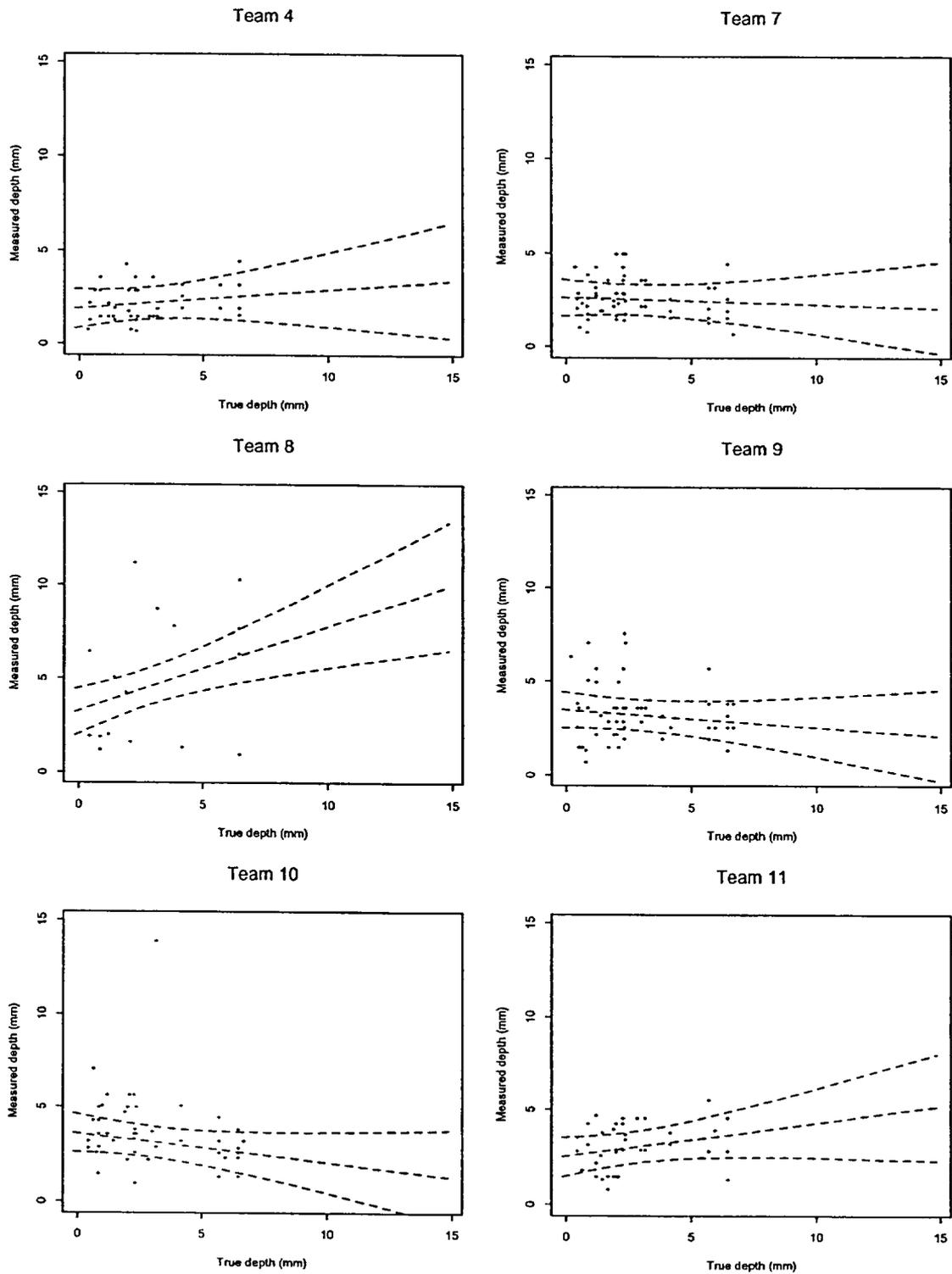


Figure 4.4 PIRR Depth Sizing Regressions for Individual Teams

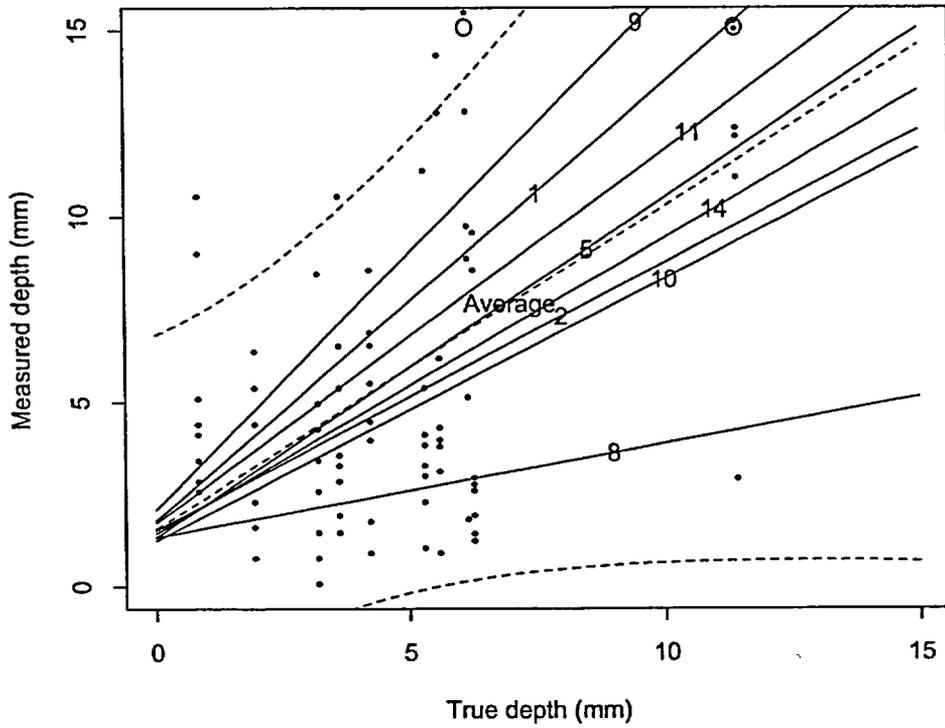


Figure 4.5 Depth Sizing Regression for MRR Round Robin Including Teams, Average, and 95% Confidence Bounds

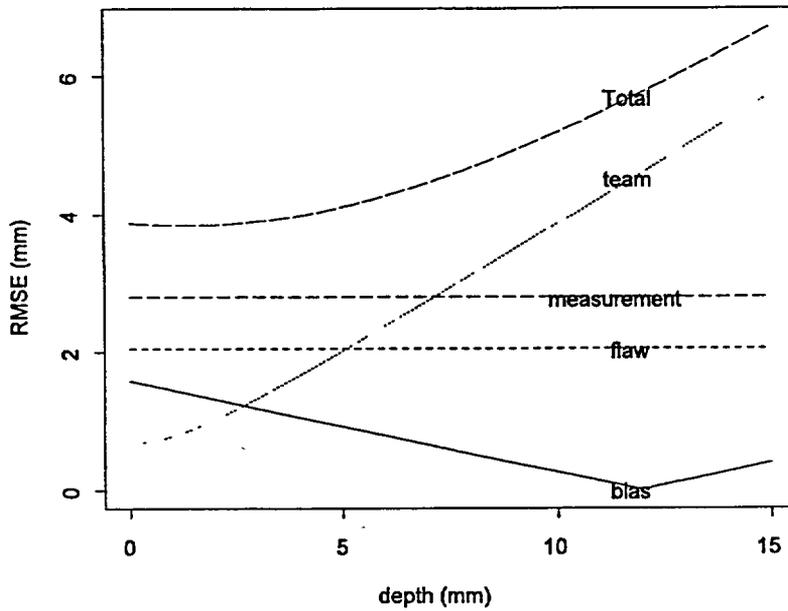


Figure 4.6 MRR RMSE Sizing Errors

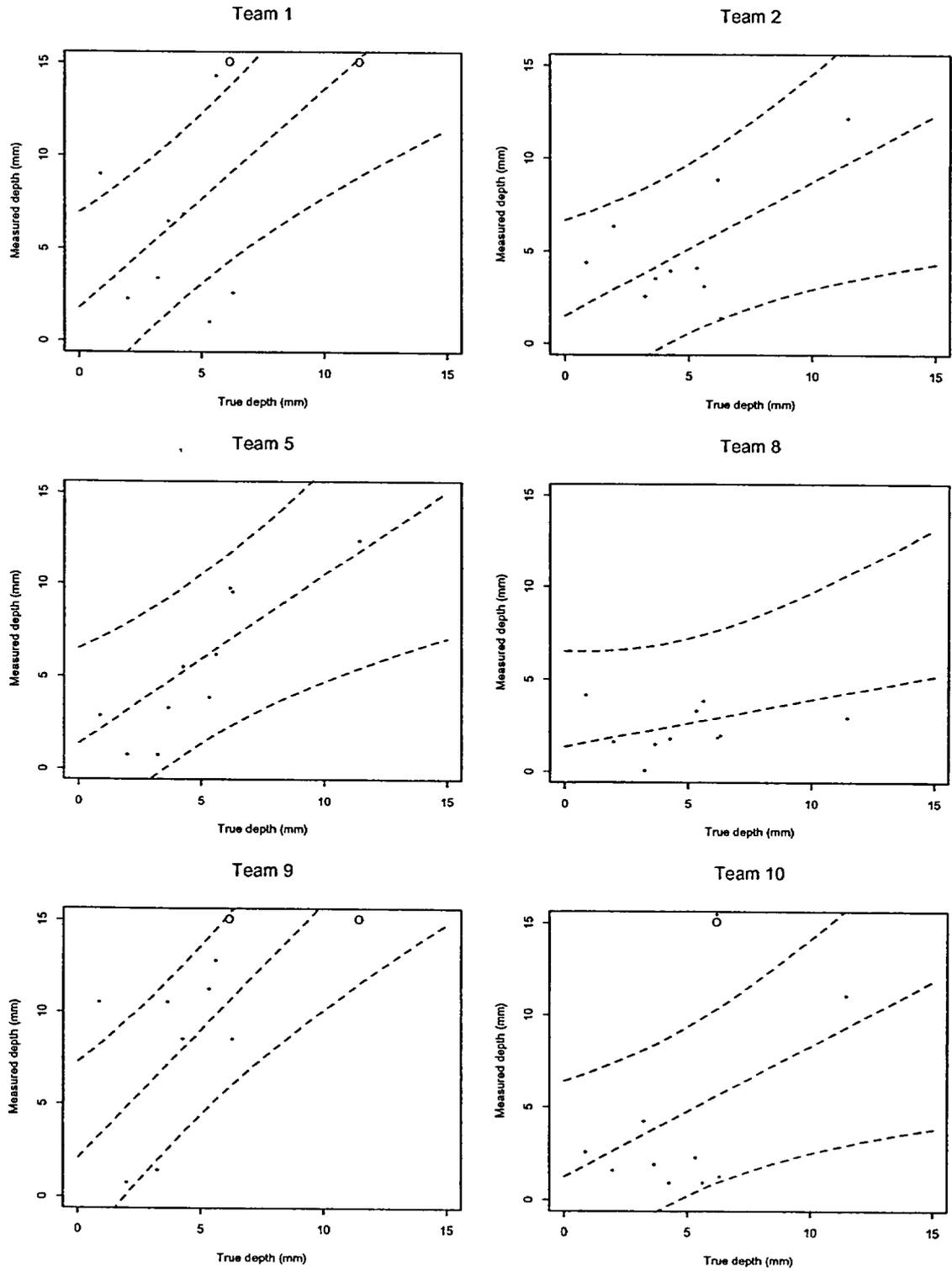


Figure 4.7 MRR Depth Sizing Regressions for Individual Teams

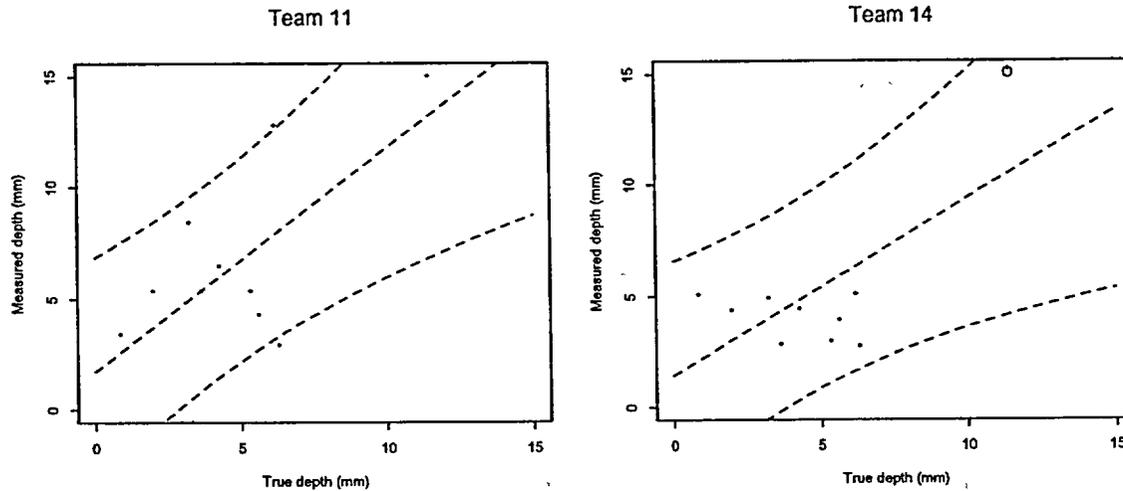


Figure 4.7 (Continued)

### 4.3 PISC-AST Depth Sizing Results

Figure 4.8 shows the regression curves for the PISC-AST teams. Most of the regression curves show a nearly flat slope of about 0.20. A line going from the lower left-hand corner to the upper right-hand corner would represent ideal performance in this figure. For true depth approaching zero, the plotted curves are all positive values, which means that they are oversizing small cracks. Likewise for large cracks approaching 15 mm, the curves are substantially less than this value, indicating that there is undersizing of large cracks. However, four of the 23 teams exhibited much higher slopes and these have been plotted in Figure 4.9 to better see them. These are teams DH, FJ, VZ, and WA, with slopes of 0.80, 0.69, 0.71, and 0.71, respectively. The regression slopes of these four best teams are sufficiently different to ask if they may represent a unique sub-population of more proficient inspectors. However, there is no strong evidence that the inspection procedures employed by these teams had any distinguishing features. Team DH employed manual scans using pulse-echo, twin crystals, and mode conversion techniques at a UT sensitivity based on the noise level. Team FJ employed twin crystal, creeping wave, and time of flight techniques. Team WA employed pulse

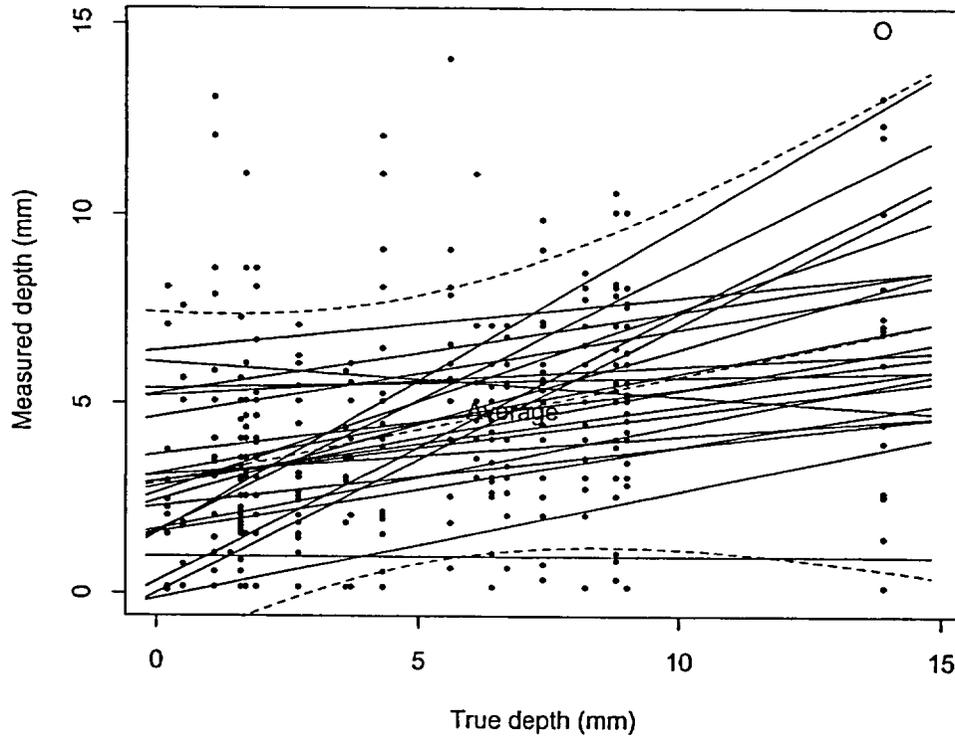
echo at 50% DAC, twin crystals, and mode conversion techniques. Finally, Team VZ employed only pulse echo at 50% DAC, and were somewhat less effective. There is nothing unique about these procedures and equipment that distinguishes these teams from others that did poorly.

Figure 4.10 presents the depth RMSE error as a function of true depth. As in the PIRR study, we see that this error is dominated by measurement bias; large flaws tend to be undersized and small ones oversized. If this measurement bias could be eliminated without inflating the other variabilities, one would achieve a RMSE of about 3 mm. Therefore, if the four best teams in PISC-AST could consistently reproduce their depth sizing performance and have a small measurement bias, they would produce a RMSE error in the 3 to 4 mm range.

Figure 4.11 contains the individual depth sizing results for the 23 teams that participated in the PISC-AST round robin. In these plots the regression fits are shown along with the 95% confidence bounds. The deepest flaw in this study is 14 mm and this flaw is about 50% larger than any other flaw in the PISC-AST study so it

has a significant influence on the overall results. However, there are another 25 flaws that aid in reducing the influence of this larger flaw. In comparing the results for PISC-AST with the individual results from the other round robins, it is seen that many of the regression slopes are

small producing a nearly horizontal line in the plots. Because of the large number of flaws in the PIRR and the PISC-AST studies, the confidence bounds are much tighter than for the MRR study.



**Figure 4.8** Depth Sizing Regression for PISC-AST Round Robin Including Teams, Average and 95% Confidence Bounds. "O" represents measurement > 15 centimeters.

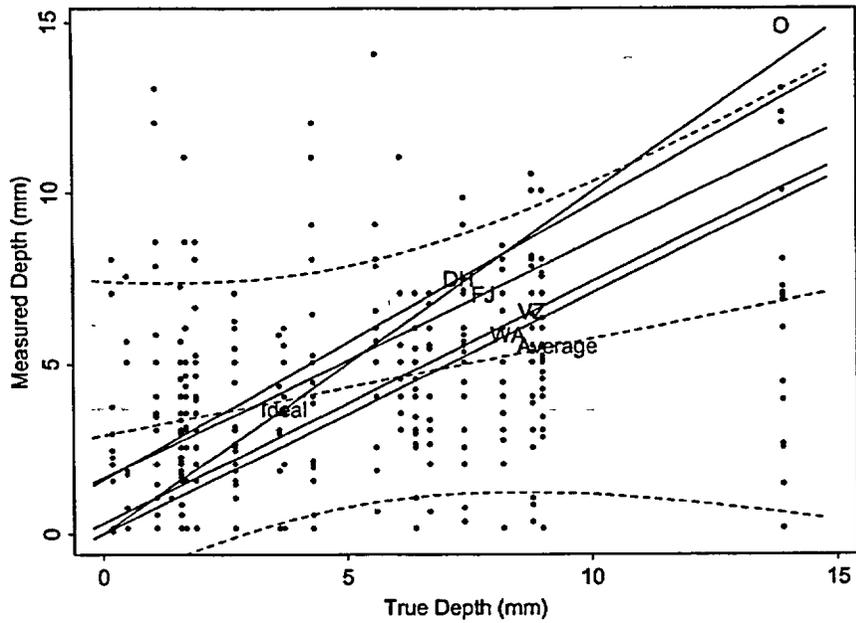


Figure 4.9 Best PISC-AST Teams and Average Performance for Depth Sizing Regression for PISC-AST. "O" represents measurement > 15 centimeters.

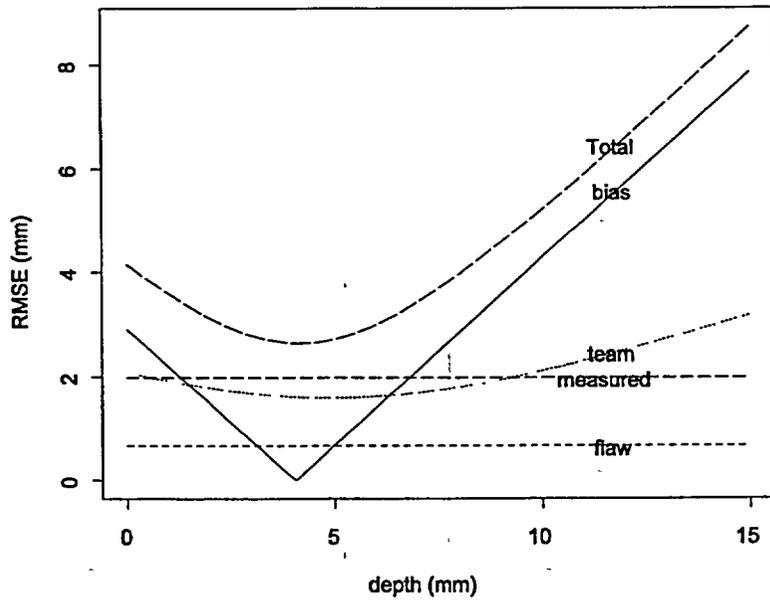


Figure 4.10 PISC-AST RMSE Sizing Errors

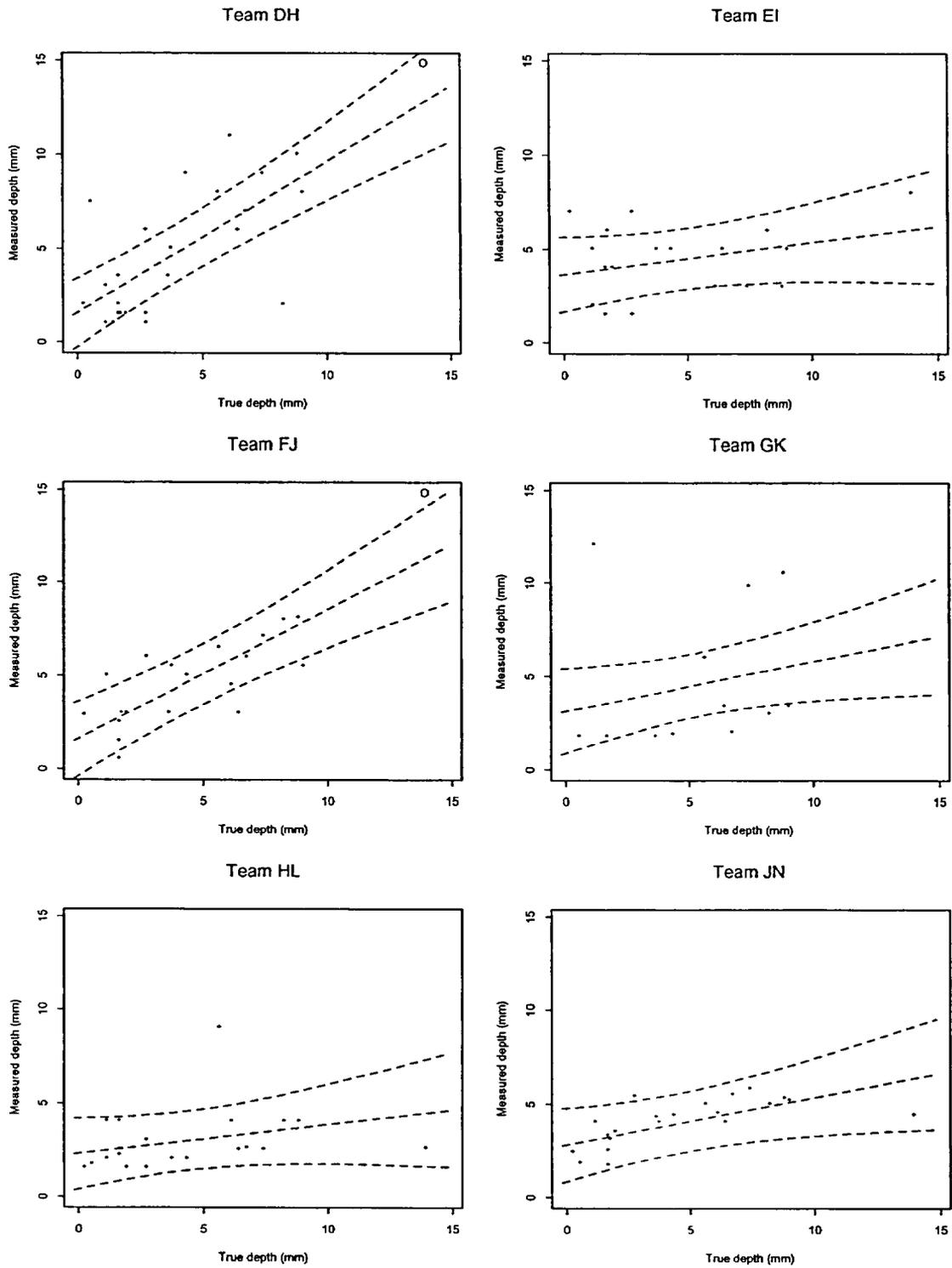


Figure 4.11 PISC-AST Depth Sizing Regressions for Individual Teams

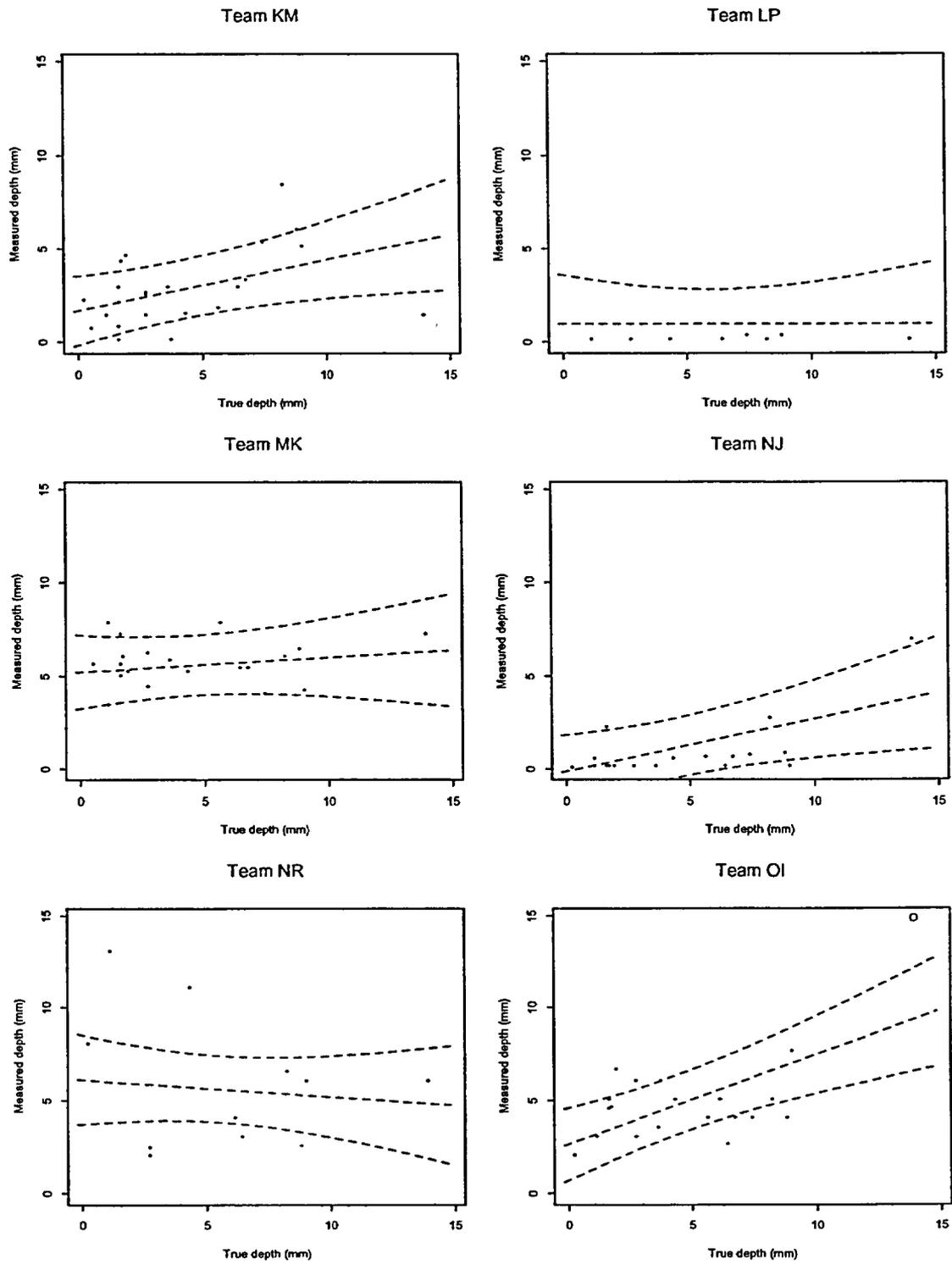


Figure 4.11 (Continued)

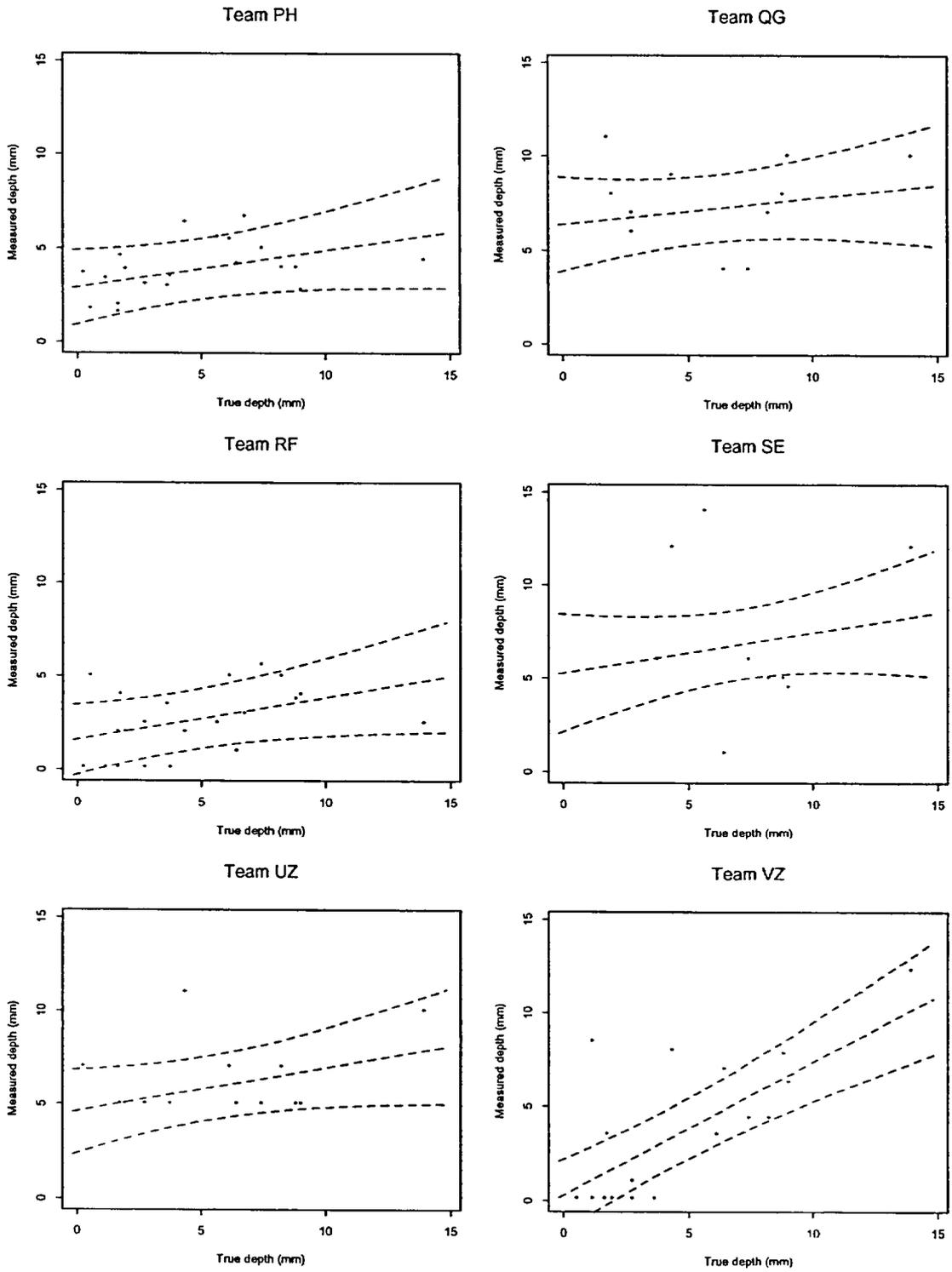


Figure 4.11 (Continued)

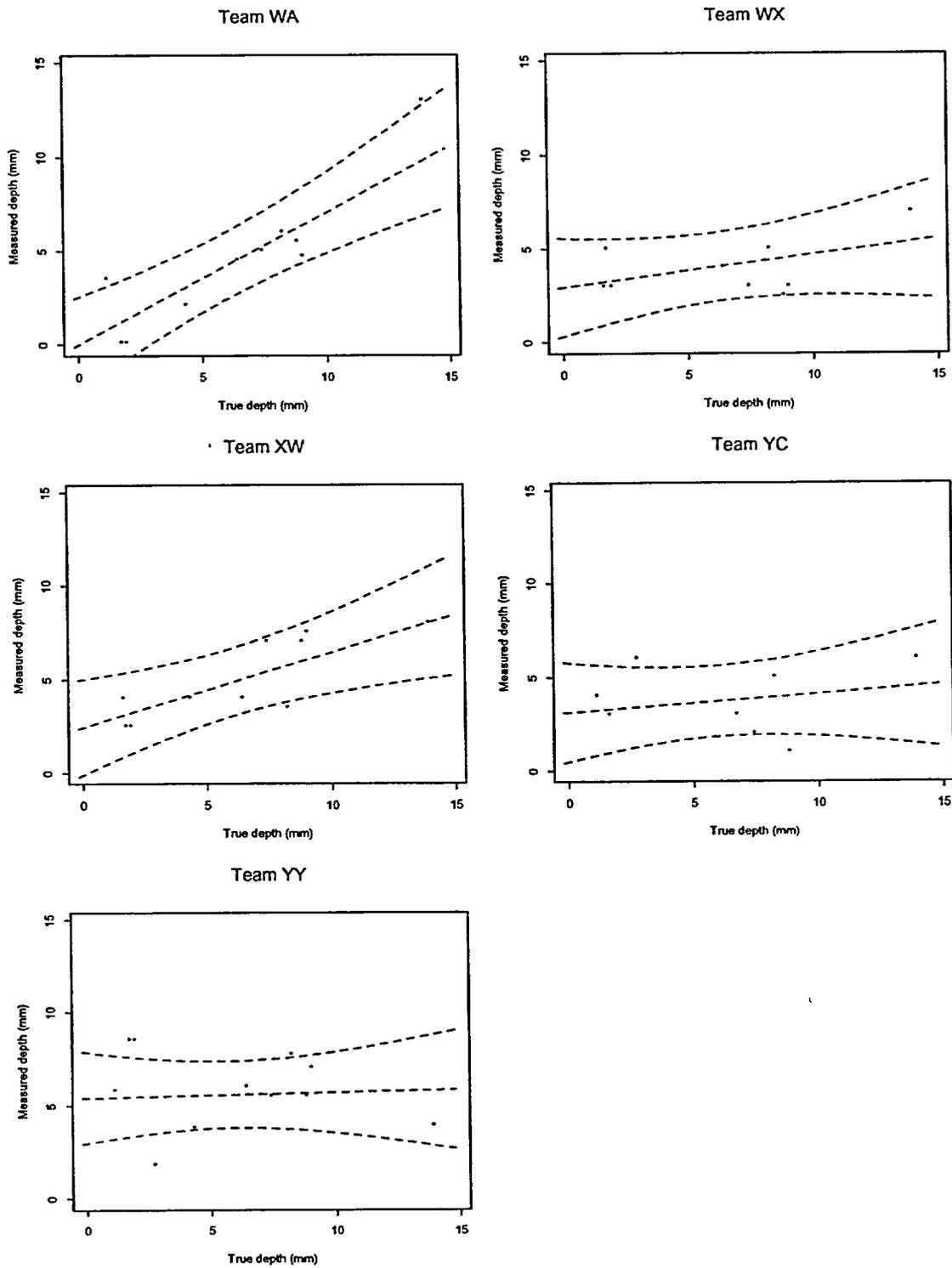


Figure 4.11 (Continued)

## 5 LENGTH SIZING CAPABILITY

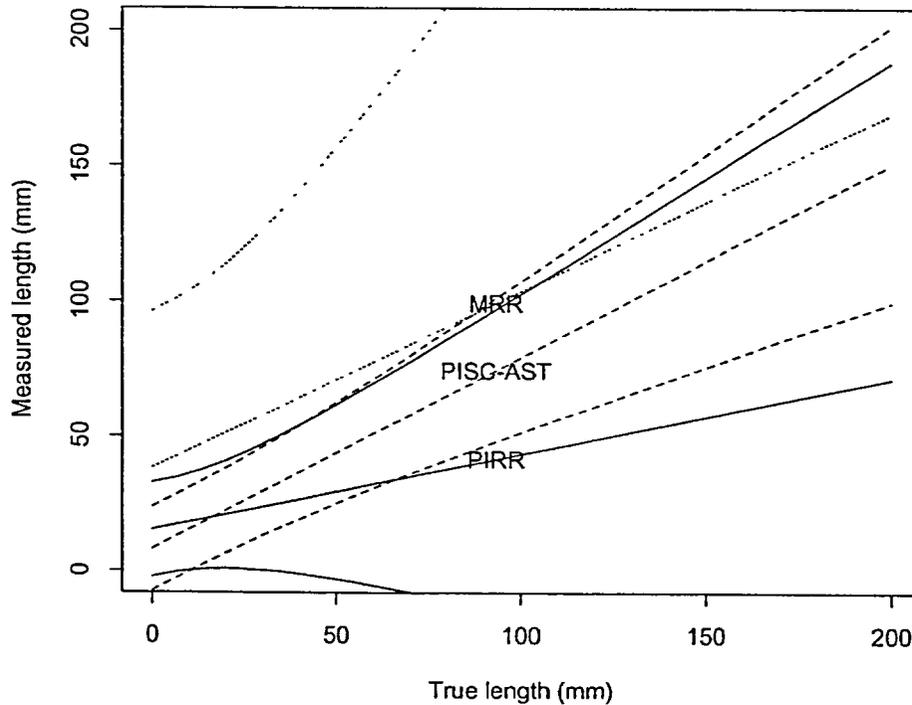
This section uses the regression model described by Equation 3.1 to evaluate length sizing for the three round robin data sets. As one would expect, all three studies showed better length sizing performance than depth sizing capability. Determining flaw lengths is an easier job than determining flaw depths, with one exception. In determining flaw lengths, one can make much larger gross errors because the upper bound on flaw length is the size of the specimen being inspected.

Table 5.1 presents a summary of the regression fits, while Figure 5.1 plots the fits. One can see that the length regression line for PISC-AST has the best slope (0.71), followed by MRR (0.65), and then by PIRR (0.28). Both the MRR and PISC-AST regression lines have a slope that is significantly different from 0, while the PIRR slope is just marginally significant.

The big difference between MRR and PISC-AST length sizing capability is with respect to the variabilities (as one can see with respect to the confidence bounds in Figure 5.1). All variabilities (measurement, flaw/flaw, and team/team) are much smaller in PISC-AST. In PISC-AST, the teams produced sizing results that were very consistent, as illustrated in Figure 5.8.

Because of the low team/team variability displayed in the PISC-AST study, the RMSE values for this study are the smallest. The PISC-AST RMSE is approximately 40% of that displayed in the MRR. The PIRR results tended to have a larger bias but were more consistent (less variability) than the MRR. The PISC-AST results tended to have a larger measurement error than the PIRR results with flaw errors that were basically the same in both studies.

| <b>Table 5.1 Summary of Length Sizing Fits</b>  |                        |                          |            |              |                 |              |
|---|------------------------|--------------------------|------------|--------------|-----------------|--------------|
|   | <b>PIRR</b>            |                          | <b>MRR</b> |              | <b>PISC-AST</b> |              |
|   | <b>Est<sup>b</sup></b> | <b>StDev<sup>c</sup></b> | <b>Est</b> | <b>StDev</b> | <b>Est</b>      | <b>StDev</b> |
| No. of teams  | 6                      |                          | 14         |              | 23              |              |
| No. of flaws  | 36                     |                          | 13         |              | 26              |              |
| No. of tests  | 267                    |                          | 123        |              | 371             |              |
| Avg. Flaw Length, mm  | 27.58                  |                          | 27.02      |              | 52.49           |              |
| $\beta_1$ mm  | 15.21                  | 3.1                      | 38.25      | 10.15        | 8.13            | 3            |
| $\beta_2$   | 0.28                   | 0.15                     | 0.65       | 0.27         | 0.71            | 0.06         |
| $\gamma_{1,1}$ mm <sup>2</sup>  | 2.05                   |                          | 491.30     |              | 6.77            |              |
| $\gamma_{1,2}$ mm   | 0                      |                          | 0          |              | 0               |              |
| $\gamma_{2,2}$  | 0.06                   |                          | 0.37       |              | 0.013           |              |
| $\sigma_F^2$ mm <sup>2</sup>  | 64.76                  |                          | 252.61     |              | 45.12           |              |
| $\sigma_E^2$ mm <sup>2</sup>  | 64.69                  |                          | 1912.58    |              | 217.77          |              |
| RMSE(5) <sup>a</sup> mm   | 16.37                  |                          | 63.24      |              | 17.74           |              |
| RMSE(30) mm   | 15.16                  |                          | 61.33      |              | 16.78           |              |
| RMSE(60) mm   | 33.82                  |                          | 65.40      |              | 20.01           |              |
| RMSE(100) mm  | 63.17                  |                          | 79.52      |              | 28.78           |              |
| <sup>a</sup> RMSE for a crack 5 mm long.<br><sup>b</sup> Estimated.<br><sup>c</sup> Standard Deviation. |                        |                          |            |              |                 |              |



**Figure 5.1** Comparison of Round Robins for Length Sizing Showing the Average and 95% Confidence Bounds

The length sizing measurements were also scanned for gross length sizing errors, with a gross error defined as a measurement differing by more than three standard deviations. All three round robins exhibited gross length sizing errors, resulting in gross error rates of:

|              |               |              |
|--------------|---------------|--------------|
| PIRR         | 4/267         | 0.015        |
| MRR          | 4/123         | 0.033        |
| PISC-AST     | 4/371         | 0.011        |
| <b>Total</b> | <b>12/761</b> | <b>0.016</b> |

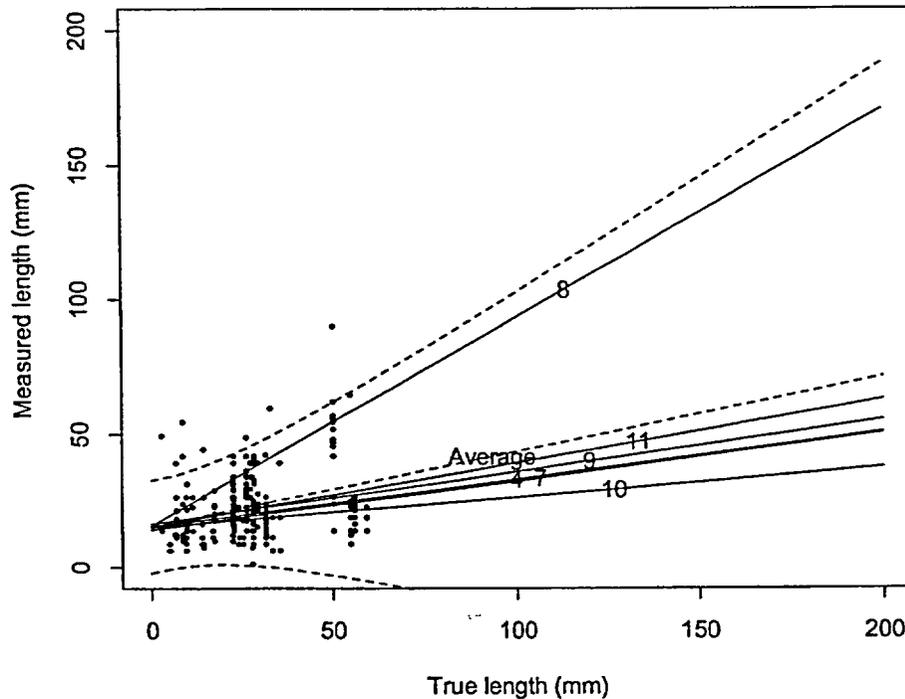
The gross sizing error seems to be fairly consistent across the PIRR and PISC-AST studies; about one and a half a percent of the measurements result in a gross sizing error. These errors can be quite large, indicating that the inspection simply missed the end of the flaw. For example, all of the gross length errors identified in the PISC-AST measurements are larger than 100 mm. These gross errors were removed before

conducting the regression analysis, so the RMSE error describes only the “normally-distributed” portion of the length measurement error. All studies indicate that 1.6 percent of the length measurements are gross oversizes.

One can expect that another half percent of the length measurements are “gross undersizes,” but unfortunately our outlier test is not sensitive enough to pick up these cases.

### 5.1 PIRR Length Sizing Results

Figure 5.2 shows the length sizing regressions for the six PIRR teams. One can see that most teams’ regressions show no significant positive slope, except one (Team 8). A line going from the lower left-hand corner to the upper right-hand corner would represent ideal performance in this figure. For true length approaching zero, the plotted curves are all positive values, which means that they are over sizing small cracks.



**Figure 5.2** Length Sizing Regression for PIRR Round Robin Including all Teams, the Average, and 95% Confidence Bounds

Likewise for large cracks approaching 200 mm, the curves are substantially less than this value indicating that there is under sizing of large cracks. Also note that the flaws used in this round robin are not as long as those used in the other round robins; the longest flaw used in the PIRR was 60 mm, while the longest flaws in the PISC-AST and MRR were twice that.

The one team (8) that did well in the PIRR correctly sized two 50 mm flaws (see Figure 5.4). Without these two correct measurements, the team would not have done so well. It appears that this team (8) sized fewer flaws than the other teams, because they did not detect a number of the flaws. It is interesting to note that this team also had the best depth sizing results. Team 8 had the lowest POD of the teams in the PIRR, but for detected flaws, they had the more accurate depth and length sizing results.

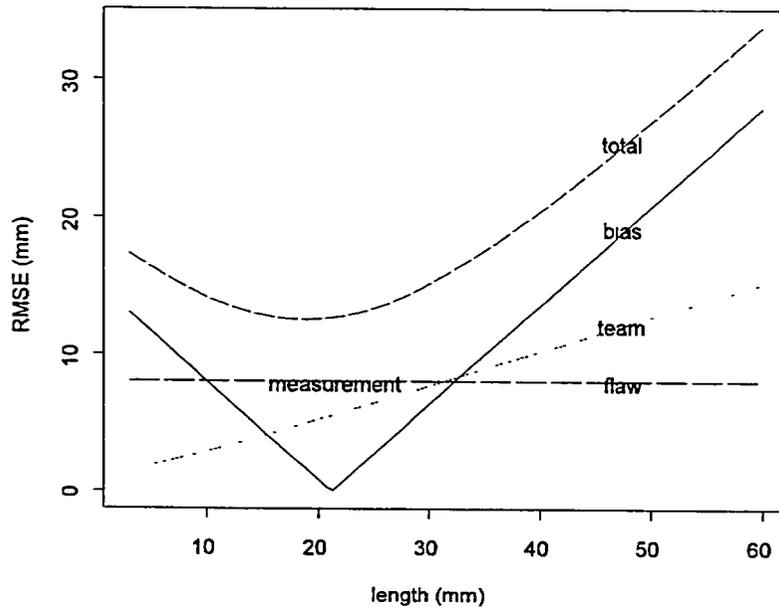
Because of bias, the length sizing RMSE is very large, as shown in Figure 5.3. If the bias could be

corrected without inflating other errors, the RMSE error would be about 15 mm. Figure 5.4 shows the PIRR length sizing regressions with 95% confidence bounds for the six individual teams.

## 5.2 MRR Length Sizing Results

Figure 5.5 shows the length sizing regressions for the MRR teams. A line going from the lower left-hand corner to the upper right-hand corner would represent ideal performance in this figure. Six of the teams (8, 9, 2, 12, 6, and 15) exhibited systematic over sizing the length of every crack. The remaining eight teams (1, 10, 11, 16, 5, 13, 3, and 14) exhibited a similar trend as found in the PIRR where they over sized the small cracks and under sized larger cracks.

As in depth sizing, MRR length sizing results show little bias, but a good deal of team/team scatter. Again, this team/team scatter may be due



**Figure 5.3** PIRR RMSE Length Sizing Errors

to the fact that the regression slopes are largely determined by the one large flaw (130 mm) present in the study. It should be noted that the team designations used in this portion of the study do not exactly match up with the designations in the depth portion of the study, because they were conducted separately. In this portion of the MRR, Level II and III team members separately performed length measurement and detection, resulting in two sets of measurements per team.

From the RMSE plot in Figure 5.6, one can see that both measurement variability and team/team variability is fairly large. In fact, measurement variability in the MRR is as large as all sources of variability in PISC-AST (see Figure 5.9). It is not readily apparent why the length sizing errors are so large for MRR versus the PISC-AST study. However, there were some significant differences in the studies that may explain these differences in errors. In the PISC-AST exercise the specimens were shipped to the teams for several weeks so that the measurements could be

performed in the vendor's facilities. In the case of the MRR, the teams traveled to PNNL and were provided the specimens in a controlled environment with time limitations. The length sizing measurements for the MRR were made by a single inspector while the results reported in PISC-AST were those obtained by a team. It is presumed that in a team activity, that all measurements are independently checked and verified by several of the team members. Thus, many errors should be caught and corrected. Consequently, the team errors are and should be significantly better than those produced by single inspectors working alone.

The typical field examination relies upon detection being made by an inspector. Once a detection has been made, other inspectors get involved to confirm the detection and to characterize the flaw. In this sense the results from the MRR are probably more realistic of what is done in the field regarding detection. The sizing results from PISC-AST are probably more

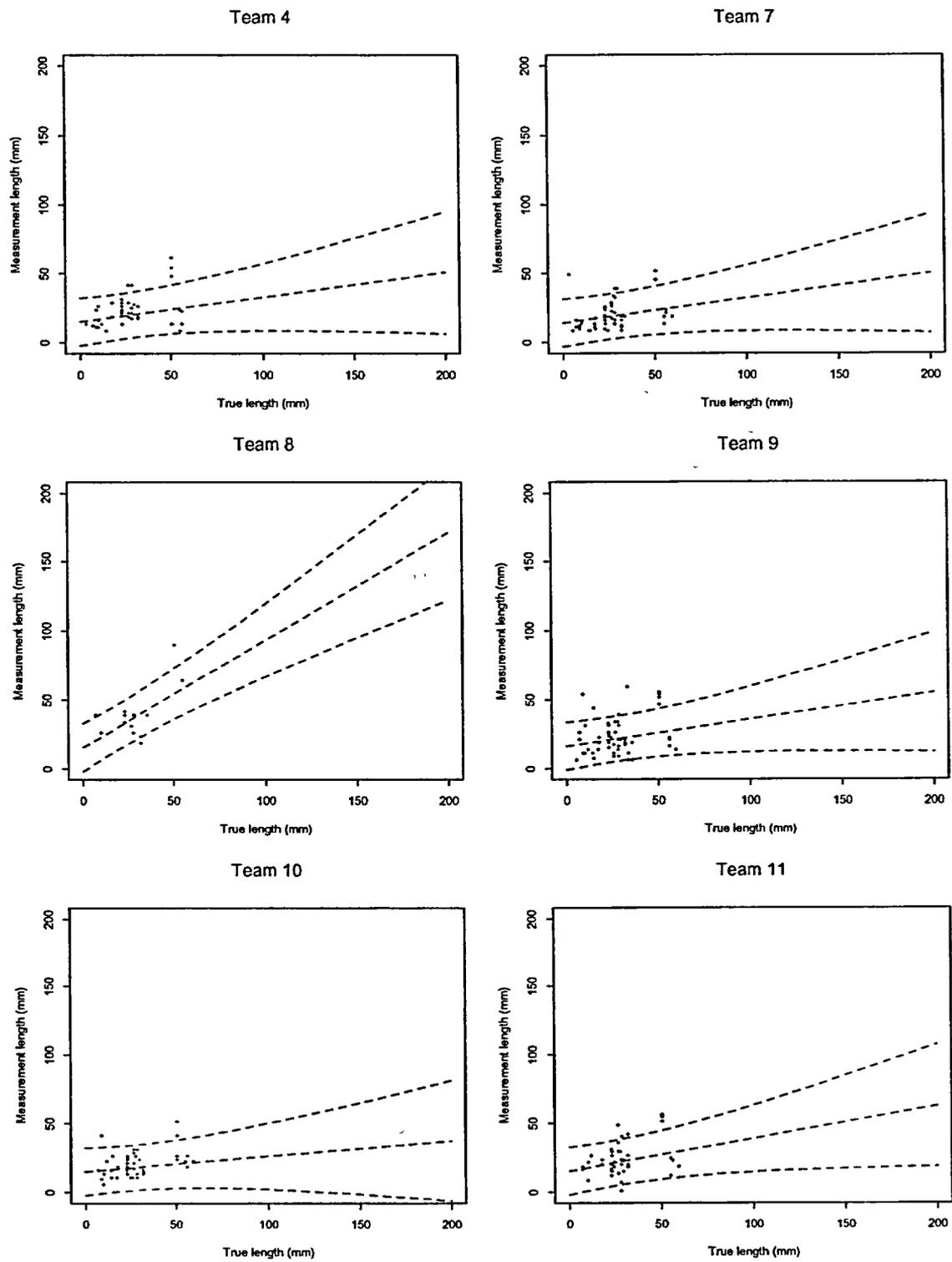
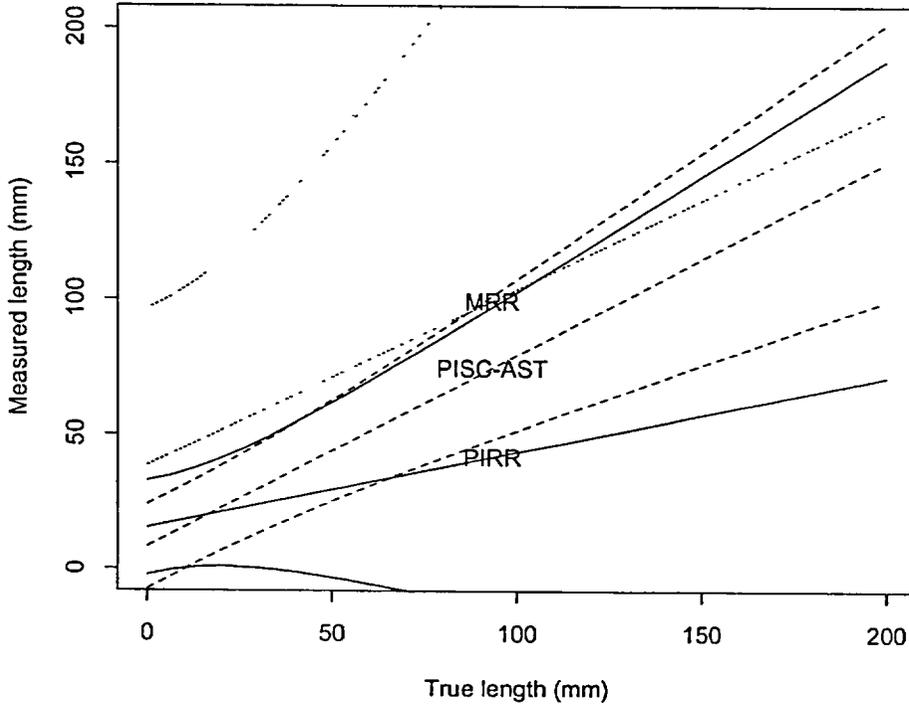
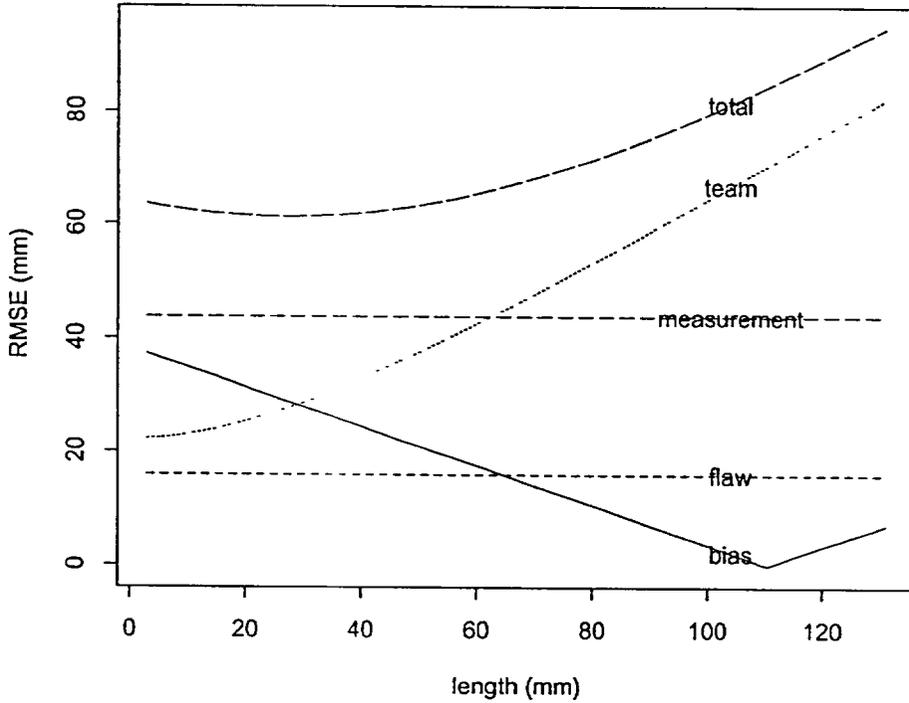


Figure 5.4 PIRR Length Sizing Regressions for Individual Teams



**Figure 5.5** Length Sizing Regression for MRR Round Robin Including all Teams, the Average, and 95% Confidence Bounds



**Figure 5.6** MRR RMSE Length Sizing Errors

realistic of sizing results produced in a field examination.

Figure 5.7 shows the MRR length sizing regressions for the 14 individual teams.

### **5.3 PISC-AST Length Sizing Results**

The PISC-AST length sizing results shown in Figure 5.8 have the smallest team/team variability exhibited in any of the three studies. A line going from the lower left-hand corner to the upper right-hand corner would represent ideal performance in this figure. Team DH exhibited systematic over sizing the length of every crack. The remaining

teams exhibited a similar trend as found in the PIRR where they over sized the small cracks and under sized larger cracks. The measurements in Figure 5.8 give some indication that length sizing errors might be proportional to the flaw length; small flaws show less scatter than the large flaws. However, mid-sized flaws (at about 60 mm) show more scatter than either small or large flaws.

Length RMSE for the PISC-AST study is about 15 mm, without consideration of bias. With bias, the RMSE is still under 30 mm for flaws less than 100 mm, as shown in Figure 5.9.

Figure 5.10 shows the PISC-AST length sizing regressions for the 23 individual teams.

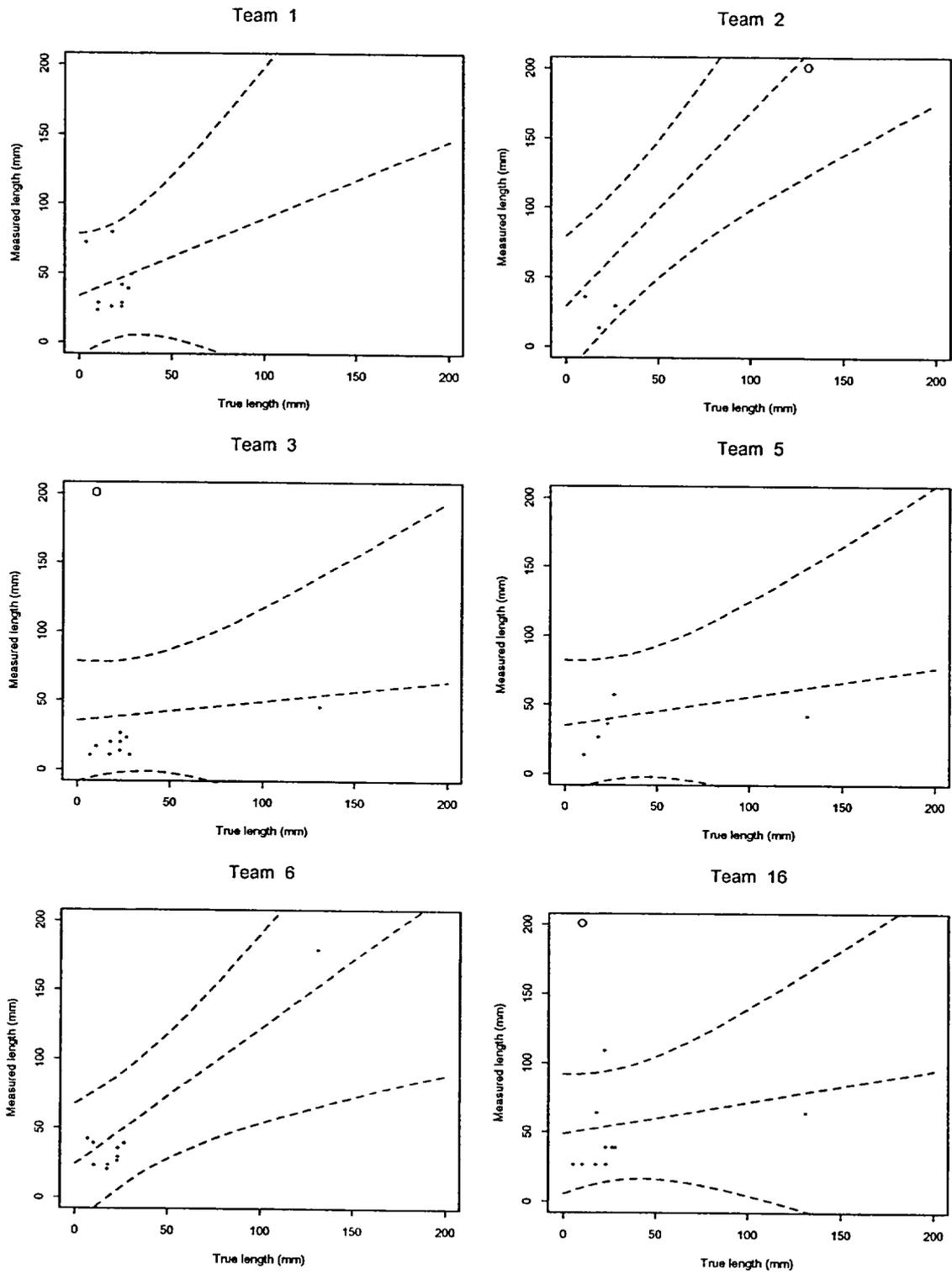


Figure 5.7 MRR Length Sizing Regressions for Individual Teams

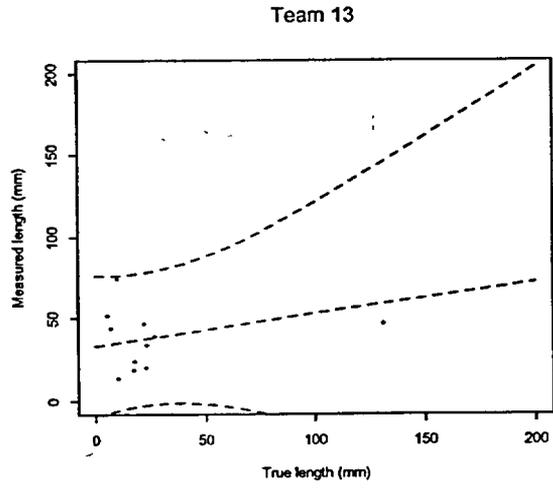
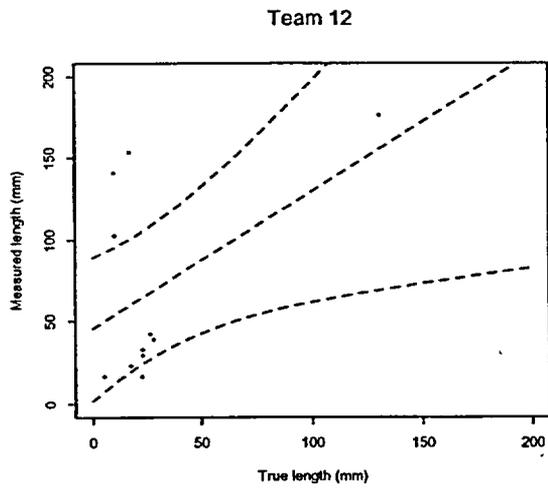
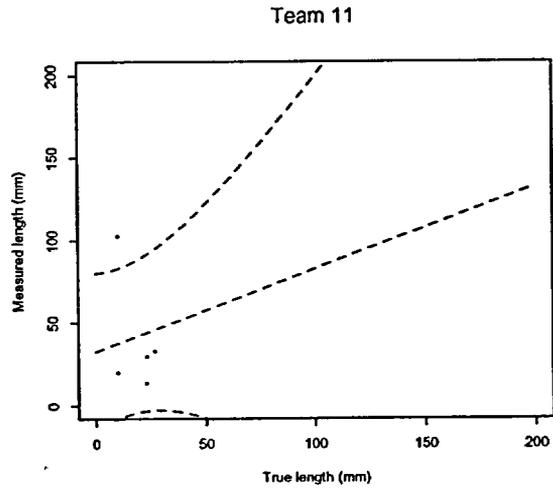
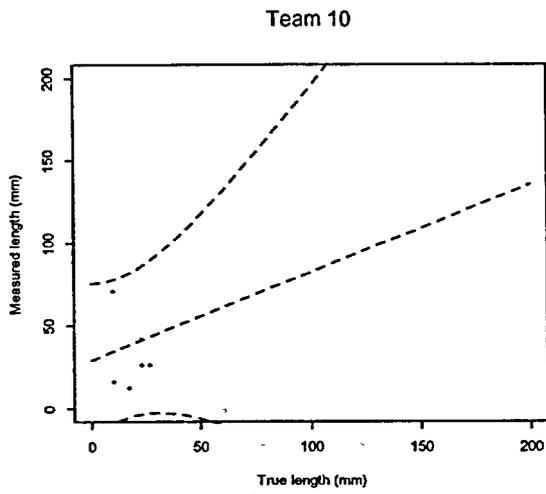
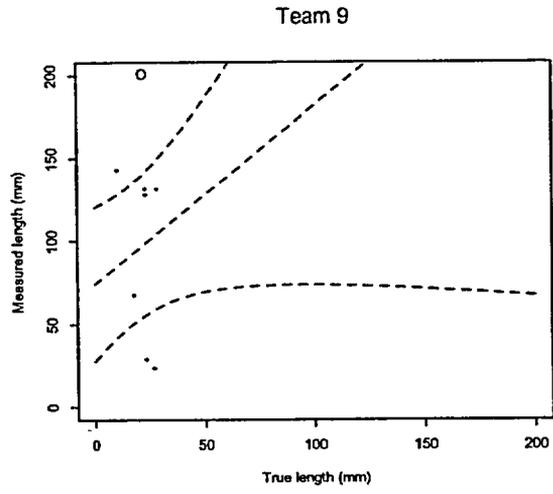
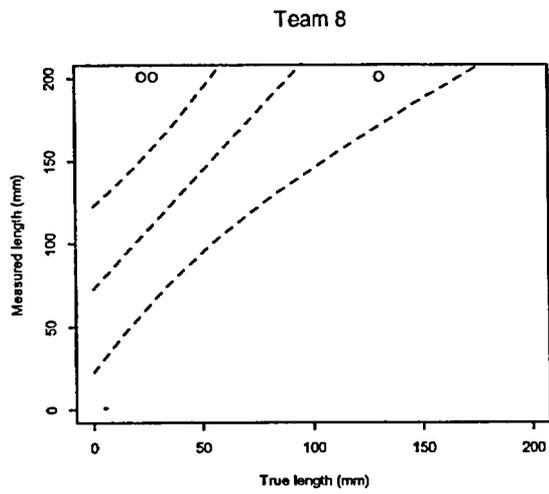


Figure 5.7 (Continued)

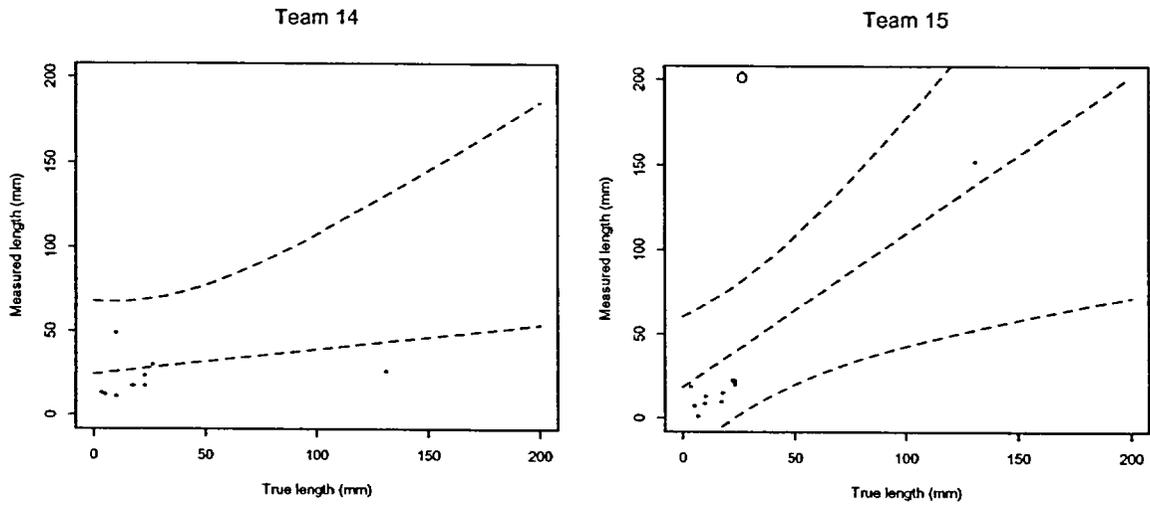


Figure 5.7 (Continued)

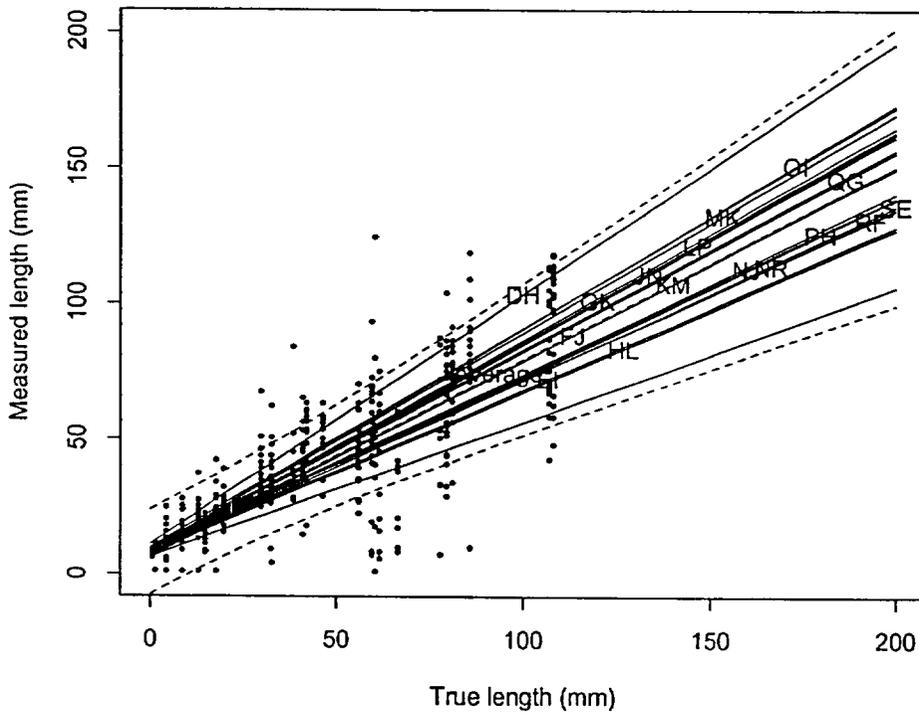


Figure 5.8 Length Sizing Regression for PISC-AST Round Robin Including all Teams, the Average, and 95% Confidence Bounds

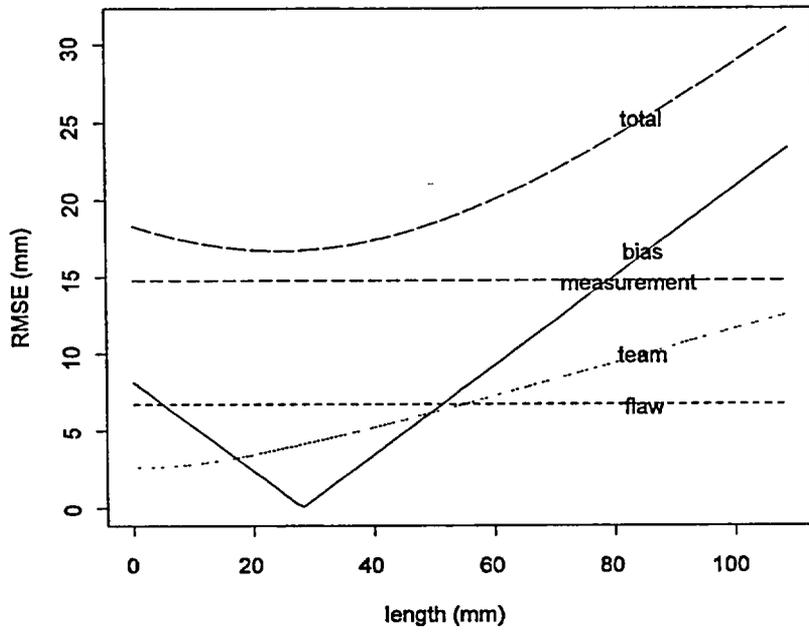


Figure 5.9 PISC-AST RMSE Length Sizing Errors

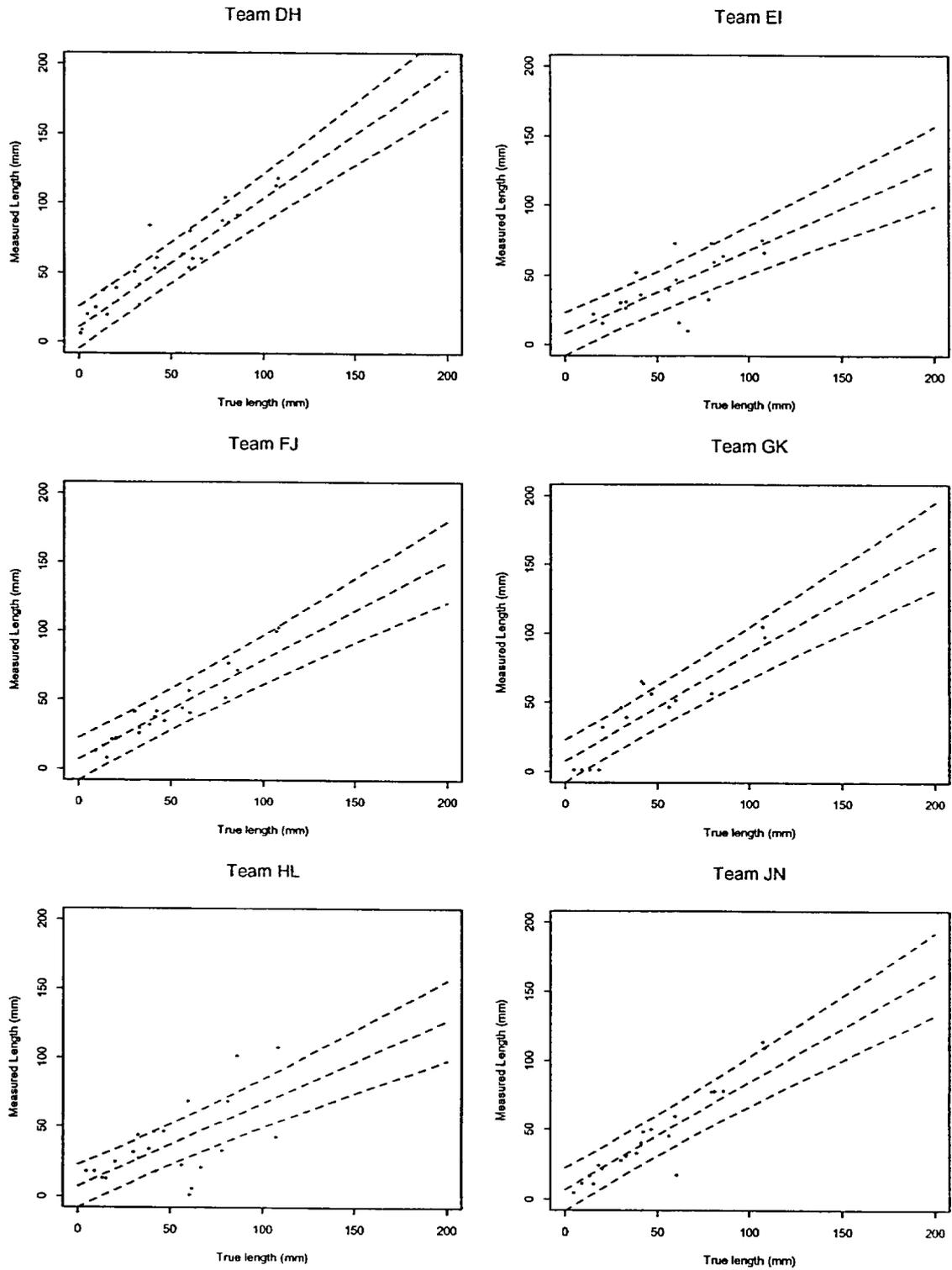


Figure 5.10 PISC-AST Length Sizing Regressions for Individual Teams

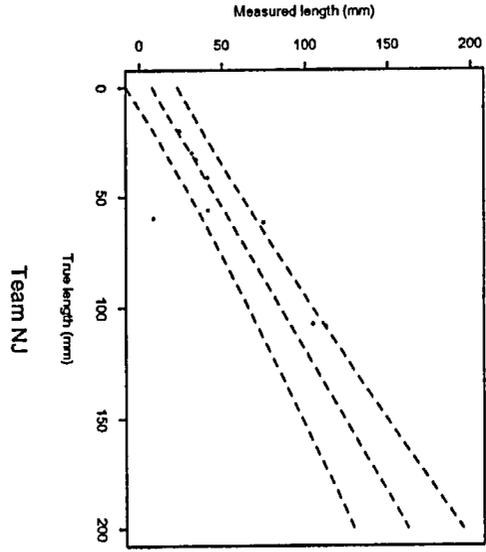
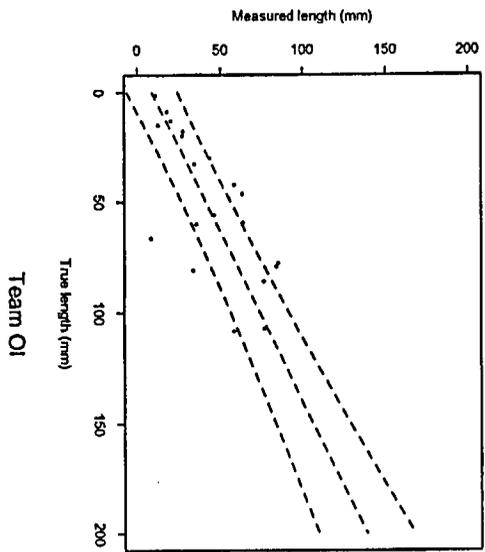
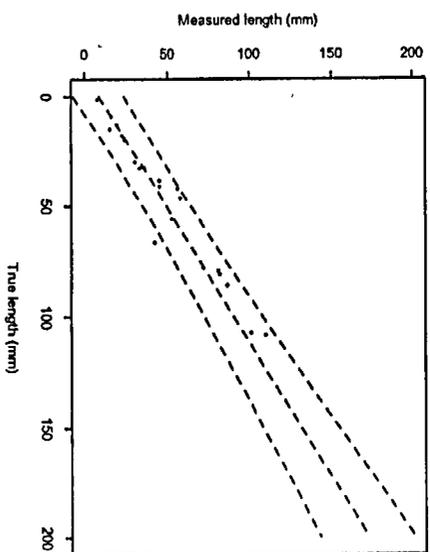
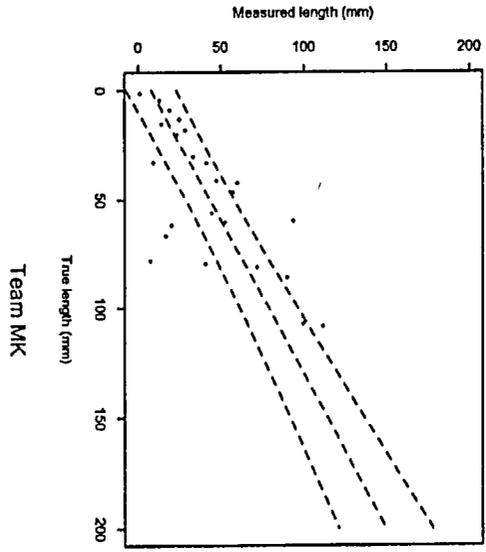
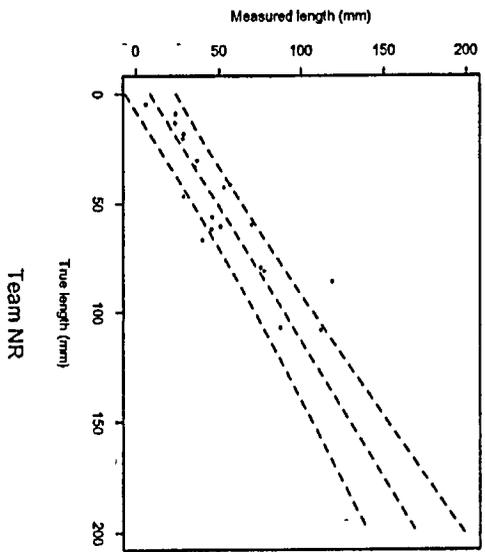
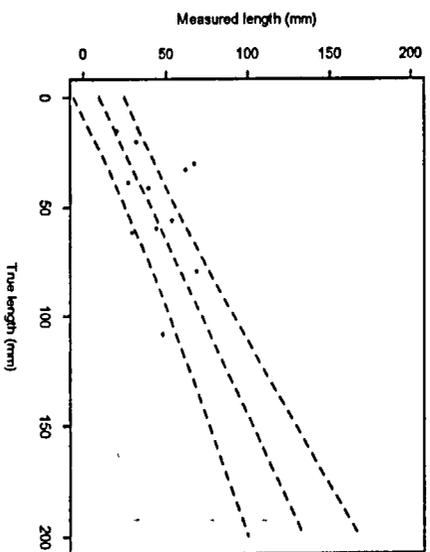


Figure 5.10 (Continued)

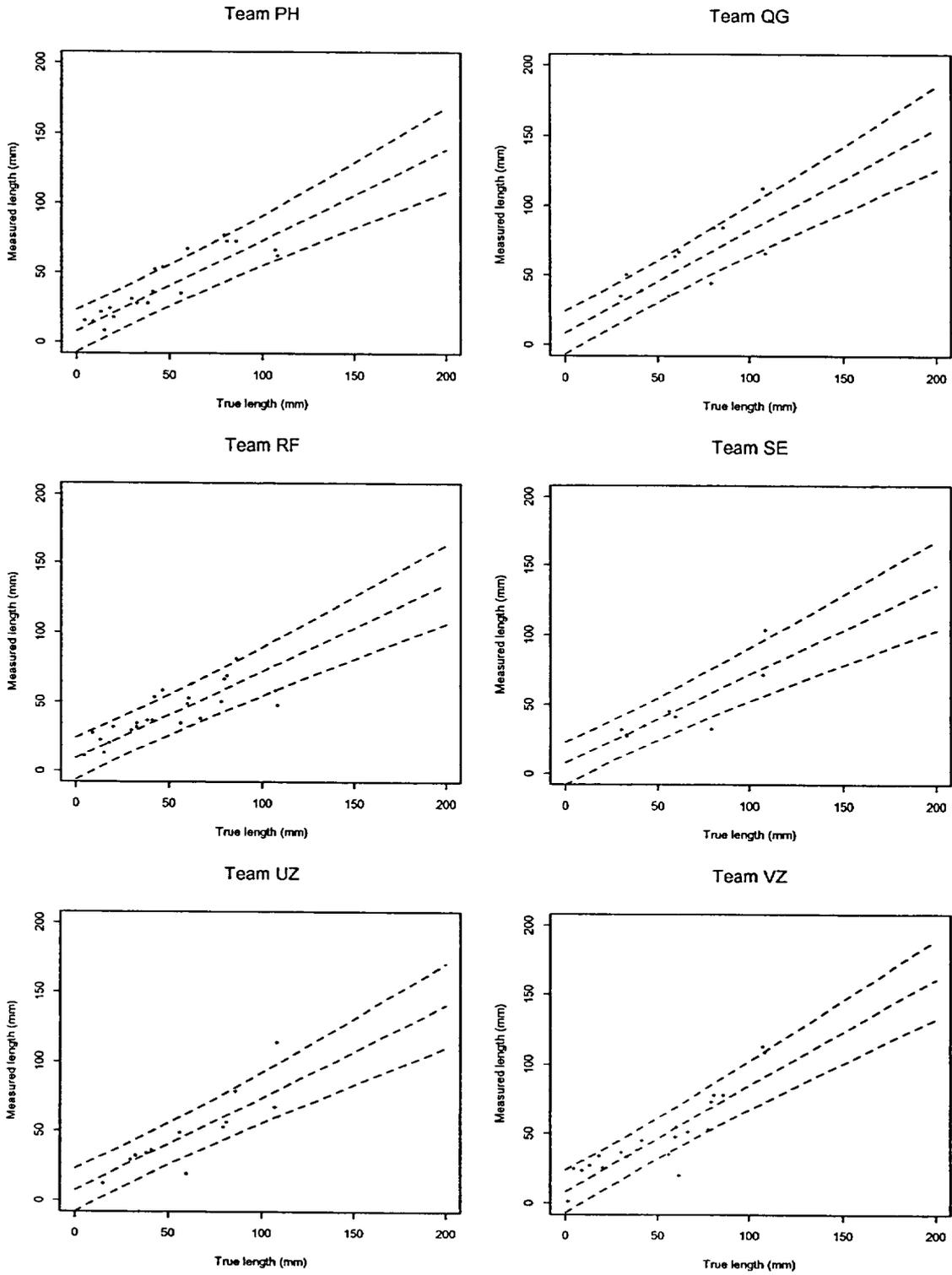


Figure 5.10 (Continued)

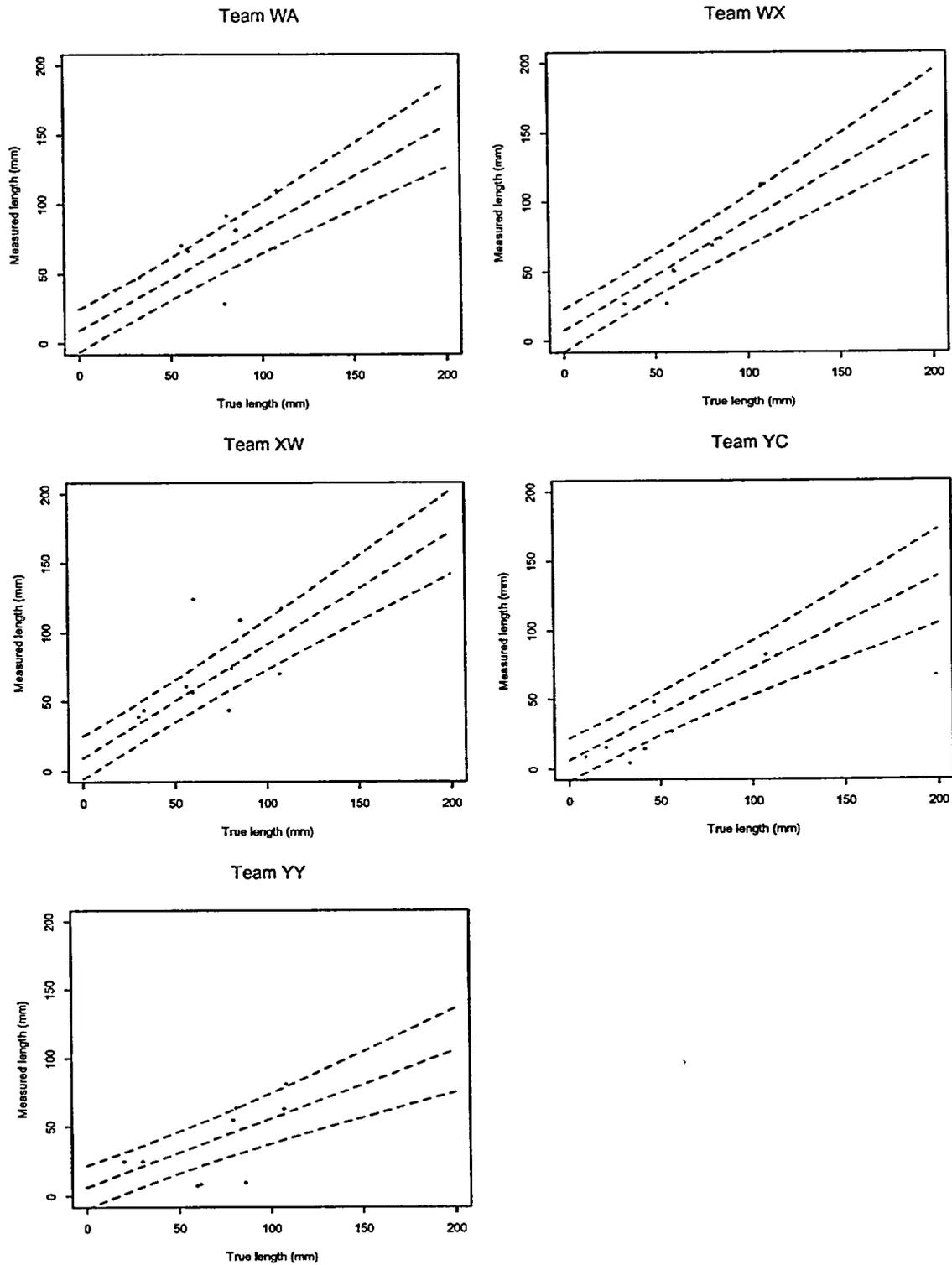


Figure 5.10 (Continued)

## 6 DETECTION CAPABILITY

Table 6.1 presents a summary of the POD regression parameters for the three studies, while Figure 6.1 provides a graphical representation. In terms of average POD performance, the studies are ordered chronologically; PISC-AST is the best, followed by the MRR and then the PIRR. There is some ambiguity in this ordering, because the slope of the MRR POD curve is less than that of PIRR. For large flaws, the PIRR POD actually exceeds the MRR, but not significantly.

From Table 6.1, one can see that the PIRR and PISC-AST results are surprisingly similar. The logistic slopes,  $\beta_2$ , are almost exactly the same at 0.45, and the team/team and flaw/flaw variabilities are also very comparable. The flaw-to-flaw variability,  $\sigma_F^2$  is about 2.5 in both studies, and the team/team covariances,  $\gamma_T$ , are similar, but the slope  $\gamma_{w,w}$  is within a factor of 2.

The essential difference between the PISC-AST and PIRR studies is an increase in the intercept term from the PIRR value of -1.54 to -0.31 in PISC-AST. This increase effectively raises average POD by 30 probability points as can be seen in Figure 6.1. This provides very unambiguous evidence that detection capability

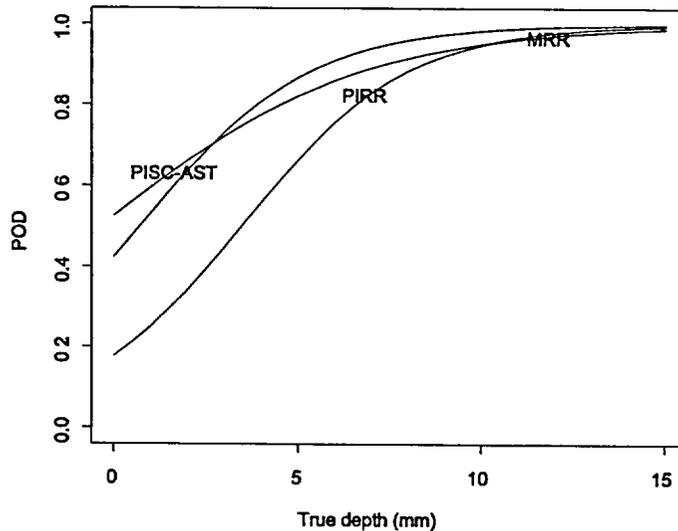
has improved in the decade between PIRR and PISC-AST and POD has been increased without increasing team/team or flaw/flaw variabilities.

Given the fact that the population of PISC-AST teams is much more diverse than the population of PIRR teams (i.e., world-wide vs. U.S. teams), the fact that team/team variability is about the same in both studies is an indication that team/team variability has probably actually decreased since the PIRR study.

It should be noted that the current POD analysis conducted on the PIRR data contains no false call information, as contrasted to the analysis presented in [4]. This results in POD curves that differ somewhat from those reported in [4]. False calls are taken into account by being the limiting value of the POD curve when the flaw depth goes to zero.

The MRR detection results show a great deal of team/team variability and because of the small number of flaws used in the round robin, do not determine the POD curve very accurately. In fact, the variability is so great that the logistic slope parameter is not significantly different

| Table 6.1 Summary of POD Fits                              |                  |                    |        |       |          |       |
|--|------------------|--------------------|--------|-------|----------|-------|
|  | PIRR             |                    | MRR    |       | PISC-AST |       |
|  | Est <sup>a</sup> | StDev <sup>b</sup> | Est    | StDev | Est      | StDev |
| No. of teams   | 6                |                    | 14     |       | 23       |       |
| No. of flaws   | 45               |                    | 13     |       | 26       |       |
| No. of tests   | 624              |                    | 190    |       | 523      |       |
| Avg. Flaw Length, mm                                       | 2.31             |                    | 2.70   |       | 4.43     |       |
| $\beta_1$  | -1.54            | 0.63               | 0.1    | 0.60  | -0.31    | 0.69  |
| $\beta_2$ mm <sup>-1</sup>                                 | 0.45             | 0.16               | 0.28   | 0.17  | 0.44     | 0.13  |
| $\gamma_{1,1}$   | 1.077            |                    | 1.788  |       | 3.151    |       |
| $\gamma_{1,2}$ mm <sup>-1</sup>                            | 0.001            |                    | -0.267 |       | -0.187   |       |
| $\gamma_{2,2}$ mm <sup>-2</sup>                            | 0.008            |                    | 0.091  |       | 0.017    |       |
| $\sigma_F^2$   | 2.97             |                    | 0.85   |       | 2.37     |       |
| <sup>a</sup> Estimate.<br><sup>b</sup> Standard Deviation. |                  |                    |        |       |          |       |



**Figure 6.1** Comparison of Round Robins

from 0. The MRR was not constructed to determine accurate detection curves, but to try to measure significant improvements in detection performance versus the PIRR study, so these results are not surprising.

It is important to note that flaw/flaw variability is large in both the PISC-AST and PIRR studies. A variability of  $\sigma_F^2 = 2.5$  translates into a sigma of about 1.5 on the logistic scale, and this can translate into 30 probability percentage points. This means, for example, that it would not be unusual for two flaws of the same size to differ by as much as 60 probability points (one flaw could have a POD of 20% and the other a POD of 80%). Such a large variability is a significant problem for POD performance. Although average performance is quite acceptable, we can't be sure of consistent POD on individual flaws.

This variability also has an important implication for "multiple inspections," which have been advocated to increase POD on important components. If inspections were independent,

multiple inspections would increase POD according to the formula:

$$POD_{\text{Multiple}} = 1 - (1 - POD_{\text{ind}})^N \quad (6.1)$$

where N is the number of multiple inspections performed. This formula shows that multiple inspections can raise the POD to any desired level (no matter how poor the individual inspection POD is), if one is willing to re-inspect enough times. However, the formula is incorrect if substantial flaw/flaw variability exists, as the present regression results indicate. When substantial differences exist between flaws, multiple inspection has limited usefulness; there will be some flaws that all teams will have difficulty detecting, and others that all will detect. The present POD regression results could be used to construct a more realistic formula for multiple inspection POD, but that is beyond the scope of this report.

Of course, both the PIRR (IGSCC, TFC) and PISC-AST (IGSCC, TFC, MFC) studies used

different types of flaws, so some of this flaw/flaw variability may be due to flaw type. While some of the variability is undoubtedly due to flaw type, differences are also observed in flaws of the same type. In fact, it was found in these studies that the variability within a flaw type was larger than the variability between types of flaws. Thus, for this analysis all flaws were combined.

## 6.1 PIRR POD Results

Figure 6.2 shows the POD curves for individual PIRR teams. As one can see from the plot, one team (8) seems to have done much worse than the other teams and might be considered an outlier.

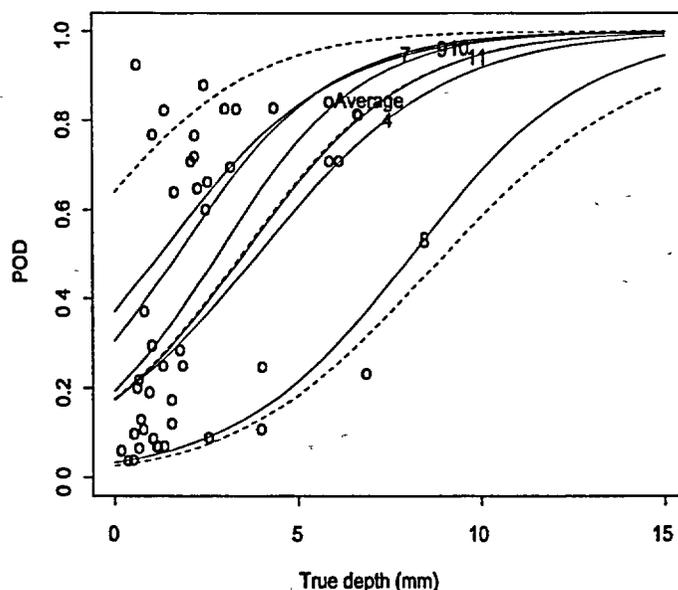
It is interesting to note that Team 8 did well on both the depth and length sizing portions of the round robin, and it is possible that its poor detection results are related to its good sizing results. If this team reported only flaws that were very easy to detect, its POD would be low, and if these flaws were “easy” to size, they would produce improved sizing results.

The average POD curve displayed in Figure 6.2 is surrounded by 95% bounds (dashed lines) that reflect sampling, flaw/flaw, and team/team variability. With all this uncertainty, the bounds are quite large. Even on a 10 mm flaw, the lower bound on POD is no more than 50%. Excluding Team 8 raises this lower bound by about 10 percentage points at the left-hand end of the plot (i.e., for small flaws).

Figure 6.3 presents individual team POD curves containing 95% confidence bounds. The 95% bounds for the individual teams are also fairly wide, due to the large flaw/flaw variability.

## 6.2 MRR POD Results

The MRR results display substantial team/team variability, but as one can see from Figure 6.4, this is due to two teams (Teams 1 and 11) that each failed to detect the one large flaw in the MRR. If these two teams are eliminated, the average POD for the MRR is very close to the PISC-AST curve.



**Figure 6.2** POD Regression for PIRR Round Robin Showing the Average Performance and 95% Confidence Bounds in Dashed Lines

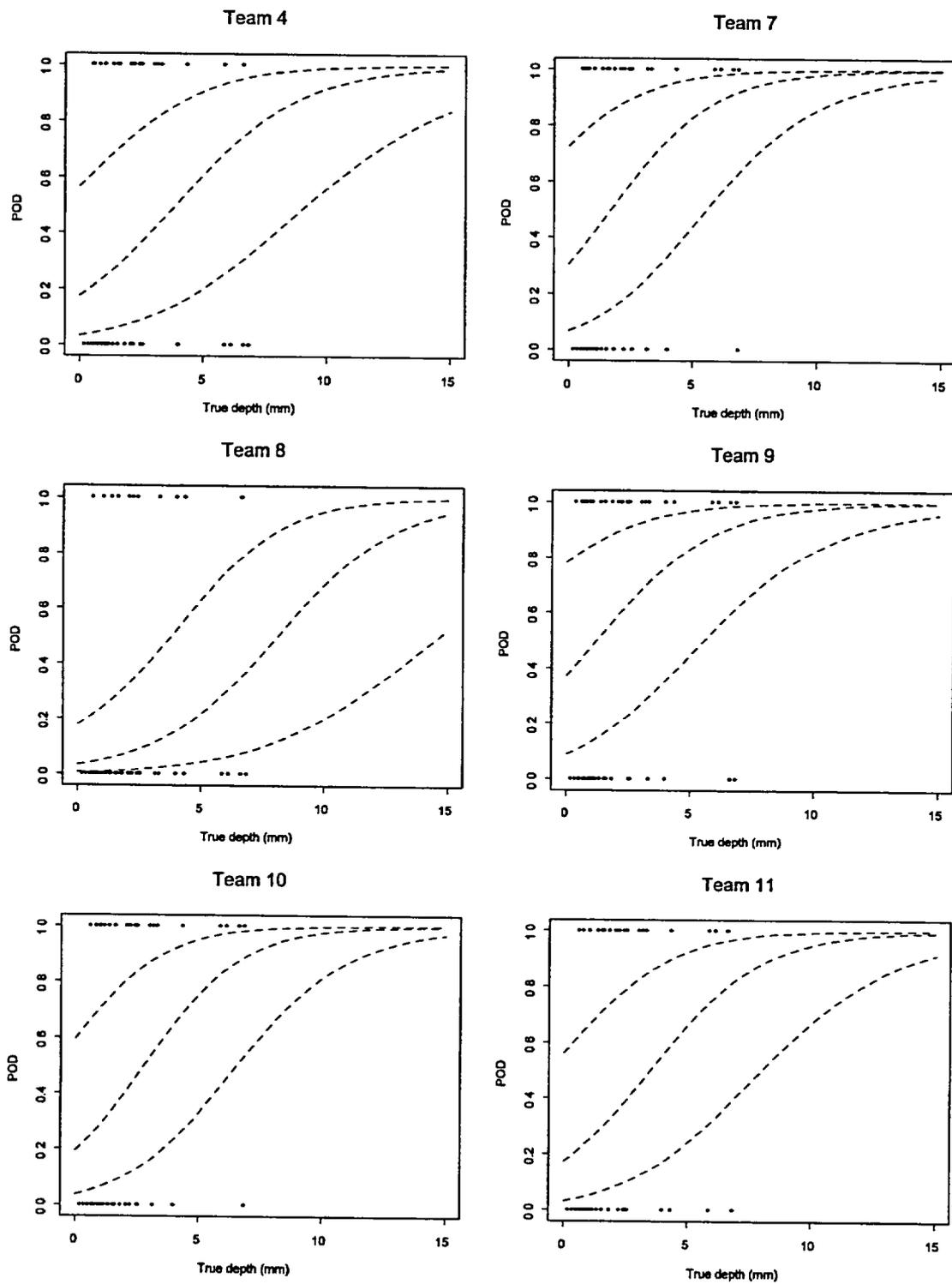
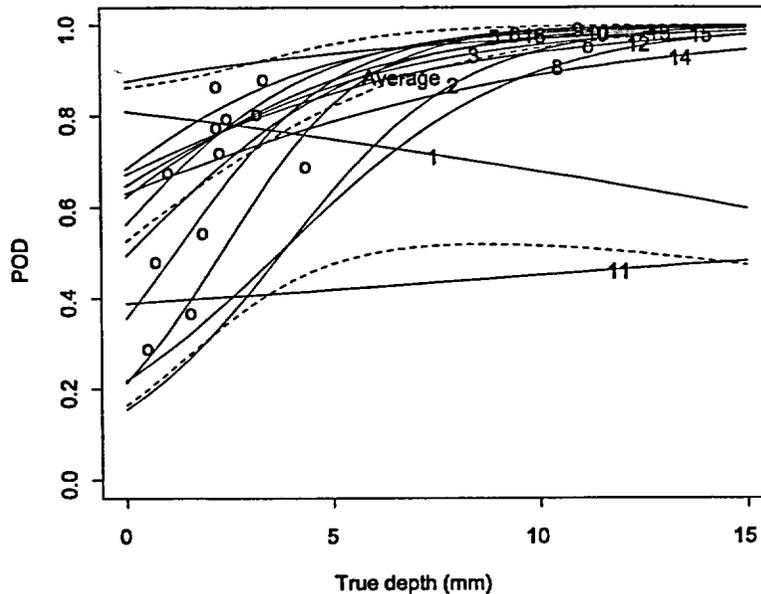


Figure 6.3 PIRR POD Regressions for Individual Teams with 95% Confidence Bounds



**Figure 6.4** POD Regression for MRR Round Robin Showing the Average Performance and 95% Confidence Bounds in Dashed Lines

One can see from Figure 6.5 that the POD curves in the MRR round robin are determined from very little data. Consequently, the 95% bounds for POD are quite large. The MRR study contains insufficient flaws to produce good POD curve estimates.

### 6.3 PISC-AST POD Results

Figure 6.6 presents the individual team POD curves along with the PISC-AST average (dashed line in figure) along with the 95% confidence bounds. An interesting feature of this population of teams is the fact that the teams cluster into two groups. One group is basically above the average POD while the other group is substantially below the average. The high group consists of teams DH, EI, FJ, HL, JN, KM, MK, NJ, OI, PH, RF, VZ, and UZ. The low group consists of teams GK, LP, NR, QG, SE, WA, WX, XW, YC, YY.

Obviously, if the “high” teams could be considered a separate population, the POD results

for this population would be substantially increased; the average POD would be increased by about 20 percentage points for small flaws, team/team variability would be substantially reduced, and the 95% confidence bounds would be much tighter.

If the “high” population results were related to some characteristic of the inspection procedures used by these teams, one would have some evidence that the “high” group of PODs is not due to chance, but reflects a definable population in its own right. There is no characteristic that exactly discriminates between these two groups, but the ordering appears to be related to the inspection sensitivity. All high teams except for three used noise level as their recording threshold; while all low teams, except for three, used 20% DAC or higher.

Figure 6.7 presents the individual team results from the PISC-AST study.

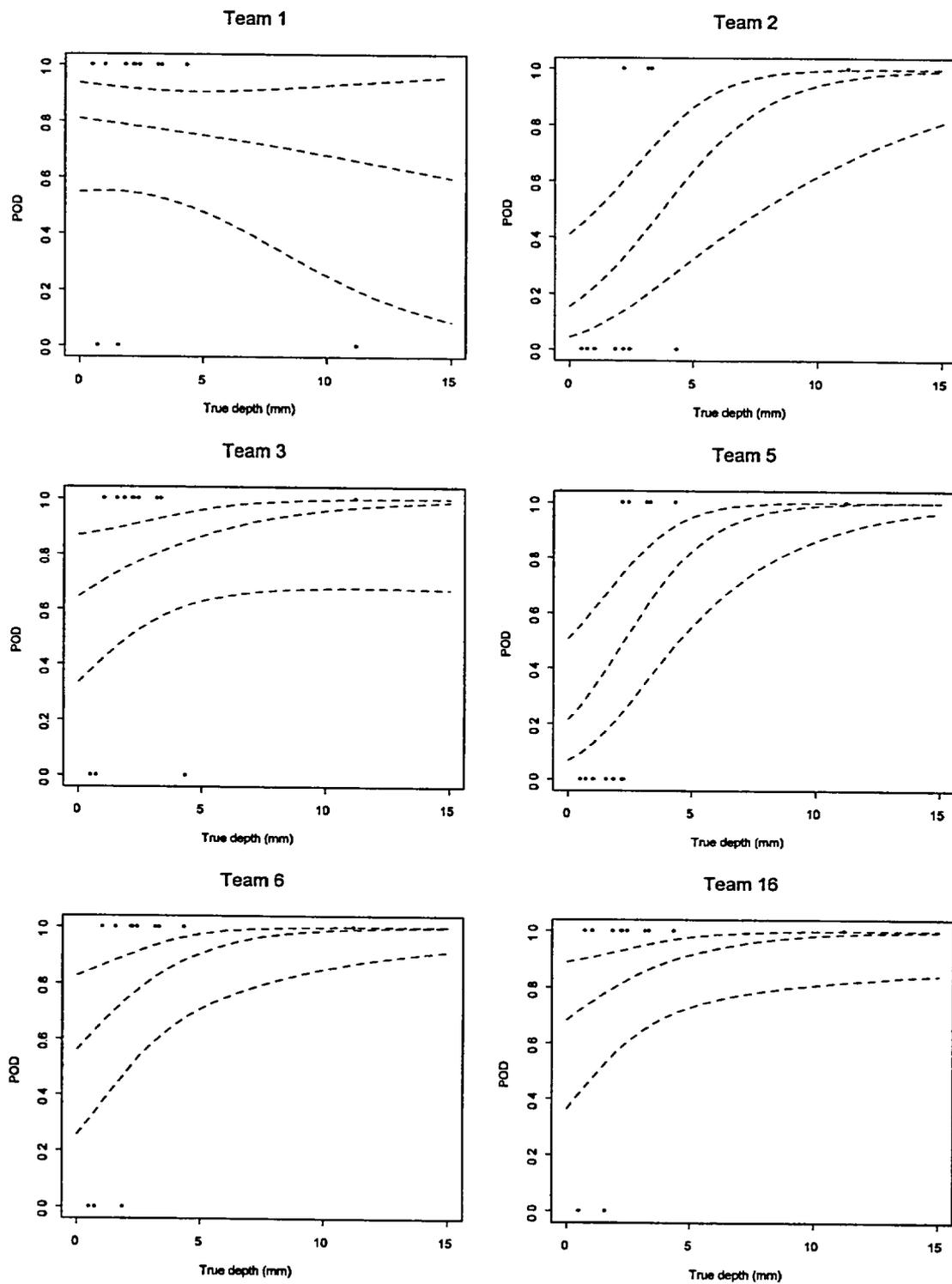


Figure 6.5 POD Regressions for Individual MRR Teams with 95% Confidence Bounds

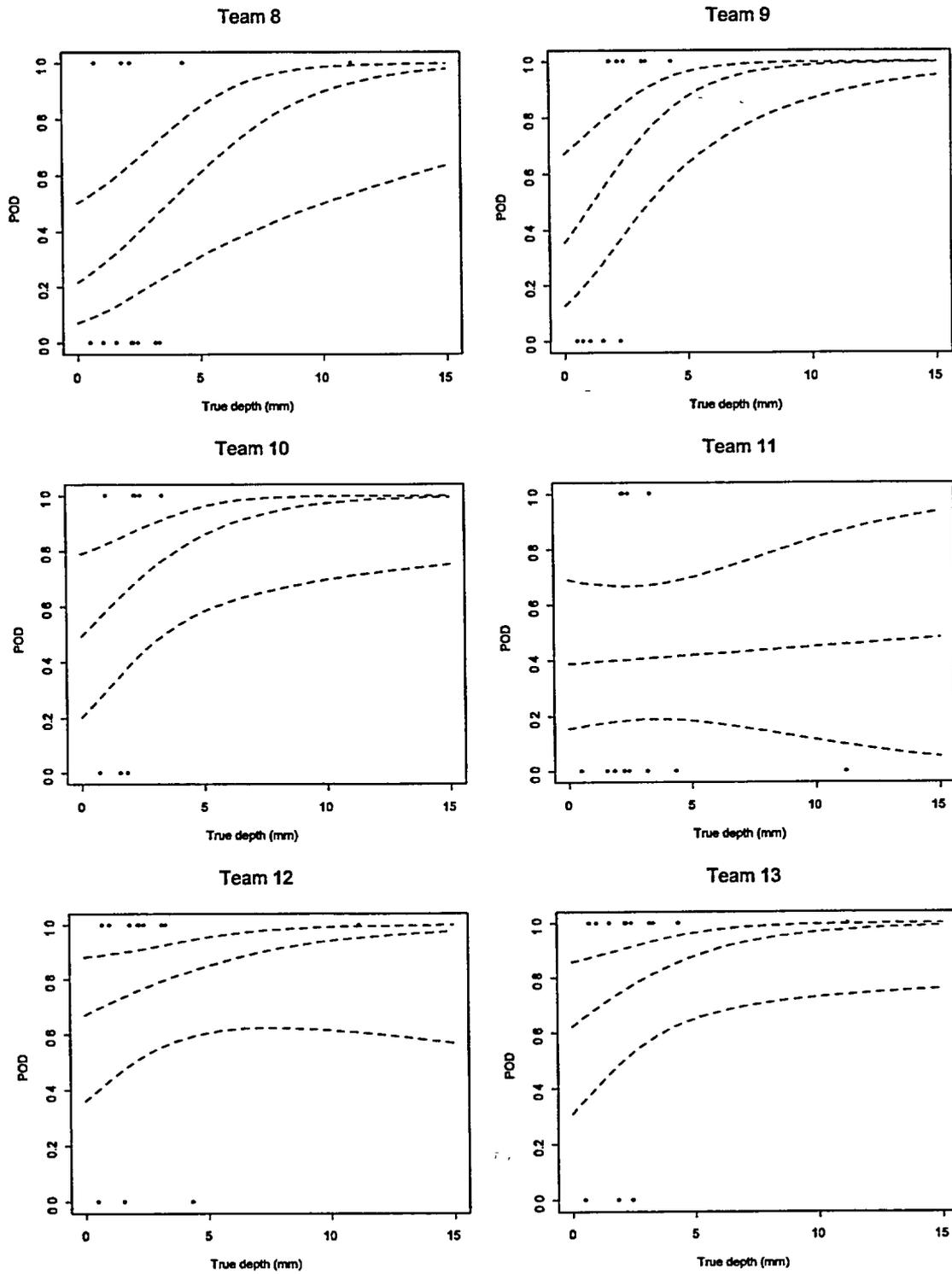


Figure 6.5 (Continued)

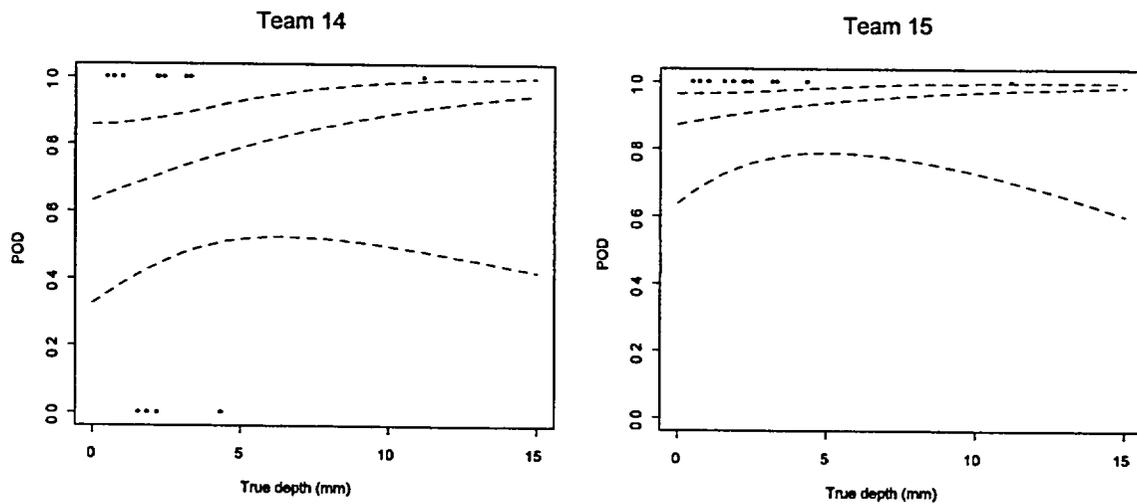


Figure 6.5 (Continued)

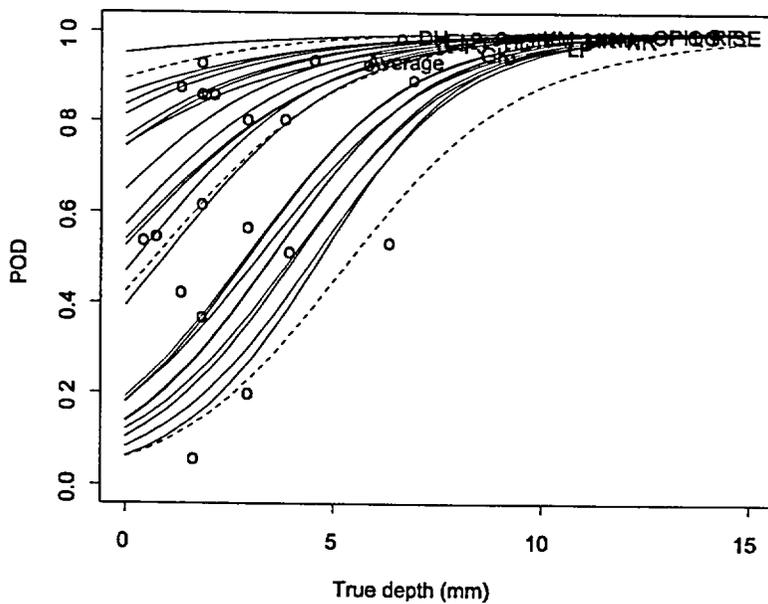


Figure 6.6 POD for PISC-AST Round Robin Showing Average Team Results (dashed line) Along with 95% Confidence Bounds

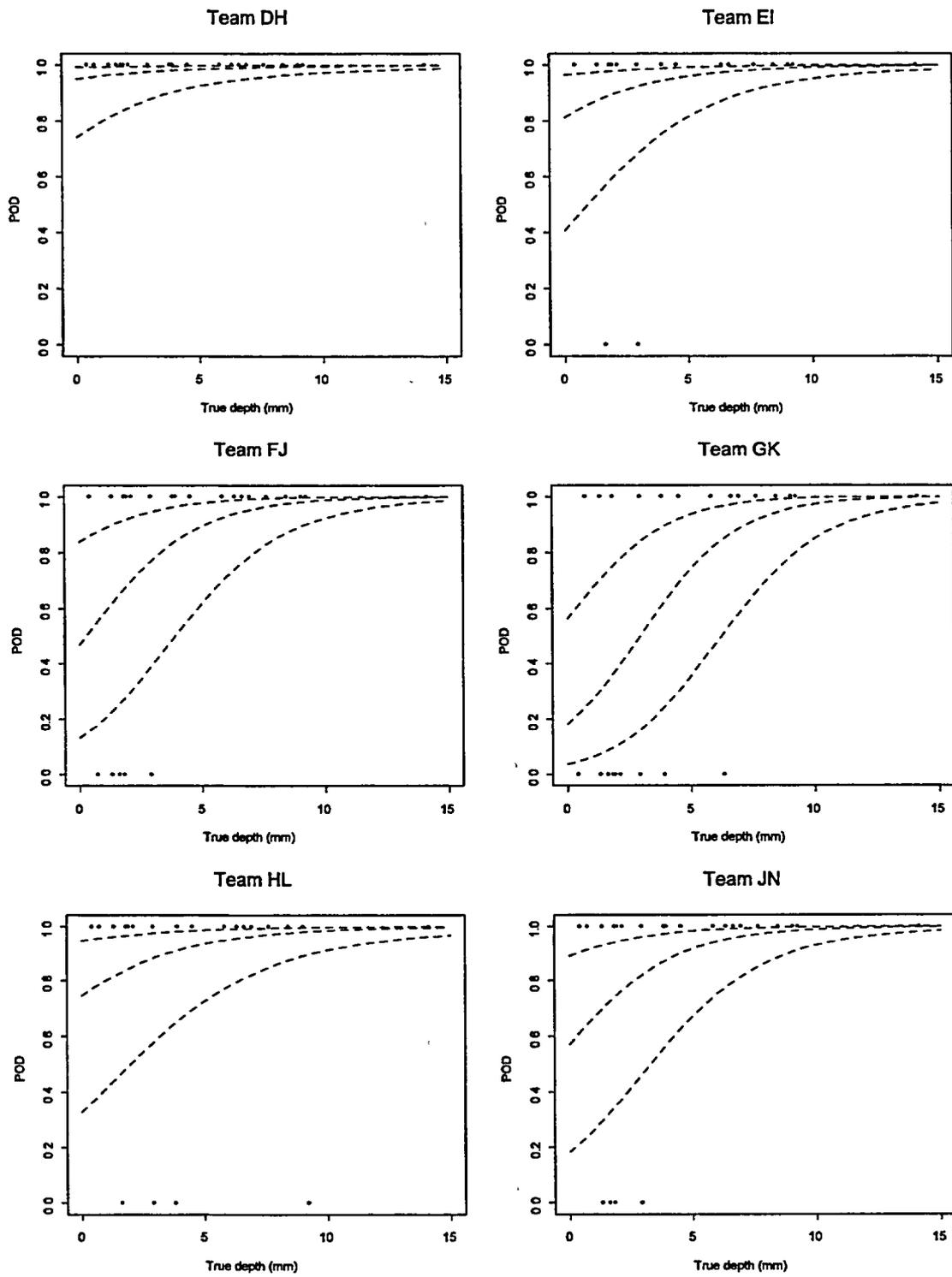


Figure 6.7 PISC-AST POD Regressions for Individual Teams with 95% Bounds

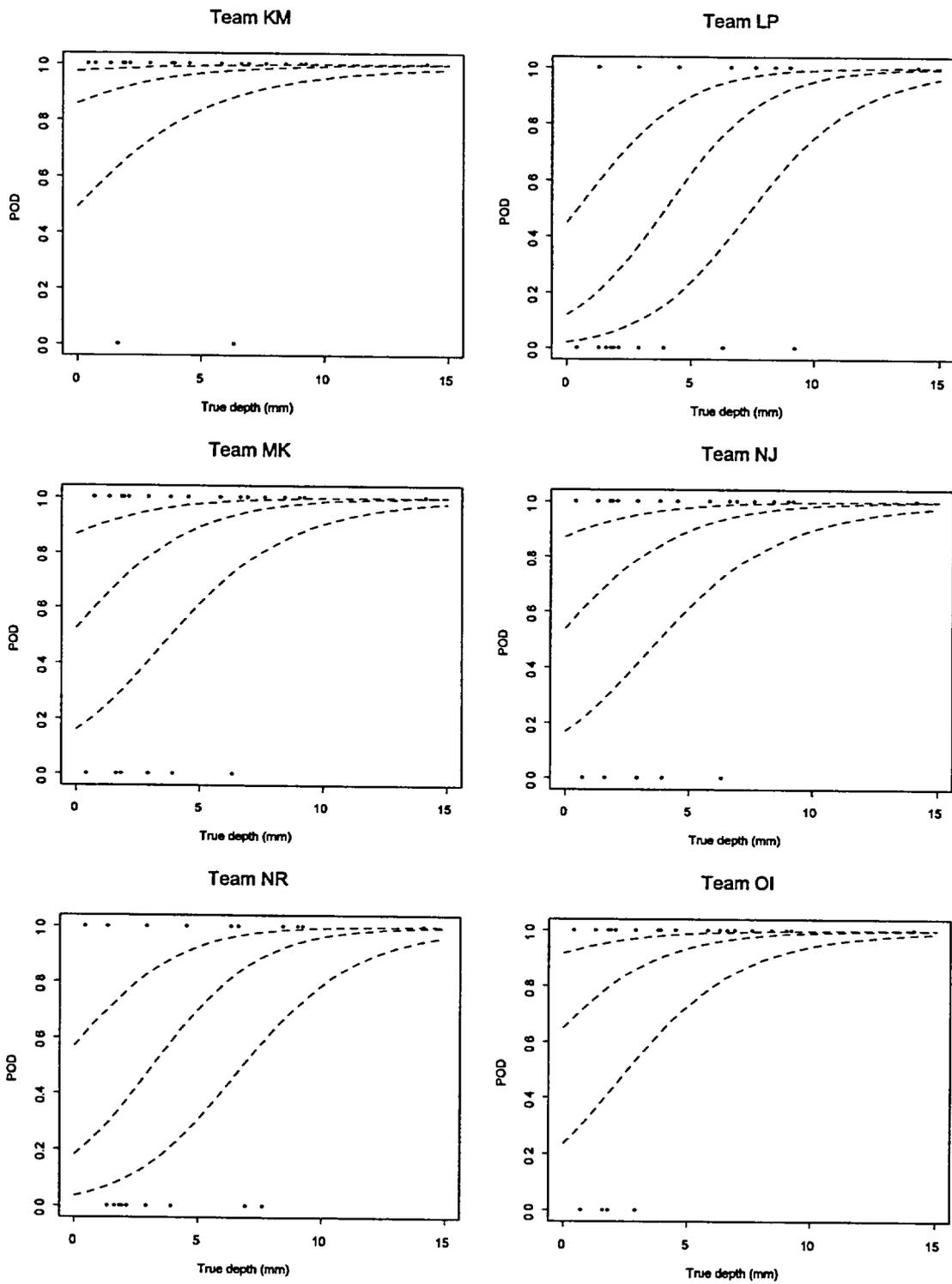


Figure 6.7 (Continued)

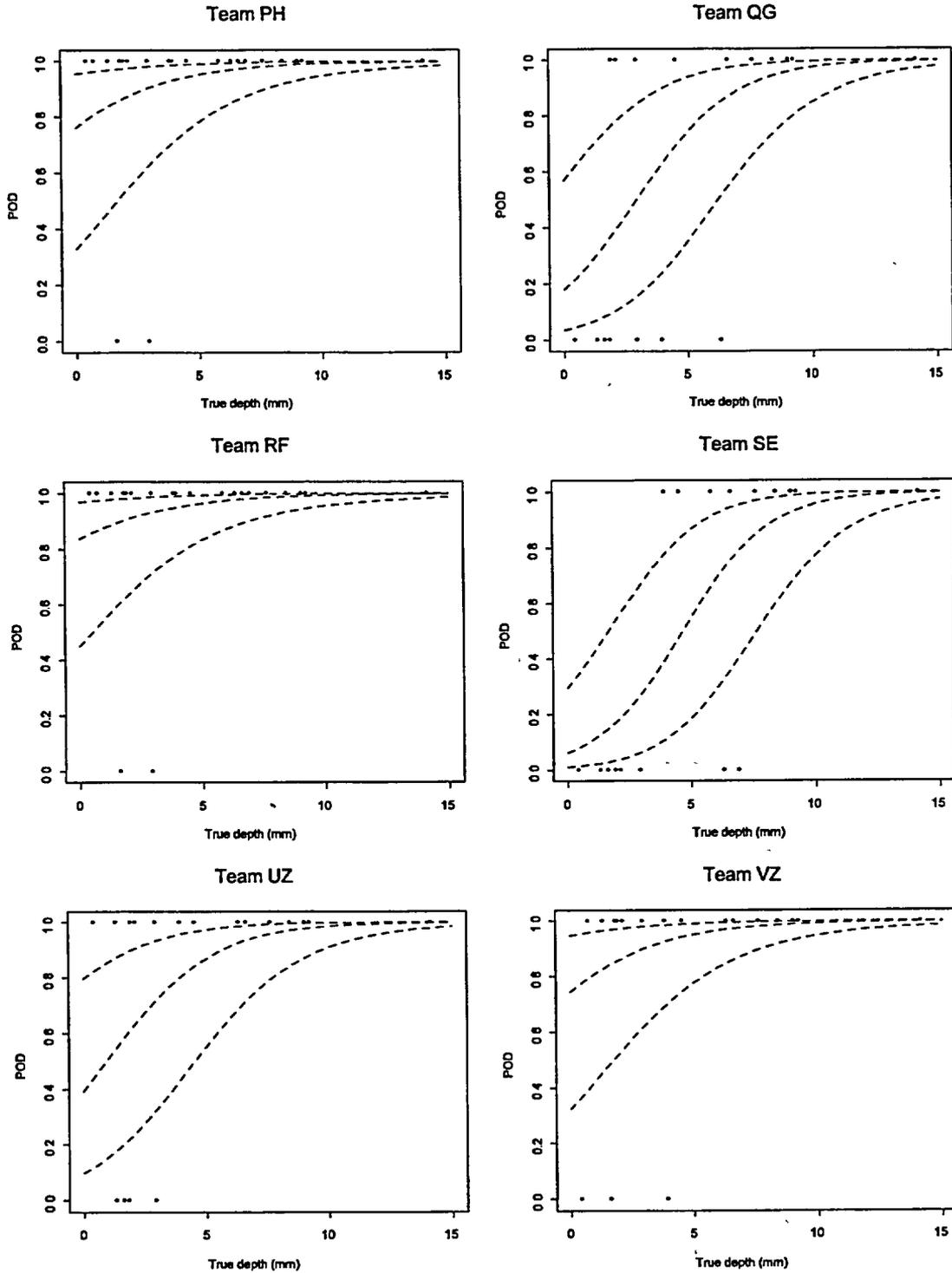


Figure 6.7 (Continued)

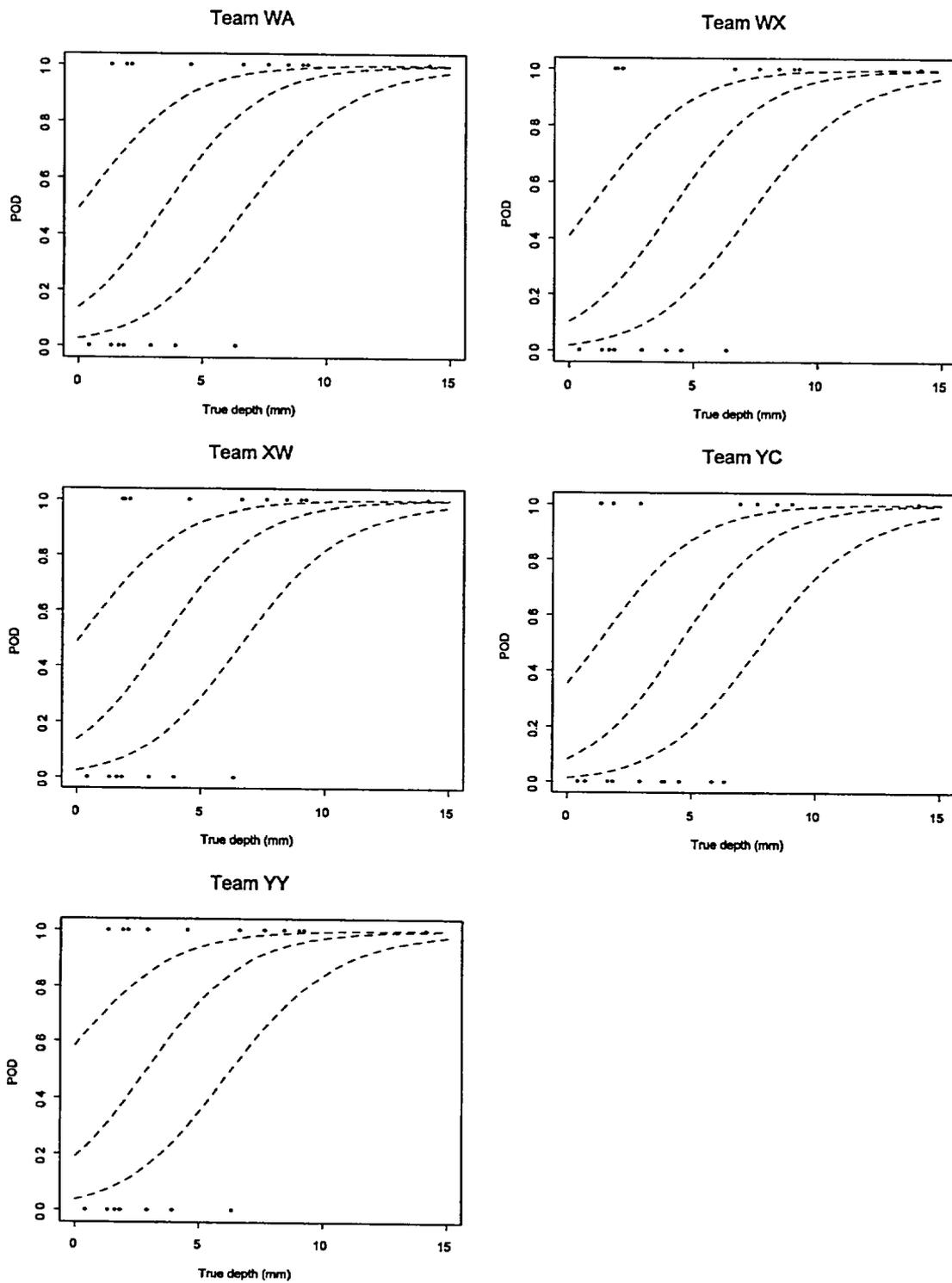


Figure 6.7 (Continued)



## 7 COMBINED RESULTS

It is reasonable to attempt to combine similar round robin results to obtain better estimates of inspection capability. From the individual round robin evaluations described in previous sections, we have evidence that inspection capability has changed over time, so we certainly would not want to combine all three round robins. However, it appears that the MRR and PISC-AST inspection capabilities are sufficiently alike to consider combining them.

We combined the MRR and PISC-AST results to form a "best" estimate of current inspection capability. Perhaps the most obvious way to combine two round robins is to combine the data, and then re-fit the regression models to the resulting data. We used an alternative method that is simpler and approximately equivalent to the first; we averaged the individual round robin regression parameters to produce a best estimate of parameters, as shown in Table 7.1.

It should be noted that the combined covariance terms are sometimes larger than either of the corresponding MRR or PISC terms. This is because the terms incorporate the "between round robin variability," something the individual round robin covariances cannot do.

It should be further noted that there is a fundamental problem associated with combining two round robins. When one combines two separate round robins together in a single data set, the resulting data is completely unbalanced. This means that the teams from the two round robins have inspected no flaws in common, so their performances cannot be directly compared. The same can be said about the flaws: since they have not been inspected by any common teams, they cannot be compared. The two round robins cannot be combined without making an assumption regarding the two flaw populations (or team populations).

A simple example may illustrate the problem more clearly. Suppose we calculate average POD from two round robins, using sets of flaws and teams that are superficially the same in both round robins (i.e., same flaw types and sizes, equivalent inspection procedures), and we find that the teams in one round robin achieve a POD of 25%, while the teams in the other round robin achieve a POD of 75%. Now the fundamental question arises: why is the POD from the second round robin so much better than the first? Is it because the teams are better or because the flaws are easier?

It is not possible to answer this question from the combined data set. The only way to obtain a definitive answer is to have some teams from one round robin inspect some flaws from the other round robin.

When two separate round robins are combined, this ambiguity will be present in the result. Our method of combination assumes that the two populations of flaws both form representative samples of flaws from the population we wish to study (flaws in wrought stainless steel piping). This is also why we have exercised some care (see Section 2) in identifying the inspections and flaws that would be included in the data sets.

Table 7.1 lists the combined regression parameters from the depth, length and POD regressions, alongside the MRR and PISC-AST parameters for comparison. The combined estimates are intermediate between the individual estimates, but are somewhat more complicated than simple averages. The combined estimate is a weighted average, designed to produce an estimate with the smallest variability. This is reflected in the parameter standard deviations.

The PISC-AST results are weighted more heavily than the MRR results, because the PISC-AST

| Table 7.1 Summary of Combined Fits (MRR and PISC-AST) |                  |                    |          |       |          |       |
|---|------------------|--------------------|----------|-------|----------|-------|
|   | MRR              |                    | PISC-AST |       | Combined |       |
|   | Est <sup>a</sup> | StDev <sup>b</sup> | Est      | StDev | Est      | StDev |
| Depth Sizing Fits                                     |                  |                    |          |       |          |       |
| $\beta_1$ mm  | 1.58             | 1.47               | 2.9      | 0.53  | 2.87     | 0.49  |
| $\beta_2$   | 0.87             | 0.29               | 0.28     | 0.08  | 0.32     | 0.07  |
| $\gamma_{1,1}$ mm <sup>2</sup>                        | 0.47             |                    | 4.36     |       | 3.83     |       |
| $\gamma_{1,2}$ mm                                     | 0                |                    | -0.37    |       | -0.46    |       |
| $\gamma_{2,2}$  | 0.14             |                    | 0.08     |       | 0.17     |       |
| $\sigma_F$ mm   | 4.18             |                    | 0.45     |       | 1.44     |       |
| $\sigma_E$ mm   | 7.78             |                    | 3.95     |       | 4.07     |       |
| RMSE(5) mm  | 4.11             |                    | 2.72     |       | 3.04     |       |
| RMSE(10) mm   | 5.18             |                    | 5.20     |       | 5.71     |       |
| RMSE(15) mm   | 6.69             |                    | 8.73     |       | 9.37     |       |
| Length Sizing Fits                                    |                  |                    |          |       |          |       |
| $\beta_1$ mm  | 38.25            | 10.15              | 8.13     | 3.00  | 10.41    | 2.87  |
| $\beta_2$   | 0.65             | 0.27               | 0.71     | 0.06  | 0.68     | 0.05  |
| $\gamma_{1,1}$ mm <sup>2</sup>                        | 491.30           |                    | 6.77     |       | 477.89   |       |
| $\gamma_{1,2}$ mm                                     | 0                |                    | 0        |       | -0.36    |       |
| $\gamma_{2,2}$  | 0.36             |                    | 0.01     |       | 0.14     |       |
| $\sigma_F$ mm   | 252.61           |                    | 45.12    |       | 112.41   |       |
| $\sigma_E$ mm   | 1912.58          |                    | 217.77   |       | 600.81   |       |
| RMSE(5) mm  | 63.24            |                    | 17.74    |       | 35.63    |       |
| RMSE(50) mm   | 63.31            |                    | 18.48    |       | 39.30    |       |
| RMSE(100) mm  | 79.52            |                    | 28.78    |       | 54.85    |       |
| POD Fits  |                  |                    |          |       |          |       |
| $\beta_1$   | 0.10             | 0.60               | -0.31    | 0.69  | -0.12    | 0.45  |
| $\beta_2$ mm <sup>-1</sup>                            | 0.28             | 0.17               | 0.44     | 0.13  | 0.38     | 0.10  |
| $\gamma_{1,1}$  | 1.79             |                    | 3.15     |       | 2.69     |       |
| $\gamma_{1,2}$ mm <sup>2</sup>                        | -0.27            |                    | -0.19    |       | -0.23    |       |
| $\gamma_{2,2}$ mm <sup>2</sup>                        | 0.09             |                    | 0.02     |       | 0.05     |       |
| $\sigma_F$  | 0.85             |                    | 2.37     |       | 1.88     |       |
| <sup>a</sup> estimated                                |                  |                    |          |       |          |       |
| <sup>b</sup> standard deviation.                      |                  |                    |          |       |          |       |

round robin contains more data and less variation. The table shows that combining the MRR results with the PISC-AST results does not change the latter dramatically. The biggest change that the MRR results cause is in the team/team variability, given by  $\gamma_{ij}$  in the table. The combined team/team variability is substantially larger than that experienced in PISC-AST.

Figure 7.1 graphically compares the combined RMSEs with the individual round robin RMSEs. From the figures, one can see that the combined depth RMSE is very close to that of PISC-AST, while the combined length RMSE is half-way between the individual RMSEs.

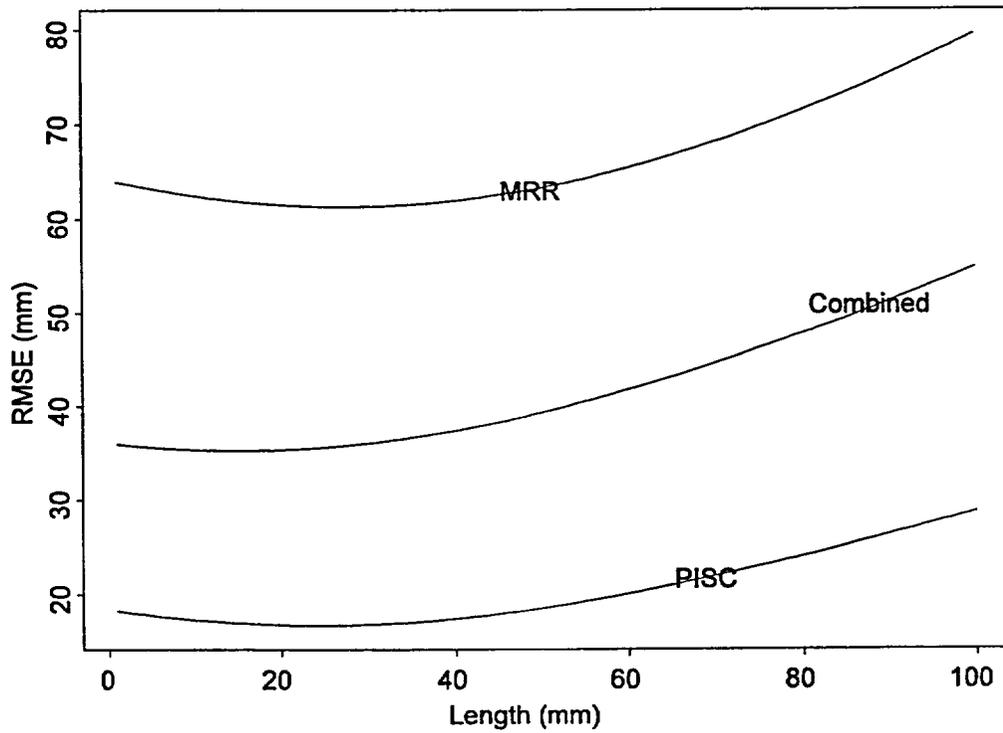
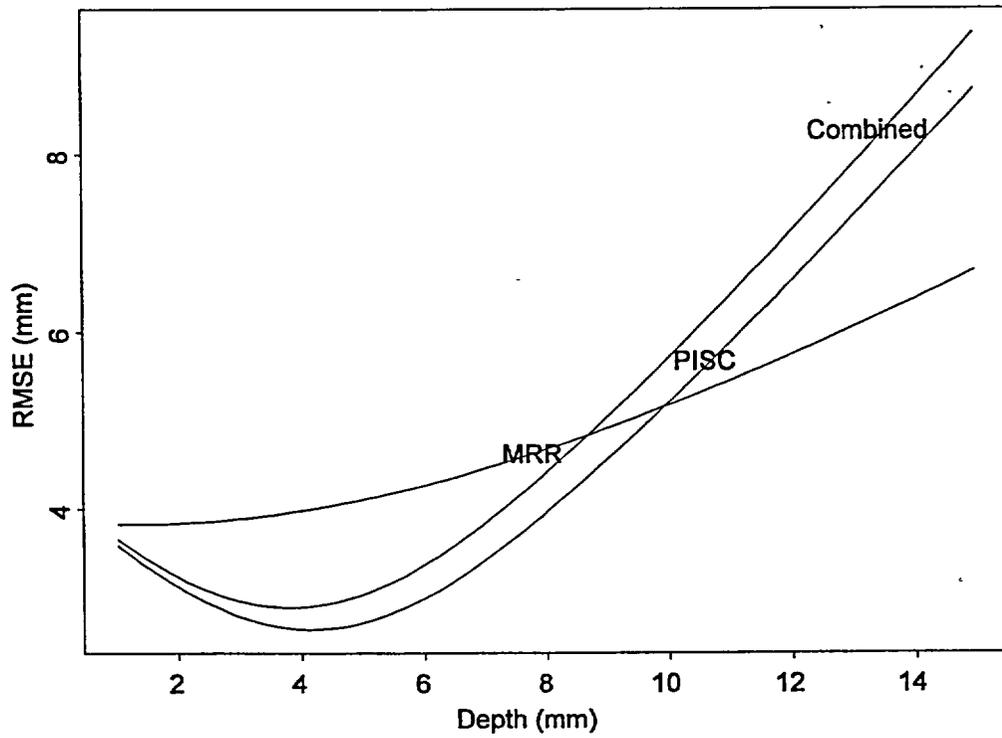
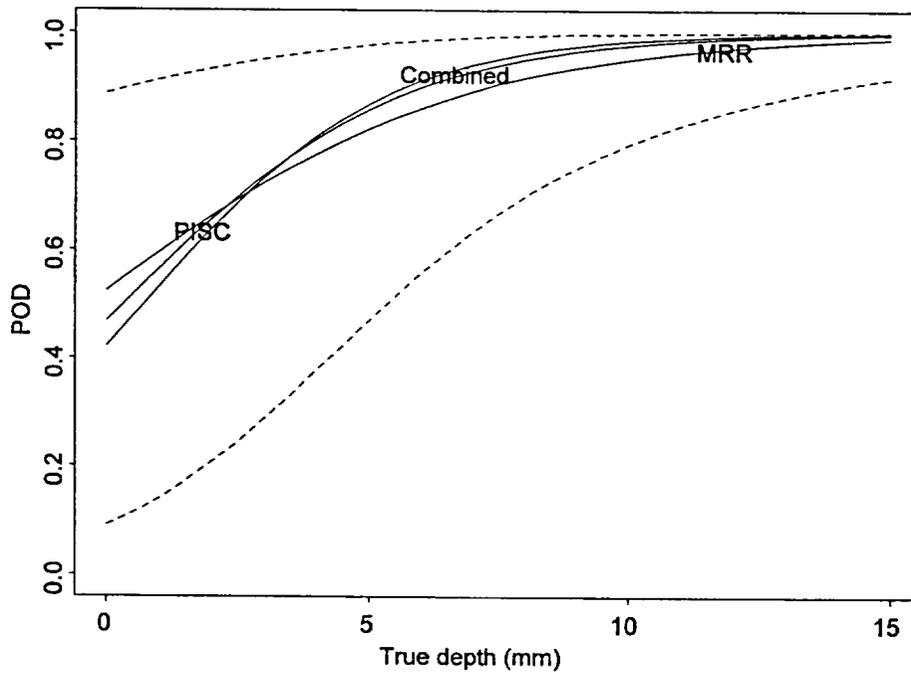


Figure 7.1 Combined MRR and PISC-AST Measurements

The combined POD curve is presented in Figure 7.2. As one can see, all three POD curves (the two original and the combined) are very close to each other. The confidence bound for the

combined result is somewhat larger than the PISC-AST confidence bound, because it includes the high team/team variability found in the MRR.



**Figure 7.2** Combined MRR and PISC-AST POD with 95% Bounds

## 8 COMPARISON OF ANOVA RESULTS TO PUBLISHED RESULTS

All three round robins have been analyzed in previous reports, and it is natural to ask how different the results presented here are from previous results. Differences could occur for two principal reasons. First of all, this analysis uses different subsets of the data from the round robins than analyzed in previous reports. Secondly, we utilize the regression model described in Section 3 of this report, which produces a more comprehensive summarization of the round robin than typically given in earlier reports.

Nevertheless, previous results should approximately correspond to the results presented in this report. This section compares the present results with those obtained from reports [1] and [4], which present earlier evaluations for the PIRR and PISC-AST round robins.

An earlier report on the MRR [3] concentrates on an ROC analysis of the data, which is not comparable to the analysis presented here, and therefore is not discussed further.

### 8.1 PIRR Evaluations

In the analysis presented for the PIRR in the previous report, [4], detection and sizing results are evaluated by individually regressing each team's results against the true sizes. The results are presented in Table 7.1, Table 5.4, and Table 5.5 of that report. In Table 8.1, we have summarized the results from those tables in terms that are comparable to the analysis in this report. The original results treated IGSCC and TFC separately, and we have averaged them together to produce the values presented in Table 8.1.

From the table, one can see that the sizing measurement errors  $\sigma_E$  are quite comparable in the studies, and the average regression lines (as described by  $\beta_1$  and  $\beta_2$ ) are also comparable. The team-to-team variabilities are also comparable, except for  $\gamma_{1,1}$  from the PIRR length sizing fit, which is inflated by length outliers.

It is interesting that the RMSE values are roughly comparable, because the RMSE calculated in the current study includes team/team variability, a source of variability not included in the previous calculations.

The POD results in this analysis do differ in a systematic way from the POD curves presented in the previous report. The previous PIRR analysis included false call data in the POD regressions; consequently, the previous POD curves are much steeper.

### 8.2 PISC-AST Evaluations

Table 8.2 compares results from the PISC-AST report [1] to the analysis presented in this report. The values presented in this table under the PISC-AST report column have been extracted from figures, and are therefore less accurate. As one can see, the sizing results from the two analyses are generally comparable.

The PISC-AST analysis did not attempt to fit curves to detection data; the POD values presented in the table are summarizations of detection scatter plots.

| <b>Table 8.1 Comparison of Present Results with PIRR Report</b> |                        |             |                    |             |
|---|------------------------|-------------|--------------------|-------------|
| <b>Parameter</b>  | <b>PIRR Report [4]</b> |             | <b>This Report</b> |             |
|   | <b>Range</b>           | <b>Mean</b> | <b>Range</b>       | <b>Mean</b> |
| <b>Depth Sizing Results</b>                                     |                        |             |                    |             |
| $\sigma_E$ mm   | 1.0 to 4.1             | 1.7         |                    | 1.4         |
| $\beta_1$ mm  | 1.5 to 4.3             | 2.6         | 1.9 to 3.6         | 2.9         |
| $\beta_2$ mm  | -0.14 to 2.71          | 0.10        | -0.15 to 0.18      | 0.07        |
| $\gamma_{1,1}$  |                        | 1.38        |                    | 0.53        |
| $\gamma_{2,2}$  |                        | 0.04        |                    | 0.06        |
| RMSE mm   | 1.3 to 5.6             | 2.8         | 2.6 to 4           | 3.3         |
| <b>Length Sizing Results</b>                                    |                        |             |                    |             |
| $\sigma_E$ mm   | 6.1 to 48              |             | 8.1                |             |
| $\beta_1$ mm  | 3.3 to 64.8            | 22          | 14 to 16           | 15.21       |
| $\beta_2$   | -0.27 to 0.80          | 0.22        | 0.17 to 0.78       | 0.28        |
| $\gamma_{1,1}$  |                        | 358         |                    | 2.05        |
| $\gamma_{2,2}$  |                        | 0.12        |                    | 0.06        |
| RMSE mm   | 10 to 56               | 23          | 16 to 34           | 25          |
| <b>POD Results</b>  |                        |             |                    |             |
| $\beta_1$   | -2.50 to -2.67         | 2.59        | -3.37 to -0.52     | -1.54       |
| $\beta_2$   | 0.52 to 1.67           | 1.1         | 0.4 to 0.51        | 0.45        |
| $\gamma_{1,1}$  |                        | 0.38        |                    | 1.08        |
| $\gamma_{2,2}$  |                        | 0.12        |                    | 0.01        |
| $\sigma_E$  |                        | NA          |                    | 1.72        |

| <b>Table 8.2 Comparison of Present Results with PISC-AST Report No. 33</b> |                            |                 |                    |                      |
|--|----------------------------|-----------------|--------------------|----------------------|
| <b>Parameter</b>   | <b>PISC-AST Report [1]</b> |                 | <b>This Report</b> |                      |
|  | <b>Team Range</b>          | <b>Estimate</b> | <b>Team Range</b>  | <b>Best Estimate</b> |
| <b>Depth Sizing Results</b>  |                            |                 |                    |                      |
| $\sigma_E$ mm  | 2 to 6                     | 3               |                    | 2                    |
| $\beta_1$ mm   | 0 to 8                     | 2               | 0 to 6.4           | 3                    |
| $\beta_2$  | 0.1 to 0.9                 | 0.3             | -0.1 to 0.8        | 0.28                 |
| RMSE mm  | 2 to 6                     | 3.3             | 2.7 to 8.7         | 5                    |
| <b>Length Sizing Results</b>   |                            |                 |                    |                      |
| $\sigma_E$   | 15 to 85                   | 20              |                    | 15                   |
| $\beta_1$ mm   | -30 to 100                 | 20              | 6 to 11            | 8                    |
| $\beta_2$  | 0.2 to 1.2                 | 0.70            | 0.50 to 0.92       | 0.71                 |
| RMSE mm  | 10 to 55                   | 30              | 18 to 30           | 25                   |
| <b>POD Results</b>   |                            |                 |                    |                      |
| POD(2 mm)  | 60 to 80%                  |                 | 64%                |                      |
| POD(5 mm)  |                            |                 | 87%                |                      |
| POD(8 mm)  | 95%                        |                 | 96%                |                      |
| POD(10 mm)   |                            |                 | 98%                |                      |

## 9 CONCLUSIONS

The three round robins evaluated in this report present a consistent story for inspection capability. As one would expect, the earliest round robin (PIRR) produced the lowest detection and sizing scores, while the latest round robin (PISC-AST) generally produced the best. The middle round robin (the MRR) produced results that are generally intermediate between the other two, but the MRR results are more uncertain, because the flaw set used in the MRR is not as extensive as that of the other two round robins.

The average POD curves from all three round robins predict a 90% POD for finding a 10 mm deep flaw, and a 70% POD for finding a 5 mm flaw. Although one might want even higher POD performance than this to satisfy a specific safety criteria, there is no doubt that a POD this high is useful from a defense in depth perspective.

The length sizing capability exhibited in the PIRR and MRR round robins exhibited large errors. In the PIRR study, large flaws are severely undersized. In the MRR, the teams exhibit too much variability (however this may be due to the small flaw sample size used in this study). In the PISC-AST study, very consistent length measurements were made, and consequently these sizing results were the best of the three round robin studies.

It should be noted that one problem associated with length measurements was present in all three studies: gross length sizing errors occurred at a rate of approximately 1 gross error per 140 measurements. This is a significant statistic to keep in mind when evaluating inspection results, but remember that these are gross oversizes.

The inspectors in all three round robins produced the least satisfactory results for depth sizing. In fact, the PIRR depth sizing results show no significant correlation between true and measured depths. The MRR displays much better depth sizing results, indicating the positive benefit of

training and testing. The MRR depth sizing results are quite comparable to the PISC-AST results and in some ways exceed them. For example, the MRR teams generally exhibit a higher slope parameter in the sizing regressions (see Figure 4.1).

Both the MRR and PISC-AST sizing results show a capability to measure the depth of flaws, but one can argue that the depth sizing RMSE is too large. For example, the ASME Section XI, Appendix VIII performance demonstration criteria is for the RMSE to be less than 3.2 mm (0.125 in). Although the RMSE in the MRR and PISC-AST round robins is close to this value for some flaw sizes, it is generally about twice as large.

Thus, the ASME Code Appendix VIII Supplement 2 demonstration requirements for crack depth sizing demands better performance than the average skills measured in these studies.

To put this important depth sizing performance into perspective, one needs to understand the impact that a 3-mm RMSE has in structural integrity calculations. For example, if the depth sizing capability truly is at an RMSE of 3 mm for a pipe having a wall thickness of 12 mm, then the inspection should only be capable of categorizing flaws into two or three rough sizing categories. The studies cited in this report show that an RMSE of 3 mm is hard to achieve. The ASME Code Appendix VIII requirements for demonstrating an RMSE depth sizing performance of 3 mm and the performance data in the three round robins (PIRR, MRR, and PISC-AST) indicates that only the best performers will become qualified to inspect nuclear components. In order to increase the number of qualified depth sizing inspectors, there would need to be significant improvements in inspector skill levels, inspection equipment, and/or the procedures.

## 10 BIBLIOGRAPHY

- [1] Lemaitre P, "Report on the evaluation of the inspection results of the wrought-to-wrought PISC III Assemblies no. 31, 32, 33, 34, 35 and 36," PISC III Report no. 33, Programme for Inspection of Steel Components, Joint Research Centre EEC, 9/94.
- [2] Lemaitre P and E Borloo. 1994. "Non-Destructive Examination Practice and Results, State of the Art and PISC III Results," *Proceedings of the Joint CEC OECD IAEA Specialists Meeting held at Petten*, Joint Research Centre, Institute for Advanced Materials, Box 2, NL-1755 ZG Petten, The Netherlands, ISBN 92-826-8813-5.
- [3] Heasler PG, TT Taylor, JC Spanner, SR Doctor, and JD 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, PNL.
- [4] Heasler PG and SR Doctor. 1996. *Piping Inspection Round Robin*, NUREG/CR-5068, PNL.
- [5] Nichools, RW and S Crutzen (ed). 1988. *Ultrasonic Inspection of Heavy Section Steel Components, The PISC II Final Report*, Elsevier Science Publishing Co., New York.
- [6] Stiratelli R, N Laird, and J Ware. 1984. "Random Effects Models for Serial Observations with Binary Response," *Biometrics* 40:961-971.
- [7] Wong GY and WM Mason. 1985. "The Hierarchical Logistic Regression Model for Multilevel Analysis," *JASA*, 80(391).
- [8] Rubin DB and RK Tsutakawa. 1981. "Estimation in Covariance Components Models," *JASA*, 75(374):341.
- [9] Corbeil RR and SR Searle. 1976. "Restricted Maximum Likelihood (REML) Estimation of Variance Components in the Mixed Model," *Technometrics*, 18(1):31.
- [10] Commission of the European Communities. 1979. *Report on Plate Inspection Steering Committee (PISCI)*, EUR 6371 EN Reports (Vols I-V), Brussels, Luxembourg.

## **APPENDIX A**

### **ROUND ROBIN DATA USED IN THIS STUDY**

## APPENDIX A

### ROUND ROBIN DATA USED IN THIS STUDY

This appendix contains the detection, depth sizing, and length sizing results used in this study. The results are presented as a flaw by team table. On the left side of the table, information about the flaw is given, while the main body of the table contains the inspection results. For detection, the inspection results are shown as a 1 (flaw detected), a 0 (flaw not detected) or a fraction, indicating that the flaw was inspected more than once.

For depth and length sizing, the main body of the table contains the team's measurements which can be compared against the true flaw dimensions listed on the left side of the table. Many measurements are frequently "not available," which can occur when a flaw is not inspected or when it is missed.

Each flaw is described by three rows in the tables. The first row describes detection statistics. In this row, under flaw characteristics, the flaw type is listed, and under truth, the flaw ID is listed. The second and third rows list depth and length sizing information, respectively. The true flaw sizes are listed under the column labeled "truth." All flaw depths and lengths in these tables are in mm.

#### A.1 PIRR Data

|                      |       | teams |     |      |      |     |      |
|----------------------|-------|-------|-----|------|------|-----|------|
| fl.char <sup>1</sup> | truth | 4     | 7   | 8    | 9    | 10  | 11   |
| TFC                  | 6.0   | 0.0   | 0.0 | 0.0  | 0.2  | 0.0 | 0.0  |
| depth                | 0.3   | NA    | NA  | NA   | 6.2  | NA  | NA   |
| length               | 7.1   | NA    | NA  | NA   | 20.3 | NA  | NA   |
| TFC                  | 9.0   | 0.0   | 0.3 | 0.0  | 0.0  | 0.0 | 0.0  |
| depth                | 1.3   | NA    | 3.1 | NA   | NA   | NA  | NA   |
| length               | 24.1  | NA    | 7.6 | NA   | NA   | NA  | NA   |
| TFC                  | 10.0  | 0.0   | 0.0 | 0.0  | 0.5  | 0.0 | 0.0  |
| depth                | 1.0   | NA    | NA  | NA   | 5.0  | NA  | NA   |
| length               | 33.0  | NA    | NA  | NA   | 58.4 | NA  | NA   |
| TFC                  | 19.0  | NA    | NA  | 0.0  | NA   | NA  | NA   |
| depth                | 0.6   | NA    | NA  | NA   | NA   | NA  | NA   |
| length               | 48.3  | NA    | NA  | NA   | NA   | NA  | NA   |
| TFC                  | 20.0  | 0.0   | 0.0 | 0.3  | 1.0  | 0.0 | 0.0  |
| depth                | 4.0   | NA    | NA  | 7.7  | 2.5  | NA  | NA   |
| length               | 35.6  | NA    | NA  | 38.1 | 11.4 | NA  | NA   |
| TFC                  | 23.0  | 0.0   | 0.0 | 0.0  | 0.0  | 0.0 | 0.0  |
| depth                | 4.0   | NA    | NA  | NA   | NA   | NA  | NA   |
| length               | 25.9  | NA    | NA  | NA   | NA   | NA  | NA   |
| TFC                  | 25.0  | 0.0   | 0.2 | 0.0  | 0.2  | 0.0 | 0.4  |
| depth                | 1.5   | NA    | 1.9 | NA   | 3.1  | NA  | 2.5  |
| length               | 24.1  | NA    | 7.6 | NA   | 21.6 | NA  | 19.5 |

|                      |       | teams |      |      |      |      |      |
|----------------------|-------|-------|------|------|------|------|------|
| fl.char <sup>1</sup> | truth | 4     | 7    | 8    | 9    | 10   | 11   |
| TFC                  | 26.0  | 0.0   | 0.0  | 0.0  | 0.8  | 0.3  | 0.0  |
| depth                | 0.9   | NA    | NA   | NA   | 1.0  | 2.5  | NA   |
| length               | 25.4  | NA    | NA   | NA   | 11.4 | 10.2 | NA   |
| TFC                  | 33.0  | 0.0   | 1.0  | 0.0  | 1.0  | 1.0  | 1.0  |
| depth                | 6.1   | NA    | 3.1  | NA   | 2.5  | 2.2  | 3.9  |
| length               | 59.2  | NA    | 17.8 | NA   | 12.7 | 21.6 | 17.5 |
| TFC                  | 37.0  | 1.0   | 1.0  | 0.0  | 1.0  | 1.0  | 1.0  |
| depth                | 5.8   | 3.1   | 2.0  | NA   | 5.6  | 3.1  | 5.5  |
| length               | 29.5  | 26.7  | 38.1 | NA   | 20.3 | 22.9 | 22.1 |
| TFC                  | 39.0  | 0.7   | 1.0  | 0.3  | 0.7  | 1.0  | 1.0  |
| depth                | 6.6   | 2.7   | 2.5  | 6.9  | 2.4  | 2.0  | 4.0  |
| length               | 31.7  | 19.0  | 12.9 | 20.3 | 17.5 | 15.0 | 24.1 |
| TFC                  | 45.0  | 1.0   | 1.0  | 1.0  | 1.0  | 1.0  | 1.0  |
| depth                | 0.6   | 1.7   | 2.2  | 4.1  | 3.1  | 2.9  | 3.1  |
| length               | 6.9   | 11.4  | 12.7 | 57.1 | 22.9 | 15.2 | 16.5 |
| TFC                  | 46.0  | 0.7   | 1.0  | 0.0  | 1.0  | 1.0  | 0.0  |
| depth                | 2.5   | 0.9   | 2.6  | NA   | 3.9  | 2.4  | NA   |
| length               | 8.6   | 16.5  | 10.2 | NA   | 43.2 | 19.5 | NA   |
| TFC                  | 54.0  | 0.0   | 0.3  | 0.0  | 0.7  | 0.3  | 0.0  |
| depth                | 6.8   | NA    | 0.6  | NA   | 3.1  | 3.1  | NA   |
| length               | 14.5  | NA    | 7.6  | NA   | 24.8 | 25.4 | NA   |
| TFC                  | 55.0  | 1.0   | 1.0  | 0.5  | 0.0  | 0.5  | 1.0  |
| depth                | 1.6   | 1.6   | 2.2  | 5.0  | NA   | 3.1  | 2.5  |
| length               | 54.9  | 15.2  | 15.2 | 63.5 | NA   | 21.6 | 17.5 |
| TFC                  | 56.0  | 0.7   | 1.0  | 0.0  | 1.0  | 1.0  | 0.3  |
| depth                | 5.8   | 1.9   | 1.9  | NA   | 2.7  | 2.7  | 2.7  |
| length               | 55.9  | 17.1  | 21.2 | NA   | 19.1 | 20.3 | 22.1 |
| TFC                  | 60.0  | 1.0   | 1.0  | 0.2  | 1.0  | 1.0  | 0.7  |
| depth                | 4.3   | 2.5   | 1.9  | 1.2  | 2.5  | 3.7  | 3.4  |
| length               | 28.2  | 20.3  | 16.1 | 38.1 | 14.4 | 11.9 | 20.7 |
| TFC                  | 62.0  | 0.0   | 0.0  | 0.0  | 0.0  | 0.3  | 0.0  |
| depth                | 1.1   | NA    | NA   | NA   | NA   | 5.0  | NA   |
| length               | 24.4  | NA    | NA   | NA   | NA   | 10.2 | NA   |
| TFC                  | 65.0  | 0.0   | 1.0  | 1.0  | 0.0  | 1.0  | 1.0  |
| depth                | 6.6   | NA    | 2.5  | 5.5  | NA   | 2.7  | 1.2  |
| length               | 27.4  | NA    | 33.0 | 31.7 | NA   | 25.4 | 28.4 |
| TFC                  | 66.0  | NA    | NA   | 0.0  | NA   | NA   | NA   |
| depth                | 1.8   | NA    | NA   | NA   | NA   | NA   | NA   |
| length               | 23.1  | NA    | NA   | NA   | NA   | NA   | NA   |
| TFC                  | 73.0  | 0.0   | 0.0  | 0.0  | 0.0  | 0.0  | 0.5  |
| depth                | 0.8   | NA    | NA   | NA   | NA   | NA   | 1.7  |
| length               | 29.2  | NA    | NA   | NA   | NA   | NA   | 14.2 |
| TFC                  | 79.0  | NA    | NA   | 0.0  | NA   | NA   | NA   |
| depth                | 1.3   | NA    | NA   | NA   | NA   | NA   | NA   |
| length               | 45.5  | NA    | NA   | NA   | NA   | NA   | NA   |

|                      |       | teams |      |       |      |      |      |
|----------------------|-------|-------|------|-------|------|------|------|
| fl.char <sup>1</sup> | truth | 4     | 7    | 8     | 9    | 10   | 11   |
| TFC                  | 80.0  | 0.0   | 0.0  | 0.0   | 0.0  | 0.0  | 0.0  |
| depth                | 1.1   | NA    | NA   | NA    | NA   | NA   | NA   |
| length               | 22.4  | NA    | NA   | NA    | NA   | NA   | NA   |
| TFC                  | 83.0  | 0.0   | 0.0  | 0.0   | 0.0  | 0.0  | 0.0  |
| depth                | 2.5   | NA    | NA   | NA    | NA   | NA   | NA   |
| length               | 35.6  | NA    | NA   | NA    | NA   | NA   | NA   |
| IGSCC                | 99.0  | 0.3   | 1.0  | 0.0   | 1.0  | 1.0  | 1.0  |
| depth                | 2.2   | 1.7   | 3.0  | NA    | 3.0  | 3.0  | 2.8  |
| length               | 22.6  | 20.3  | 14.1 | NA    | 27.5 | 17.8 | 19.1 |
| IGSCC                | 100.0 | 0.0   | 0.0  | 0.0   | 0.0  | 0.0  | 0.0  |
| depth                | 0.6   | NA    | NA   | NA    | NA   | NA   | NA   |
| length               | 2.5   | NA    | NA   | NA    | NA   | NA   | NA   |
| IGSCC                | 101.0 | 0.0   | 0.3  | 0.0   | 0.7  | 0.0  | 0.7  |
| depth                | 1.8   | NA    | 3.5  | NA    | 2.1  | NA   | 1.0  |
| length               | 27.9  | NA    | 12.7 | NA    | 22.9 | NA   | 3.2  |
| IGSCC                | 105.0 | 0.8   | 1.0  | 0.4   | 1.0  | 0.8  | 0.6  |
| depth                | 1.0   | 2.3   | 2.3  | 1.5   | 5.6  | 3.8  | 3.6  |
| length               | 50.3  | 43.5  | 49.2 | 88.9  | 51.1 | 34.9 | 53.4 |
| IGSCC                | 109.0 | 0.3   | 0.2  | 0.0   | 0.0  | 0.0  | 0.0  |
| depth                | 0.5   | 0.7   | 4.2  | NA    | NA   | NA   | NA   |
| length               | 3.0   | 12.7  | 48.3 | NA    | NA   | NA   | NA   |
| IGSCC                | 110.0 | 0.0   | 0.5  | 0.0   | 1.0  | 0.0  | 0.0  |
| depth                | 0.6   | NA    | 2.8  | NA    | 2.4  | NA   | NA   |
| length               | 33.5  | NA    | 17.8 | NA    | 7.6  | NA   | NA   |
| IGSCC                | 194.0 | 0.0   | 0.0  | 0.0   | 0.0  | 0.0  | 0.0  |
| depth                | 0.2   | NA    | NA   | NA    | NA   | NA   | NA   |
| length               | 6.3   | NA    | NA   | NA    | NA   | NA   | NA   |
| IGSCC                | 200.0 | 0.5   | 1.0  | 0.3   | 1.0  | 0.5  | 1.0  |
| depth                | 2.1   | 4.2   | 2.2  | 4.2   | 2.8  | 4.6  | 2.7  |
| length               | 28.4  | 40.6  | 34.9 | 101.6 | 34.3 | 30.5 | 30.9 |
| IGSCC                | 202.0 | 1.0   | 1.0  | 0.5   | 1.0  | 1.0  | 0.8  |
| depth                | 2.4   | 3.1   | 3.8  | 12.6  | 3.7  | 3.4  | 3.8  |
| length               | 22.9  | 23.2  | 16.2 | 35.6  | 23.5 | 22.5 | 27.4 |
| IGSCC                | 209.0 | 0.0   | 0.2  | 0.0   | 0.5  | 0.0  | 0.0  |
| depth                | 0.7   | NA    | 1.0  | NA    | 1.4  | NA   | NA   |
| length               | 5.1   | NA    | 7.6  | NA    | 5.1  | NA   | NA   |
| IGSCC                | 210.0 | 0.0   | 0.0  | 0.0   | 0.0  | 0.0  | 0.0  |
| depth                | 0.5   | NA    | NA   | NA    | NA   | NA   | NA   |
| length               | 3.3   | NA    | NA   | NA    | NA   | NA   | NA   |
| IGSCC                | 212.0 | 1.0   | 0.7  | 0.0   | 1.0  | 0.5  | 1.0  |
| depth                | 3.1   | 2.4   | 2.8  | NA    | 3.1  | 3.6  | 2.8  |
| length               | 17.5  | 22.9  | 12.7 | NA    | 19.7 | 17.8 | 20.7 |
| IGSCC                | 214.0 | 1.0   | 1.0  | 0.5   | 0.5  | 1.0  | 1.0  |
| depth                | 3.3   | 1.6   | 2.8  | 8.7   | 3.5  | 2.8  | 3.6  |
| length               | 26.4  | 34.3  | 26.7 | 30.5  | 25.4 | 29.2 | 41.4 |

|                      |       | teams |      |      |      |      |      |
|----------------------|-------|-------|------|------|------|------|------|
| fl.char <sup>1</sup> | truth | 4     | 7    | 8    | 9    | 10   | 11   |
| IGSCC                | 216.0 | 1.0   | 1.0  | 0.0  | 1.0  | 1.0  | 1.0  |
| depth                | 2.2   | 2.8   | 1.6  | NA   | 2.8  | 2.1  | 2.7  |
| length               | 9.9   | 15.2  | 14.0 | NA   | 30.5 | 12.7 | 20.6 |
| IGSCC                | 220.0 | 1.0   | NA   | NA   | 1.0  | 1.0  | 1.0  |
| depth                | 3.0   | 1.4   | NA   | NA   | 3.5  | 2.1  | 4.5  |
| length               | 11.4  | 12.7  | NA   | NA   | 12.7 | 21.6 | 25.4 |
| IGSCC                | 231.0 | 1.0   | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
| depth                | 1.5   | 1.4   | NA   | NA   | NA   | NA   | NA   |
| length               | 6.6   | 11.4  | NA   | NA   | NA   | NA   | NA   |
| IGSCC                | 239.0 | 0.0   | 1.0  | 0.0  | 0.5  | 0.5  | 0.0  |
| depth                | 1.0   | NA    | 1.4  | NA   | 3.5  | 1.4  | NA   |
| length               | 17.0  | NA    | 11.4 | NA   | 12.7 | 10.2 | NA   |
| IGSCC                | 240.0 | 1.0   | 0.5  | 0.5  | 1.0  | 0.5  | 0.5  |
| depth                | 2.2   | 1.0   | 2.2  | 1.5  | 3.1  | 5.6  | 1.4  |
| length               | 9.7   | 17.8  | 12.7 | 25.4 | 96.5 | 5.1  | 7.9  |
| IGSCC                | 242.0 | 1.0   | 1.0  | 0.0  | 0.7  | 0.7  | 0.7  |
| depth                | 2.5   | 2.3   | 3.0  | NA   | 5.2  | 4.2  | 3.5  |
| length               | 26.4  | 19.5  | 21.6 | NA   | 29.2 | 18.4 | 20.7 |
| IGSCC                | 243.0 | 0.5   | 0.5  | 0.0  | 0.5  | 1.0  | 0.0  |
| depth                | 0.8   | 2.8   | 2.2  | NA   | 1.4  | 5.6  | NA   |
| length               | 14.0  | 7.6   | 7.6  | NA   | 10.2 | 12.7 | NA   |
| IGSCC                | 244.0 | 1.0   | 1.0  | 0.3  | 1.0  | 0.7  | 1.0  |
| depth                | 1.3   | 1.9   | 3.2  | 2.0  | 4.2  | 4.5  | 2.7  |
| length               | 22.9  | 22.9  | 24.0 | 40.6 | 19.5 | 16.5 | 15.9 |

<sup>1</sup>flaw characteristic

## A.2 MRR Data

### A.2.1 MRR POD and Length Data

|         |       | teams |     |      |      |      |      |       |
|---------|-------|-------|-----|------|------|------|------|-------|
| fl.char | truth | 1     | 2   | 3    | 5    | 6    | 16   | 8     |
| IGSCC   | 99.0  | 1.0   | 0.0 | 1.0  | 0.0  | 1.0  | 1.0  | 1.0   |
| length  | 22.6  | 25.4  | NA  | 12.7 | NA   | 25.4 | 38.1 | 210.8 |
| IGSCC   | 101.0 | 1.0   | 0.0 | 1.0  | 0.0  | 0.0  | 1.0  | 1.0   |
| length  | 27.9  | 48.3  | NA  | 9.7  | NA   | NA   | 38.1 | 210.8 |
| IGSCC   | 202.0 | 1.0   | 0.0 | 1.0  | 1.0  | 1.0  | 1.0  | 0.0   |
| length  | 22.9  | 34.3  | NA  | 22.2 | 36.6 | 31.6 | 25.4 | NA    |
| IGSCC   | 209.0 | 0.0   | 0.0 | 0.0  | 0.0  | 0.0  | 1.0  | 1.0   |
| length  | 5.1   | NA    | NA  | NA   | NA   | NA   | 25.4 | 0.0   |
| IGSCC   | 210.0 | 1.0   | 0.0 | 0.0  | 0.0  | 0.0  | 0.0  | 0.0   |
| length  | 3.3   | 71.1  | NA  | NA   | NA   | NA   | NA   | NA    |
| IGSCC   | 212.0 | 1.0   | 1.0 | 1.0  | 1.0  | 1.0  | 1.0  | 0.0   |

|         |       | teams |       |       |       |       |       |       |
|---------|-------|-------|-------|-------|-------|-------|-------|-------|
| fl.char | truth | 1     | 2     | 3     | 5     | 6     | 16    | 8     |
| length  | 17.5  | 78.7  | 12.7  | 19.0  | 25.4  | 22.4  | 63.5  | NA    |
| IGSCC   | 214.0 | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   | 0.0   |
| length  | 26.4  | 38.1  | 28.4  | 22.4  | 55.9  | 38.1  | 38.1  | NA    |
| IGSCC   | 216.0 | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   | 0.0   |
| length  | 9.9   | 27.9  | 34.8  | 15.7  | 12.7  | 22.1  | 25.4  | NA    |
| IGSCC   | 231.0 | 0.0   | NA    | 1.0   | 0.0   | 1.0   | 0.0   | 0.0   |
| length  | 6.6   | NA    | NA    | 9.4   | NA    | 41.1  | NA    | NA    |
| IGSCC   | 239.0 | 1.0   | 0.0   | 1.0   | 0.0   | 1.0   | 1.0   | 0.0   |
| length  | 17.0  | 25.4  | NA    | 9.7   | NA    | 19.0  | 25.4  | NA    |
| IGSCC   | 240.0 | 1.0   | NA    | 1.0   | 0.0   | 1.0   | 1.0   | 0.0   |
| length  | 9.7   | 22.9  | NA    | 213.4 | NA    | 38.1  | 215.9 | NA    |
| IGSCC   | 274.0 | 1.0   | 0.0   | 0.0   | 1.0   | 1.0   | 1.0   | 1.0   |
| length  | 21.6  | 254.0 | NA    | NA    | 254.0 | 254.0 | 108.0 | 254.0 |
| IGSCC   | 305.0 | 0.0   | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   |
| length  | 130.8 | NA    | 254.0 | 44.4  | 40.6  | 177.8 | 63.5  | 254.0 |

|         |       | teams |      |       |       |      |      |       |
|---------|-------|-------|------|-------|-------|------|------|-------|
| fl.char | truth | 9     | 10   | 11    | 12    | 13   | 14   | 15    |
| IGSCC   | 99.0  | 1.0   | 1.0  | 1.0   | 1.0   | 1.0  | 0.0  | 1.0   |
| length  | 22.6  | 130.6 | 40.6 | 12.7  | 15.7  | 19.0 | NA   | 20.3  |
| IGSCC   | 101.0 | 1.0   | 0.0  | 0.0   | 1.0   | 0.0  | 0.0  | 1.0   |
| length  | 27.9  | 130.6 | NA   | NA    | 38.1  | NA   | NA   | -19.0 |
| IGSCC   | 202.0 | 1.0   | 1.0  | 0.5   | 1.0   | 0.5  | 1.0  | 1.0   |
| length  | 22.9  | 77.5  | 33.0 | 28.7  | 30.1  | 33.0 | 19.7 | 20.3  |
| IGSCC   | 209.0 | 0.0   | 0.0  | NA    | 1.0   | 1.0  | 1.0  | 1.0   |
| length  | 5.1   | NA    | NA   | NA    | 15.7  | 50.8 | 11.4 | 6.4   |
| IGSCC   | 210.0 | 0.0   | NA   | 0.0   | 0.0   | 0.0  | 1.0  | 1.0   |
| length  | 3.3   | NA    | NA   | NA    | NA    | NA   | 12.7 | 17.8  |
| IGSCC   | 212.0 | 1.0   | NA   | 0.0   | 1.0   | 1.0  | 1.0  | 1.0   |
| length  | 17.5  | 66.5  | NA   | NA    | 22.1  | 22.9 | 16.5 | 14.0  |
| IGSCC   | 214.0 | 1.0   | 1.0  | 1.0   | 1.0   | 1.0  | 1.0  | 1.0   |
| length  | 26.4  | 22.9  | 25.4 | 31.8  | 41.4  | 38.1 | 29.2 | 203.2 |
| IGSCC   | 216.0 | 1.0   | 1.0  | 1.0   | 1.0   | 1.0  | 1.0  | 1.0   |
| length  | 9.9   | 142.2 | 15.2 | 101.6 | 101.6 | 12.7 | 10.2 | 11.9  |
| IGSCC   | 231.0 | 0.0   | 0.0  | 0.0   | 0.0   | 1.0  | 0.0  | 1.0   |
| length  | 6.6   | NA    | NA   | NA    | NA    | 43.2 | NA   | 0.0   |
| IGSCC   | 239.0 | 0.0   | 1.0  | NA    | 1.0   | 1.0  | 1.0  | 1.0   |
| length  | 17.0  | NA    | 11.4 | NA    | 152.4 | 17.8 | 16.5 | 8.9   |
| IGSCC   | 240.0 | 0.0   | 1.0  | 0.5   | 1.0   | 1.0  | 1.0  | 1.0   |
| length  | 9.7   | NA    | 69.8 | 19.0  | 139.7 | 73.7 | 48.3 | 7.6   |
| IGSCC   | 274.0 | 1.0   | NA   | 0.0   | 0.0   | 1.0  | 0.0  | 1.0   |
| length  | 21.6  | 246.4 | NA   | NA    | NA    | 45.7 | NA   | 21.6  |
| IGSCC   | 305.0 | NA    | NA   | 0.0   | 1.0   | 1.0  | 1.0  | 1.0   |
| length  | 130.8 | NA    | NA   | NA    | 174.8 | 45.7 | 25.4 | 151.1 |

### A.2.2 MRR Depth Data

| fl.id <sup>1</sup> | fl.type <sup>2</sup> | fl.depth <sup>3</sup><br>(mm) | teams |      |      |     |      |      |      |      |
|--------------------|----------------------|-------------------------------|-------|------|------|-----|------|------|------|------|
|                    |                      |                               | 1     | 2    | 5    | 8   | 9    | 10   | 11   | 14   |
| 102                | IGSCC                | 0.8                           | 8.9   | 4.3  | 2.8  | 4.1 | 10.5 | 2.5  | 3.4  | 5.0  |
| 148                | IGSCC                | 2.0                           | 2.2   | 6.3  | 0.7  | 1.5 | 0.7  | 1.5  | 5.3  | 4.3  |
| 149                | IGSCC                | 3.6                           | 6.4   | 3.5  | 3.2  | 1.4 | 10.5 | 1.9  | 5.3  | 2.8  |
| 154                | IGSCC                | 3.2                           | 3.4   | 2.5  | 0.7  | 0.0 | 1.4  | 4.2  | 8.4  | 4.9  |
| 156                | IGSCC                | 5.3                           | 1.0   | 4.1  | 3.8  | 3.2 | 11.2 | 2.2  | 5.3  | 2.9  |
| 160                | IGSCC                | 5.6                           | 14.3  | 3.1  | 6.1  | 3.7 | 12.7 | 0.8  | 4.2  | 3.9  |
| 161                | IGSCC                | 4.2                           | 6.8   | 3.9  | 5.4  | 1.7 | 8.5  | 0.8  | 6.4  | 4.4  |
| 162                | IGSCC                | 6.2                           | 16.5  | 8.8  | 9.7  | 1.8 | 16.5 | 15.4 | 12.8 | 5.1  |
| 163                | IGSCC                | 11.4                          | 17.2  | 12.1 | 12.3 | 2.9 | 16.5 | 11.0 | 15.0 | 15.2 |
| 1003               | IGSCC                | 6.3                           | 2.5   | 1.4  | 9.5  | 1.9 | 8.5  | 1.2  | 2.9  | 2.7  |

<sup>1</sup> flaw identification  
<sup>2</sup> flaw type  
<sup>3</sup> flaw depth

### A.3 PISC-AST Data

| fl.char | truth | teams |    |     |     |     |     |     |     |     |    |    |     |
|---------|-------|-------|----|-----|-----|-----|-----|-----|-----|-----|----|----|-----|
|         |       | DH    | EI | FJ  | GK  | HL  | JN  | KM  | LP  | MK  | NJ | NR | OI  |
| IGSCC   | 1     | 1     | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 0  | 1  |     |
| depth   | 8     | 9     | 3  | 7   | 10  | 2   | 6   | 5   | 0   | 4   | 1  | NA | 4   |
| length  | 107   | 112   | 74 | 99  | 104 | 42  | 113 | 99  | 104 | 86  | 75 | NA | 100 |
| IGSCC   | 2     | 1     | 0  | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 1  | 0  | 1   |
| depth   | 3     | 2     | NA | NA  | NA  | NA  | NA  | 1   | NA  | NA  | 0  | NA | 3   |
| length  | 1     | 8     | NA | NA  | NA  | NA  | NA  | 0   | NA  | NA  | 10 | NA | 7   |
| IGSCC   | 3     | 1     | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1  | 1  | 1   |
| depth   | 9     | 10    | 3  | 8   | 10  | 4   | 5   | 6   | 0   | 6   | 1  | 2  | 4   |
| length  | 108   | 117   | 65 | 101 | 96  | 107 | 108 | 111 | 112 | 111 | 57 | 47 | 109 |
| IGSCC   | 4     | 1     | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0  | 0  | 0   |
| depth   | 2     | 1     | NA | NA  | NA  | NA  | NA  | NA  | NA  | NA  | NA | NA | NA  |
| length  | 1     | 5     | NA | NA  | NA  | NA  | NA  | NA  | NA  | NA  | NA | NA | NA  |
| TFC     | 9     | 1     | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1  | 1  | 1   |
| Depth   | 8     | 2     | 6  | 8   | 3   | 4   | 5   | 8   | 0   | 6   | 3  | 6  | 5   |
| Length  | 33    | 41    | 30 | 29  | 38  | 26  | 31  | 40  | 33  | 34  | 33 | 61 | 32  |
| TFC     | 10    | 1     | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1  | 1  | 1   |
| Depth   | 14    | 16    | 8  | 16  | 7   | 3   | 4   | 1   | 0   | 7   | 7  | 6  | 16  |
| Length  | 56    | 62    | 38 | 43  | 46  | 22  | 44  | 44  | 41  | 45  | 45 | 53 | 52  |
| MFC     | 11    | 1     | NA | 1   | 1   | 1   | 1   | 1   | NA  | 1   | 1  | 0  | 1   |
| depth   | 7     | 7     | NA | 6   | 2   | 3   | 6   | 3   | NA  | 5   | 1  | NA | 4   |
| length  | 46    | 53    | NA | 33  | 55  | 45  | 49  | 56  | NA  | 27  | 62 | NA | 57  |

|         |       | teams |    |    |    |     |    |    |    |     |    |    |     |
|---------|-------|-------|----|----|----|-----|----|----|----|-----|----|----|-----|
| fl.char | truth | DH    | EI | FJ | GK | HL  | JN | KM | LP | MK  | NJ | NR | OI  |
| MFC     | 12    | 1     | NA | 1  | 1  | 1   | 1  | 1  | NA | 1   | 1  | NA | 1   |
| depth   | 6     | 8     | NA | 6  | 6  | 9   | 5  | 2  | NA | 8   | 1  | NA | 4   |
| length  | 42    | 60    | NA | 41 | 62 | 17  | 47 | 59 | NA | 52  | 57 | NA | 56  |
| EDM     | 20    | 1     | NA | 0  | 1  | 1   | 1  | 1  | NA | 1   | 0  | NA | 0   |
| depth   | 1     | 8     | NA | NA | 2  | 2   | 2  | 1  | NA | 6   | NA | NA | NA  |
| length  | 4     | 19    | NA | NA | 0  | 17  | 4  | 12 | NA | 5   | NA | NA | NA  |
| EDM     | 21    | 1     | NA | 1  | 1  | 1   | 1  | 1  | NA | 1   | 1  | NA | 1   |
| depth   | 2     | 2     | NA | 2  | 2  | 2   | 2  | 2  | NA | 5   | 0  | NA | 4   |
| length  | 9     | 24    | NA | 12 | 0  | 17  | 10 | 18 | NA | 23  | 17 | NA | 15  |
| EDM     | 22    | 1     | NA | 1  | 1  | 1   | 1  | 1  | NA | 1   | 1  | NA | 1   |
| depth   | 2     | 4     | NA | 2  | 2  | 2   | 3  | 3  | NA | 6   | 0  | NA | 4   |
| length  | 13    | 36    | NA | 16 | 0  | 12  | 16 | 24 | NA | 22  | 19 | NA | 19  |
| EDM     | 23    | 1     | NA | 1  | 1  | 0   | 1  | 1  | NA | 1   | 1  | NA | 1   |
| depth   | 4     | 4     | NA | 3  | 2  | NA  | 4  | 3  | NA | 6   | 0  | NA | 4   |
| length  | 18    | 41    | NA | 21 | 0  | NA  | 23 | 28 | NA | 28  | 27 | NA | 22  |
| IGSCC   | 33    | 1     | 1  | 1  | 1  | 0   | 1  | 1  | 0  | 1   | 1  | 1  | 1   |
| depth   | 9     | 8     | 5  | 6  | 3  | NA  | 5  | 5  | NA | 4   | 0  | 6  | 8   |
| length  | 79    | 103   | 72 | 50 | 55 | NA  | 76 | 40 | NA | 74  | 83 | 68 | 80  |
| IGSCC   | 34    | 1     | 1  | 1  | 0  | 1   | 1  | 1  | 0  | 1   | 1  | 0  | 1   |
| depth   | 2     | 2     | 4  | 3  | NA | 2   | 4  | 5  | NA | 5   | 0  | NA | 7   |
| length  | 81    | 85    | 59 | 75 | NA | 68  | 77 | 71 | NA | 76  | 33 | NA | 81  |
| IGSCC   | 35    | 1     | 1  | 1  | 1  | 1   | 1  | 1  | 1  | 1   | 1  | 1  | 1   |
| depth   | 7     | 6     | 5  | 3  | 3  | 2   | 4  | 3  | 0  | 5   | 0  | 3  | 3   |
| length  | 60    | 53    | 72 | 55 | 50 | 68  | 59 | 92 | 8  | 69  | 63 | 43 | 186 |
| IGSCC   | 36    | 1     | 1  | 1  | 0  | 1   | 1  | 1  | 0  | 1   | 1  | 0  | 1   |
| depth   | 2     | 2     | 6  | 3  | NA | 2   | 3  | 4  | NA | 6   | 0  | NA | 5   |
| length  | 86    | 90    | 63 | 70 | NA | 101 | 77 | 89 | NA | 118 | 75 | NA | 86  |
| IGSCC   | 37    | 1     | 1  | 0  | 0  | 1   | 0  | 1  | 0  | 0   | 1  | 0  | 1   |
| depth   | 2     | 2     | 4  | NA | NA | 4   | NA | 1  | NA | NA  | 2  | NA | 5   |
| length  | 78    | 86    | 32 | NA | NA | 32  | NA | 6  | NA | NA  | 84 | NA | 71  |
| IGSCC   | 38    | 1     | 1  | 1  | 0  | 1   | 1  | 1  | 0  | 1   | 1  | 0  | 0   |
| depth   | 2     | 2     | 2  | 0  | NA | 2   | 2  | 0  | NA | 7   | 2  | NA | NA  |
| length  | 60    | 79    | 46 | 40 | NA | 0   | 17 | 52 | NA | 50  | 35 | NA | NA  |
| IGSCC   | 39    | 1     | 1  | 0  | 0  | 1   | 0  | 1  | 0  | 1   | 1  | 0  | 1   |
| depth   | 1     | 1     | 2  | NA | NA | 4   | NA | 1  | NA | 8   | 0  | NA | 3   |
| length  | 66    | 59    | 9  | NA | NA | 20  | NA | 16 | NA | 39  | 7  | NA | 41  |
| IGSCC   | 40    | 1     | 1  | 0  | 0  | 1   | 0  | 1  | 1  | 1   | 0  | 1  | 0   |
| depth   | 3     | 1     | 2  | NA | NA | 2   | NA | 3  | 3  | 4   | NA | 2  | NA  |
| length  | 61    | 59    | 15 | NA | NA | 5   | NA | 20 | 74 | 44  | NA | 28 | NA  |
| MFC     | 45    | 1     | 1  | 1  | 1  | 1   | 1  | 1  | 1  | 1   | 0  | 1  | 1   |
| depth   | 3     | 6     | 7  | 6  | 21 | 3   | 5  | 2  | 0  | 6   | NA | 2  | 6   |
| length  | 41    | 52    | 35 | 37 | 64 | 45  | 39 | 47 | 40 | 56  | NA | 38 | 44  |
| EN      | 46    | 1     | 1  | 1  | 1  | 1   | 1  | 1  | 1  | 1   | 1  | 1  | 1   |
| depth   | 4     | 9     | 5  | 5  | 2  | 2   | 4  | 2  | 0  | 5   | 0  | 11 | 5   |
| length  | 30    | 50    | 29 | 40 | 45 | 31  | 27 | 32 | 31 | 36  | 43 | 66 | 30  |

|         |       | teams |    |    |    |    |    |    |    |    |    |    |    |
|---------|-------|-------|----|----|----|----|----|----|----|----|----|----|----|
| fl.char | truth | DH    | EI | FJ | GK | HL | JN | KM | LP | MK | NJ | NR | OI |
| EDM     | 47    | 1     | 1  | 1  | 0  | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 1  |
| depth   | 4     | 5     | 5  | 6  | NA | 2  | 4  | 0  | NA | NA | NA | NA | 6  |
| length  | 33    | 26    | 26 | 25 | NA | 43 | 30 | 8  | NA | NA | NA | NA | 35 |
| EN      | 48    | 1     | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  |
| depth   | 1     | 3     | 5  | 5  | 12 | 2  | 4  | 1  | 0  | 3  | 0  | 13 | 3  |
| length  | 20    | 38    | 14 | 21 | 31 | 24 | 21 | 23 | 24 | 27 | 26 | 31 | 23 |
| EDM     | 49    | 1     | 1  | 1  | 0  | 1  | 1  | 0  | 0  | 0  | 0  | 1  | 1  |
| depth   | 6     | 11    | 3  | 4  | NA | 4  | 4  | NA | NA | NA | NA | 4  | 5  |
| length  | 38    | 83    | 51 | 31 | NA | 33 | 32 | NA | NA | NA | NA | 26 | 44 |
| EDM     | 50    | 1     | 1  | 1  | 0  | 1  | 1  | 1  | 0  | 0  | 1  | 1  | 1  |
| depth   | 0     | 2     | 7  | 3  | NA | 2  | 2  | 2  | NA | NA | 0  | 8  | 2  |
| length  | 15    | 19    | 21 | 7  | NA | 12 | 10 | 14 | NA | NA | 12 | 19 | 14 |

|         |       | teams |     |    |     |     |     |     |     |     |    |     |  |
|---------|-------|-------|-----|----|-----|-----|-----|-----|-----|-----|----|-----|--|
| fl.char | truth | PH    | QG  | RF | SE  | UZ  | VZ  | WA  | WX  | XW  | YC | YY  |  |
| IGSCC   | 1     | 1     | 1   | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1  | 1   |  |
| depth   | 8     | 5     | 4   | 6  | 6   | 5   | 4   | 5   | 3   | 7   | 2  | 6   |  |
| length  | 107   | 66    | 112 | 57 | 70  | 67  | 112 | 68  | 112 | 69  | 81 | 62  |  |
| IGSCC   | 2     | 0     | 0   | 0  | 0   | 0   | 1   | 0   | 0   | 0   | 0  | 0   |  |
| depth   | 3     | NA    | NA  | NA | NA  | NA  | 1   | NA  | NA  | NA  | NA | NA  |  |
| length  | 1     | NA    | NA  | NA | NA  | NA  | 0   | NA  | NA  | NA  | NA | NA  |  |
| IGSCC   | 3     | 1     | 1   | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1  | 1   |  |
| depth   | 9     | 4     | 8   | 4  | 5   | 5   | 8   | 6   | 2   | 7   | 1  | 6   |  |
| length  | 108   | 61    | 65  | 47 | 103 | 113 | 108 | 109 | 112 | 117 | 97 | 81  |  |
| IGSCC   | 4     | 0     | 0   | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 0  | 0   |  |
| depth   | 2     | NA    | NA  | NA | NA  | NA  | NA  | NA  | NA  | NA  | NA | NA  |  |
| length  | 1     | NA    | NA  | NA | NA  | NA  | NA  | NA  | NA  | NA  | NA | NA  |  |
| TFC     | 9     | 1     | 1   | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1  | 1   |  |
| depth   | 8     | 4     | 7   | 5  | 5   | 7   | 4   | 6   | 5   | 4   | 5  | 8   |  |
| length  | 33    | 27    | 49  | 33 | 26  | 31  | 33  | 47  | 26  | 43  | 3  | 196 |  |
| TFC     | 10    | 1     | 1   | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1  | 1   |  |
| depth   | 14    | 4     | 10  | 2  | 12  | 10  | 12  | 13  | 7   | 8   | 6  | 4   |  |
| length  | 56    | 34    | 34  | 33 | 44  | 48  | 34  | 69  | 26  | 60  | 26 | 213 |  |
| MFC     | 11    | 1     | NA  | 1  | 0   | NA  | NA  | NA  | NA  | NA  | 1  | NA  |  |
| depth   | 7     | 7     | NA  | 3  | NA  | NA  | NA  | NA  | NA  | NA  | 3  | NA  |  |
| length  | 46    | 53    | NA  | 57 | NA  | NA  | NA  | NA  | NA  | NA  | 47 | NA  |  |
| MFC     | 12    | 1     | NA  | 1  | 1   | NA  | NA  | NA  | NA  | NA  | 0  | NA  |  |
| depth   | 6     | 6     | NA  | 2  | 14  | NA  | NA  | NA  | NA  | NA  | NA | NA  |  |
| length  | 42    | 51    | NA  | 53 | 33  | NA  | NA  | NA  | NA  | NA  | NA | NA  |  |
| EDM     | 20    | 1     | NA  | 1  | NA  | NA  | 1   | NA  | NA  | NA  | 0  | NA  |  |
| depth   | 1     | 2     | NA  | 5  | NA  | NA  | 0   | NA  | NA  | NA  | NA | NA  |  |
| length  | 4     | 15    | NA  | 10 | NA  | NA  | 24  | NA  | NA  | NA  | NA | NA  |  |

|         |       | teams |    |    |    |    |    |    |    |     |    |    |
|---------|-------|-------|----|----|----|----|----|----|----|-----|----|----|
| fl.char | truth | PH    | QG | RF | SE | UZ | VZ | WA | WX | XW  | YC | YY |
| EDM     | 21    | 1     | NA | 1  | NA | NA | 1  | NA | NA | NA  | 1  | NA |
| depth   | 2     | 2     | NA | 2  | NA | NA | 0  | NA | NA | NA  | 3  | NA |
| length  | 9     | 14    | NA | 27 | NA | NA | 23 | NA | NA | NA  | 8  | NA |
| EDM     | 22    | 1     | NA | 1  | NA | NA | 1  | NA | NA | NA  | 0  | NA |
| depth   | 2     | 2     | NA | 2  | NA | NA | 0  | NA | NA | NA  | NA | NA |
| length  | 13    | 21    | NA | 22 | NA | NA | 26 | NA | NA | NA  | NA | NA |
| EDM     | 23    | 1     | NA | 1  | NA | NA | 1  | NA | NA | NA  | 0  | NA |
| depth   | 4     | 3     | NA | 4  | NA | NA | 0  | NA | NA | NA  | NA | NA |
| length  | 18    | 23    | NA | 19 | NA | NA | 33 | NA | NA | NA  | NA | NA |
| IGSCC   | 33    | 1     | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1   | NA | 1  |
| depth   | 9     | 3     | 10 | 4  | 4  | 5  | 6  | 5  | 3  | 8   | NA | 7  |
| length  | 79    | 76    | 43 | 65 | 31 | 52 | 72 | 28 | 85 | 43  | NA | 54 |
| IGSCC   | 34    | 1     | 1  | 1  | 0  | 1  | 1  | 1  | 1  | 1   | NA | 1  |
| depth   | 2     | 4     | 8  | 2  | NA | 5  | 0  | 0  | 3  | 2   | NA | 8  |
| length  | 81    | 72    | 83 | 68 | NA | 55 | 77 | 90 | 68 | 73  | NA | 63 |
| IGSCC   | 35    | 1     | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1   | NA | 1  |
| depth   | 7     | 4     | 4  | 1  | 1  | 5  | 7  | 4  | 4  | 4   | NA | 6  |
| length  | 60    | 67    | 63 | 48 | 40 | 18 | 47 | 65 | 50 | 55  | NA | 6  |
| IGSCC   | 36    | 1     | 1  | 1  | 0  | 1  | 1  | 1  | 1  | 1   | NA | 1  |
| depth   | 2     | 5     | 11 | 4  | NA | 5  | 4  | 0  | 5  | 2   | NA | 8  |
| length  | 86    | 72    | 83 | 80 | NA | 78 | 77 | 80 | 73 | 108 | NA | 9  |
| IGSCC   | 37    | NA    | 0  | 1  | NA | 0  | 1  | 0  | 0  | 0   | NA | 0  |
| depth   | 2     | NA    | NA | 0  | NA | NA | 0  | NA | NA | NA  | NA | NA |
| length  | 78    | NA    | NA | 49 | NA | NA | 52 | NA | NA | NA  | NA | NA |
| IGSCC   | 38    | NA    | 0  | 1  | NA | 0  | 1  | 0  | 1  | 1   | NA | 0  |
| depth   | 2     | NA    | NA | 0  | NA | NA | 0  | NA | 3  | 4   | NA | NA |
| length  | 60    | NA    | NA | 52 | NA | NA | 54 | NA | 49 | 123 | NA | NA |
| IGSCC   | 39    | NA    | 0  | 1  | NA | 0  | 1  | 0  | 0  | 0   | NA | 0  |
| depth   | 1     | NA    | NA | 0  | NA | NA | 0  | NA | NA | NA  | NA | NA |
| length  | 66    | NA    | NA | 37 | NA | NA | 50 | NA | NA | NA  | NA | NA |
| IGSCC   | 40    | NA    | 1  | 1  | NA | 0  | 1  | 0  | 0  | 0   | NA | 1  |
| depth   | 3     | NA    | 6  | 0  | NA | NA | 0  | NA | NA | NA  | NA | 2  |
| length  | 61    | NA    | 66 | 62 | NA | NA | 19 | NA | NA | NA  | NA | 7  |
| MFC     | 45    | 1     | 1  | 1  | 0  | 1  | 1  | 0  | 0  | 0   | 1  | 0  |
| depth   | 3     | 3     | 7  | 2  | NA | 5  | 0  | NA | NA | NA  | 6  | NA |
| length  | 41    | 35    | 38 | 36 | NA | 35 | 44 | NA | NA | NA  | 14 | NA |
| EN      | 46    | 1     | 1  | 1  | 1  | 1  | 1  | 1  | 0  | 1   | 0  | 1  |
| depth   | 4     | 6     | 9  | 2  | 12 | 11 | 8  | 2  | NA | 4   | NA | 4  |
| length  | 30    | 30    | 34 | 28 | 31 | 28 | 36 | 45 | NA | 38  | NA | 24 |
| EDM     | 47    | 1     | 0  | 1  | 1  | 1  | 0  | 0  | 0  | 0   | 0  | 0  |
| depth   | 4     | 4     | NA | 0  | 6  | 5  | NA | NA | NA | NA  | NA | NA |
| length  | 33    | 27    | NA | 31 | 27 | 32 | NA | NA | NA | NA  | NA | NA |

|         |       | teams |    |    |    |     |    |    |    |    |    |    |
|---------|-------|-------|----|----|----|-----|----|----|----|----|----|----|
| fl.char | truth | PH    | QG | RF | SE | UZ  | VZ | WA | WX | XW | YC | YY |
| EN      | 48    | 1     | 0  | 1  | 0  | 1   | 1  | 1  | 0  | 0  | 1  | 1  |
| depth   | 1     | 3     | NA | 4  | NA | 5   | 8  | 4  | NA | NA | 4  | 6  |
| length  | 20    | 17    | NA | 31 | NA | 188 | 25 | 38 | NA | NA | 15 | 24 |
| EDM     | 49    | 1     | 0  | 1  | 0  | 1   | 1  | 0  | 0  | 0  | 0  | 0  |
| depth   | 6     | 6     | NA | 5  | NA | 7   | 4  | NA | NA | NA | NA | NA |
| length  | 38    | 27    | NA | 36 | NA | 33  | 37 | NA | NA | NA | NA | NA |
| EDM     | 50    | 1     | 0  | 1  | 0  | 1   | 0  | 0  | 0  | 0  | 0  | 0  |
| depth   | 0     | 4     | NA | 0  | NA | 7   | NA | NA | NA | NA | NA | NA |
| length  | 15    | 8     | NA | 12 | NA | 12  | NA | NA | NA | NA | NA | NA |

**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- 6795

PNNL - 13873

2. TITLE AND SUBTITLE

A Comparison of Three Round Robin Studies on ISI Reliability  
of Wrought Stainless Steel Piping

3. DATE REPORT PUBLISHED

MONTH | YEAR

February | 2003

4. FIN OR GRANT NUMBER

Y6604

5. AUTHOR(S)

P.G. Heasler and 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  
P.O. Box 999  
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

D.A. Jackson, NRC Project Manager

11. ABSTRACT (200 words or less)

The Pacific Northwest National Laboratory (PNNL) is conducting a multi-year program sponsored by the Nuclear Regulatory Commission (NRC) to address issues related to the reliability of ultrasonic testing (UT) and the development of improved programs for inservice inspection (ISI). This includes establishing the accuracy and reliability of UT methods for ISI of light water reactor components. From 1981 through 1990, three major round robin studies were conducted to quantify the capability of ISI inspectors to detect cracks in wrought stainless steel, and their accuracy in crack sizing. This report concentrates on analysis techniques to estimate comparable ISI detection and sizing statistics from the three round robins. This analysis provides a tool for evaluating the effect of technological advances in UT capability during the 1980s, and to assess combining the data from the three studies to form a better overview of inspection performance.

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

Inservice Inspection, Ultrasonic Testing, Round Robin Studies,  
Probability of Detection

13. AVAILABILITY STATEMENT

unlimited

14. SECURITY CLASSIFICATION

(This Page)

unclassified

(This Report)

unclassified

15. NUMBER OF PAGES

16. PRICE



Federal Recycling Program

NUREG/CR-6795

A COMPARISON OF THREE ROUND ROBIN STUDIES ON ISI  
RELIABILITY OF WROUGHT STAINLESS STEEL PIPING

FEBRUARY 2003

UNITED STATES  
NUCLEAR REGULATORY COMMISSION  
WASHINGTON, DC 20555-0001

---

OFFICIAL BUSINESS  
PENALTY FOR PRIVATE USE, \$300