U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

NUREG-2228

Weld Residual Stress Finite Element Analysis Validation

Part II—Proposed Validation Procedure

Draft Report for Comment

Office of Nuclear Regulatory Research

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 Library at 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 Publishing Office
   Washington, DC 20402-0001
   Internet: bookstore.gpo.gov
   Telephone: (202) 512-1800
   Fax: (202) 512-2104

2. **The National Technical Information Service**
   5301 Shawnee Road
   Alexandria, VA 22312-0002
   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: **U.S. Nuclear Regulatory Commission**
Office of Administration
Multimedia, Graphics, and Storage & Distribution Branch
Washington, DC 20555-0001
E-mail: distribution.resource@nrc.gov
Facsimile: (301) 415-2289

Some publications in the NUREG series that are posted at NRC's Web site address 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, 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 42nd Street
New York, NY 10036-8002
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 contractorprepared 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).

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

NUREG-2228

Weld Residual Stress Finite Element Analysis Validation

Part II—Proposed Validation Procedure

Draft Report for Comment

Manuscript Completed: August 2018
Date Published: August 2018

Prepared by:
M. Benson
P. Raynaud
J. Wallace

Michael Benson, NRC Project Manager

Office of Nuclear Regulatory Research

COMMENTS ON DRAFT REPORT

Any interested party may submit comments on this report for consideration by the NRC staff. Comments may be accompanied by additional relevant information or supporting data. Please specify the report number **NUREG-2228** in your comments, and send them by the end of the comment period specified in the *Federal Register* notice announcing the availability of this report.

Addresses: You may submit comments by any one of the following methods. Please include Docket ID **NRC-2018-0168** in the subject line of your comments. Comments submitted in writing or in electronic form will be posted on the NRC website and on the Federal rulemaking website http://www.regulations.gov.

Federal Rulemaking Website: Go to http://www.regulations.gov and search for documents filed under Docket ID **NRC-2018-0168**.

Mail comments to: May Ma, Director, Program Management, Announcements and Editing Branch (PMAE), Office of Administration, Mail Stop: TWFN-7-A-60M, U.S. Nuclear Regulatory Commission, Washington, DC 20555-0001.

For any questions about the material in this report, please contact: Michael Benson, Materials Engineer, and 301-415-2425 or by e-mail at michael.benson@nrc.gov.

Please be aware that any comments that you submit to the NRC will be considered a public record and entered into the Agencywide Documents Access and Management System (ADAMS). Do not provide information you would not want to be publicly available.

ABSTRACT

Under a Memorandum of Understanding, the U.S. Nuclear Regulatory Commission and the Electric Power Research Institute conducted a research program aimed at gathering data on weld residual stress modeling. As described in NUREG-2162, "Weld Residual Stress Finite Element Analysis Validation: Part I—Data Development Effort," issued March 2014, this program consisted of round robin measurement and modeling studies on various mockups. At that time, the assessment of the data was qualitative. This report describes an additional residual stress round robin study and a methodology for capturing residual stress uncertainties. This quantitative approach informed the development of guidelines and a validation methodology for finite element prediction of weld residual stress. For example, comparisons of modeling results to measurements provided a basis for establishing guidance on a material hardening approach for residual stress models. The proposed validation procedure involves an analyst modeling a known case (the Phase 2b round robin mockup) and comparing results to three proposed quality metrics. These recommendations provide a potential method by which analysts can bolster confidence in their modeling practices for regulatory applications.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Weld residual stress (WRS) is known to be an important driver of primary water stress corrosion cracking in safety-related nuclear piping. For this reason, it is desirable to formalize finite element modeling procedures for residual stress prediction. The U.S. Nuclear Regulatory Commission (NRC) and the Electric Power Research Institute have conducted joint research programs on residual stress prediction under a memorandum of understanding. These studies have involved modeling and measurement of WRS in various mockups. The latest of these studies, Phase 2b, is discussed in this document.

The Phase 2b mockup was prototypic of a pressurizer surge nozzle dissimilar metal weld, which forms part of the reactor coolant pressure boundary. Two sets of residual stress measurement data were obtained on the Phase 2b mockup: deep hole drilling and contour. Both these methods are strain-relief techniques. In addition to the measurements, 10 independent analysts submitted finite element modeling results of the residual stresses in the mockup. Each participant was provided the same set of modeling guidelines, with the aim of reducing analyst-to-analyst scatter as much as possible. These measurement and modeling data were then used to develop an uncertainty quantification methodology.

The residual stress uncertainty methodology consisted of constructing a statistical model of the data and using bootstrapping methods to calculate relevant 95/95 tolerance bounds and confidence bounds. This methodology improves on past work (e.g., NUREG-2162, "Weld Residual Stress Finite Element Analysis Validation: Part I—Data Development Effort," issued March 2014), which described uncertainty in WRS predictions only in qualitative terms. Furthermore, the results of the uncertainty quantification effort informed the development of a validation approach of residual stress finite element models.

The uncertainty quantification work provided methods to compare measurements to models, which in turn led to recommendations on hardening law (see Section 5.2). The validation method is a step-by-step procedure for comparing independent finite element modeling results of the Phase 2b mockup to the acceptance measures. If an analyst meets the criteria, then the modeling procedure may be applied with greater confidence to a real case. This procedure is intended as a recommendation rather than a regulatory requirement. It provides a means to demonstrate proficiency in finite element modeling of WRS.

The validation methodology is aimed at WRS predictions for deterministic flaw growth evaluations. The nuclear industry often performs flaw evaluations when seeking alternatives to established inspection and repair/replacement rules. These evaluations require a WRS assumption. If that assumption is based on finite element results, then following the validation procedure offers the industry one method to strengthen its case when seeking NRC approval. This document also investigated how differences in residual stress can affect these flaw evaluations. Important features of the stress profiles include the inner diameter stress, the stress magnitude at the initial flaw depth, and the depths at which the stress profile crosses zero. Decision-makers can review these aspects of submitted stress profiles as another option for gaining confidence in residual stress predictions.

ACKNOWLEDGMENTS

The authors would like to thank the following.

- John Broussard of Dominion Engineering, Inc., Paul Crooker of the Electric Power Research Institute (EPRI), and Michael Hill of University of California, Davis for technical cooperation in the joint NRC-EPRI research program.
- Dusty Brooks, Remy Dingreville, and John Lewis of Sandia National Laboratory for their excellent work on developing an uncertainty quantification scheme for the round robin dataset (see Chapter 3).
- The round robin modeling participants for contributing their work to this effort, as described in Chapter 2.

ABBREVIATIONS AND ACRONYMS

ASME Code	American Society of Mechanical Engineers *Boiler and Pressure Vessel Code*
DHD	deep hole drilling
EPRI	Electric Power Research Institute
EWR	excavate and weld repair
FE	finite element
fPCA	functional principal components analysis
ID	inner diameter
mm	millimeter
MPa	megapascal
NDE	nondestructive examination
NRC	U.S. Nuclear Regulatory Commission
OD	outer diameter
PWSCC	primary water stress corrosion cracking
RMSE	root mean square error
SIF	stress intensity factor
WRS	weld residual stress
i	reference index
k	reference index
μ_k	mean at the k^{th} position through the wall thickness
σ_k	standard deviation at the k^{th} position through the wall thickness
wi	weighting factor for the i^{th} WRS profile
$x_{i,k}{}^{WRS}$	stress magnitude of the ith profile at the kth position through the thickness
f	a function
r	radial position through the wall thickness
t	wall thickness of the weld or pipe
d	normalized distance through the wall thickness, $d = r/t$
γ	warping function
T	operating temperature
P	operating pressure
τ	time
K_I	mode I stress intensity factor
a	half-depth of a flaw

c	half-length of a flaw
σ_m	membrane stress
σ_b	bending stress
σ_{cfp}	crack face pressure stress
G_b	influence coefficient for global bending
Q	flaw shape parameter
$h(x,a)$	weight function for the Universal Weight Function Method
$s(x)$	stress variation along the crack face
$da/d\tau$	flaw growth with respect to time
K_{Ith}	stress intensity factor threshold
Q_g	activation energy
R_g	ideal gas constant
T_{abs}	absolute operating temperature
T_{ref}	empirical reference temperature
ϕ	tabulated crack growth coefficient
η	tabulated crack growth coefficient
K_{90}	SIF at the deepest point along the crack front
K_0	SIF at the surface point along the crack front
g	a function
L	number of locations through the wall thickness where a WRS magnitude is known
n_e	number of sampled measurement WRS profiles
n_p	number of sampled prediction WRS profiles
s	reference index
$h(d)$	difference in means function
$RMSE_{WRS}$	quality metric on the root mean square error of stress magnitude
WRS_{mean}	benchmark WRS (the mean of the isotropic predictions from the Phase 2b study)
D_1	first derivative of the WRS magnitude with respect to through-wall position
h	interval between two positions through the wall thickness
$RMSE_{D1}$	quality metric on the root mean square error of D1
$diff_{avg}$	quality metric on the average difference between the prediction WRS and the benchmark value

1 INTRODUCTION

The U.S. Nuclear Regulatory Commission (NRC) and the Electric Power Research Institute (EPRI) initiated a long-term research program on understanding and reducing uncertainty in the numerical prediction of weld residual stress (WRS) in safety-related nuclear components. Part 1 of this report [1] discusses the background and past work of the program in detail (also see [2]). The through-wall WRS profile is an important input to deterministic flaw growth calculations. These calculations may form the technical basis for regulatory relief requests to modify repair/replacement or nondestructive examination requirements in nuclear components subject to primary water stress corrosion cracking (PWSCC). Probabilistic fracture mechanics calculations [3]–[6] rely on well characterized uncertainty for important inputs. For these reasons, it is important to develop sound approaches for reaching best estimates of WRS and the associated uncertainty.

The past research was categorized according to four phases:

(1) Phase 1: small-scale scientific specimens
(2) Phase 2a: fabricated prototypic pressurizer surge line nozzle
(3) Phase 3: pressurizer surge line nozzles from a canceled plant
(4) Phase 4: optimized weld overlay on a prototypic cold-leg nozzle

These four phases consisted of double-blind measurement and modeling studies on the mockups of varying geometry. In general, the work showed that axisymmetric finite element (FE) models provided reasonable estimations of the measurements, but that relatively large analyst-to-analyst uncertainty existed in the predictions. NUREG-2162 [1] lists the following recommendations for future work in the WRS Validation Program:

- Develop specific validation criteria for comparing WRS measurement and modeling results.
- Establish guidelines for WRS input development for deterministic flaw evaluations, including FE best practices.
- Develop additional guidance for accounting for uncertainty in WRS inputs for flaw evaluations.
- Focus future FE round robin studies on reducing model-to-model variability, given lessons learned in FE modeling best practices.
- Apply more robust methods to quantify modeling uncertainty in future round robin efforts.

The final phase of this research, dubbed Phase 2b, aimed at addressing these issues.

1.1 Phase 2b Effort

Phase 2b was a second double-blind round robin measurement and FE modeling study involving a pressurizer surge line nozzle mockup. This mockup was similar to, but not exactly the same as, the Phase 2a mockup discussed in [1]–[2]. Modeling guidelines [7], based on lessons learned from the previous research phases, were developed with the intent of reducing the uncertainty observed in the past work. The measurement program consisted of deep hole drilling (DHD) and the contour methods. Ten international participants submitted independent modeling results.

The dataset from the Phase 2b study was intended to address the items for future work identified in NUREG-2162. An unbiased view of expected modeling uncertainty is important for developing acceptance measures and WRS input guidelines. These efforts also require more quantitative approaches to describing the data.

1.2 Scope of This Report

This report is intended to document the development of an FE validation approach for prediction of WRS. The validation scheme proposed here draws on the results of the Phase 2b round robin and an uncertainty quantification methodology. Chapter 2 summarizes the Phase 2b round robin study and the resulting dataset. Chapter 3 discusses the mathematical methods developed to characterize uncertainty in the Phase 2b dataset. Chapter 4 presents the impacts of WRS assumptions on flaw growth calculations. Chapter 5 develops guidelines for WRS inputs for flaw growth calculations and the proposed validation approach. Finally, Chapter 6 contains overall conclusions of the work.

2 PHASE 2B ROUND ROBIN STUDY

2.1 Purpose

The purpose of this effort was to conduct a second FE round robin, similar to the Phase 2a study [1], [8]–[10], with improved FE modeling guidance. The modeling guidance aimed at reducing analyst-to-analyst scatter. Determining appropriate scatter bands for FE predictions is important for formulating acceptance measures and modeling guidelines. WRS measurements have uncertainties as well. Chapter 3 discusses quantification of both measurement and modeling uncertainties. Both uncertainties must be accounted for when deciding what constitutes an appropriate FE prediction. This chapter discusses the research effort designed to collect the measurement and modeling data. An NRC technical letter report [7] presents more detailed information about the Phase 2b effort.

2.2 Mockup Fabrication

The geometry chosen for the Phase 2b round robin study was representative of a pressurizer surge nozzle. Figure 2-1 shows the overall geometry of the mockup.

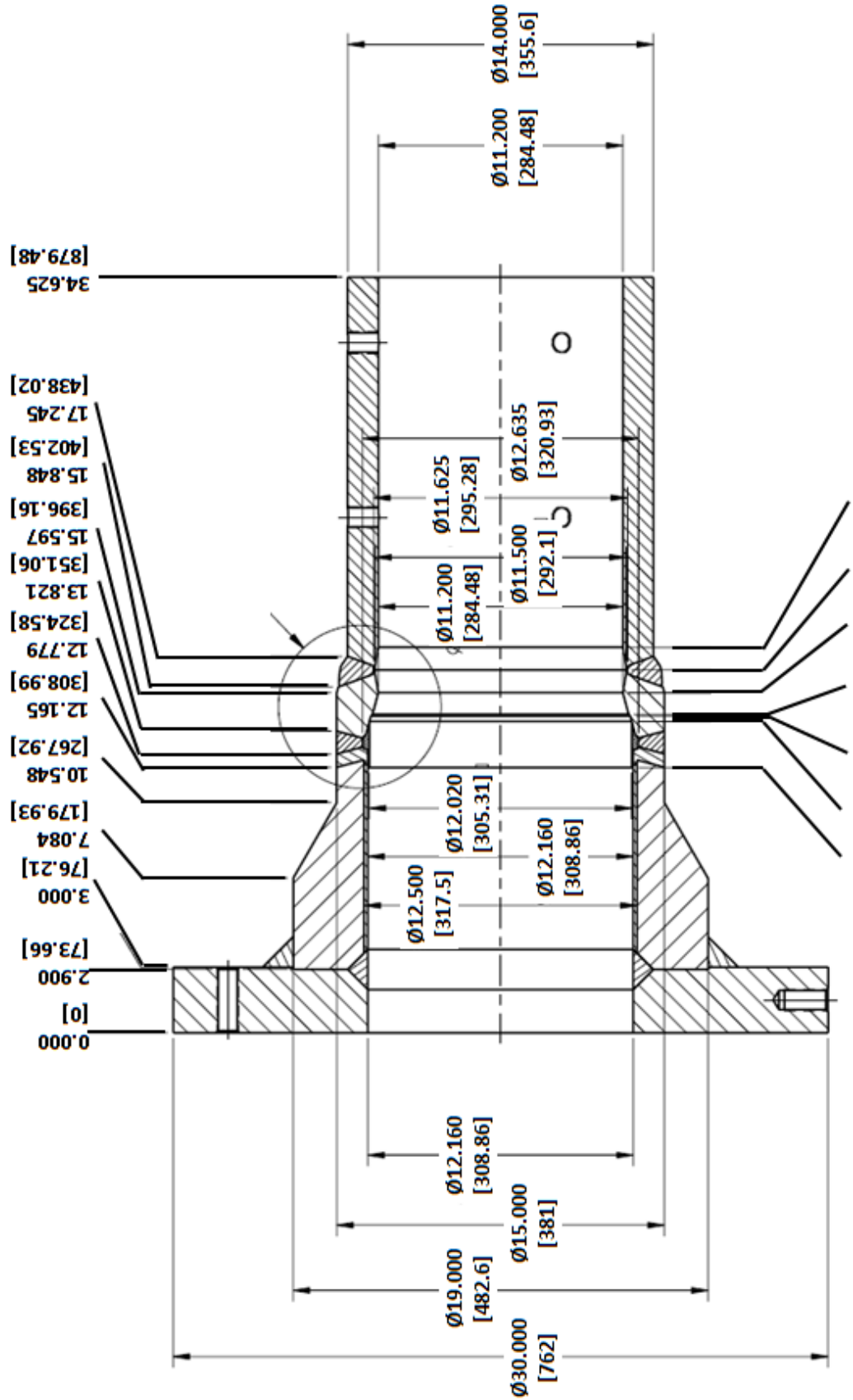


Figure 2-1: Phase 2b Mockup Geometry (Dimensions in inches [millimeters])

Table 2-1 provides an overview of the fabrication process. More detailed fabrication information, including welding parameters and bead map drawings, is found in [7].

Table 2-1: Mockup Fabrication Steps

Step	Description	Purpose
1	A36 flange welded to SA182 nozzle	Simulates nozzle stiffness in service; not modeled
2	Alloy 82 buttering applied to nozzle	Allows for post-weld heat treat of low alloy steel and prepares dissimilar metal weld
3	Post weld heat treatment	Tempers martensite in low alloy steel and relieves residual stress
4	Buttered nozzle welded to F316L safe end with Alloy 182 filler metal	Simulates shop weld
5	Backchip and reweld	Simulates repair weld at inner diameter
6	Safe end welded to TP316 pipe	Simulates field closure weld

2.3 Round Robin Participants

Ten participants representing 12 organizations submitted FE results to the round robin study, as represented in Figure 2-2.

Figure 2-2: Participating Organizations

These participants represent a cross section of international industry, government, academic, and private contractor organizations. The study was double blind, so the modelers did not have access to the measurement data. Likewise, the measurement practitioners did not have access to the modelers' results.

2.4 Weld Residual Stress Measurements

Additional background on residual stress measurement is given in Section 2.2 of [1]. VEQTER, Ltd., in Bristol, United Kingdom, and Hill Engineering, LLC, in Rancho Cordova, CA, performed

the Phase 2b residual stress measurements. Two sets of measurements were carried out: hole drilling and contour (see Section 2.2.2 of [1]). The hole drilling measurements consisted of a combination of DHD and incremental DHD. Figure 2-3 shows the experimental setup of the hole drilling measurements. Four hole drilling measurements were taken starting at location B shown in Figure 2-3. Location B was located 22° from the weld start location. The other three measurements were made 90° apart from one another (see Figure 2-4). Care was taken to avoid weld start/stop locations around the circumference.

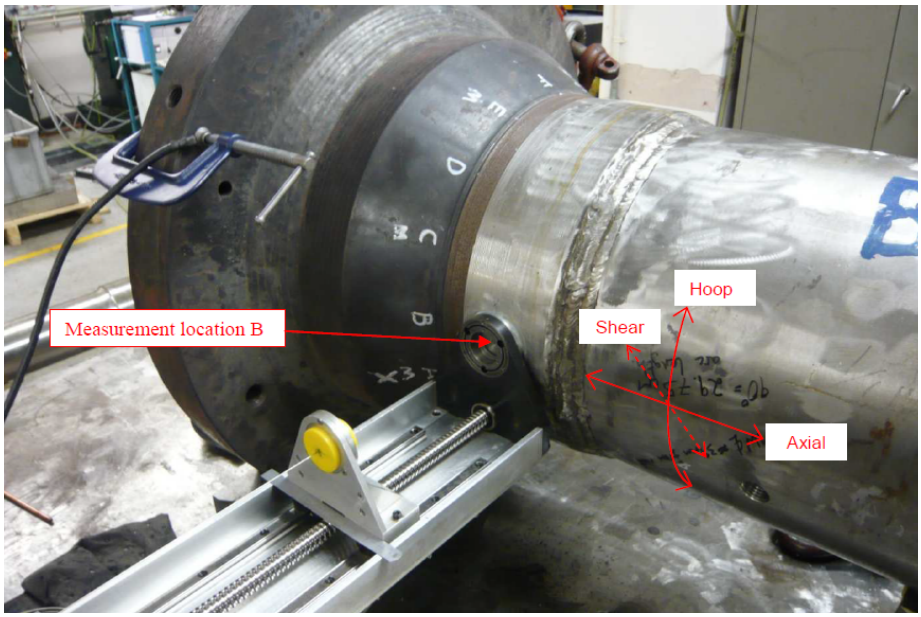


Figure 2-3: Deep Hole Drilling Measurement Setup

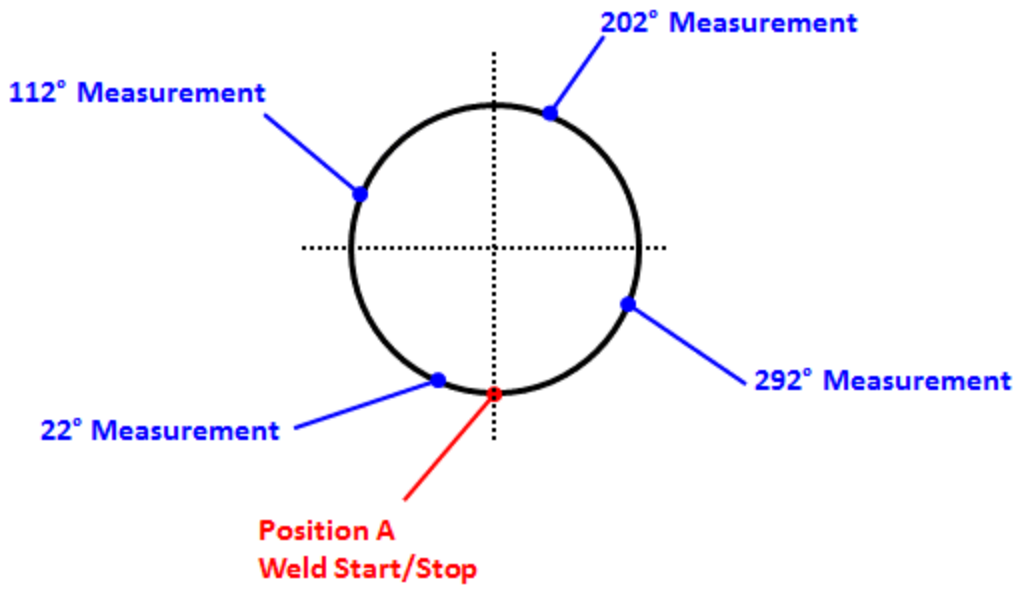


Figure 2-4: Hole Drilling Measurements around Circumference

The contour measurements involved several cuts, including one cut each for the axial residual stress measurement and the hoop residual stress measurement (see Figure 2-5).

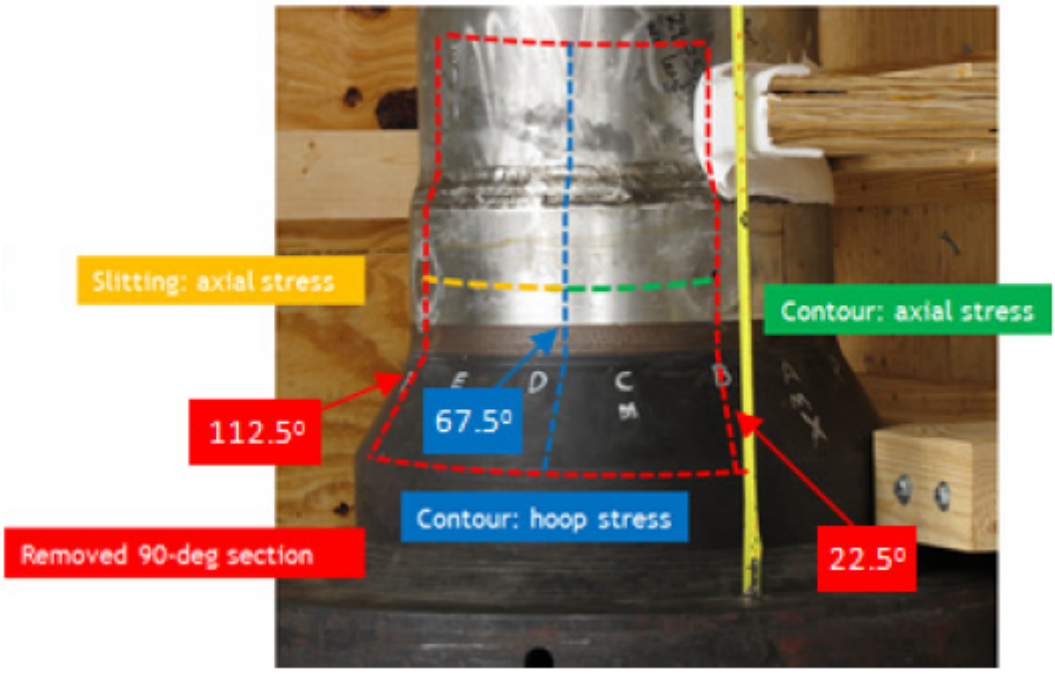


Figure 2-5: Contour Measurement Setup

The final calculation of residual stress accounted for the release of stress at each sectioning operation. The hole drilling measurements were made before the destructive contour

measurements. Each of the required cuts was made with the hole drilling measurements in mind, as shown in Figure 2-6.

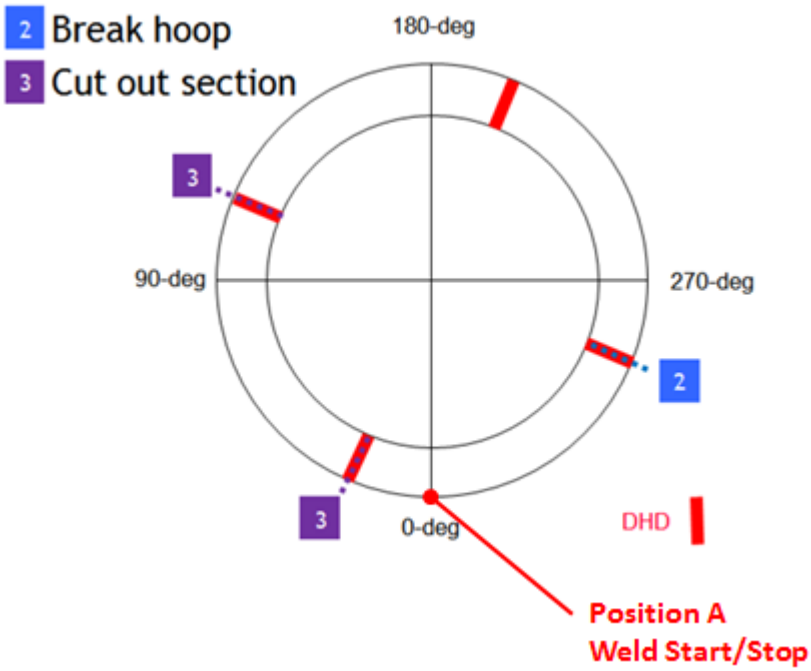


Figure 2-6: Cuts to Extract Contour Specimen

The section outlined in red in Figure 2-5 is represented by the "Cut out section" cuts shown in Figure 2-6 (i.e., Cut 3). At each of these cuts, strain gauge measurements are made for the final stress calculation. The cuts represented by the blue, green, and yellow lines in Figure 2-5 were then made. A laser profilometer measured displacements along the relevant cross sections.

2.5 Modeling Guidance

The round robin participants were tasked with creating an axisymmetric FE model to predict the residual stress distribution of the Phase 2b mockup. The written problem statement provided to the round robin participants is provided in [7]. Table 2-2 summarizes this guidance, which was based on modeling experience gained in previous work [1], [2], [11].

Table 2-2: Model Guidance

Modeling Topic	Guidance Description
Hardening Law	Participants to complete two models: one assuming isotropic hardening, one assuming kinematic hardening. Material properties for each hardening law provided to the participants.
Weld Bead Geometry	Participants should model the specified number of weld passes and layers provided in the problem statement. Precise use of profilometry data was not required. Participants can use trapezoidal beads of approximately equal area.
Thermal Model Tuning	Material properties for heat transfer calculation provided to the particpants. Participants free to choose heat input model. Precise tuning of thermal model to thermocouple data optional, due to weak sensitivity on heat input. Participants should tune thermal model to approximate expected melt zone area.
Structural Boundary Conditions	Mock-up was not extensively constrained during fabrication. Participants should fix one single node (located away from welding areas) from displacement along the axial direction of the pipe.
Material Properties	Material properties for both the heat transfer and static stress analysis were provided to the particpants.
Post Processing	Particpants requested to define a path through the centerline of the dissimilar metal weld and extract data at 24 equally-spaced points along the path.
Pass Lumping and Bead Sequence	Participants requested not to lump weld beads. Participants requested to model the bead sequence provided in the problem statement.
Miscellaneous	Fine mesh of linear elements recommended. Mesh size of approximately 1.25 mm square in weld beads recommended. No triangular elements. Mesh density may coarsen away from weld areas.

2.6 Results

This section reports the basic set of results from the Phase 2b FE round robin study. The raw data are reported in both graphs and tables in [7].

2.6.1 Measurement Results

Figure 2-7 shows the hole drilling measurement results. In this report, r/t=0 represents the inner surface of the pipe wall, and r/t=1 represents the outer surface of the pipe wall.

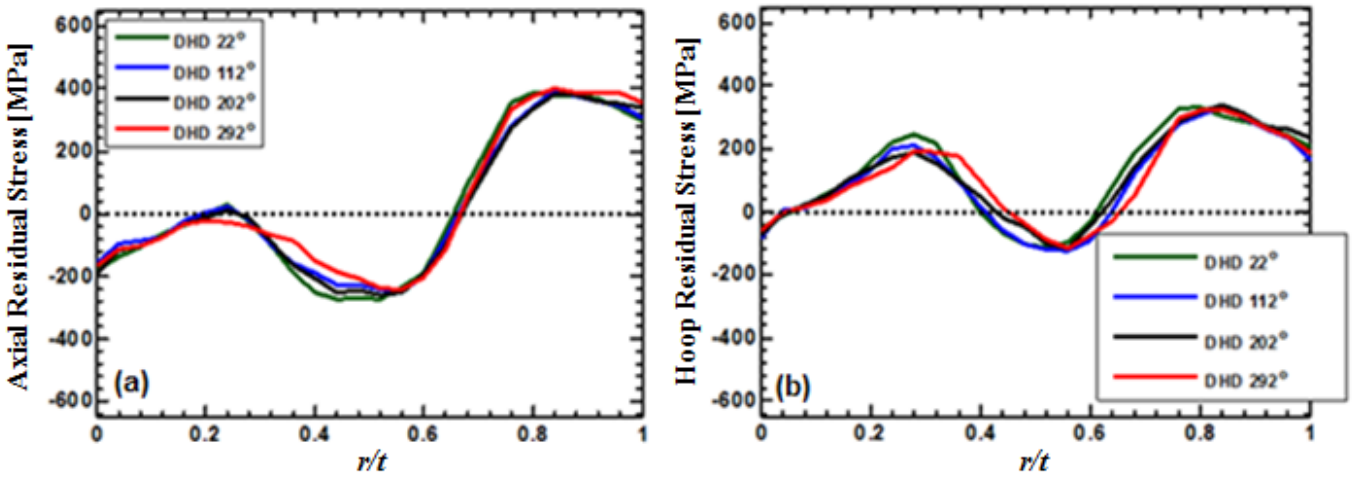


Figure 2-7: Hole Drilling Measurement: (a) Axial, (b) Hoop

Figure 2-8 and Figure 2-9, respectively, show the hoop and axial stress measurements for the contour method.

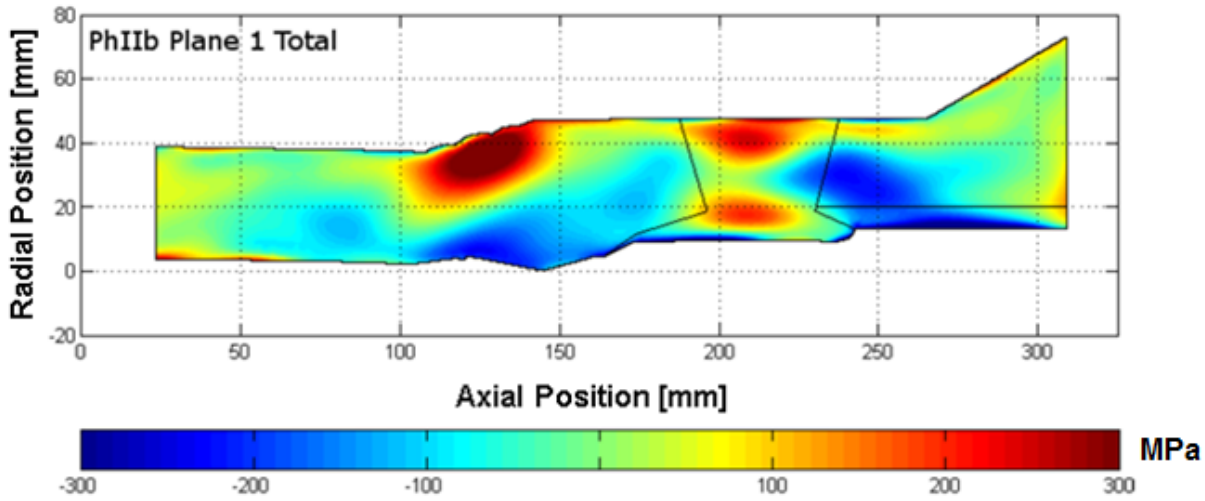


Figure 2-8: Hoop Stress—Contour Measurement

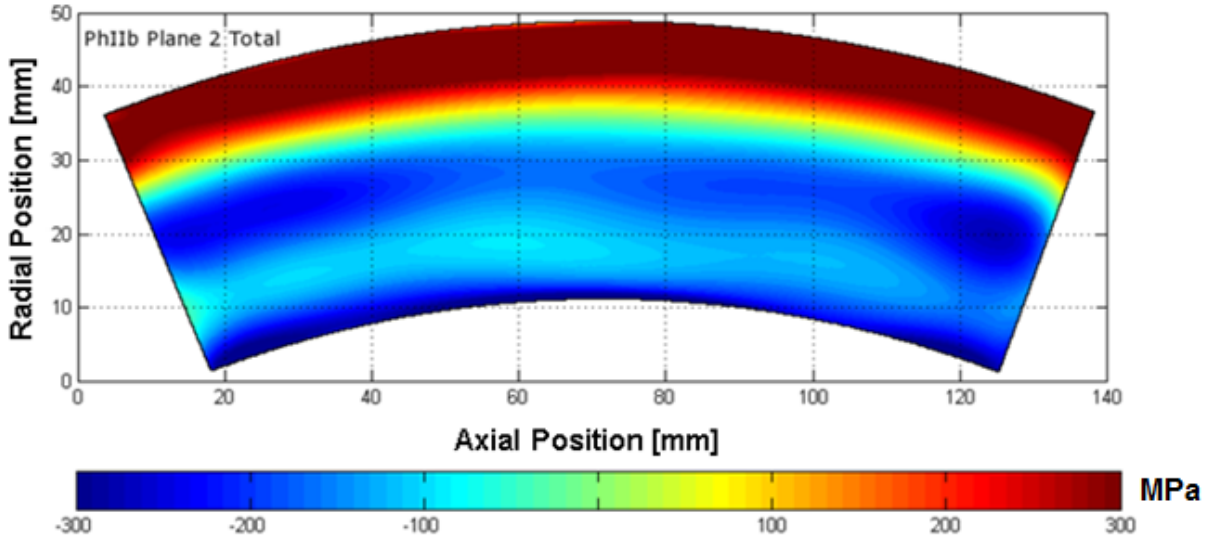


Figure 2-9: Axial Stress—Contour Measurement

The DHD and contour datasets are, by nature, different. The DHD data are a one-dimensional profile of the stress variation along a straight path through the weld centerline. In contrast, the contour method gives a two-dimensional representation of the stress variation on an entire cross section. As described in Table 2-2, the modeling data were collected as one-dimensional path data extracted from the FE results along the centerline of the weld. Therefore, the contour data must be processed to extract appropriate path data to compare to the modeling data. The extracted contour data should be a one-dimensional stress profile and represent the stress along a straight path through the weld centerline, normal to the inside surface.

2.6.2 Modeling Results

Figure 2-10 shows an example mesh from one of the FE round robin participants.

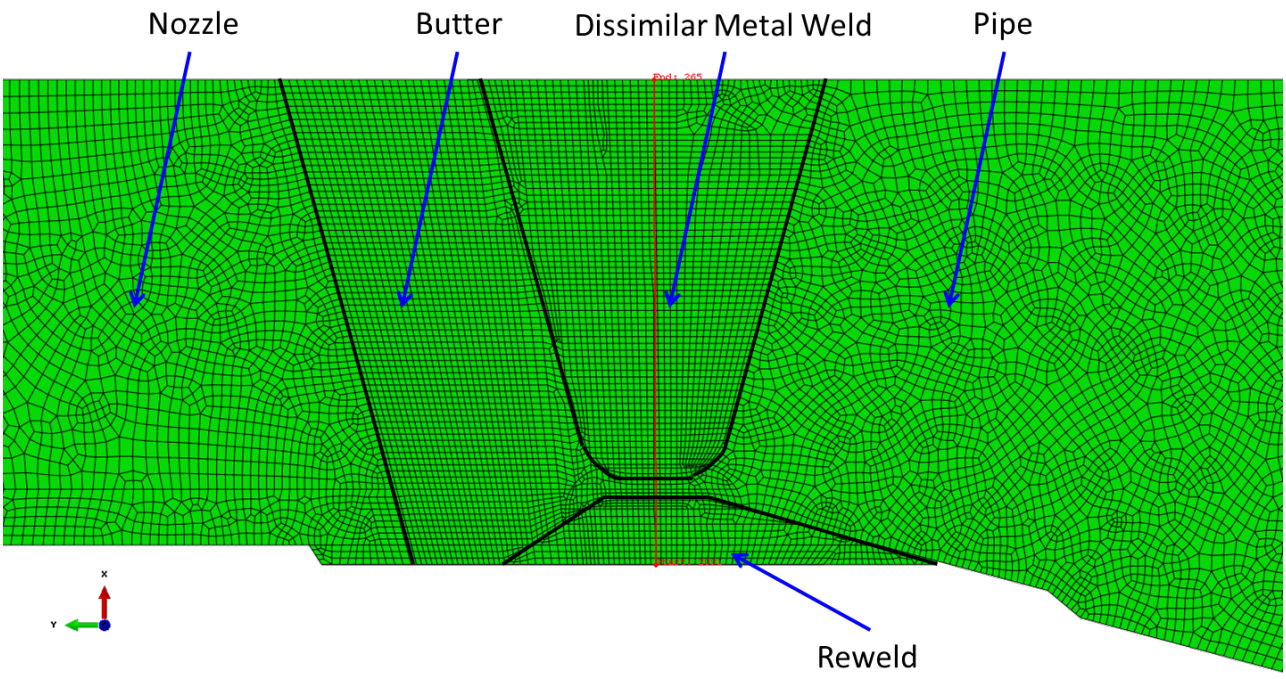


Figure 2-10: Example Mesh

The red line in Figure 2-10 represents the path along which the participant extracted the data (i.e., the weld centerline). The figure also illustrates major geometry features modeled by the participants. Figure 2-11 and Figure 2-12 show the isotropic and nonlinear kinematic hardening results of the Phase 2b round robin study, respectively. A qualitative look at the modeling results reveals the following observations:

- Individual predictions may potentially be considered outliers (e.g., participant J), when compared to the rest of the sample.
- Nonlinear kinematic and isotropic results show different through-wall trends.
- The nonlinear kinematic results show smaller stress magnitudes than the isotropic results.
- The nonlinear kinematic results generally exhibit less scatter than the isotropic results.

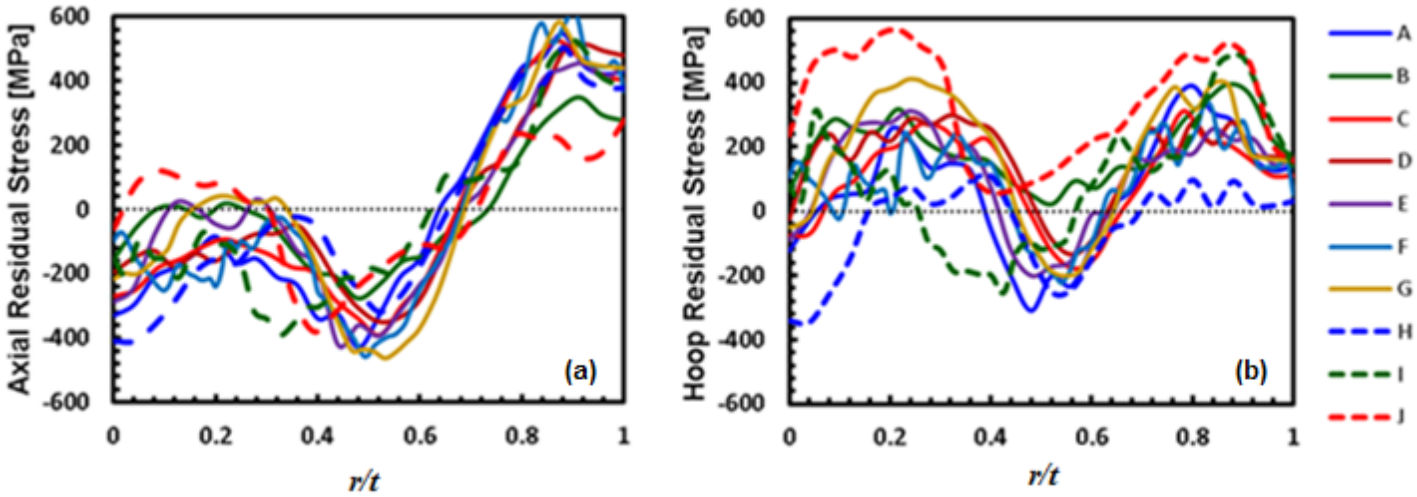


Figure 2-11: Processed Isotropic Hardening Results: (a) Axial, (b) Hoop

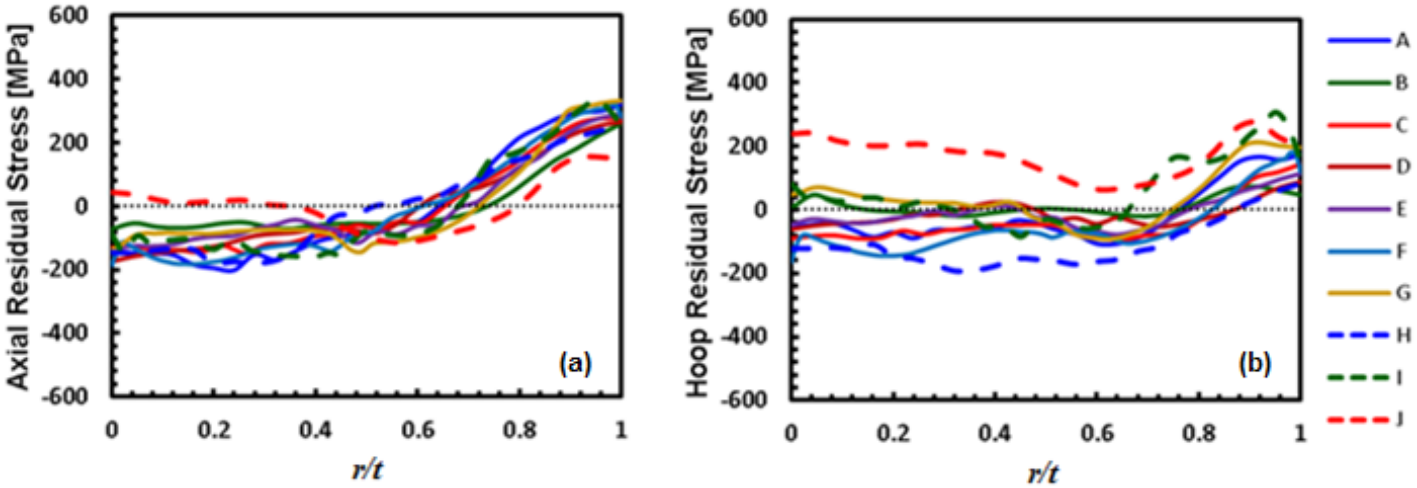


Figure 2-12: Processed Nonlinear Kinematic Hardening Results: (a) Axial, (b) Hoop

2.6.3 Discussion

Comparison of the measurement and modeling data requires careful thought. Both the modeling data and the measurement data exhibit uncertainties. Chapter 3 of this document focuses on quantitatively evaluating both modeling and measurement uncertainty. The end goal is to develop a procedure to objectively judge FE models of WRS, as discussed in Chapter 5.

One final technical topic to be resolved is the choice of hardening law. When providing guidance on hardening law, it is important to avoid biases in the model prediction in addition to minimizing prediction errors. One simplified approach suggested elsewhere [12] involves use of the average of the nonlinear kinematic and isotropic predictions. Section 5.2 discusses the choice of hardening law.

2.7 Conclusions

This chapter summarizes the Phase 2b round robin study. The WRS modeling and measurement results from this study constitute the dataset analyzed in the uncertainty analysis in Chapter 3.

Ten analysts participated in the modeling portion of the round robin, resulting in 10 isotropic and 10 nonlinear kinematic WRS predictions for both axial and hoop stresses. Two measurement vendors performed strain relief-based WRS measurements on the mockup. This resulted in four DHD measurements of axial and hoop stresses along the weld centerline. The contour measurement resulted in a two-dimensional representation of the WRS along a given cross section. This dataset, viewed in the context of the Chapter 3 analysis, will be used to develop the guidelines and validation scheme presented in Chapter 5.

3 UNCERTAINTY QUANTIFICATION METHODOLOGY

3.1 Motivation

In [1], the NRC documented the need to apply more sophisticated data analysis techniques to residual stress measurement and modeling data. At that stage, only qualitative judgments were applied to describe modeling uncertainty and measurement-to-model comparisons. To define an objective validation process for WRS predictions, it is necessary to use a quantitative analysis.

Previous work in this area includes the development of a sampling scheme for WRS in a probabilistic fracture mechanics code [5]–[6]. The baseline dataset was four FE WRS profiles obtained by different analysts for a given weld configuration. Estimates of skewness and kurtosis were used to assign an appropriate uncertainty distribution type for the WRS FE data at each point through the pipe thickness. In determining the final estimates for the mean WRS and standard deviation, the analysts introduced a weighting approach that decreased the importance of a particular WRS prediction the further away it was from the other predictions. This is represented mathematically as:

$$\mu_k = \frac{\sum_{i=1}^{n} w_i x_{i,k}^{WRS}}{\sum_{i=1}^{n} w_i} \text{ and}$$

Equation 3-1

$$\sigma_k = \sqrt{\frac{\sum_{i=1}^{n} w_i \left(x_{i,k}^{WRS} - \mu_k\right)^2}{\sum_{i=1}^{n} w_i}},$$

where μ_k and σ_k are the mean and standard deviation of the distribution at the k^{th} position through the weld thickness, respectively; w_i is the weight for the i^{th} WRS profile based on the differences in stress predictions between two WRS profiles; and $x_{i,k}$ is the value of the stress prediction for the i^{th} profile at the k^{th} position through the wall thickness. With the uncertainty in WRS thus defined, Kurth *et al.* [5]–[6] conceived a sampling strategy that accounted for point-to-point smoothness and static equilibrium requirements. In this way, the probabilistic fracture mechanics analysis may account for the uncertainty in the WRS profile for a given weld configuration.

The approach presented here is aimed at deterministic fitness-for-service calculations, where a residual stress assumption is required (see Chapter 4). This approach involves a range of mathematical tools aimed at defining uncertainty bands on the round robin measurement and modeling results. The results then form the basis for recommended modeling practices and model validation approaches. The methodology described here is documented in greater detail in [13], and Section 3.2 provides only a summary.

3.2 Methodology

3.2.1 Functional Data

The round robin WRS modeling and measurement dataset is discrete by nature. Even the contour measurement is based on a finite number of measurements along the surface of the part.

It is also functional data, in that the WRS magnitude for a given stress component, WRS, depends on the spatial location along the pipe cross section, as:

$$WRS = f(d) \qquad \textbf{Equation 3-2}$$

where $d=r/t$ is the normalized distance from the inside surface to the outside surface of the weld. References [15] and [16] point out that it is useful to consider such data as a continuous function. They introduce the mathematical methods that can be applied to functional data. This section describes a statistical model constructed to represent the round robin dataset, based on the methods of [15] and [16]. This approach enables bootstrapping to estimate confidence bounds and tolerance bounds on both the measurement and modeling data [17]–[18]. The final outcome of the work is an objective process for validating WRS FE modeling (see Chapter 5).

3.2.2 Screening of Outlier Predictions

As described in Chapter 2, the idea behind the Phase 2b round robin study was to assess the prediction uncertainty of a group of analysts modeling the same problem under a given set of guidelines. The 10 submissions were screened for potential outlying results that may not have been obtained in strict accordance with the modeling guidance. Tran et al. [19] described two outlier predictions in the round robin dataset and the reasons behind them. One participant used incorrect material property data, and the other incorrectly modeled the heat input of the stainless steel closure weld. A third outlier involving incorrect weld thickness was identified in [13]. These three predictions were screened out for the purposes of uncertainty quantification. As discussed in Section 3.3.1, one other prediction was removed from the hoop stress profiles because of the undue influence it had on the bootstrap tolerance bound results.

3.2.3 Data Smoothing

The WRS measurements were reported at discrete spatial locations. Similarly, the round robin modeling participants provided stress magnitudes at discrete depths through the wall thickness. The actual WRS distribution is expected to be a continuous function of spatial location. Data smoothing was applied here to arrive at a smooth, continuous representation of the WRS profile. The smoothing was accomplished via cubic splines, which are a series of third degree polynomials connected together at a given number of nodes [20]. The optimal number of nodes to achieve an acceptable fit was determined by an algorithm discussed in [13]. As an example, the axial stress predictions assuming isotropic hardening are shown after smoothing in Figure 3-1 (compare with Figure 2-11a; outliers removed).

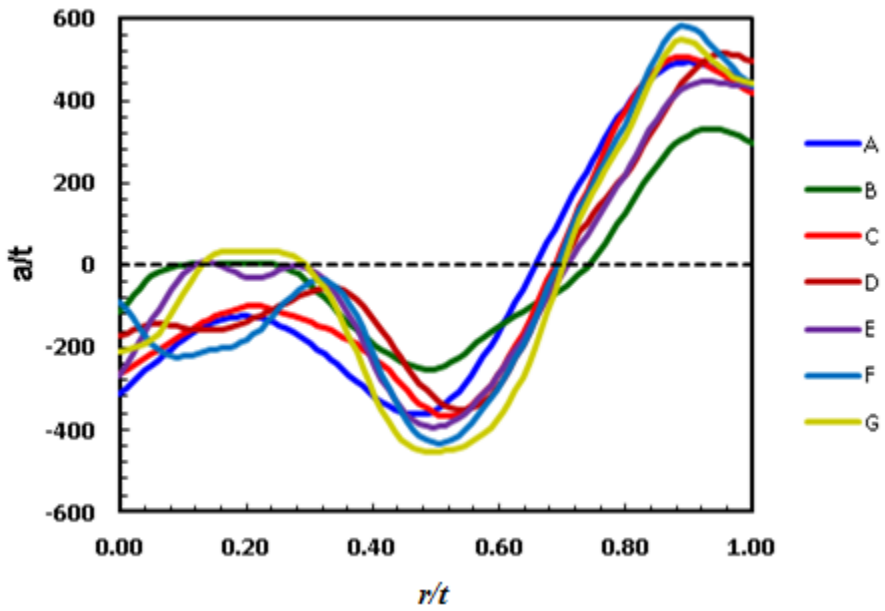


Figure 3-1: Axial Isotropic Data after Smoothing

3.2.4 Amplitude and Phase Variability

Functional data can exhibit two types of variability: amplitude and phase variability. Considering a sinusoidal function, the amplitude variability is the result of differences in peak height of two curves, and phase variability is the result of a horizontal shift of one curve relative to the other (see Figure 3-2).

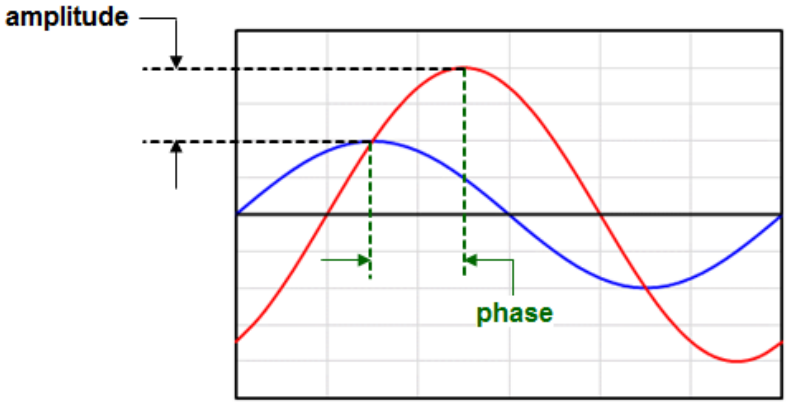


Figure 3-2: Amplitude and Phase Variability

Both types of uncertainty are present in the round robin dataset, as shown in Figure 3-1. For the case of WRS, amplitude variability is equivalent to variability in stress magnitude at corresponding local extrema. Phase variability has a spatial context (e.g., how the depth at the local maximum differs among various predictions). The methodology proposed in this section accounts for these two types of variability.

3.2.5 Modeling Amplitude and Phase Variability

Registration is the process of aligning the data horizontally and thus removing phase variability. Specifically, this process aligns the local extrema. Registration is accomplished through the use of warping functions, γ [13]. Warping functions are chosen such that the boundaries of the original functions are preserved. For the case of WRS, this means that the inner diameter (ID) and outer diameter (OD) stresses of the smoothed data are retained in the transformed functions. The other requirement for warping functions is that both γ and γ^{-1} are differentiable, so that the transformed function is smooth and can be mapped back to the original function [13]. Removing the phase variability via the warping functions allows characterization of the amplitude variability. The warping functions themselves provide a useful characterization of the phase variability.

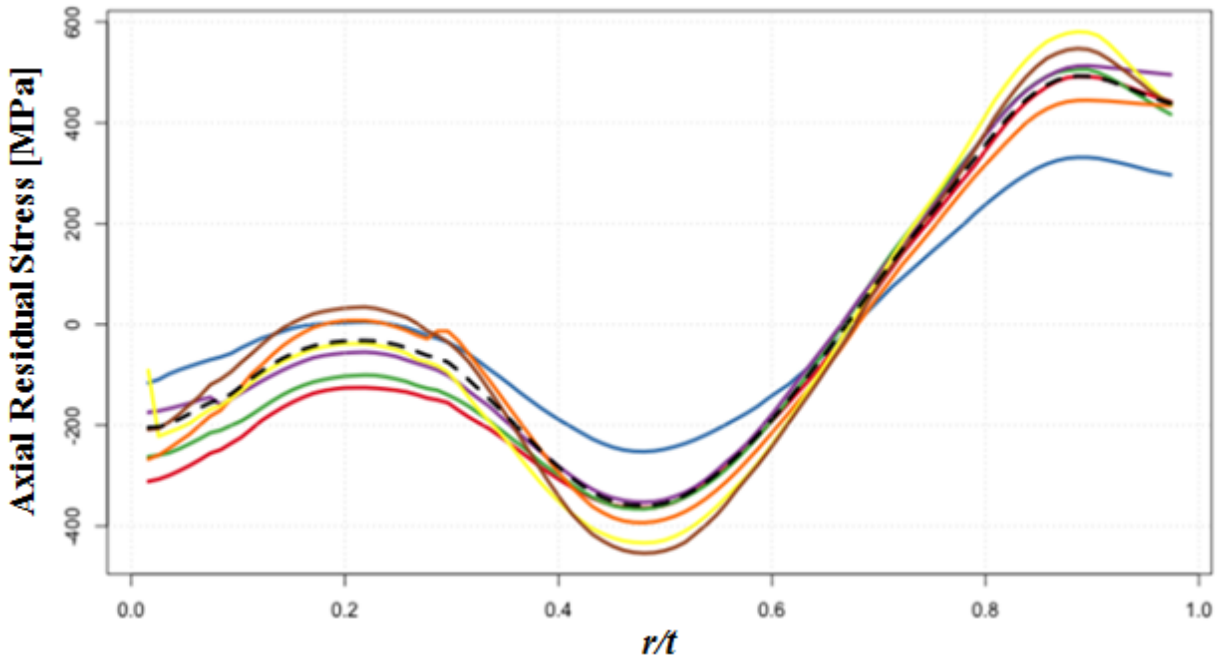


Figure 3-3: Axial Isotropic Data after Alignment

Functional principal components analysis (fPCA) is a dimension reduction technique to model the dominant modes of variation in the aligned data. The mathematical details and fundamental concepts of fPCA are better described elsewhere [13]–[16]. fPCA was applied to both the registered data and the warping functions to construct a statistical model of the residual stress data. The model allows for statistical sampling. Figure 3-4 shows 100 sampled profiles from the model constructed from the seven isotropic hardening axial stress predictions. The black curves in Figure 3-4 are the smoothed WRS predictions from the Phase 2b study. The modeled profiles demonstrate amplitude and phase variability similar to those of the original sample.

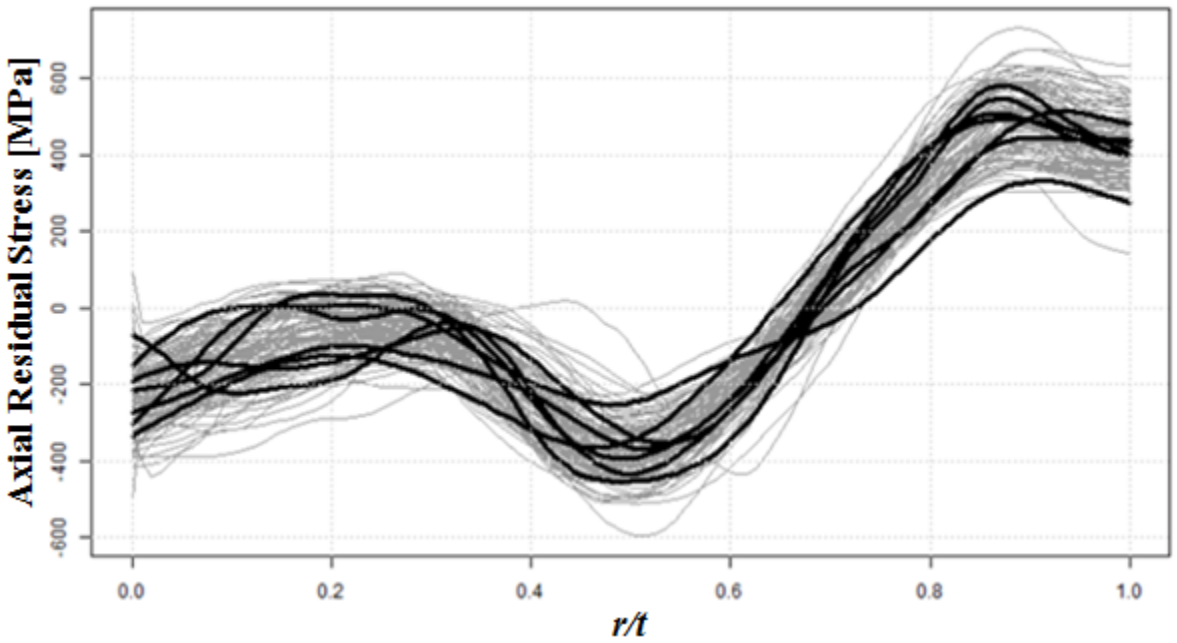


Figure 3-4: 100 Sampled WRS Curves Based upon Round Robin Modeling Data

3.2.6 Bootstrapping

Bootstrapping is a statistical sampling technique that provides a method to estimate uncertainty in distribution parameters, such as the mean. Further details of bootstrapping are described in [13]–[16]. This technique is applied here, along with the model described in Sections 3.2.1–3.2.5, to determine confidence bounds and tolerance bounds related to the Phase 2b round robin dataset, with mathematical details provided in [13]. The results are applied in Chapter 5 to draw conclusions about modeling recommendations and to inform development of a validation procedure for FE prediction of WRS.

3.2.7 Uncertainty Characterization of the Measurement Data

The statistical model summarized in Sections 3.2.1–3.2.5, while presented in the context of the modeling data, can also be applied to the measurement data. The DHD data consisted of four measurements around the circumference of the mockup, 90° apart from one another. The statistical model for bootstrapping was constructed from the four measured stress profiles.

Figure 3-5 shows the axial contour data again. As the figure suggests, many one-dimensional stress profiles through the weld centerline can be extracted from the axial contour dataset. Because of this unique feature of the axial contour data, it was not necessary to construct a statistical model or to perform bootstrapping.

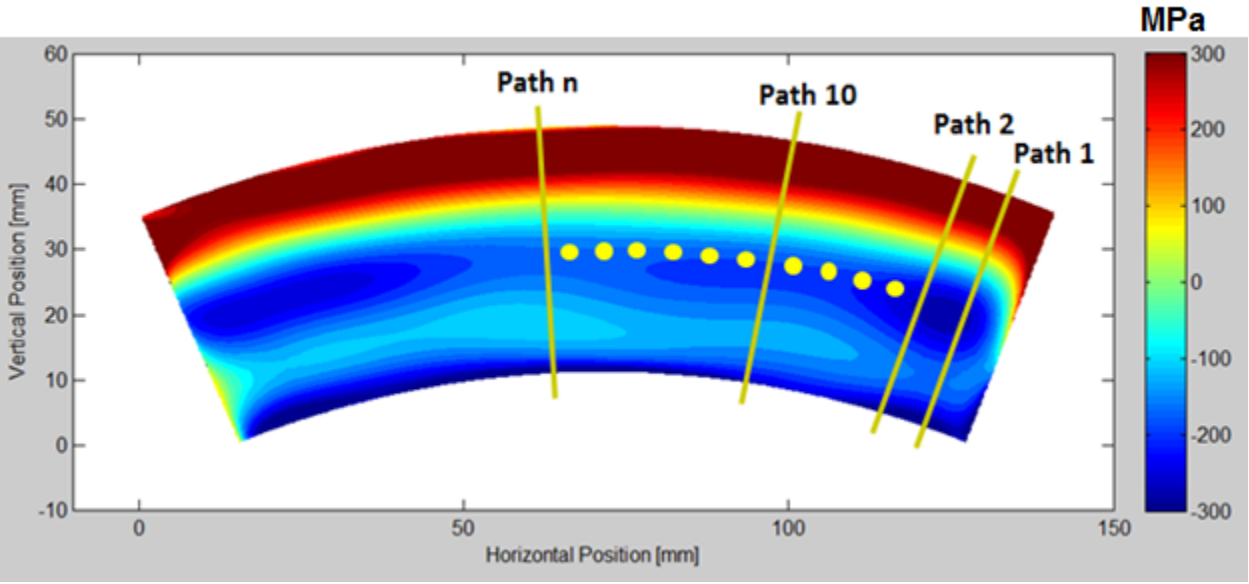


Figure 3-5: Contour Axial Stress Data

3.2.8 Tolerance Bounds versus Confidence Bounds

Bootstrapping was employed to construct both tolerance bounds and confidence bounds related to the round robin dataset. A more rigorous treatment of these statistical bounds is found elsewhere [21]. For the purposes of this document, a high-level definition of the statistical bounds determined in this work will suffice. Confidence bounds provide intervals within which a particular statistic is expected to lie. For instance, the confidence bounds on the mean prediction indicate the interval within which the true mean lies, with 95 % statistical confidence. Figure 3-6 gives examples of bootstrapped confidence bounds on the mean. Confidence bounds can be constructed on other statistics, such as quantiles.

Tolerance bounds are intervals within which 95 % of the residual stress data (either measurement or prediction) are expected to lie. The upper tolerance bound is constructed using the upper confidence bound on the 0.975 quantile. The lower tolerance bound is constructed using the lower confidence bound on the 0.025 quantile. The choice of the 0.975 and 0.025 quantiles leads to a coverage level of 95 %. Figure 3-7 shows examples of bootstrapped tolerance bounds.

3.3 Results

This section presents example results that illustrate the outcomes of the methods described in Section 3.2. Comprehensive results are presented in [13].

3.3.1 Uncertainty Quantification for the Prediction Data

Figure 3-6(a) shows bootstrap sample means compared against the smoothed axial stress FE prediction data for isotropic hardening. These results lead to 95 % confidence bounds on the prediction mean for these data, which are shown in Figure 3-6(b). Figure 3-7 shows similar plots for constructing the 95/95 tolerance bounds.

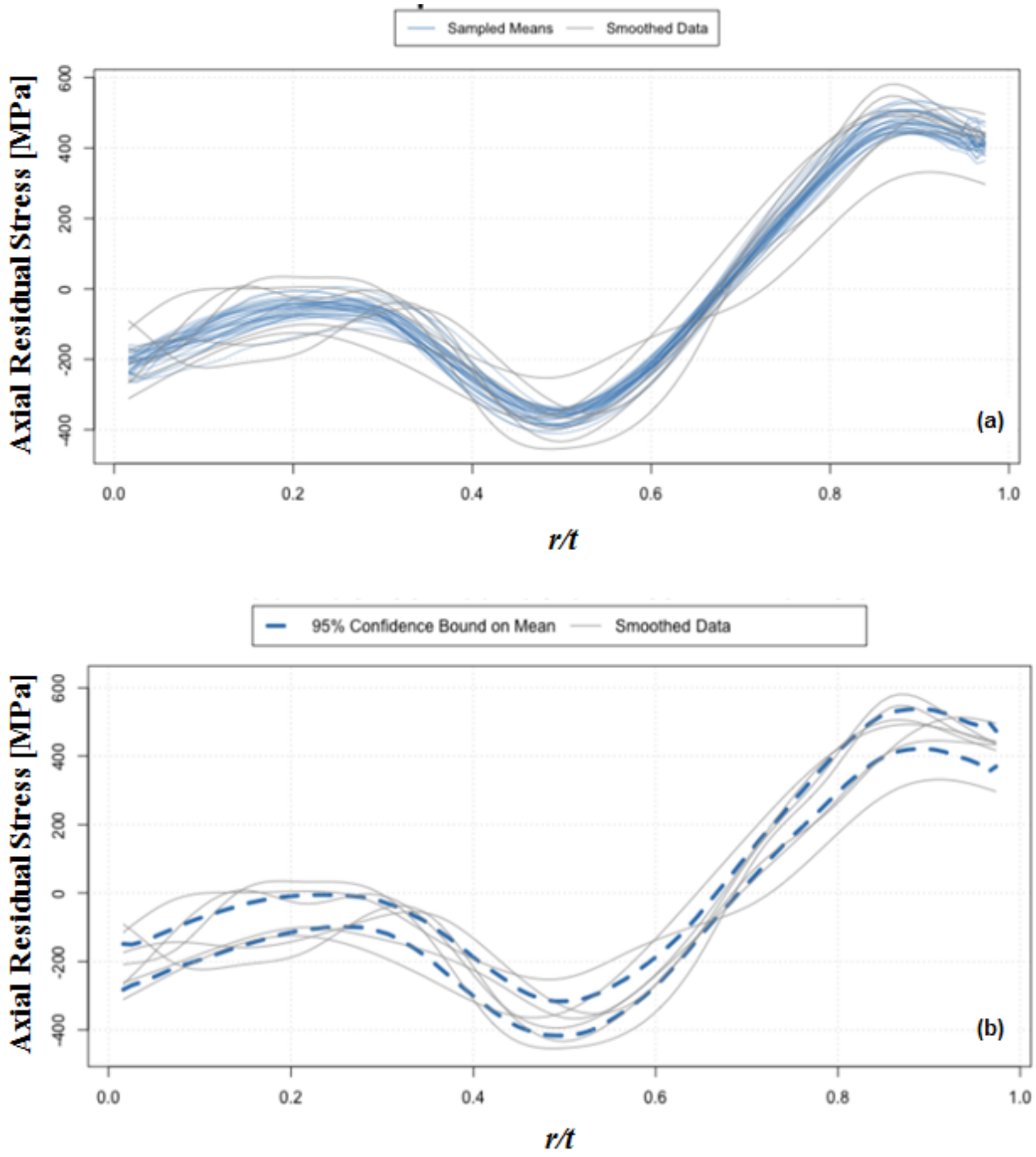


Figure 3-6: Constructing Confidence Bounds on the Mean (Axial, Isotropic Case) (a) 30 of the 1,000 Bootstrap Sample Means and (b) Resulting Confidence Bounds

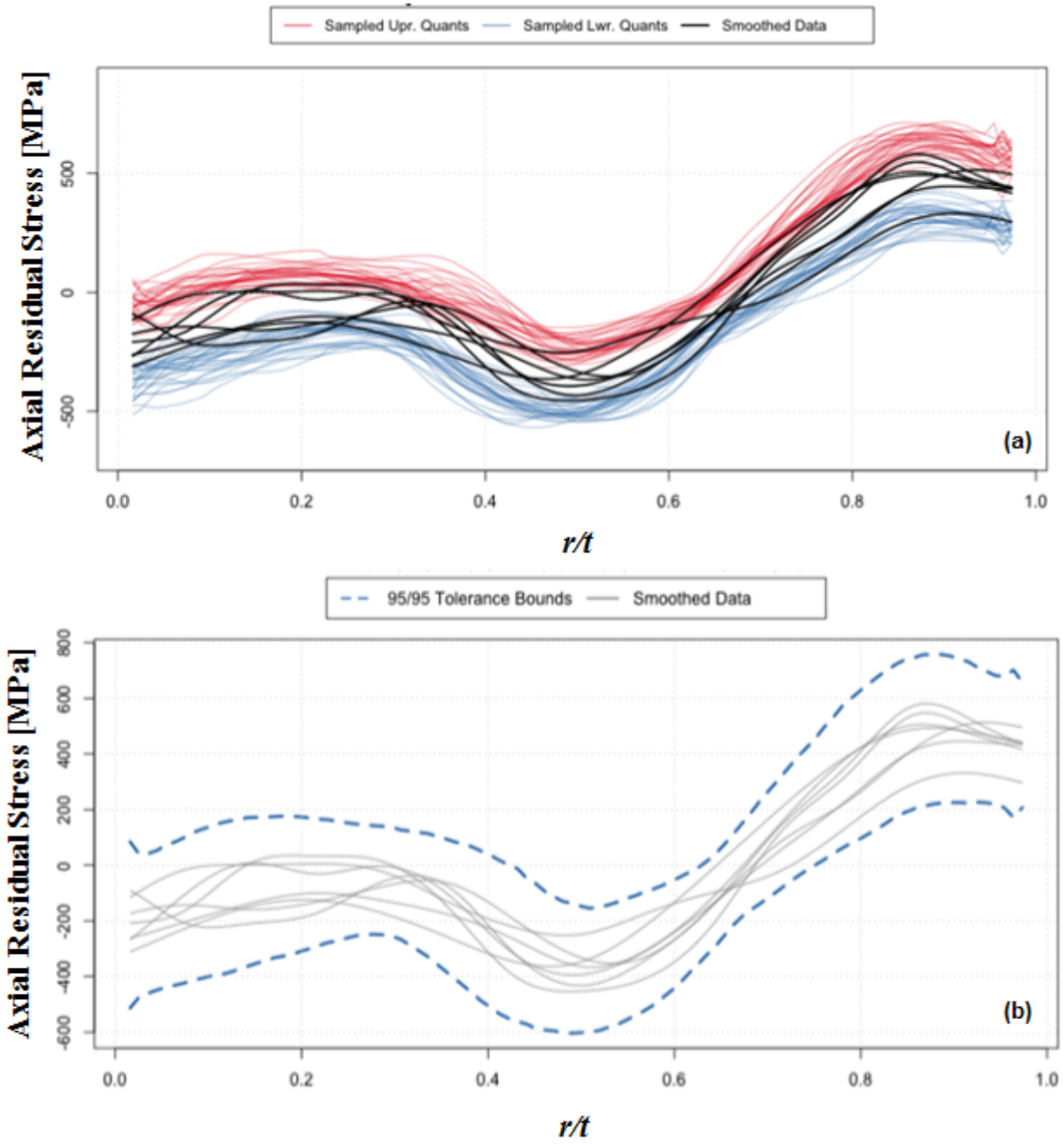


Figure 3-7: Constructing Tolerance Bounds (Axial, Isotropic Case)
(a) 30 of the 1,000 Bootstrap 2.5th and 97.5th Quantiles and (b) Resulting Tolerance Bounds

As introduced in Section 3.2.2, the analysis of the seven isotropic hoop stress predictions that passed the initial screening revealed that one prediction was strongly influencing the upper tolerance bound results. Figure 3-8 shows the potential outlier and demonstrates the significant impact it has on the 95/95 tolerance bounds. Specifically, the potential outlier suggests a roughly constant hoop stress prediction through the thickness. Since a majority of this dataset, including measurements (see Figure 2-7 and Figure 2-8) and modeling results, indicates some variation of hoop stress through the weld thickness, it may be appropriate to screen out this potential outlier when determining tolerance bounds for validation purposes.

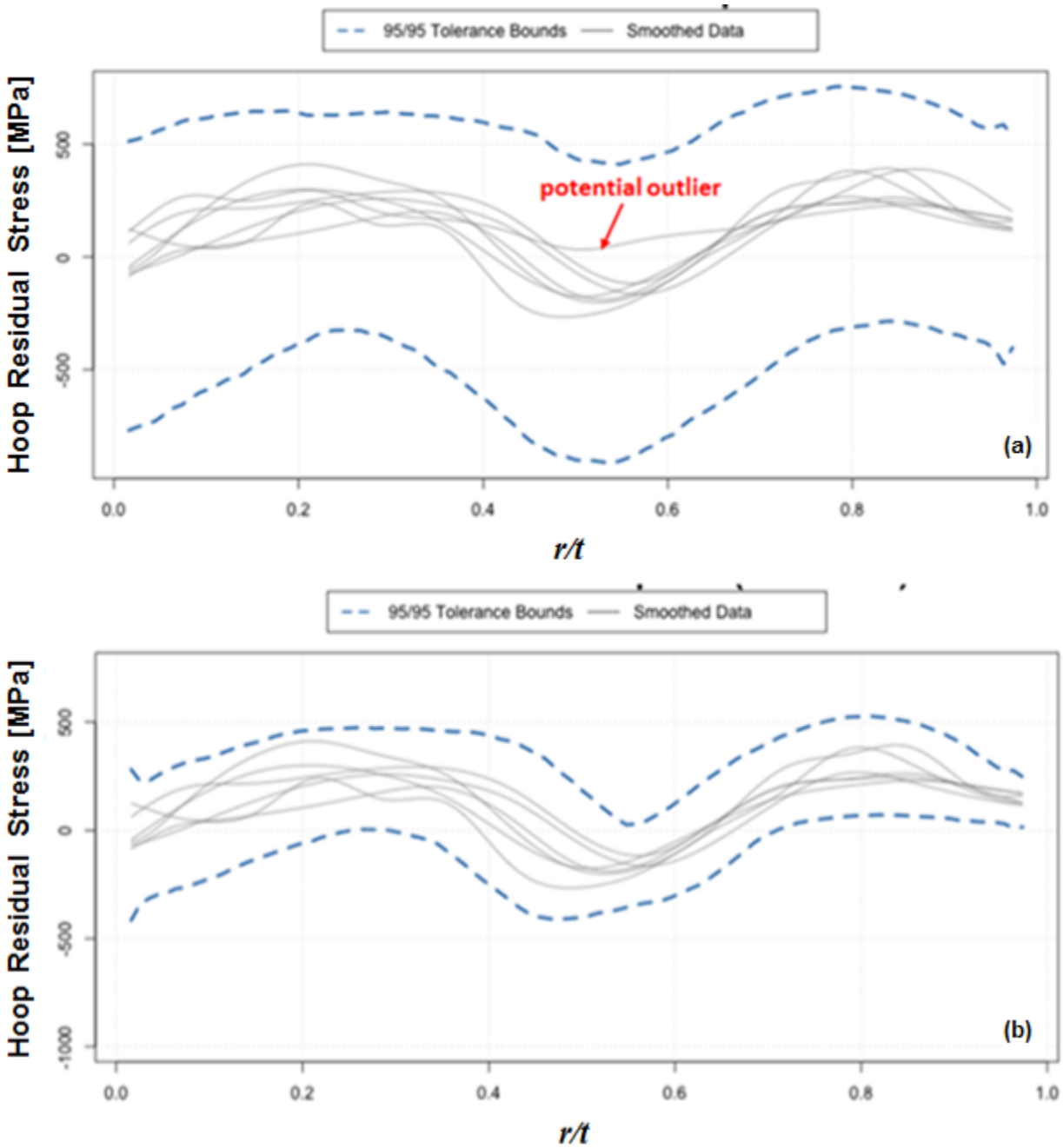


Figure 3-8: Bootstrap Tolerance Bounds on Isotropic Hoop Stress Predictions (a) With Potential Outlier and (b) Without Potential Outlier

3.3.2 Uncertainty Quantification for the Deep Hole Drilling Measurement Data

Figure 3-9 demonstrates the data smoothing process for axial DHD data. It shows that the smoothing residuals increase beyond a normalized depth of 0.6. This is because the measurement data were obtained at relatively coarse spatial increments near the OD. This data fitting issue may lead to less confidence in the bootstrap quantities determined for all DHD data beyond r/t=0.6.

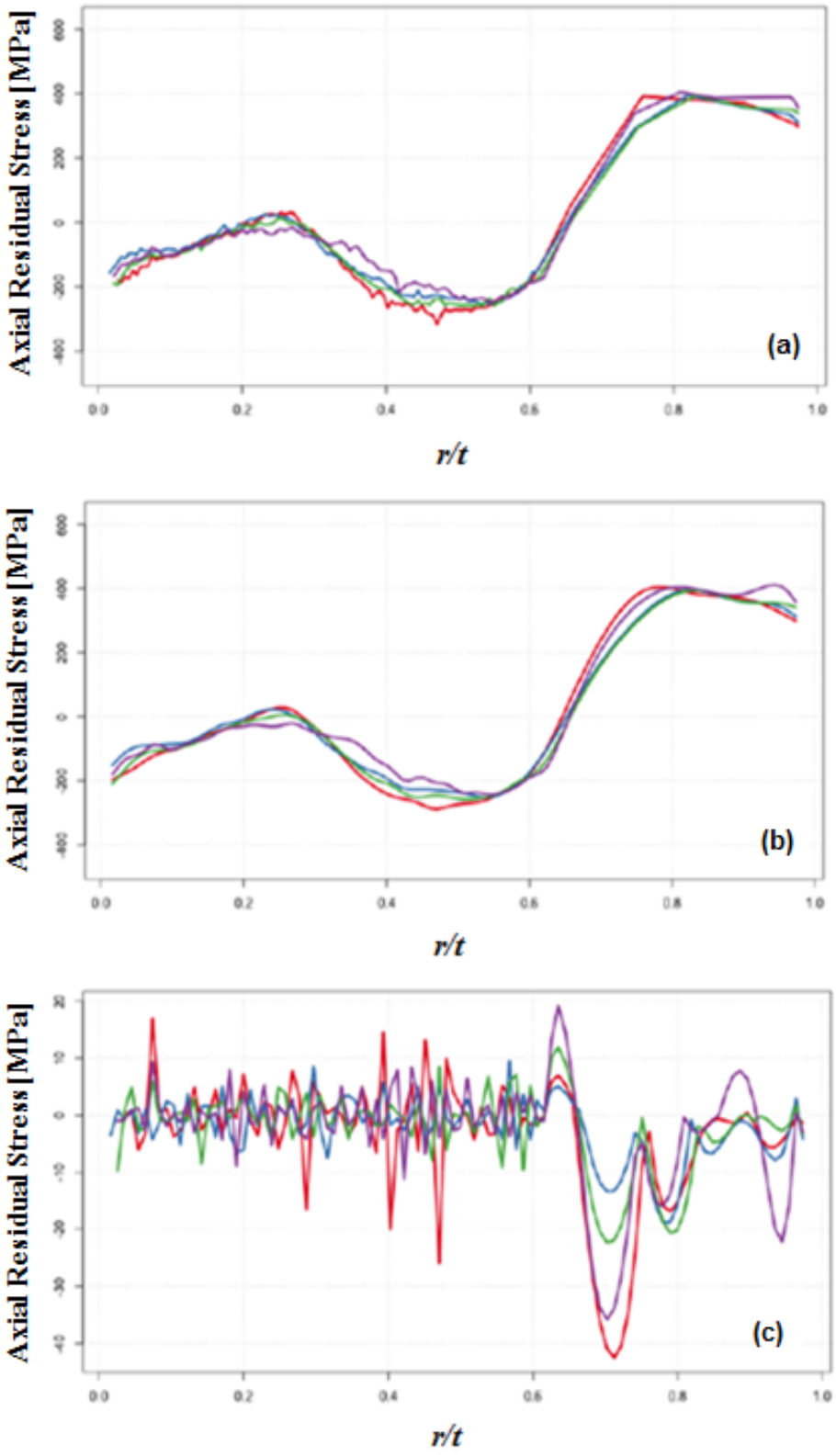


Figure 3-9: Data Smoothing for Axial DHD Data
(a) Raw Data, (b) Smoothed Data, and (c) Residuals

Figure 3-10 and Figure 3-11 show the bootstrap confidence bounds and tolerance bounds for the DHD axial stress measurement data, respectively.

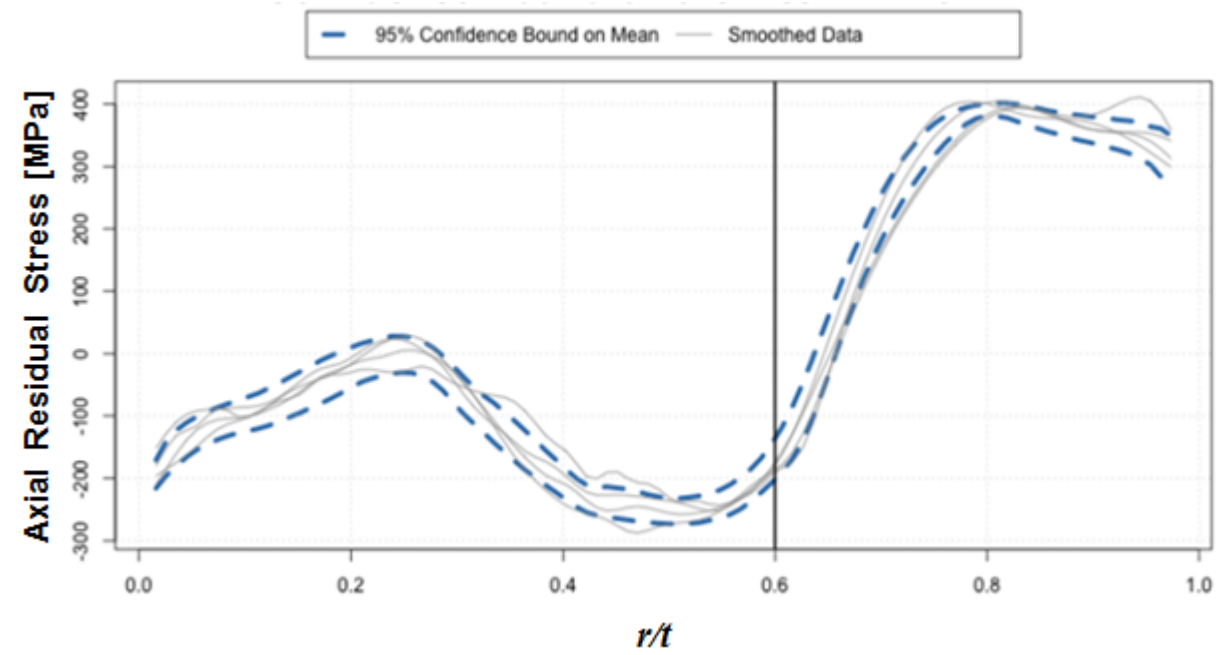


Figure 3-10: Confidence Bounds on the Mean (Axial DHD Data)

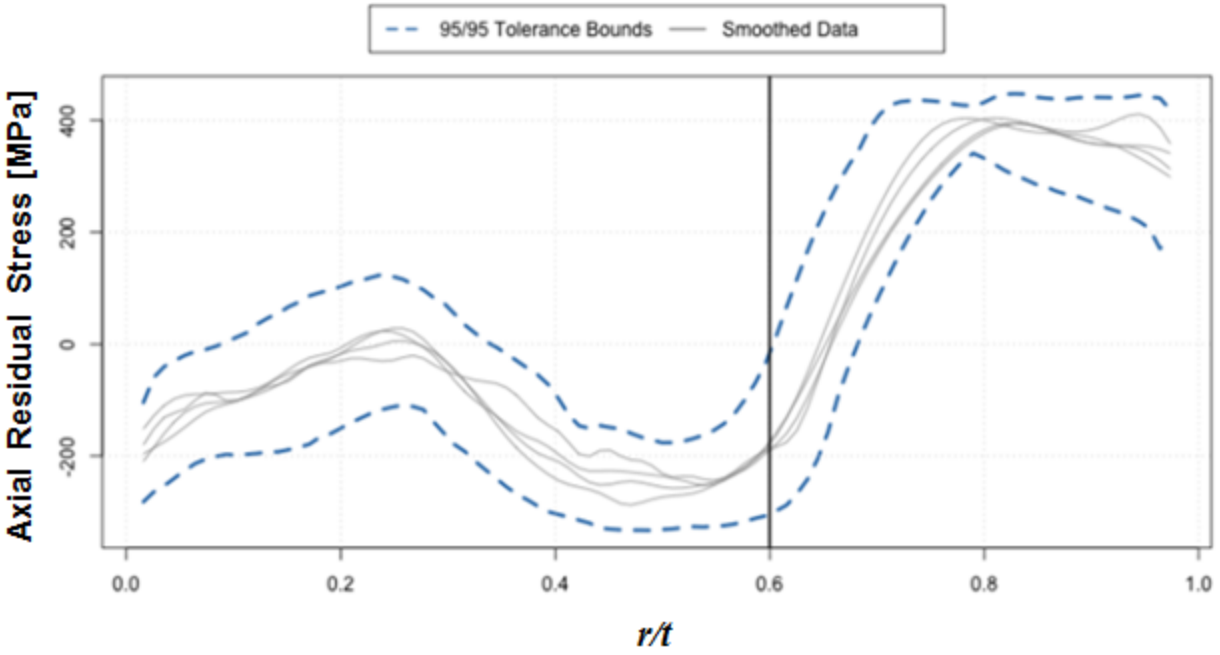


Figure 3-11: Tolerance Bounds (Axial DHD Data)

Figure 3-12 and Figure 3-13 show the bootstrap confidence bounds and tolerance bounds for the DHD hoop stress measurement data, respectively.

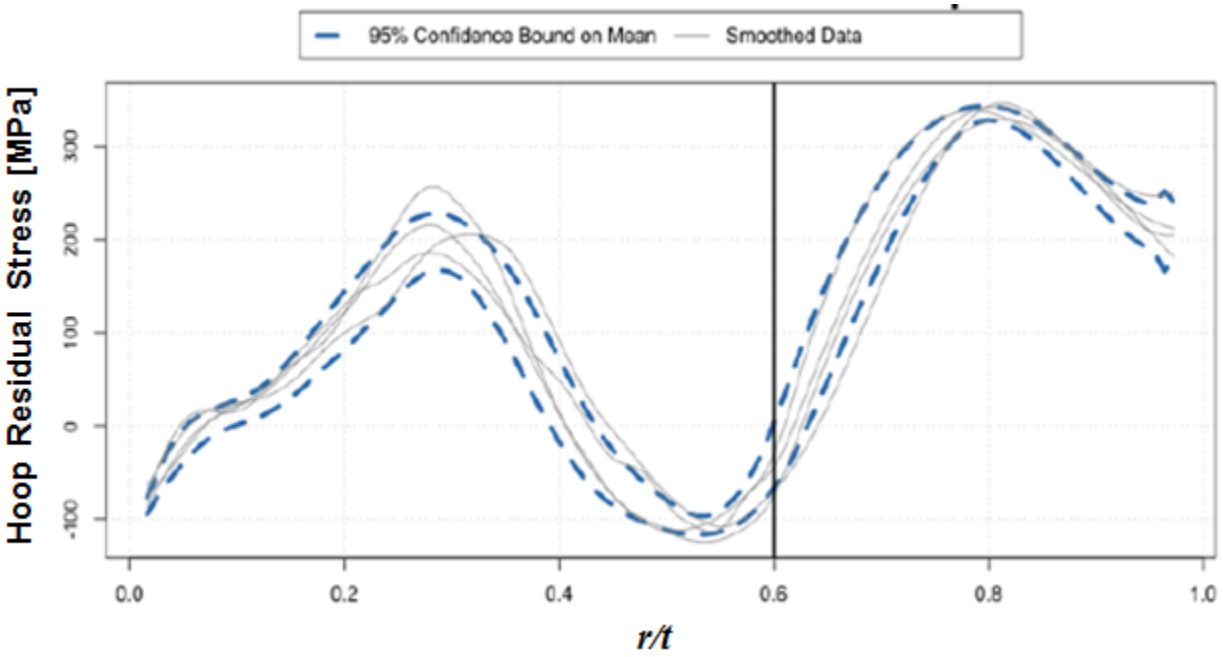


Figure 3-12: Confidence Bounds on Mean (Hoop DHD Data)

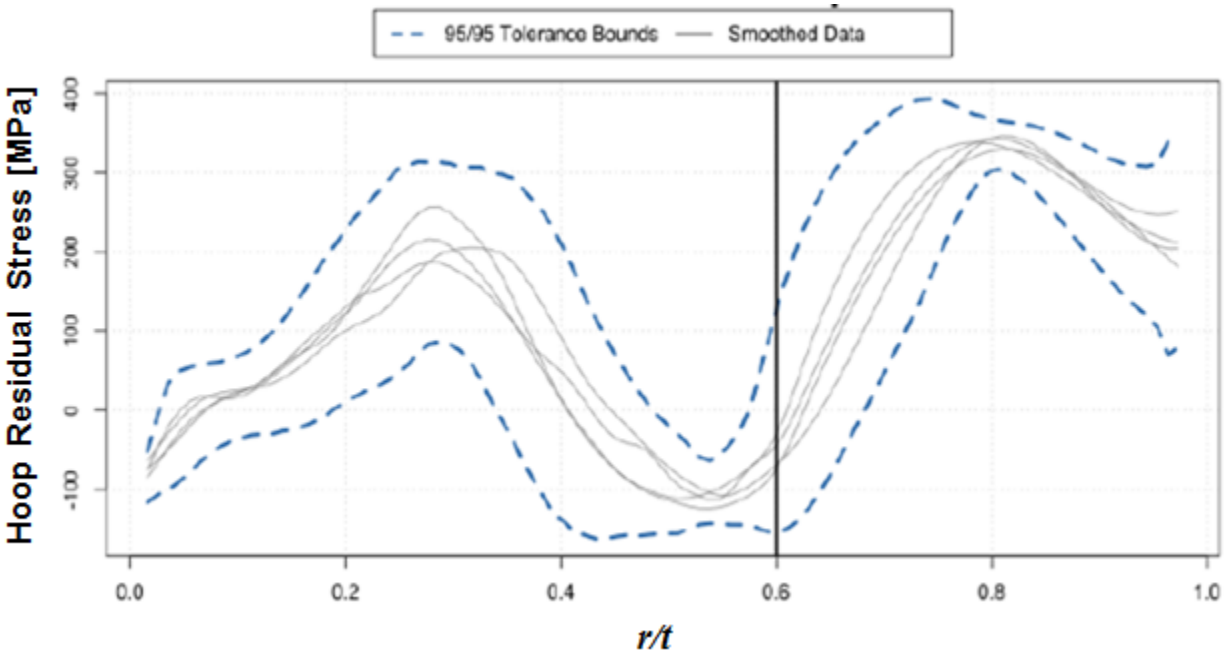


Figure 3-13: Tolerance Bounds (Hoop DHD Data)

3.3.3 Uncertainty Quantification for the Contour Measurement Data

As mentioned in Section 3.2.7, the axial stress contour data did not require construction of a statistical model and bootstrapping to quantify the uncertainty in the data. The cross section shown in Figure 3-5 is located entirely along the dissimilar metal weld centerline (also see the illustration in Figure 2-5). This centerline data can be used to directly compare to the prediction stress profiles by extracting path data, as illustrated in Figure 3-5. The linear paths defined for data extraction should be perpendicular to the boundaries of the data. Figure 3-14 shows example stress profiles extracted from the contour data, along with the profiles provided by the measurement vendor.

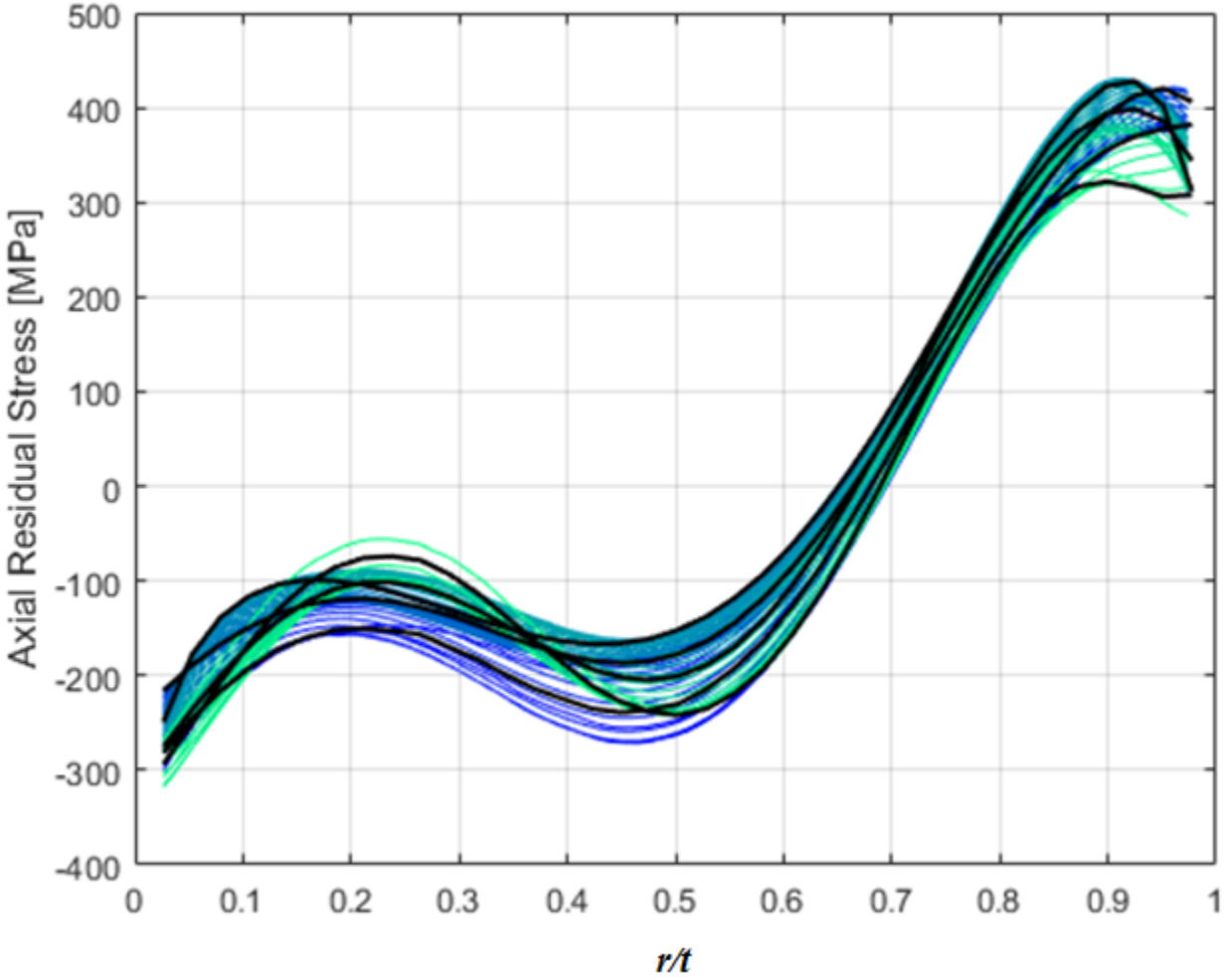


Figure 3-14: 50 Extracted Stress Profiles

While Figure 3-14 shows 50 example stress profiles, the analysis procedure used 500 extracted profiles in order to have adequate statistics to determine confidence bounds and tolerance bounds. Repeating the analysis with 5,000 profiles did not change the results. Figure 3-15 compares the tolerance bounds based on two different methods. One method is based on the five stress profiles provided by the contour measurement vendor. The other method is based on extracting 500 stress profiles without modeling and bootstrapping. Figure 3-15 shows that using the extracted 500 stress profiles leads to tighter tolerance bounds.

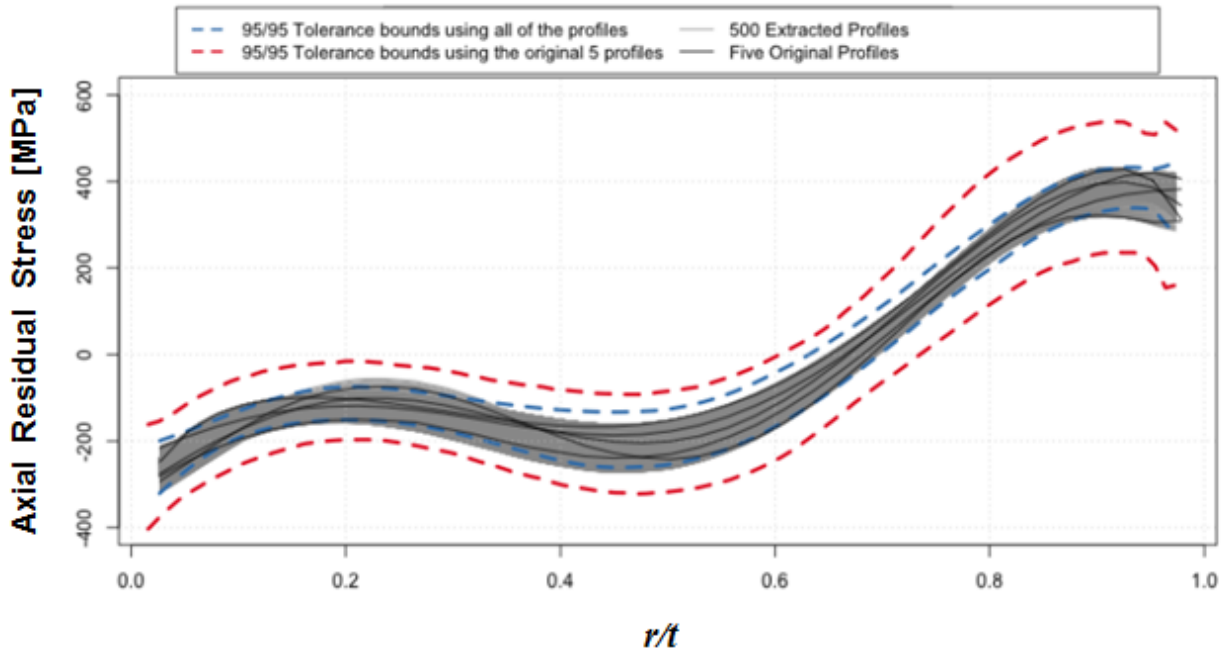


Figure 3-15: Tolerance Bounds for Axial Contour Data.

3.4 Conclusions

This chapter presented a methodology for quantifying uncertainty in the round robin WRS dataset. A statistical model was constructed for the modeling data and the DHD measurement data. The statistical model enabled bootstrap estimates of confidence bounds and tolerance bounds. For the axial contour measurement data, it was not necessary to employ bootstrap techniques. Instead, 500 curvilinear stress profiles were extracted from the contour measurements. Use of these profiles allowed direct determination of the mean and tolerance bounds. This method improves on the previous work [1], where uncertainty was described only in subjective terms. The results of this analysis will help inform the development of a validation process in Chapter 5.

4 WRS IMPACT ON FLAW GROWTH CALCULATIONS

4.1 Regulatory Application

NRC regulations require owners of nuclear power plants to periodically perform nondestructive examinations of safety-related piping according to the American Society of Mechanical Engineers *Boiler and Pressure Vessel Code* (ASME Code), Section XI [22], and ASME Code Case N-770 [23]. If the exam discovers an indication, then the geometry of the potential flaw is compared to the acceptance standards of Section XI, IWB-3500. If the flaw is not allowable, then the licensee must either repair or replace the piping or perform an analytical evaluation according to IWB-3600 for temporary acceptance of the flaw. Connected flaws on the inner surface are generally not allowed in service because of PWSCC [22]. The NRC has granted short-term regulatory relief to licensees, usually in cases where inservice inspection requirements present a demonstrated hardship to the plant owner (see [24] as one example). Figure 4-1 summarizes this process.

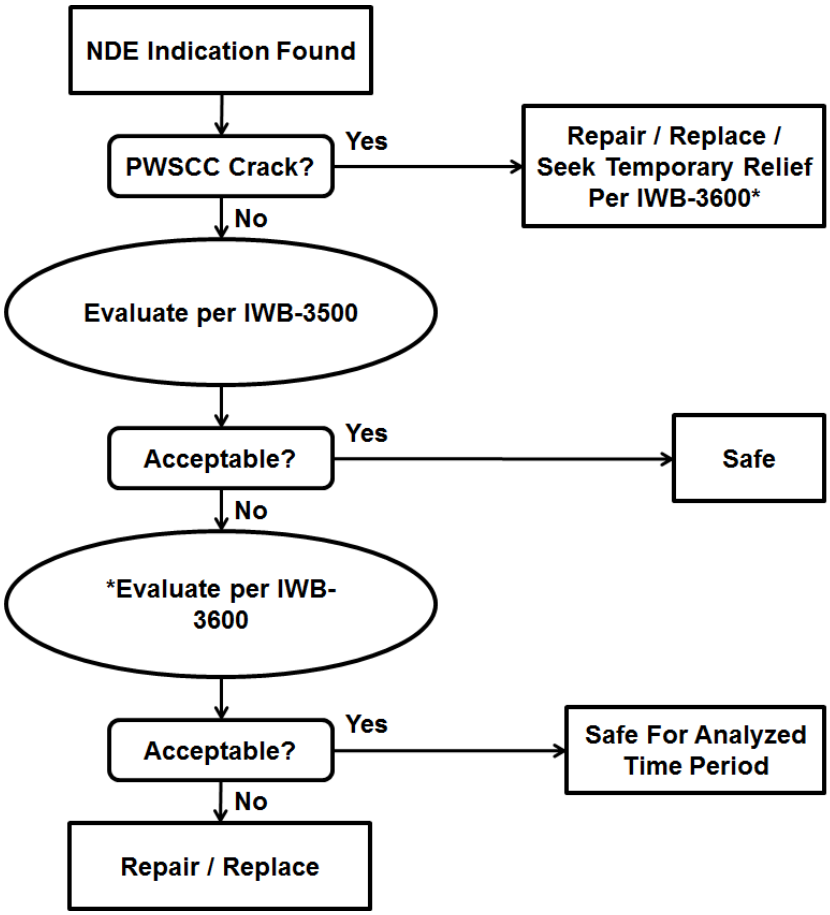


Figure 4-1: ASME Code Flaw Disposition Procedure

An analytical evaluation per ASME Code, Section XI, IWB-3600, involves an engineering estimate of the growth of the flaw within an established timeframe. ASME Code, Section XI, Nonmandatory Appendix C, provides guidance and equations for many aspects of a flaw evaluation in piping, including the following:

- pipe stress
- acceptance measures
- screening for failure mode
- flaw stability
- crack growth rate laws [25]

Nonmandatory Appendix A of Section XI, paragraph A-3000, contains stress intensity factor (SIF) solutions that may be applied to piping. Figure 4-2 outlines the basic procedure of a flaw growth evaluation.

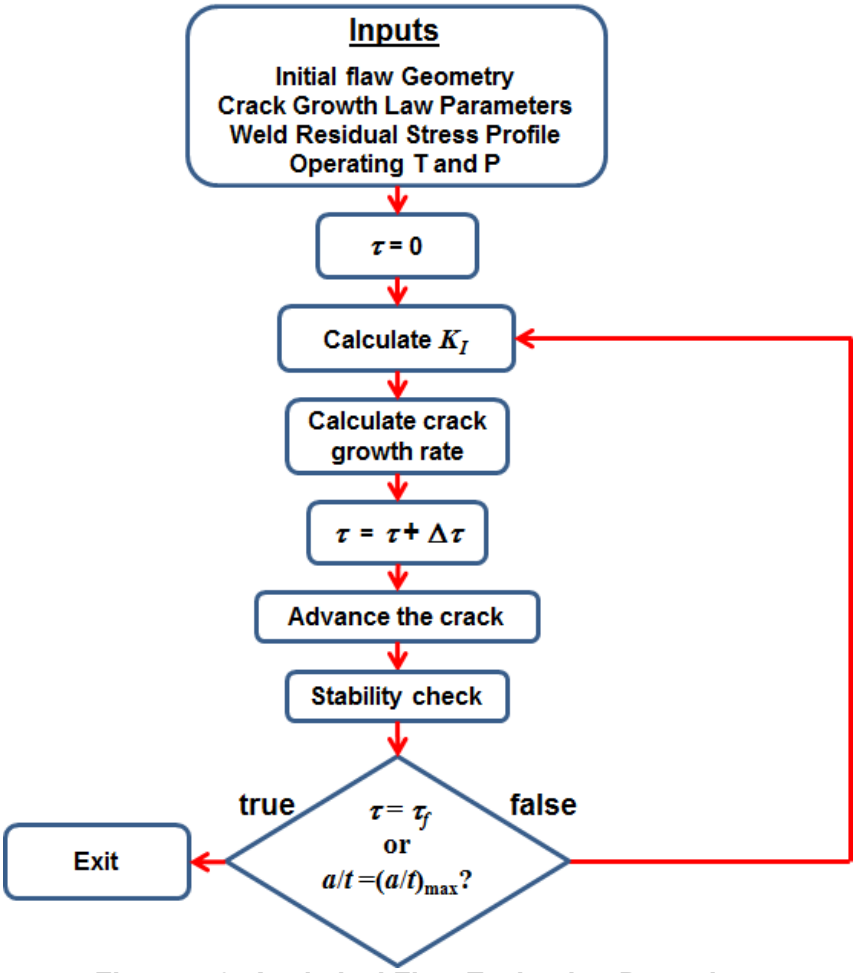


Figure 4-2: Analytical Flaw Evaluation Procedure

As Figure 4-2 shows, the analytical evaluation requires an assumption about WRS. In many licensee submittals, an FE model serves as the basis behind the assumed WRS. The focus of this chapter is to examine how the WRS input affects the flaw growth calculation.

4.2 Inputs

This work draws on a series of flaw growth calculations documented in the technical letter report on the Phase 2b study [7]. Table 4-1 shows the inputs for this work.

Table 4-1: Inputs for Flaw Growth Calculations

OD [mm]	t [mm]	Weld Width [mm]	a_0 [mm]	$2c_0$ [mm]	T [°C]	p [MPa]	σ_m [MPa]	σ_b [MPa]
381	36.07	26.48	3.607	7.214	315.6	15.5	60	100

OD – outer diameter
t – pipe wall thickness
a_0 – initial flaw depth
$2c_0$ – initial flaw length
T – operating temperature
p – operating pressure
σ_m – operating membrane stress
σ_b – operating bending stress

The dimensions in Table 4-1 are consistent with the Phase 2b mockup geometry given in Chapter 2. This chapter will examine the case of a circumferential flaw subjected to the WRS profiles determined by the axial WRS measurements in the Phase 2b study, as shown in Figure 2-7a and Figure 2-9.

4.3 Superposition of Stresses

Figure 4-3 shows the residual stress profiles, with σ_m and σ_{cfp} (the crack face pressure stress) overlayed on the figure. A representative contour measurement of axial residual stress is also included on the figure.

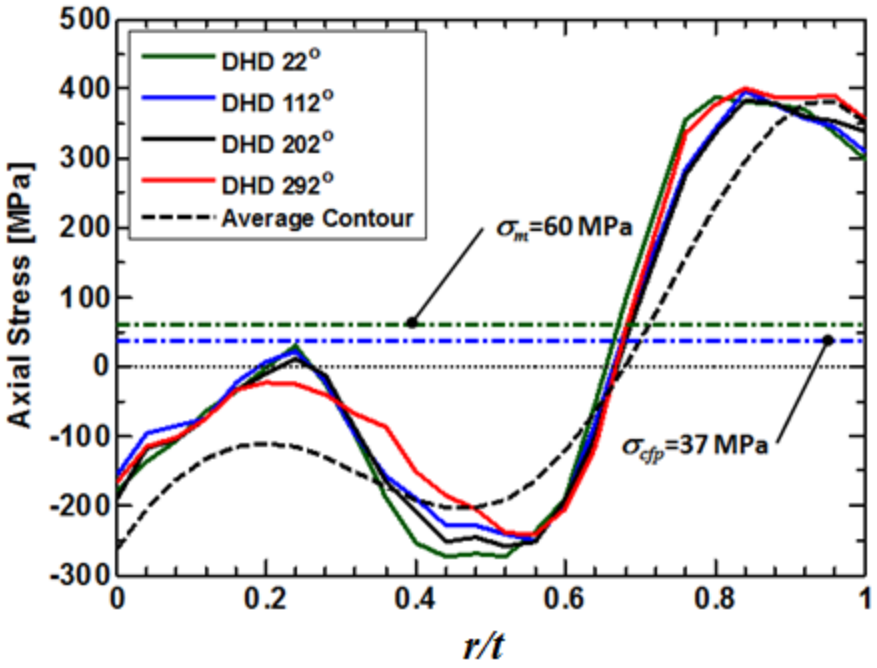


Figure 4-3: Loads from Various Sources

In the method applied here, σ_m, σ_{cfp}, and WRS were superimposed for the purpose of calculating the SIF. Figure 4-4 shows the results of this superposition, along with annotations of salient features of the curves.

The ID stresses (i.e., at r/t=0) were compressive for each curve, with the contour measurement being the most compressive at -166 MPa. The stress at the initial crack depth (i.e., r/t=0.1) was tensile for the DHD curves and compressive for the contour measurement. The highest stress at r/t=0.1 was observed in the 112° curve at 18 MPa.

Three of the four DHD curves crossed zero for the second time at r/t=0.32, with very little spread. The 292° curve deviated from this trend by crossing zero at r/t = 0.37. The local maximum around r/t=0.2 for the 292° curve, however, was less than that of the other three DHD curves. The DHD curves crossed zero a third time at roughly r/t=0.64, although there was noticeable spread about this value. Finally, the contour curve crossed zero only one time, beyond the mid-thickness of the pipe and just ahead of the DHD curves.

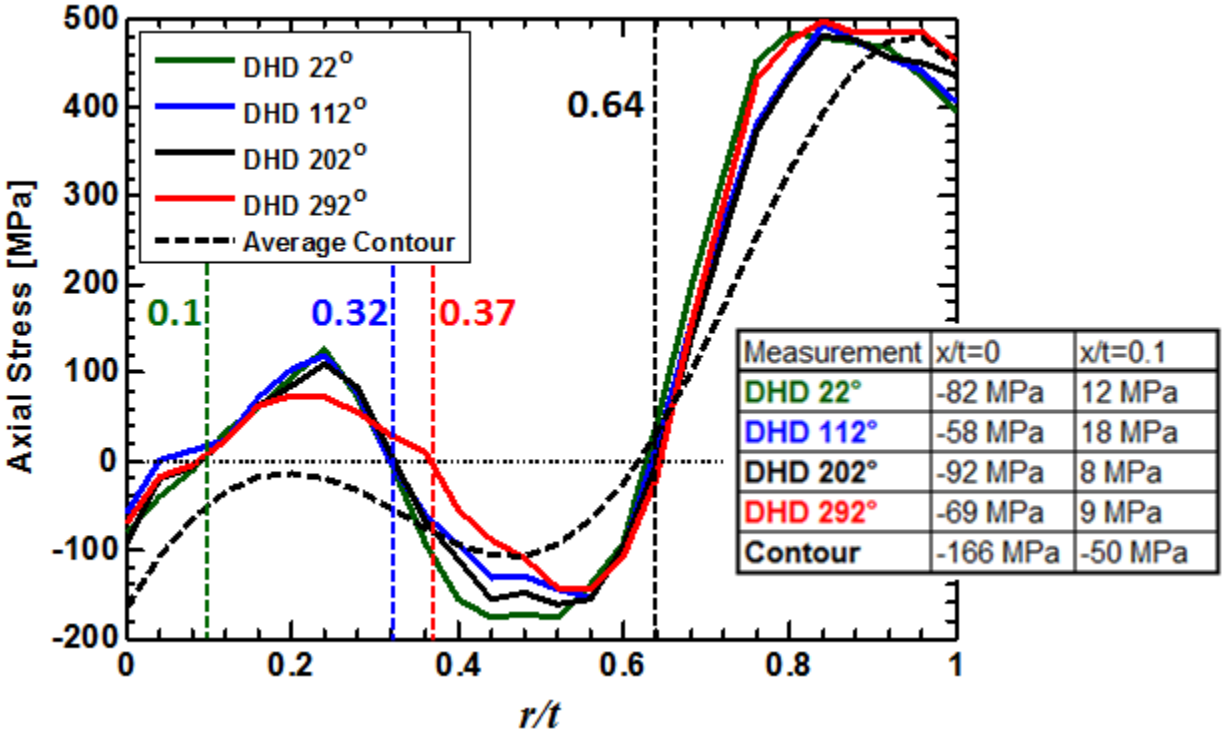


Figure 4-4: Superposition of Membrane, Crack-Face Pressure, and Weld Residual Stresses

4.4 Stress Intensity Factor and Crack Growth

The total SIF was the sum of that stemming from global bending stress and that stemming from the remaining stresses (i.e., σ_m, σ_{cfp}, and WRS). The SIF for σ_b was calculated with influence coefficients for global bending [26], according to Equation 4-1.

$$K_I = \sigma_b G_b \sqrt{\frac{\pi a}{Q}} \qquad \textbf{Equation 4-1}$$

where G_b is the influence coefficient and Q is the flaw shape parameter. In this case, σ_b was considered to be the maximum bending stress occurring at the top dead-center location of the pipe as a result of the applied bending moment.

The SIF for the remaining stresses was determined with the Universal Weight Function Method [27]-[28]. The superimposed stress profiles shown in Figure 4-4 were input into the calculation as discrete arrays. As such, there was no need for a polynomial fit of the stress profile, as is sometimes the practice. The basic form for this SIF solution is shown in Equation 4-2.

$$K_I = \int_0^a h(x,a)\sigma(x)dx \qquad \textbf{Equation 4-2}$$

where $h(x,a)$ is the weight function. In this work the SIF was calculated at both the deepest point (K_{90}) and the surface point (K_0) of the flaw, as shown in Figure 4-5.

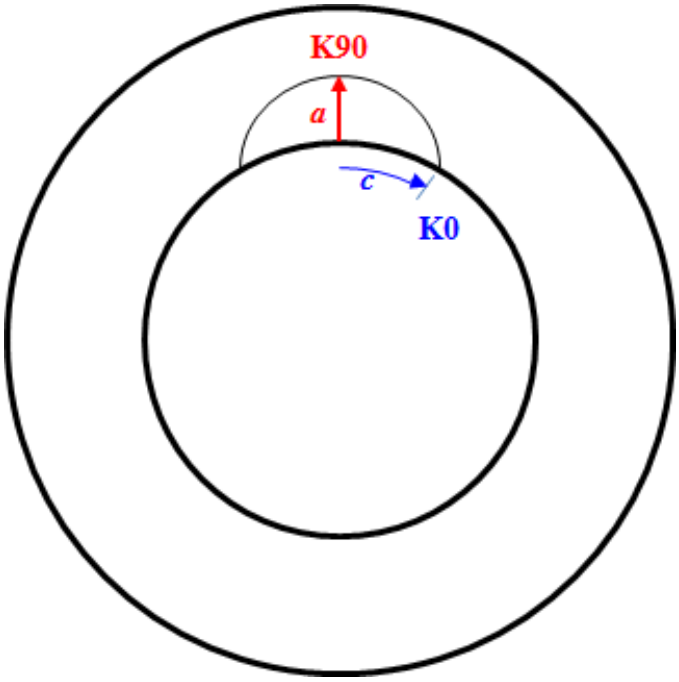


Figure 4-5: SIF at Two Locations along Crack Front

SIF calculations are necessary to use the established crack growth law, which is based on laboratory crack growth experiments on compact tension fracture mechanics specimens [25]. The equation describing the crack growth rate is given in Equation 4-3.

$$\frac{da}{d\tau} = \exp\left[-\frac{Q_g}{R_g}\left(\frac{1}{T_{abs}} - \frac{1}{T_{ref}}\right)\right]\phi\left(K_I - K_{Ith}\right)^{\eta} \qquad \textbf{Equation 4-3}$$

Table 4-2 defines the symbols in Equation 4-3.

Table 4-2: Symbol Definition for Equation 4-3

$da/d\tau$	crack growth rate
K_I	stress intensity factor
K_{Ith}	stress intensity factor threshold
Q_g	activation energy
R_g	ideal gas constant
T_{abs}	absolute operating temperature
T_{ref}	emperical reference temperature
ϕ	tabulated crack growth coefficient (see ASME Section XI, Appendix C)
η	tabulated crack growth exponent (see ASME Section XI, Appendix C)

4.5 Flaw Growth Results

Figure 4-6 shows K_{90} and a/t versus time for the flaw growth calculation. No growth resulted from the calculation based on the contour measurement. The DHD SIF curves exhibited similar trends of increasing to a peak early in time and subsequently decreasing to a plateau. The 112° and 292° curves showed a rapid increase in SIF later in time. Correspondingly, the 112° and 292° curves showed through-wall crack growth, while the 22° and 202° curves demonstrated crack arrest for the time period analyzed. However, a typical relief request submitted to the NRC is only concerned with timeframes of less than 20 years or 240 months, as discussed further in Section 4.6.

The 112° curve started at the highest K_{90} value and peaked the earliest in time at about 100 months and 14 MPa√m. It plateaued at a value of roughly 3.5 MPa√m, which was noticeably higher than the other three curves. After 400 months, the SIF steadily increased, followed by a rapid increase at 480 months. The flaw growth in the depth direction responded to the trends in SIF just described. Hence, when the SIF plateaued at 3.5 MPa√m, the flaw grew linearly in time.

The other three DHD curves exhibited similar trends, but with a few differences. The peak in SIF was shifted to later times relative to the 112° measurement. This peak was also of a slightly lower magnitude, but still in the range of 12–13 MPa√m. The 22° and 202° curves did not exhibit the sharp increase in SIF in the time period analyzed here.

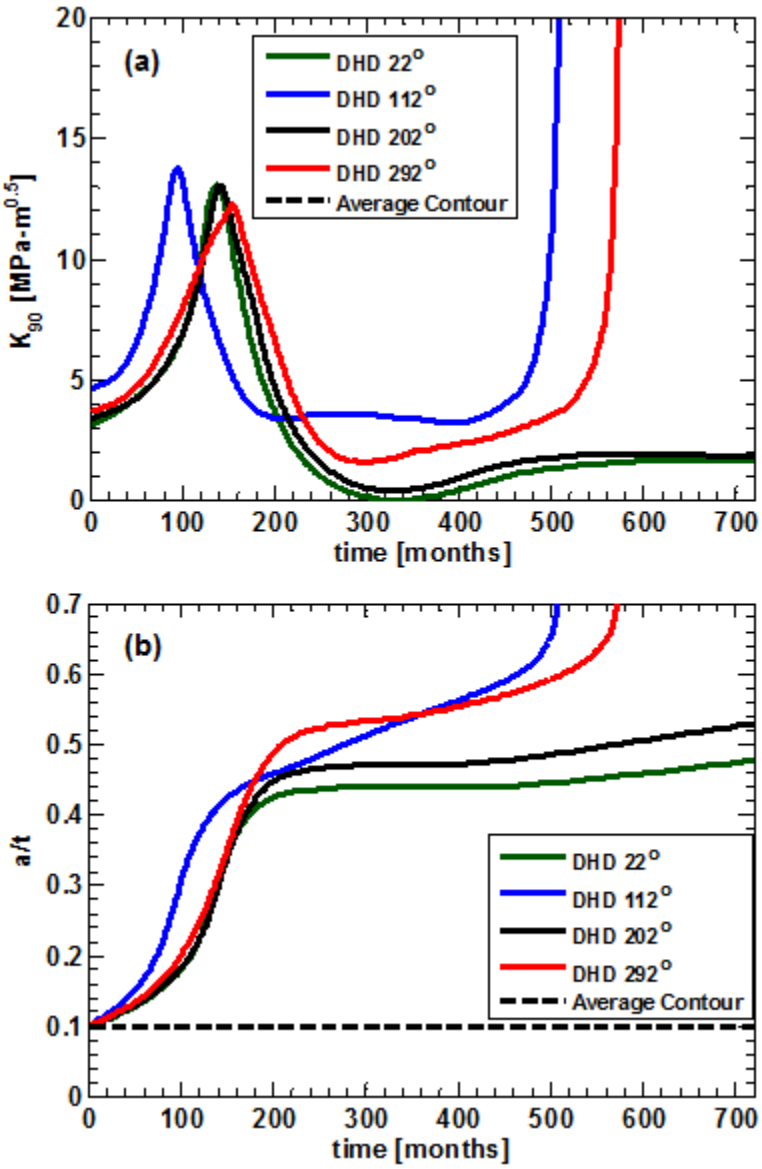


Figure 4-6: (a) K_{90} and (b) Growth in Depth Direction

Figure 4-7 shows K_0 and $2c/C$, where C is the inner circumference of the pipe, versus time for the analytical flaw evaluation. The SIF value was highest for the 112° measurement, followed by 292°, 202°, and 22°. The crack length grew relatively slowly during the first 100 months for the 112° measurement. After this time, the length steadily grew to 6 % of the circumference. The other curves exhibited similar trends but did not grow to the same extent.

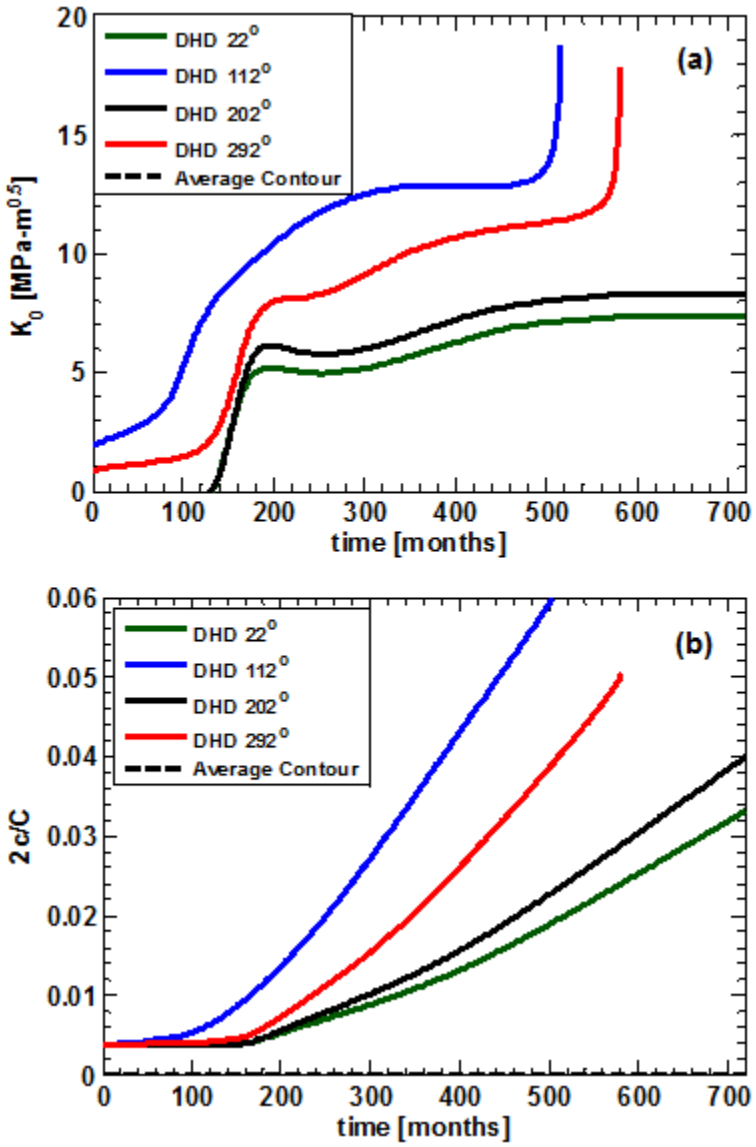


Figure 4-7: (a) K_0 and (b) growth in length direction

4.6 Discussion

Figure 4-6 shows the apparent uncertainty in flaw growth calculations resulting from the residual stress assumption. However, regulatory relief submittals to the NRC do not evaluate 720 months of operation. In fact, evaluation periods may extend only one or two refueling outages (1.5 to 3 years). Figure 4-8 shows that the uncertainty in the results decreases for shorter evaluation periods.

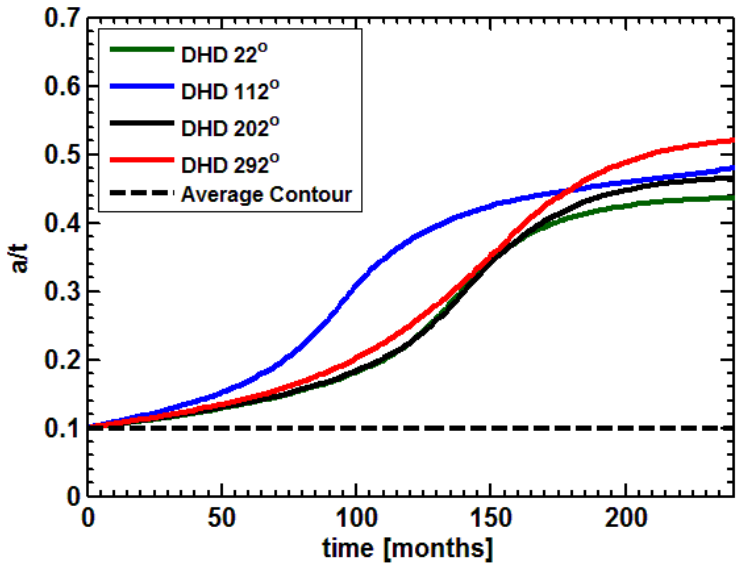


Figure 4-8: Flaw Growth after 20 Years

However, it is still useful to examine the reasons behind the apparent sensitivity to residual stress assumption evident in Figure 4-6. As Figure 4-4 shows, the 112° case remained within 50 MPa of the other DHD measurements for the first 30 % of the wall thickness. Even so, the calculated flaw growth roughly doubled the others with a/t=0.3 at 100 months for the 112° curve. Only the 112° and 292° cases showed through-wall crack growth. The 292° residual stress led to through-wall growth despite peaking below 100 MPa at r/t=0.2, which is in stark contrast to the remaining curves. This section will seek to explain how features of the assumed residual stress profile may affect the calculated flaw growth behavior.

The SIF for membrane and residual stresses is given by the integral shown in Equation 4-2. As a first approximation, this integral is similar to the area under the curves of Figure 4-4. Figure 4-9 shows the area under these curves as calculated by the trapezoidal rule.

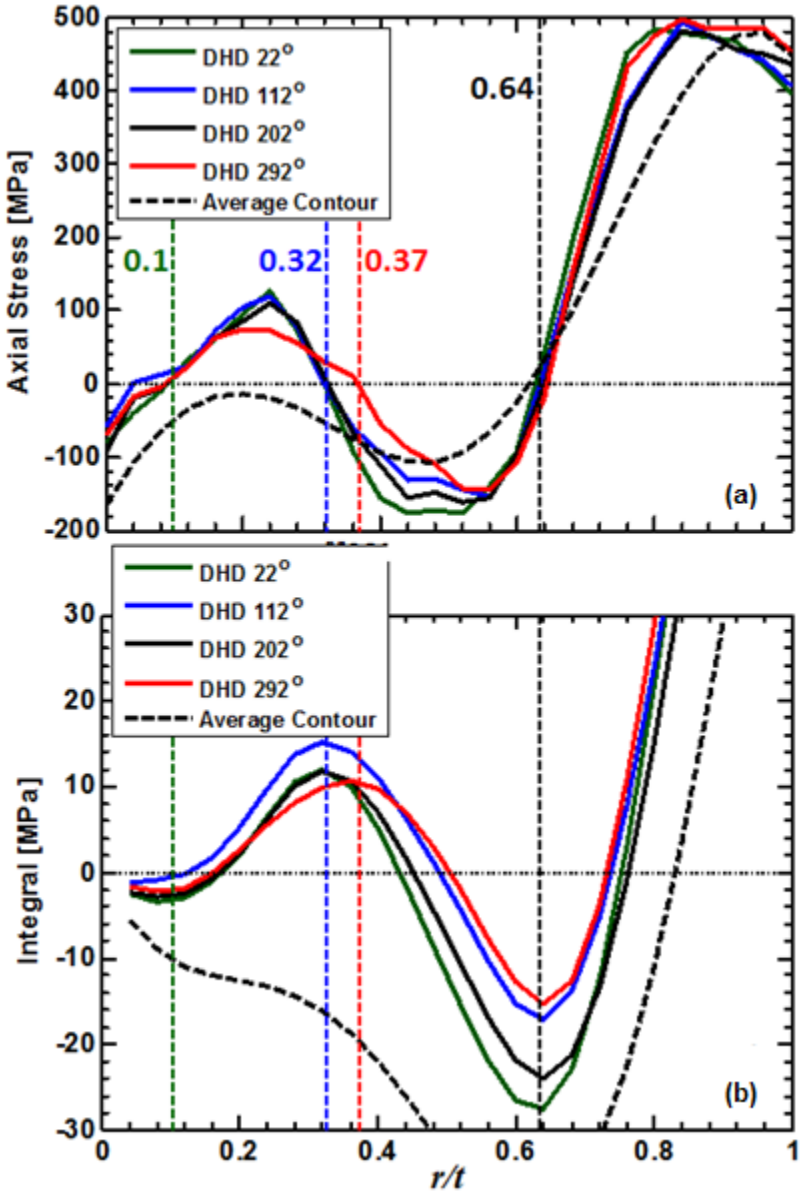


Figure 4-9: (a) Membrane Stresses, (b) Area under the Curve

The area under the curve in Figure 4-9(b) reveals greater contrasts among the various measurements than the stress profiles in Figure 4-9(a). The area under the 112° curve remains clearly highest until about half way through the wall thickness, where the 292° curve becomes slightly dominant. The 292° curve, despite peaking at a lower stress magnitude than the other measurements, remains tensile for the greatest depth. This fact keeps the SIF high enough through the calculation to allow for through-wall growth by 720 months. Once the crack tip reaches the compressive zone (e.g., r/t=0.32), the 22° and 202° measurements show a steeper drop in both stress magnitude and area under the curve. This leads to crack arrest in these two cases.

4.7 Conclusion

This chapter confirms previous work showing that flaw growth calculations are sensitive to the residual stress input [29]. Features of the residual stress curves that affected the results included the following:

- the ID stress magnitude
- the stress magnitude at initial flaw depth
- the location through the wall thickness where the compressive zone started
- the slope of the curve in the compressive zone

While Chapter 5 presents a validation methodology for WRS FE models, this chapter demonstrates that simplified approaches for judging the adequacy of residual stress inputs may also be applied. Where confirmatory analyses of residual stress predictions are practical, the listed features can be compared and contrasted to provide confidence in assumed inputs. Estimating the area under the curve may also provide additional insights. While current approaches to flaw evaluation appear to be adequate, a validation methodology for WRS FE models may ease regulatory uncertainty and review times for relief requests.

5 VALIDATION PROCEDURE AND FINITE ELEMENT GUIDELINES

5.1 Introduction

This chapter draws on the uncertainty quantification methodology described in Chapter 3 to develop a validation procedure for FE predictions of WRS. Accompanying the validation procedure are guidelines for creating FE models of WRS. Together, the validation method and guidelines may increase confidence in WRS inputs in relief requests.

The procedure proposed here involves two aspects: (1) establishing and justifying modeling guidelines (Sections 5.2 and 5.3) and (2) proposing a series of quality metrics an analyst can calculate to objectively assess the quality of an individual FE prediction of WRS (Section 5.4). The discussion in Sections 5.2 and 5.3 assesses the Phase 2b dataset as a whole, while Section 5.4 focuses on a particular analyst seeking to validate an FE methodology.

The approach presented here requires an individual analyst to construct an FE model of the Phase 2b mockup, according to the guidelines given in [7]. A series of quality metrics and acceptance measures are proposed to validate the analyst's prediction. Since this process is aimed at deterministic fitness-for-service calculations (see Chapter 4), the metrics were designed to ensure acceptable predictions of flaw growth.

The Phase 2b dataset is directly applicable to axisymmetric (i.e., two-dimensional) models of dissimilar metal butt weld geometry. Extending this procedure to other geometries may require additional work, such as fabrication of a mockup and measurement of residual stress. Section 5.6 discusses further the validation of residual stress predictions in other applications.

5.2 Material Hardening Law

One topic identified in earlier work was the need to establish guidance on hardening law choice because of the significant impact this assumption has on the FE results [1]. This section describes measurement-model comparisons with the goal of making informed judgments about the appropriate approach to modeling material hardening during thermal cycling that occurs during welding operations. This section first describes a methodology to account for modeling and measurement uncertainty when making measurement-model comparisons, as developed in [13]. Sections 5.2.3 and 5.2.4 compare and contrast three approaches (isotropic, nonlinear kinematic, and the average of isotropic and kinematic) to hardening law.

5.2.1 Difference in Means and Root Mean Square Error Functions

One method used here to investigate the performance of various hardening law approaches is to examine the difference in means between the predictions and the measurements in the Phase 2b dataset. The methodology, described in [13], is summarized as follows.

1. Sample n_e measurement WRS functions and n_p prediction WRS functions from the model described in Chapter 3 on a fine grid of L values of d, where d is the normalized distance through the pipe wall thickness. This results in samples of WRS functions from both measurements and predictions representing uncertainty in both. Let f be a function representing sampled measurement WRS profiles and g be a function representing sampled prediction WRS profiles, as follows.

$f_i(d_k)$, k = 1,2,…,L and i = 1,2,…,n_e

$g_i(d_k)$, k = 1,2,…,L and i = 1,2,…,n_p

where i represents the i^{th} sample. The two functions are sampled independently.

2. Compute $h_s = (h_s(d_1), h_s(d_2), ..., h_s(d_k))$ where

$$h_s(d_k) = \frac{1}{n_e}\sum_{i=1}^{n_e} f_i(d_k) - \frac{1}{n_p}\sum_{i=1}^{n_p} g_i(d_k) \qquad \textbf{Equation 5-1}$$

   The subscript s represents the s^{th} difference in means calculation.

3. As a measure of prediction quality, calculate the root mean square error (RMSE) of h_s, which is defined as

$$RMSE_s = \sqrt{\frac{1}{L}\sum_{k=1}^{L} h(d_k)^2} \qquad \textbf{Equation 5-2}$$

4. Repeat steps 1, 2, and 3 S times. In this study, S = 1,000.

5. Compute the pointwise 0.975 and 0.025 quantiles of $h_s(d_k)$ and $RMSE$ over the S samples for each k. These quantiles form a pointwise 95 % bootstrap confidence bound for the population difference of means function and the $RMSE$.

This methodology allows consideration of both measurement and modeling uncertainty when assessing the quality of predictions.

5.2.2 Assessment of Prediction Trends

WRS predictions should capture variations in stress magnitude with spatial position. Within the context of the procedure outlined in Section 5.2.1, this means that the estimated confidence bounds on the mean difference function, $h_s(d_k)$, should encompass zero. If the confidence bounds on the mean difference function do not encompass zero, this implies a prediction bias. Figure 5-1 shows how well the nonlinear kinematic hardening FE results for axial stress predict the DHD measurements. Figure 5-1 clearly illustrates where the nonlinear kinematic predictions systematically over- and under-predict the DHD data in a statistically significant manner. In Figure 5-1(b), a positive mean difference implies an underprediction of the measurements.

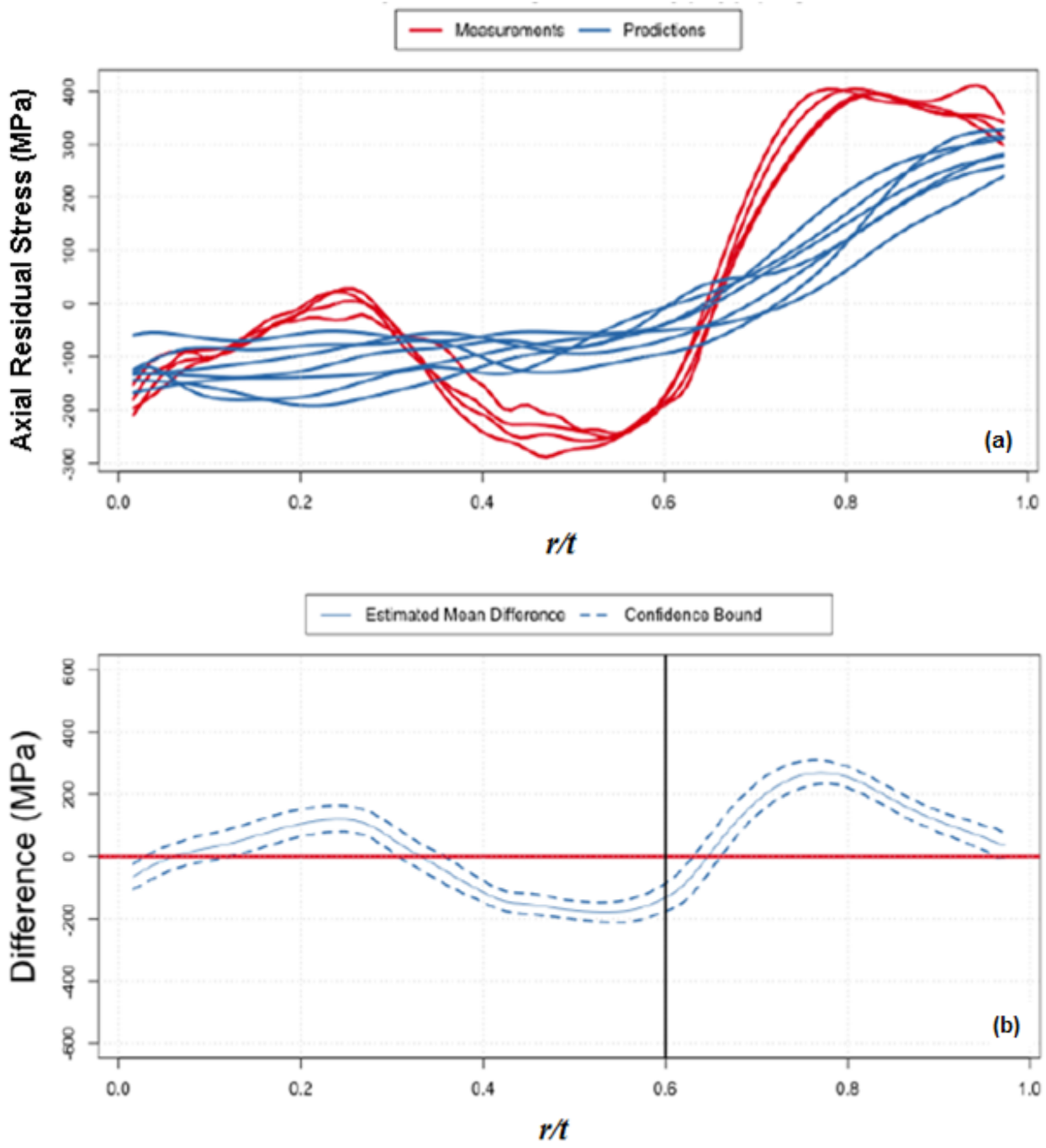


Figure 5-1: Nonlinear Kinematic Hardening Predictions against the DHD Measurements (a) Actual Data and (b) Mean Difference Function and Tolerance Bounds

Figure 5-2 and Figure 5-3 show similar figures for isotropic hardening and average hardening, respectively. Overall, the mean difference functions for these two cases remain closer to zero throughout the wall thickness than was observed for the nonlinear kinematic predictions. There are locations through the wall thickness where the tolerance bounds do not encompass zero, indicating certain trends that are not captured by the FE. In Figure 5-2(b) and Figure 5-3(b), for instance, the isotropic and average hardening results consistently underpredict the DHD measurements around r/t=0.75. A result such as that in Figure 5-2(b), while not perfect, indicates that the predictions are reasonable in a qualitative sense.

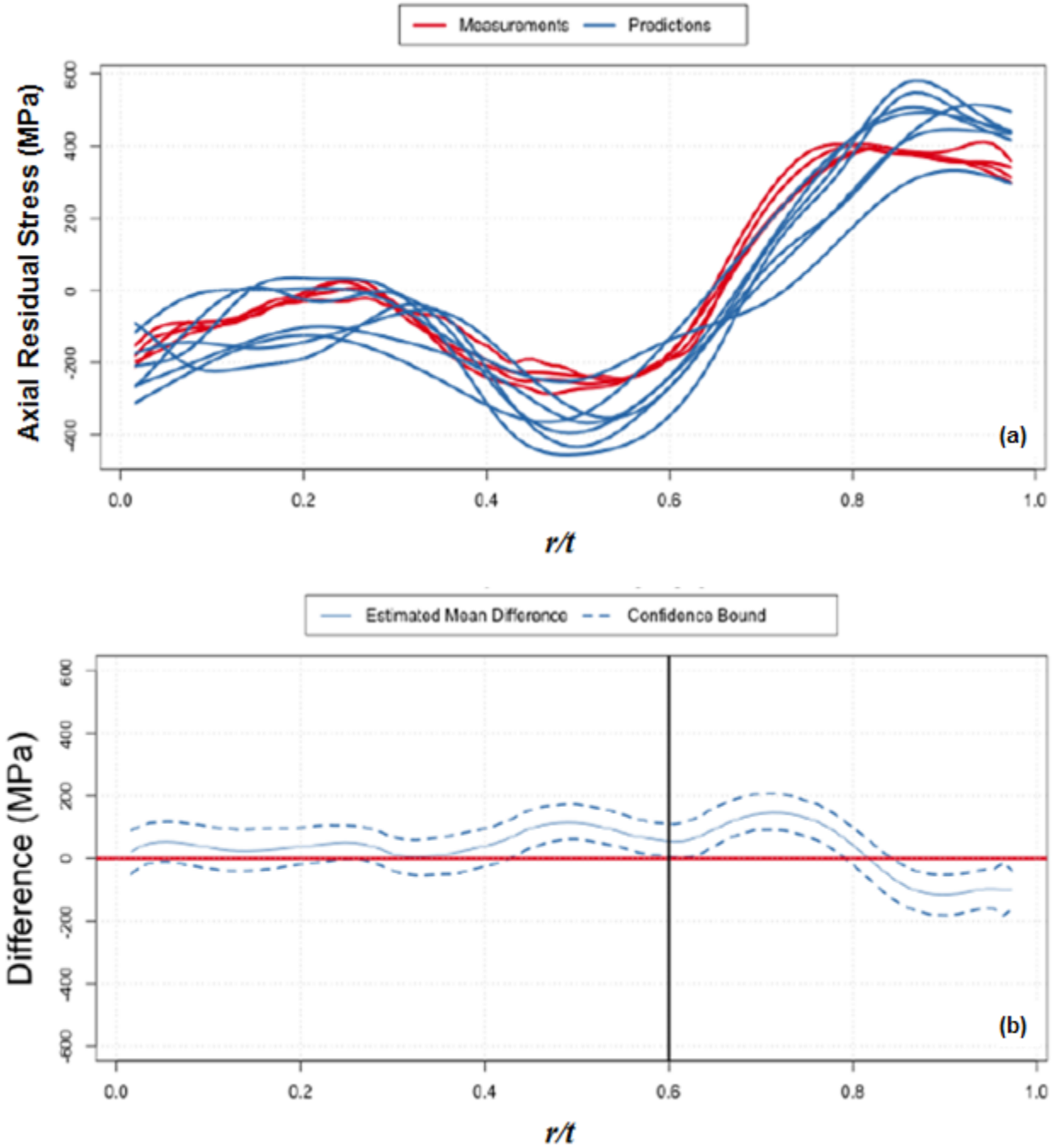


Figure 5-2: Isotropic Hardening Predictions against the DHD Measurements (a) Actual Data and (b) Mean Difference Function and Tolerance Bounds

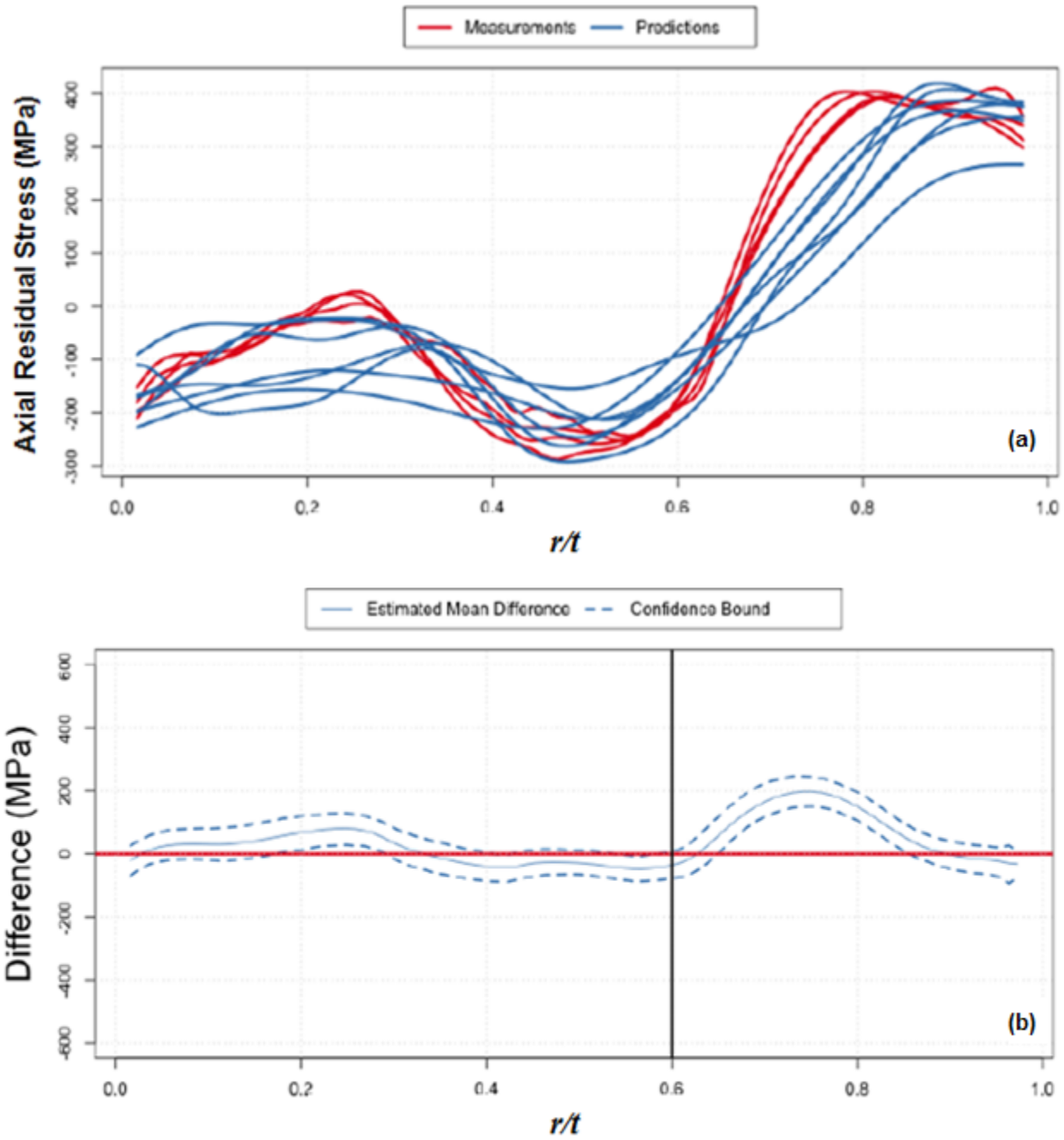


Figure 5-3: Average Hardening Predictions against the DHD Measurements (a) Actual Data and (b) Mean Difference Function and Tolerance Bounds

Appendix A shows other plots similar to Figure 5-1, Figure 5-2, and Figure 5-3 for all. Table 5-1 lists all the relevant cases and the corresponding location of the data plots.

Table 5-1: Benchmark Cases and Their Location in Appendix A

Case	Stress Component	Hardening Law	Benchmark	Figure in Appendix A
1	Axial	kinematic	DHD	Figure A-1
2	Axial	isotropic	DHD	Figure A-2
3	Axial	average	DHD	Figure A-3
4	Hoop	kinematic	DHD	Figure A-4
5	Hoop	isotropic	DHD	Figure A-5
6	Hoop	average	DHD	Figure A-6
7	Axial	kinematic	Contour	Figure A-7
8	Axial	isotropic	Contour	Figure A-8
9	Axial	average	Contour	Figure A-9
10	Hoop	kinematic	Contour	Figure A-10
11	Hoop	isotropic	Contour	Figure A-11
12	Hoop	average	Contour	Figure A-12

Table 5-2 contains a qualitative description of the prediction quality for each of the cases in Table 5-1 along with a description of where measurement trends are not well captured by the model predictions. The mean RMSE is included in Table 5-2 for a quantitative point of reference. Table 5-2 indicates that RMSE alone may not be an adequate indicator of prediction quality. For case 3 (axial stress, average hardening, DHD benchmark) RSME=48.14 MPa, which is a relatively low value for this dataset. However, the average hardening approach consistently underpredicted the DHD measurements at r/t=0.25 and 0.75 for this case. In contrast, case 5 (hoop stress, isotropic hardening, DHD benchmark) had a relatively high RSME but no evident prediction bias. In other words, the confidence bounds encompassed zero throughout the entire thickness for case 5 (Figure A-5). Case 9 (axial stress, average hardening, contour benchmark) is perhaps the best agreement between measurements and models obtained in this study. Figure A-9 shows that, where systematic prediction biases existed, the tolerance bounds were very close to encompassing zero. This case also had the lowest mean RMSE (36.77 MPa). Overall, this work (and past work [1]-[2]) demonstrates that FE provides reasonable predictions of residual stress. Although, it is evident that established guidelines on hardening law approach are needed.

Table 5-2: Qualitative Assessment of Prediction Bias

Case	Description	Prediction Bias	Mean RMSE	Qualitative Assessment
1	Axial Nonlinear Kinematic DHD	- overprediction at ID - underpredition at x/t=0.25 - overprediction at x/t=0.5 - underprediction at x/t=0.75 - underprediction at OD	112.57	Trends not predicted well
2	Axial Isotropic DHD	- underprediction at x/t=0.5 - underprediction at x/t=0.7 - overprediction at OD	64.64	Trends predicted well
3	Axial Average DHD	- underprediction at x/t=0.25 - underprediction at x/t=0.75	48.14	Trends predicted marginally
4	Hoop Nonlinear Kinematic DHD	- underprediction at x/t=0.3 - overprediction at x/t=0.5 - underprediction at x/t=0.75	128.97	Trends not predicted well
5	Hoop Isotropic DHD	No evident bias	83.54	Trends predicted well
6	Hoop Average DHD	- underprediction at x/t=0.3 - underprediction at x/t>0.65	63.25	Trends predicted marginally
7	Axial Nonlinear Kinematic Contour	- overprediction at ID - overprediction at x/t=0.5 - underprediction at x/t>0.7	83.43	Trends not predicted well
8	Axial Isotropic Contour	- overprediction at x/t=0.25 - underprediction at x/t=0.5 - overprediction at x/t=0.85	93.99	Trends predicted marginally
9	Axial Average Contour	- overprediction at ID - overprediction at x/t=0.3 - underprediction at x/t=0.55	36.77	Trends predicted well
10	Hoop Nonlinear Kinematic Contour	- overprediction at ID - underprediction at 0.1<x/t<1	133.22	Trends not predicted well
11	Hoop Isotropic Contour	- overprediction at ID - underprediction at x/t=0.55 - overprediction at x/t=0.8	109.62	Trends not predicted well
12	Hoop Average Contour	- overprediction at ID - underprediction at x/t=0.55	78.33	Trends predicted marginally

5.2.3 Assessment of Root Mean Square Error

RMSE (Equation 5-2) may be one potential indicator of prediction quality. This value provides a general measure of how well stress magnitudes are predicted. Table 5-3 and Table 5-4 show RMSE for the various comparison cases for the DHD benchmark and the contour benchmark, respectively.

Table 5-3: RMSE for DHD Benchmark

RMSE	Axial(MPa)	Hoop(MPa)
KIN	112.57 (94.21,128.78)	128.97 (104.2,153.42)
ISO	64.64 (37.54,90.82)	83.54 (44.88,131.92)
AVE	48.14 (24.29,69.98)	63.25 (40.37,99)

Table 5-4: RMSE for Contour Benchmark

RMSE	Axial(MPa)	Hoop(MPa)
KIN	83.43 (69.61,95.11)	133.22 (116.83,154.31)
ISO	93.99 (69.42,117.84)	109.62 (86.22,142.57)
AVE	36.77 (23.29,51.76)	78.33 (56.01,120.01)

Figure 5-4 and Figure 5-5 show these values plotted as bar charts, including the appropriate confidence bounds, for the axial stress predictions and the hoop stress predictions, respectively. In most cases, the models that used the nonlinear kinematic assumption demonstrated the highest RSME. For the axial stress predictions with contour benchmark [Figure 5-4(b)], the nonlinear kinematic and isotropic models were indistinguishable. This is also the only case where the average hardening approach showed clearly superior predictions than the isotropic models. In all other cases, the apparent improvement in prediction agreement of average over isotropic was within the confidence bounds on RMSE.

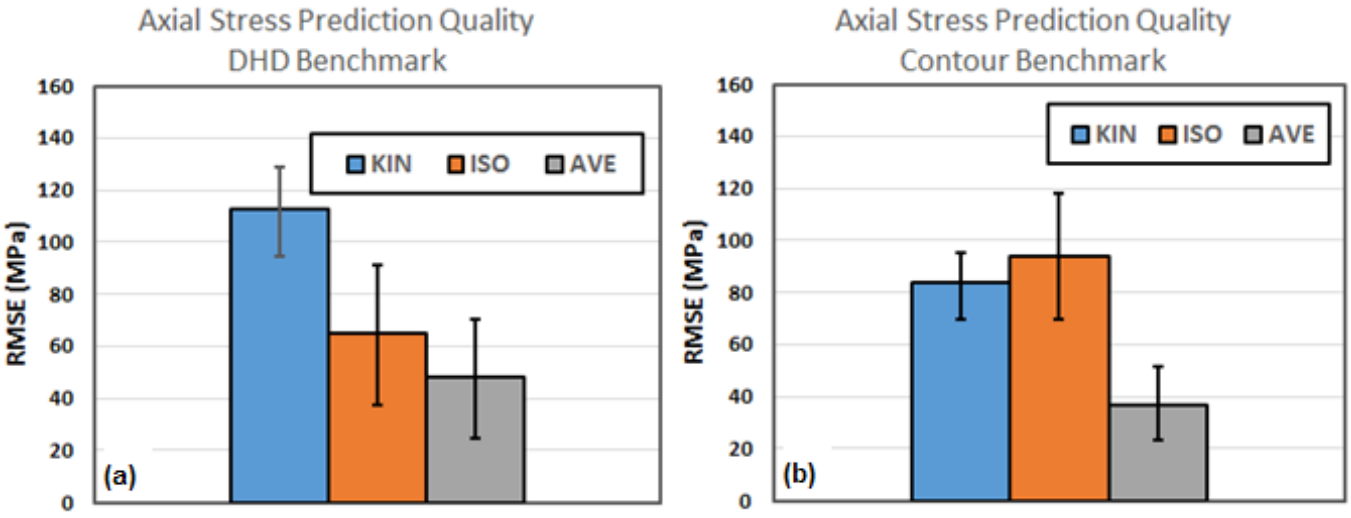


Figure 5-4: Root Mean Square Error for Axial Stress Predictions (a) DHD Benchmark and (b) Contour Benchmark

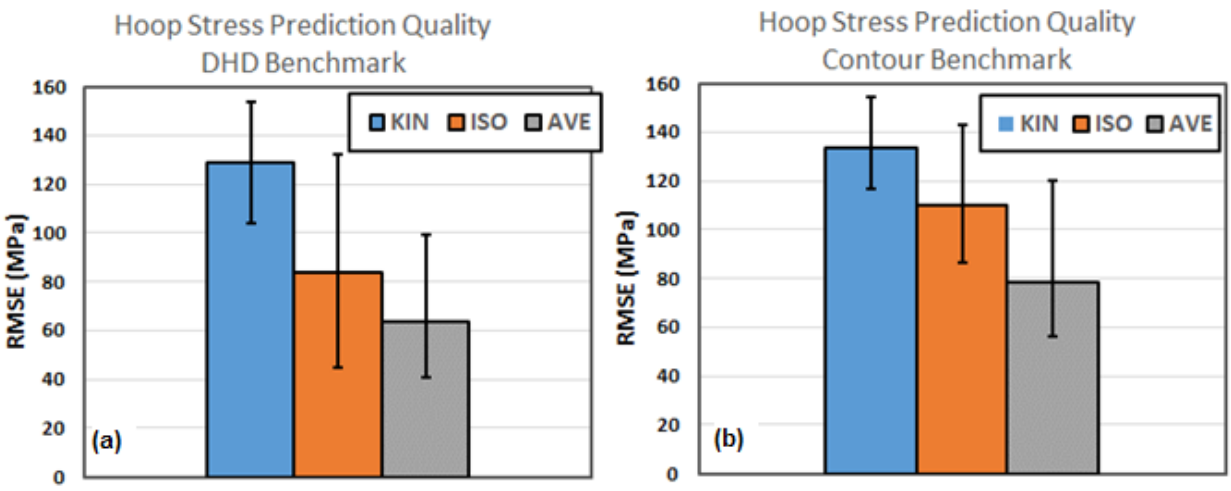


Figure 5-5: Root Mean Square Error for Hoop Stress Predictions (a) DHD Benchmark and (b) Contour Benchmark

5.2.4 Hardening Law Recommendation

This study investigated three hardening law approaches: nonlinear kinematic, isotropic, and the average of the kinematic result and isotropic result. Two aspects that should be assessed when evaluating different hardening law approaches are prediction bias and RMSE. Table 5-2 indicates that the through-wall trends were predicted well for three cases: 2, 5, and 9. Overall, the analysis in this section indicates that the nonlinear kinematic hardening models were the least accurate. While the uncertainties were large, there was some indication that the averaging approach provides better predictions than isotropic. Given these considerations, the authors recommend use of the averaging approach.

This study did not consider the Lemaitre-Chaboche hardening law, also known as mixed hardening [34]. Real materials exhibit both isotropic and kinematic hardening characteristics. The experimental material data needed to develop a mixed hardening law were not available at the time the Phase 2b round robin was began. In the future, it may be valuable to repeat a round robin modeling study with mixed hardening to investigate how well this hardening law compares with the measurement data presented here. The same statistical methods should be applied to account for both modeling and measurement uncertainty and develop tolerance bounds on the mean difference function.

5.3 Modeling Guidelines

While hardening law may be the most important factor, guidelines regarding other modeler choices have been developed through industry experience [11] and through the round robin studies [1]. These guidelines were provided to the Phase 2b modelers as instructions to participate in the round robin analysis effort, as described in [7]. As indicated by the comparisons of measurements and models in Section 5.2.2 and Appendix A, there is a general, qualitative sense that the FE predictions of residual stress from the Phase 2b study are reasonable (e.g., see Figure A-5 and Figure A-9). Thus, the modeling guidelines of [11] are adopted here. The recommendations for axisymmetric models are shown in the following list.

- Weld Bead Geometry Definition
  - Modeling the precise bead shape is unnecessary.

- Weld beads may be approximated as trapezoids.
- The total number of weld beads, the number of weld layers, and the number of beads in each layer should approximate the real weld configuration as closely as possible.
- Given current computational capabilities, modeling a realistic number of weld passes in an axisymmetric model is feasible. Approximations resulting from pass-lumping should be avoided.
- Where fabrication records are lacking, the assumed weld configuration should be based on common industry practice and knowledge of the component (e.g., nominal wall thickness).
- The cross-sectional area of each weld bead should be approximately equal to aid in heat input tuning.

- Bead and Process Sequence
  - Residual stress FE models are path-dependent. The analyst should model the actual fabrication process as closely as possible, including bead sequencing, number and size of weld beads, and other relevant processes (e.g., repairs and PWSCC mitigation).
  - The analyst should explicitly model the application of the butter and associated postweld heat treatment.
- Heat Input Model Tuning
  - Tuning the heat input to match known quantities, such as expected interpass temperature, is acceptable but not required.
  - In all cases, the analyst should visually confirm that the entire weld bead reaches the melting temperature. The material surrounding the weld bead (approximately one element in size) should also reach the melting temperature.
- Structural and Thermal Boundary Conditions
  - In general, boundary conditions should represent the physical situation being modeled. As such, they can change from application to application.
  - For a typical dissimilar metal butt weld, axial displacement in one node, located away from the weld at the edge of the model, should be constrained (see [11] for more information).
  - For a weld to infinitely-long straight pipe, the modeled pipe length should be 4 times the ID to avoid edge effects.
  - For the nozzle and pipe cross-sections typically of interest in reactor coolant pressure boundary welds, heat convection at the surface is negligible compared to conduction through the part.
- Material Properties
  - The average of isotropic and nonlinear kinematic is the recommended hardening approach.
  - The temperature-dependent material properties provided to the Phase 2b round-robin participants (see Appendix B) may be used, subject to the following constraints.
    - The material property inputs should accurately reflect the materials in the real situation, including temperature dependence.

        - The scope of material property inputs depends on the material behavior required to successfully execute all aspects of the model. For example, creep properties are required to approximate stress relaxation during postweld heat treatment.
- Element Selection and Mesh
    - Quadrilateral, linear elements should be applied.
    - Triangular elements should be avoided in all regions of the model.
    - A fine mesh is recommended in the weld regions of the model. Approximate element size for the weld passes should be 1.25 mm^2. This corresponds to 20-25 elements in a typical weld pass in reactor coolant pressure boundary nozzle welds.
    - The mesh may coarsen away from the weld regions.

5.4 Proposed Validation Scheme

The purpose of this Section is to develop a process to judge the quality of a particular WRS prediction. To apply the method proposed here, an analyst must create an FE model of the Phase 2b mockup according to the guidelines of Section 5.3 and reference [7]. Then, the analyst calculates a series of quality metrics to judge how well the analyst's prediction agrees with the round robin dataset. This validation methodology was developed assuming that the end application is a deterministic flaw growth calculation. As such, flaw growth studies were performed here to inform development of the metrics and acceptance measures. Section 5.5 summarizes the entire validation process, while Section 5.6 gives additional recommendations on modeling a real application.

5.4.1 Overview of Approach

As discussed in [19], validation of a model determines how well the model reflects the physical system being approximated. A validation approach requires a benchmark, a set of metrics, and acceptance measures. Section 5.4.2 describes the recommended benchmark and associated justification. The recommended metrics and acceptance measures were based on a flaw growth argument. The concept applied here was to find a set of metrics that were relevant to flaw growth predictions. The metrics should, therefore, interrogate features of a residual stress curve that are important to flaw growth. Three metrics are proposed here for validating FE predictions of residual stress:

1. RMSE on WRS magnitude through the entire wall thickness
2. RMSE on the first derivative of the WRS curve
3. Average difference up to the initial crack depth

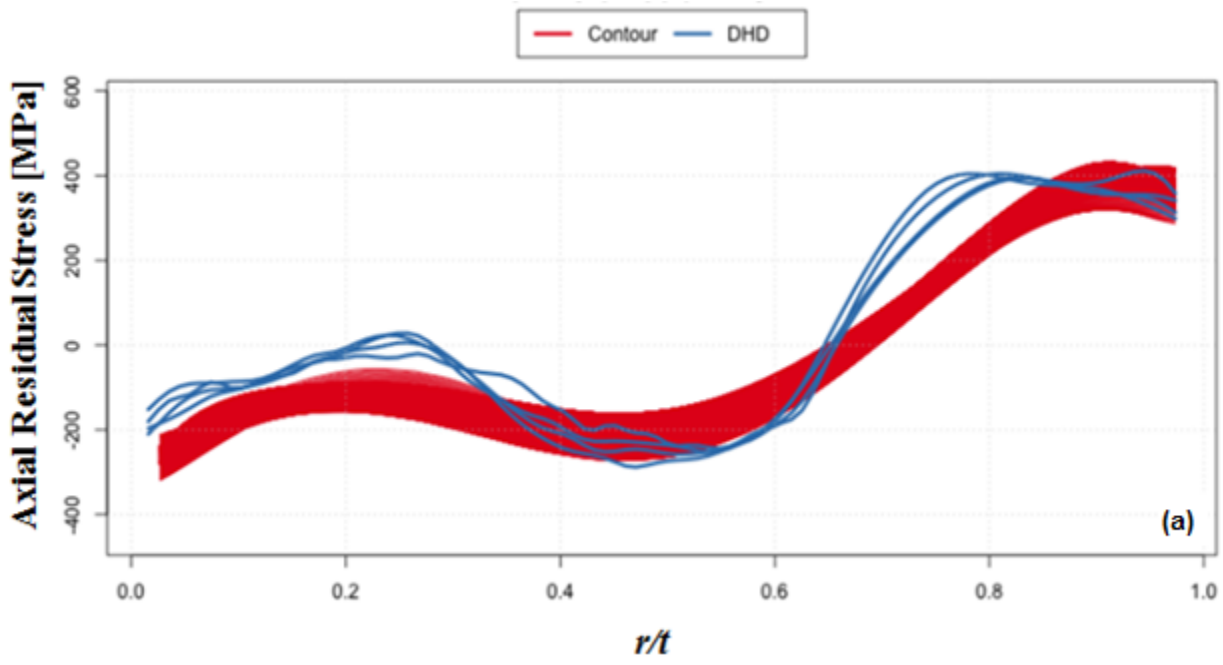
Sections 5.4.7, 5.4.8, and 5.4.10 describe and develop these metrics in detail. Each metric will have associated acceptance measures. The proposed acceptance measures are based on values of the metric that lead to a reasonable crack growth prediction, given the chosen benchmark. Sections 5.4.9 and 5.4.11 develop these acceptance measures. Additional analysis of potential metrics and acceptance measures appears in Appendix D. While the recommended acceptance measures were developed assuming an average hardening benchmark, corresponding criteria assuming an isotropic hardening benchmark are included in Sections 5.4.8 and 5.4.10 for illustration purposes. Additional analysis of potential metrics and acceptance measures for isotropic hardening appears in Appendix E.

5.4.2 Benchmark

Validation first requires choice of a benchmark that reflects the real world. The predictions of the model can then be quantitatively compared to the benchmark to assess the robustness of the modeling approach. Often, physical measurements are a natural choice of benchmark. In the application of concern here, there are four possibilities for a benchmark.

1. DHD measurement data
2. contour measurement data
3. average of 1 and 2
4. mean of the Phase 2b models

In the case of WRS, the "measurement data" is part physical measurement and part model, since residual stress cannot be directly measured. This fact complicates the rigorous selection of a benchmark for validating residual stress predictions. Lewis and Brooks [13] compared the Phase 2b DHD and contour measurement results to each other using the same methodology outlined in Section 5.2.1 for comparing models to measurements. Figure 5-6 demonstrates that the two measurements are significantly different than each other, especially near the inner surface. The flaw growth calculations in Chapter 4 show that the differences between the contour and DHD measurements lead to different flaw growth results. Averaging the two measurements (option 3) may not be a valid option, since there is no reason to believe that the two datasets belong to the same population. While a benchmark based upon the measurements may be ideal, parsing out which measurement is most correct (option 1 or option 2) requires a more thorough investigation than was performed in this work. Given these complications, the mean of the average hardening models from the Phase 2b study was chosen as a benchmark for demonstrating how a validation process may be developed.



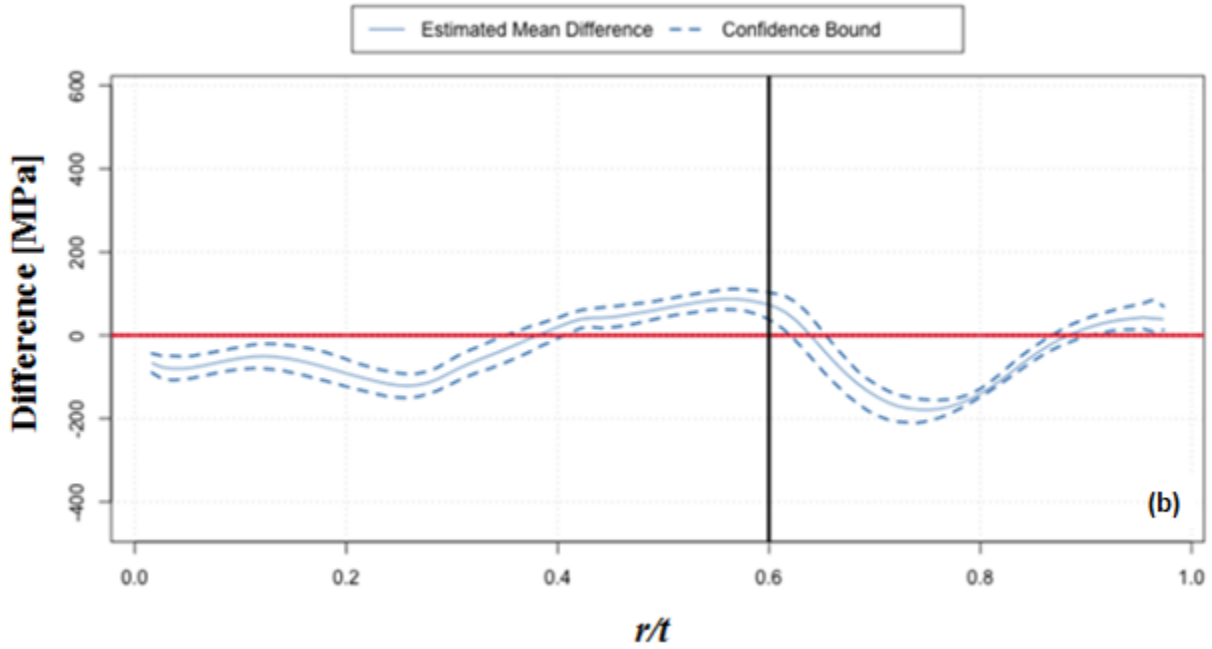


Figure 5-6: Comparison of DHD and Contour Axial Stress Predictions (a) Raw Data and (b) Difference in Means.

5.4.3 Circumferential Flaw Growth – Isotropic Hardening

Figure 5-7 shows the seven smoothed isotropic WRS profiles that passed the initial outlier screening discussed in Section 3.2.2. These profiles represent the axial WRS predictions from the Phase 2b study, along with the mean WRS profile. The mean profile is the mean of six of the predictions in Figure 5-7. Although prediction B was screened out of certain calculations, as described in Section 3.2.8, it is included in this discussion for illustrative purposes. The mean profile in Figure 5-7 is the cross-sectional mean, as described in [13].

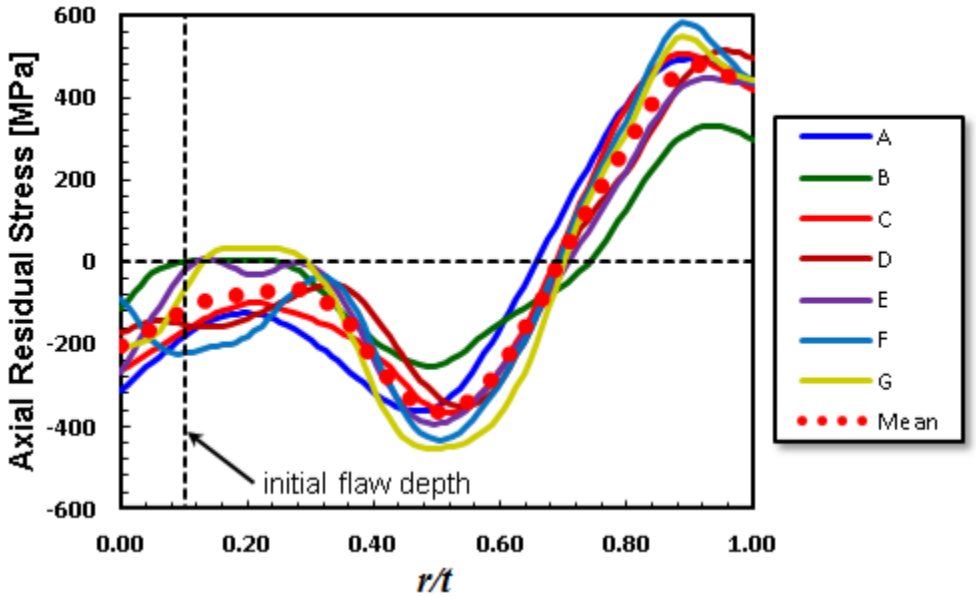


Figure 5-7: Smoothed Axial WRS Profiles and Mean, Isotropic Hardening

The flaw growth study in Chapter 4 (see Table 4-1) was repeated assuming a circumferential flaw and the residual stress profiles in Figure 5-7. The results are shown in Figure 5-8.

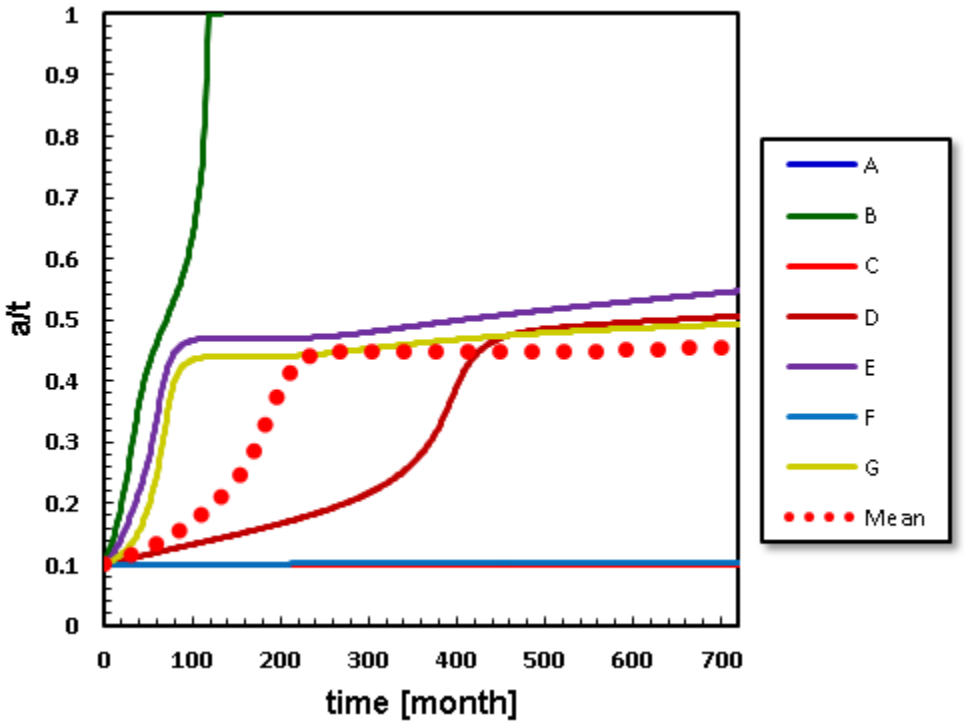


Figure 5-8: Circumferential Flaw Growth, Isotropic Hardening

As mentioned in Chapter 4, 720 months is much longer than a typical industry flaw analysis. However, decreasing the evaluation period did not affect the metrics and acceptance measures (see Appendix E). The mean WRS profile exhibited flaw growth to 45 % through-wall in 250 months, followed by flaw arrest. Three other residual stress predictions (D, E, and G) led to arrest at 45-50 % through-wall. The other four WRS profiles (A, B, C, and F) led to different crack growth behavior. Using participant B's WRS profile, the flaw grew through-wall in under 200 months. The remaining calculations showed negligible growth throughout the evaluation period.

This qualitative discussion of the different flaw growth predictions aids in establishing quantitative acceptance measures for the three quality metrics proposed in Sections 5.4.8 and 5.4.10. WRS predictions D, E, and G are considered reasonable, given that they result in similar end-of-life flaws as the proposed benchmark.

5.4.4 Circumferential Flaw Growth – Average Hardening

Figure 5-9 shows the seven smoothed average hardening WRS profiles that passed the initial outlier screening discussed in Section 3.2.2. These profiles represent the axial WRS predictions from the Phase 2b study, along with the cross-sectional mean WRS profile.

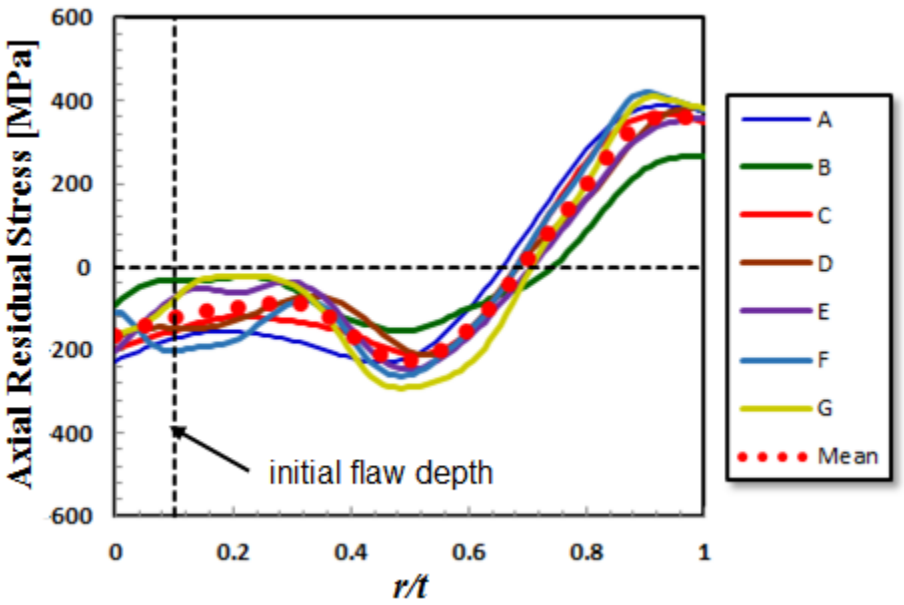


Figure 5-9: Smoothed Axial WRS Profiles, Average Hardening

The flaw growth results are shown in Figure 5-10. Once again, as mentioned in Chapter 4, 720 months is much longer than a typical industry flaw analysis. However, decreasing the evaluation period did not affect the metrics and acceptance measures (see Appendix D).

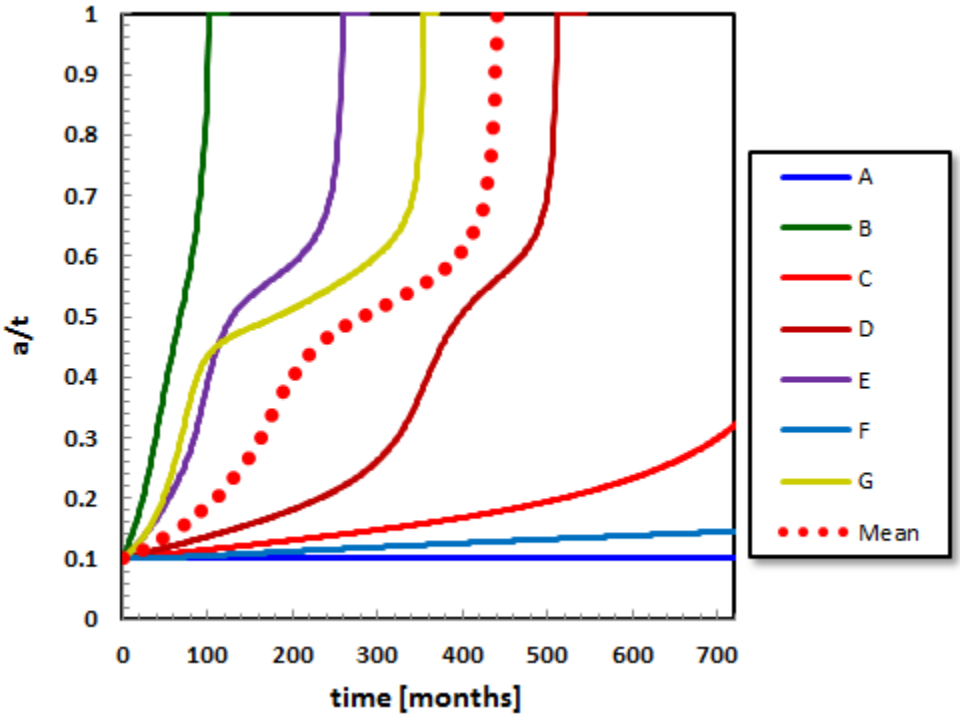


Figure 5-10: Circumferential Flaw Growth, Average Hardening

In comparison to the isotropic case in Figure 5-8, there is more uncertainty in the flaw growth prediction. It is also apparent that the average hardening WRS predictions led to through-wall flaw growth, while three cases arrested for the isotropic curves. The averaging approach tends to

decrease the magnitude of tensile and compressive stress peaks, relative to isotropic hardening. On the one hand, the shallower compressive troughs of average hardening speeds up crack growth and leads to more cases that go through-wall. On the other hand, arresting of the flaws with isotropic hardening decreases uncertainty, since the exact magnitude of the compressive trough is only important in cases where the flaw has not arrested. As will be shown in Section 5.4.9, these variances lead to different approaches to determining appropriate acceptance measures. Therefore, the exact values of the acceptance measures are dependent upon the choice of the benchmark in the approach adopted here.

5.4.5 Axial Flaw Growth – Isotropic Hardening

Figure 5-11 shows the seven smoothed hoop WRS profiles that passed the initial screening described in Section 3.2.2, assuming isotropic hardening. The hoop stresses are distinct from the axial stresses in that there is no force balance requirement through the wall thickness, given path data extracted from an axisymmetric analysis. While the area under the curve of an axial stress profile extracted from an axisymmetric FE model will be roughly zero, that of a hoop stress profile may be non-zero. These differences are apparent in the smoothed curves of Figure 5-7 and Figure 5-11.

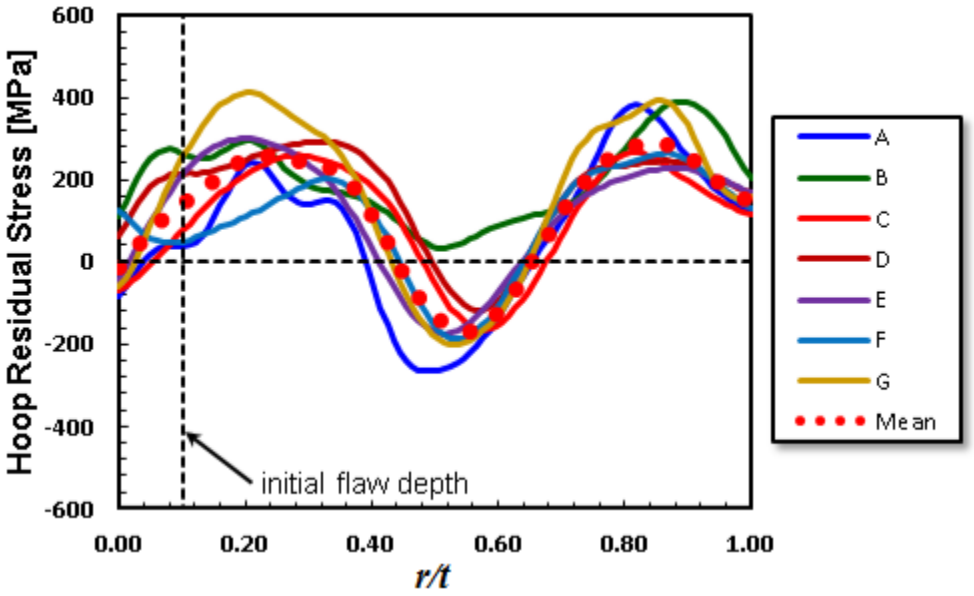


Figure 5-11: Smoothed Hoop WRS Profiles, Isotropic Hardening

Figure 5-12 shows the axial flaw growth results, assuming the residual stress profiles of Figure 5-11. The results appear much more consistent among the various stress inputs than was observed for the circumferential flaw growth study in Figure 5-8. This result is a consequence of the hoop residual stresses being either positive or only slightly negative in the initial flaw depth zone. With the axial stresses being compressive in the initial flaw depth zone, the exact magnitude of the compressive stress can have a large impact on the flaw growth early in time. Whereas the axial flaw will always exhibit early growth in the depth direction, the circumferential flaw may or may not, depending on the interaction of the residual stresses with operating loads.

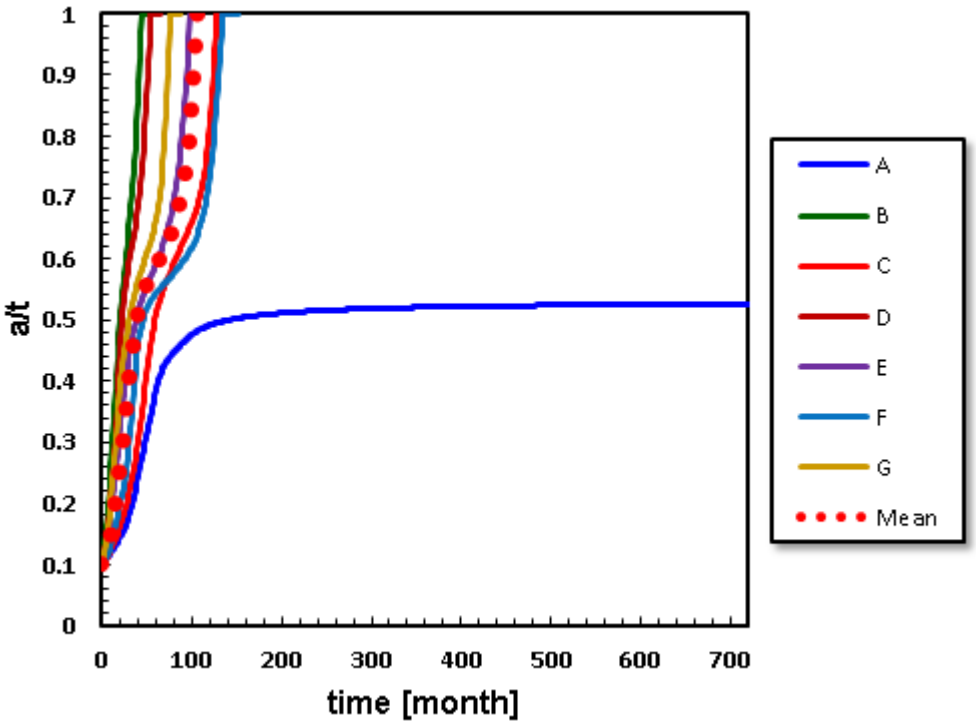


Figure 5-12: Axial Flaw Growth, Isotropic Hardening

Prediction A is unique in that it led to crack arrest at 50 % through-wall for an axial flaw. All other residual stress predictions led to through-wall growth in under 200 months. Table 5-5 shows the time to through-wall for each prediction, with a value of 1,000 months assigned to prediction A to represent "essentially infinite." These observations will inform quality metrics for hoop stress predictions using isotropic hardening.

Table 5-5: Time to Through-Wall

Stress Profile	Time to Through-wall [months]
A	1000
B	44
C	128
D	54
E	98
F	138
G	76
Mean WRS	105
min	44
1st quartile	65
median	101.5
3rd quartile	133
max	1000

5.4.6 Axial Flaw Growth – Average Hardening

Figure 5-13 shows the seven smoothed hoop WRS profiles that passed the initial screening described in Section 3.2.2, assuming average hardening.

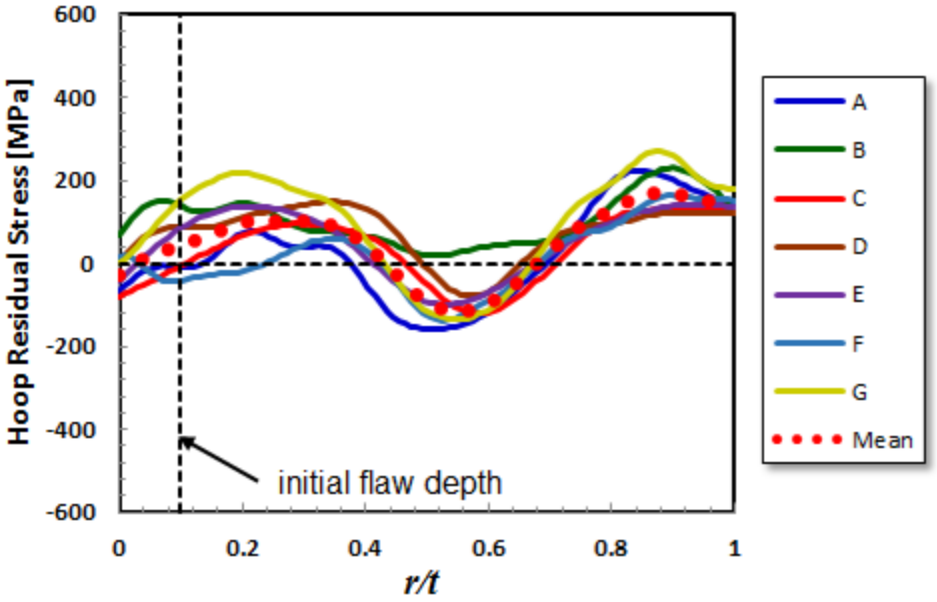


Figure 5-13: Hoop WRS Profiles, Average Hardening

Figure 5-14 shows the corresponding crack growth results.

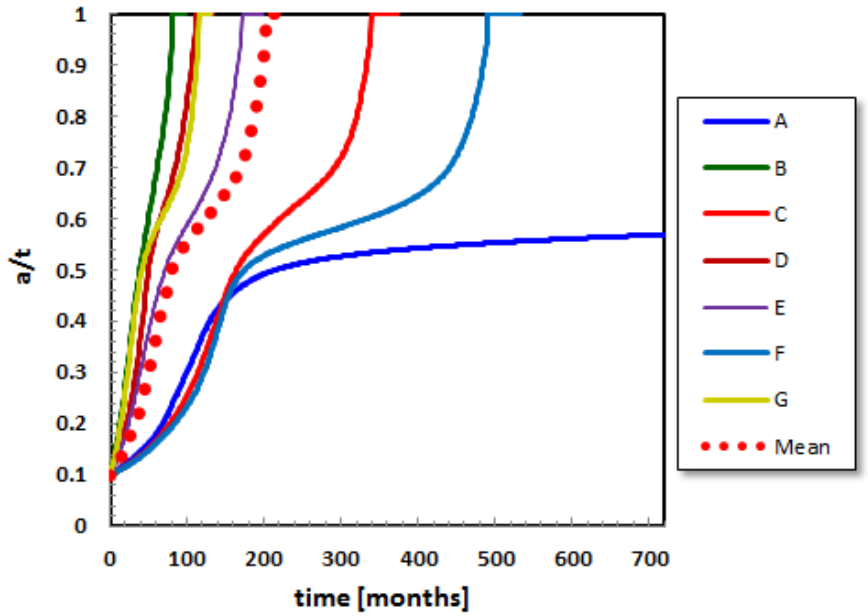


Figure 5-14: Axial Flaw Growth, Average Hardening

As was seen for the circumferential crack case, the uncertainty in time to through-wall increases for average hardening relative to the isotropic results shown in Figure 5-12. This increase in uncertainty is due to the fact that the averaging process broadened the tensile concave-down region in the first 50 % of the wall thickness. The effect is illustrated more clearly in Figure 5-15,

which shows the stress intensity factors as a function of time for the first 200 months. The time of the initial peak in stress intensity factor is much more variable for the average hardening law case.

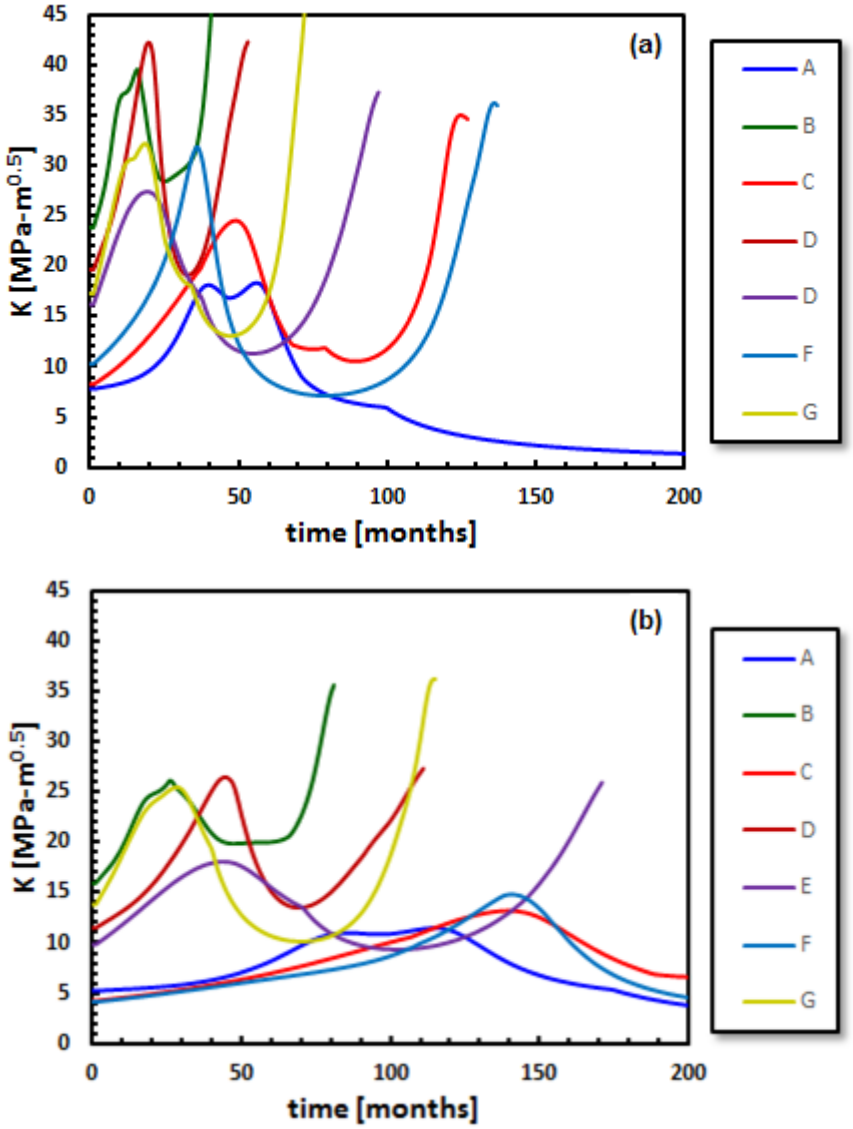


Figure 5-15: Stress Intensity Factor: (a) Isotropic Hardening and (b) Average Hardening

5.4.7 Overview of Quality Metrics

To complete the validation procedure, an analyst must complete an FE model of the Phase 2b mockup with isotropic hardening. The analyst then extracts WRS data from a weld centerline path. The analyst may also smooth the extracted profile, provided that the smoothed WRS profile is representative of the original. The extracted profile and the cross-sectional mean profiles shown in Figure 5-7 and Figure 5-11 are then used to calculate three quality metrics. This section develops the recommended metrics.

The first quality metric involves RMSE averaged through the entire wall thickness, as described in Equation 5-3.

$$RMSE_{WRS} = \sqrt{\frac{1}{L}\sum_{k=1}^{L}\left(WRS_k - WRS_k^{mean}\right)^2} \qquad \text{Equation 5-3}$$

where k and L are the same as in Equation 5-2, WRS_k is the analyst's predicted stress magnitude at the k^{th} position through the wall thickness, and WRS_k^{mean} is the cross-sectional mean prediction (see [13]). For the analyst to complete this step, the WRS values from the analyst's FE calculation should be extracted at the same spatial intervals as those shown in Appendix C (i.e., every 1 % of the thickness). This quality metric provides an overall measure of how close the predicted stress magnitudes are to the mean of the Phase 2b isotropic hardening predictions.

The second quality metric is RMSE on the first derivative of WRS with respect to through-wall position. This quality metric is related to through-wall WRS trends and concavity. A true concavity test (i.e., the second derivative) was investigated, but it did not provide improved screening of FE results (see Appendix D for average hardening and Appendix E for isotropic hardening). Here, the three-point formula was used to estimate the first derivative (Equation 5-4) [20],[35].

$$D1_k = \left.\frac{dWRS}{dx_{norm}}\right|_k \approx \frac{1}{2h}\left(-WRS_{k-1} + WRS_{k+1}\right) \qquad \text{Equation 5-4}$$

where $x_{norm} = \frac{r}{t}$ is the through-wall position normalized to the wall thickness and h is the distance between the k^{th} and k+1th data point (h should not vary from interval to interval). The RMSE on first derivative was calculated according to Equation 5-5.

$$RMSE_{D1} = \sqrt{\frac{1}{L-2}\sum_{k=2}^{L-1}\left(D1_k - D1_k^{mean}\right)} \qquad \text{Equation 5-5}$$

Note that the derivative cannot be calculated at k=1 and k=L using Equation 5-4.

The final quality metric is average difference over the initial flaw depth. This metric is related to the importance of the stress prediction over the distance corresponding to the initial flaw depth (the first 10 % of wall thickness, in this case). The stress magnitude and trends in this area have a profound impact on circumferential flaw growth behavior, as discussed further in Section 5.4.8. This metric is calculated according to Equation 5-6.

$$diff_{avg} = \frac{1}{L_{0.1}}\sum_{k=1}^{L_{0.1}}\left(WRS_k - WRS_k^{mean}\right) \qquad \text{Equation 5-6}$$

where $L_{0.1}$ is the number of spatial locations where WRS is determined up to x_{norm}=0.1.

5.4.8 Quality Metrics for Axial Stress Predictions

This section explores how the proposed metrics can be applied to discriminate between axial residual stress predictions, with reference to the circumferential flaw growth studies presented in Sections 5.4.3 and 5.4.4. As such, two approaches are considered here: one using the isotropic hardening predictions and the other using the average hardening predictions. This discussion

leads to proposing acceptance measures for each of the metrics. Final recommendations on acceptance measures are presented in Section 5.4.9.

Table 5-6 shows the three quality metrics applied to the seven isotropic predictions. The entries are sorted according to how the corresponding circumferential crack growth prediction compared with the mean prediction (see Figure 5-8). The prediction was deemed to yield "similar" flaw growth results as the benchmark if the final flaw depth was within 10 % of the wall thickness (i.e., within a/t=0.1) of the benchmark case. Thus, predictions D, E, and G were deemed to yield acceptable flaw growth results.

Table 5-6: Quality Metrics Applied to Phase 2b Axial Isotropic Predictions

Participant	$RMSE_{WRS}$ [MPa]	$RMSE_{D1}$ [MPa]	$diff_{avg}$ [MPa]	Crack Growth
D	52	768	8	Similar
E	48	660	22	Similar
G	74	865	-1	Similar
B	109	839	120	Too Fast
A	78	605	-84	Too Slow
C	43	521	-55	Too Slow
F	67	1055	-16	Too Slow
Min	43	521	-84	
25th percentile	50	632	-36	
Median	67	768	-1	
75th percentile	76	852	15	
Max	109	1055	120	

Table 5-6 demonstrates that $RMSE_{WRS}$ alone is not a sufficient measure of the quality of the prediction. For example, $RMSE_{WRS}$ for prediction C was the minimum of the seven predictions at 43 MPa. In fact, $RMSE_{D1}$ for prediction C was also the minimum of the sample at 521 MPa. However, $diff_{avg}$ for prediction C was -55 MPa. Figure 5-16 shows a plot of prediction C compared against the mean WRS. The fact that prediction C was below the mean for the first 10 % of the wall thickness severely impacted the flaw growth calculation, such that the flaw never grew. Therefore, it is important that $diff_{avg}$ be closer to zero in order to obtain the expected crack growth behavior.

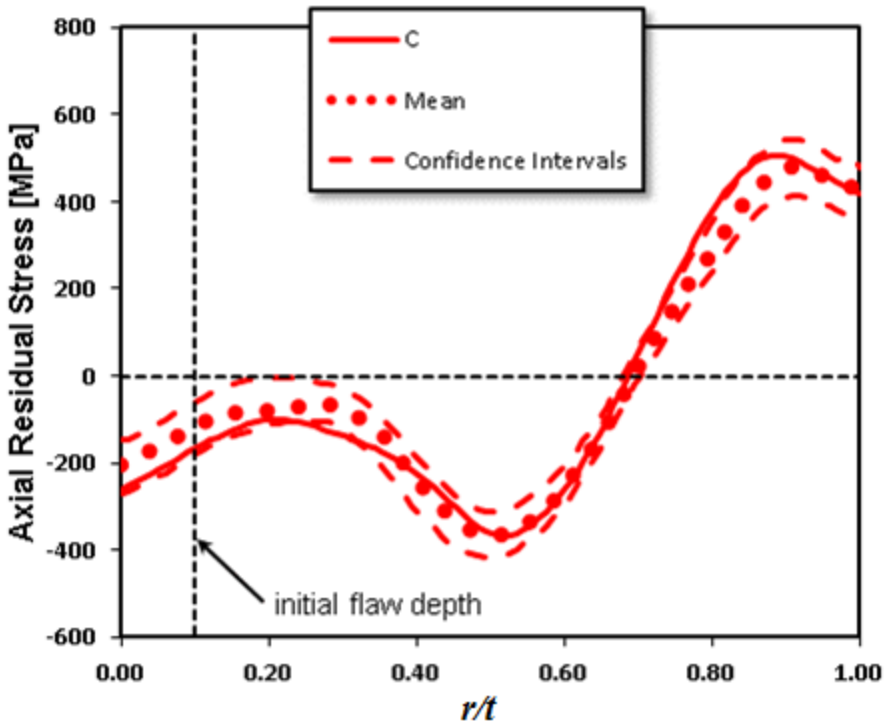


Figure 5-16: Prediction C (Isotropic) against the Mean Prediction

The importance of $RMSE_{D1}$ is best demonstrated by considering prediction F, which is shown in Figure 5-17. Figure 5-17 shows that prediction F is concave up in the first 35 % of wall thickness, while the mean is concave down. This behavior led to $diff_{avg}$ being relatively low in magnitude, -16 MPa. $RMSE_{WRS}$ for prediction G was relatively low as well at 67 MPa, which is the median of the seven predictions. However, the concave down characteristic led to immediate crack arrest, which does not appear to be reasonable crack growth behavior given the mean of the dataset. Figure 5-18 compares the first derivatives of the two curves in Figure 5-17. The plot shows the difference in signs in the first 10 % of the wall thickness.

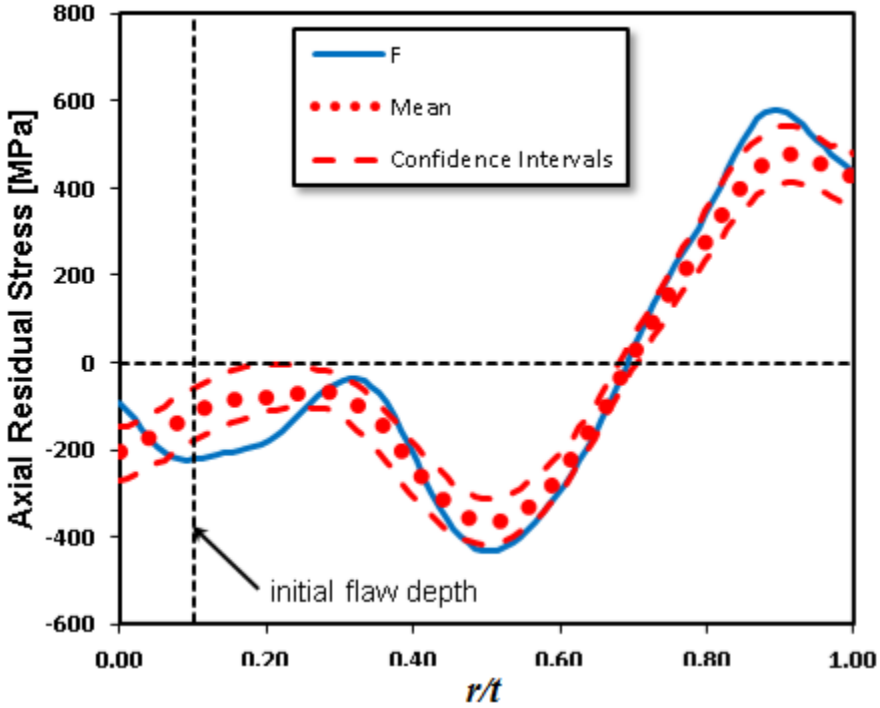


Figure 5-17: Prediction F (Isotropic) against the Mean Prediction

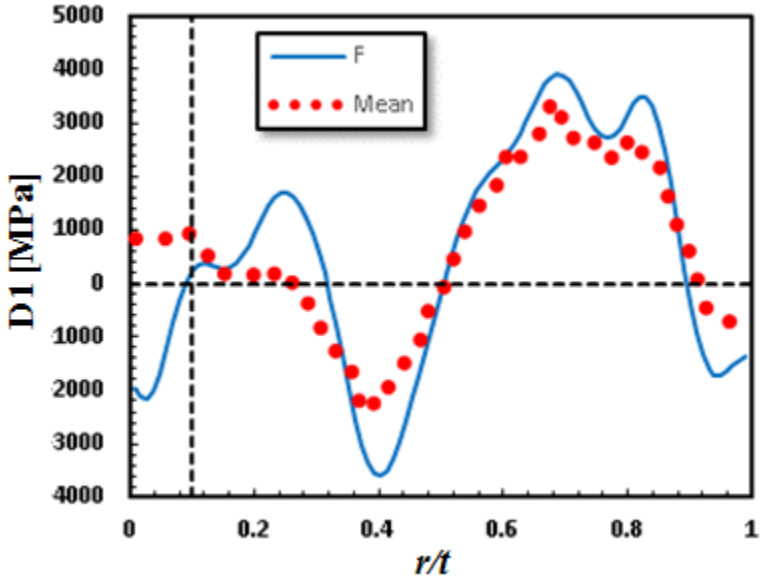


Figure 5-18: Comparison of First Derivatives

Since the flaw growth results were more variable for average hardening (see Figure 5-10), discriminating between good and bad predictions requires a different thought process. In the case of average hardening, the authors chose the 3 predictions (roughly, the top 50 %) that matched closest to the benchmark as acceptable WRS predictions. For this case, predictions D, E, and G were again considered acceptable. Table 5-7 shows the values of the quality metrics for each of the average hardening predictions.

Table 5-7: Quality Metrics Applied to the Phase 2b Axial Average Hardening Predictions

Participant	$RMSE_{WRS}$ [MPa]	$RMSE_{D1}$ [MPa]	$diff_{avg}$ [MPa]	Delta from Mean	Time to leakage
D	29	409	-10	71	512
E	32	366	11	-181	260
G	52	512	14	-86	355
B	79	523	92	-338	103
A	59	374	-55	559	1000
C	28	300	-31	408	849
F	46	624	-20	559	1000
Min	28	300	-55		103
25th percentile	30	370	-26		308
Median	46	409	-10		512
75th percentile	56	518	12		925
Max	79	624	92		1000

5.4.9 Recommended Acceptance Measures – Axial Residual Stress

While the analysis in Section 5.4.8 was applied to both the isotropic hardening and average hardening datasets, the final recommended acceptance measures should be based on one or the other. The discussion in Section 5.2 indicated that the averaging approach may be preferable to isotropic hardening. Therefore, the recommend acceptance measures is based upon the analysis of the average hardening dataset in Table 5-7. In general, the recommended acceptance measures are based upon values of the metric that (1) screen out those WRS curves that lead to unreasonable crack growth predictions and (2) screen in those curves that lead to reasonable crack growth predictions. For $RMSE_{WRS}$ and $RMSE_{D1}$, the recommended criteria are $RMSE_{WRS}$≤55 MPa and $RMSE_{D1}$≤520 MPa. It is also recommended that the analyst plot the first derivative of the WRS prediction along with that of the Phase 2b mean (as in Figure 5-18) to show that the derivatives are of the same sign up to the initial flaw depth. The recommended acceptance criterion for $diff_{avg}$ is -15 MPa≤ $diff_{avg}$≤15 MPa.

5.4.10 Quality Metrics for Hoop Stress Predictions

Table 5-8 shows the three quality metrics for the seven isotropic hoop stress predictions, sorted according to time to through-wall (see Figure 5-12).

Table 5-8: Quality Metrics Applied to Phase 2b Hoop Isotropic Predictions

Participant	$RMSE_{WRS}$ [MPa]	$RMSE_{D1}$ [MPa]	$diff_{avg}$ [MPa]	Delta from Mean	Time to leakage
D	70	698	95	-51	54
E	49	713	23	-7	98
G	82	939	23	-29	76
B	114	1234	156	-61	44
A	86	1079	-68	895	1000
C	53	627	-71	23	128
F	59	914	1	33	138
Min	49	627	-71		44
25th percentile	56	706	-34		65
Median	70	914	23		98
75th percentile	84	1009	59		133
Max	114	1234	156		1000

The time to leakage assuming the mean hoop residual stress profile in Figure 5-11 was 105 months. This value was considered as a benchmark for this discussion. The times to leakage assuming Participant G, E, and C's residual stress profiles were each within 36 months of this value. Thirty-six months corresponds to the length of time for two refueling outages. Participant E's case is straightforward, in that all three quality metrics were relatively low in magnitude. Participant G, on the other hand, had a high $RMSE_{D1}$ value of 939 MPa. Figure 5-19 compares Participant G's prediction against the mean curve. The high value of the initial slope is apparent. However, the relatively low $diff_{avg}$ and close agreement of the curves in other spatial locations led to an acceptable flaw growth result.

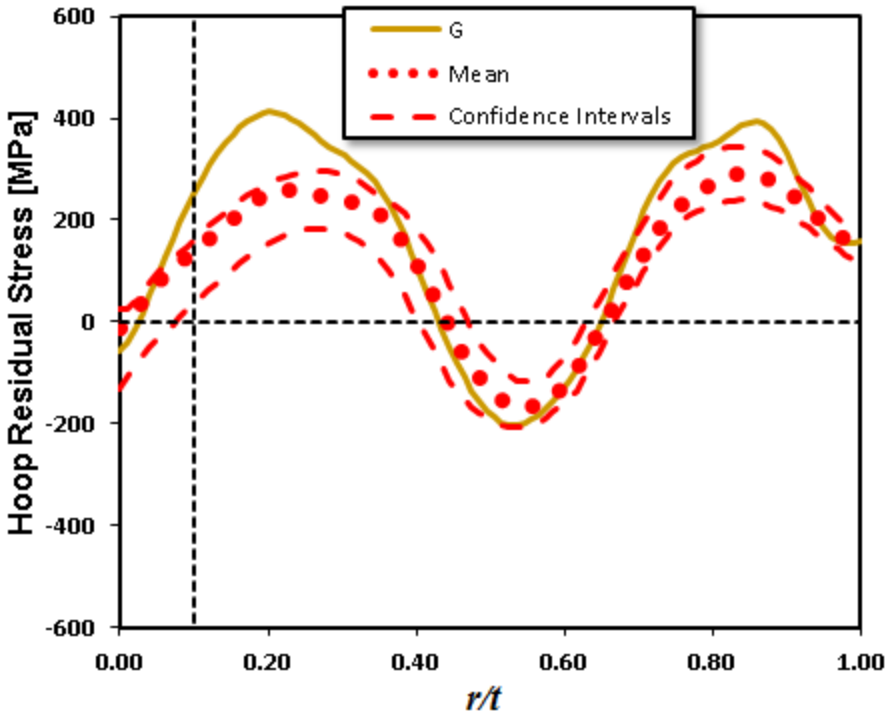


Figure 5-19: Hoop Stress Prediction from Participant G

Prediction C is shown against the mean curve in Figure 5-20. For the circumferential flaw growth, a negative $diff_{avg}$ could easily lead to complete arrest of the flaw. This is not the case for axial flaw growth. When operating loads are superimposed on the hoop stresses, the flaw still grows in the depth direction. Even though Participant C's $diff_{avg}$ was -71 MPa, a reasonable time to leakage of 128 months was still possible. However, $diff_{avg}$ remains an important metric for the hoop stresses, as demonstrated by Participant D (Figure 5-21). Participant D's $diff_{avg}$ was highly positive, and this led to a short time to leakage of 54 months, as compared to the assumed benchmark of 105 months.

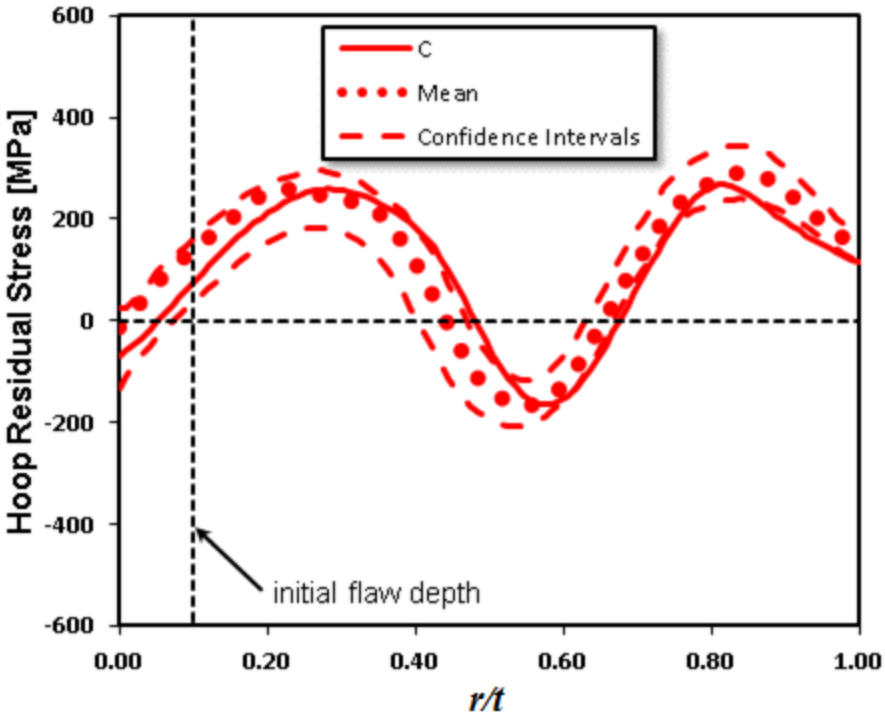


Figure 5-20: Hoop Stress Prediction from Participant C

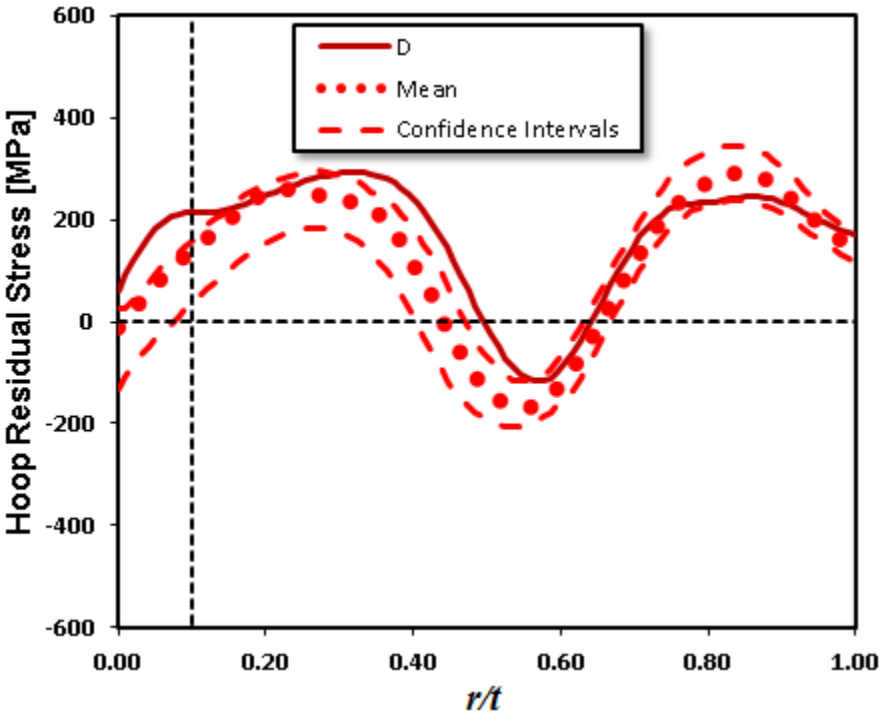


Figure 5-21: Hoop Stress Prediction from Participant D

Since the flaw growth results were more variable for average hardening (see Figure 5-14), discriminating between good and bad predictions requires a different thought process. As was proposed in Section 5.4.8, the authors chose the 3 predictions (roughly, the top 50 %) that matched closest to the benchmark as acceptable WRS predictions. For this case, predictions D, E, and G were again considered acceptable. Table 5-9 shows the values of the quality metrics for each of the average hardening predictions.

Table 5-9: Quality Metrics Applied to Phase 2b Hoop Average Hardening Predictions

Participant	$RMSE_{WRS}$ [MPa]	$RMSE_{D1}$ [MPa]	$diff_{avg}$ [MPa]	Delta from Mean	Time to leakage
D	49	417	47	-91	112
E	26	330	13	-31	172
G	68	525	63	-87	116
B	74	717	114	-121	82
A	51	524	-37	517	720
C	34	331	-56	137	340
F	50	591	-30	288	491
Min	26	330	-56		82
25th percentile	42	374	-33		114
Median	50	524	13		172
75th percentile	60	558	55		416
Max	74	717	114		720

5.4.11 Recommended Acceptance Measures – Hoop Residual Stress

While the analysis in Section 5.4.10 was applied to both the isotropic hardening and average hardening datasets, the final recommended acceptance measures should be based upon one or

the other. The discussion in Section 5.2 indicated that the averaging approach may be preferable to isotropic hardening. Therefore, the recommend acceptance measures is based upon the analysis of the average hardening dataset in Table 5-9. In general, the recommended acceptance measures are based on values of the metric that (1) screen out those WRS curves that lead to unreasonable crack growth predictions and (2) screen in those curves that lead to reasonable crack growth predictions. For $RMSE_{WRS}$ and $RMSE_{D1}$, the recommended criteria are $RMSE_{WRS}$≤70 MPa and $RMSE_{D1}$≤550 MPa. It is also recommended that the analyst plot the first derivative of the WRS prediction along with that of the Phase 2b mean (as in Figure 5-18) to show that the derivatives are of the same sign up to the initial flaw depth. The recommended acceptance criterion for $diff_{avg}$ is 0≤$diff_{avg}$≤65 MPa.

5.5 Summary of Validation Procedure

This section is a final summary of the validation procedure developed in this Chapter.

1. The analyst executes an axisymmetric FE model of the Phase 2b mockup, according to the guidelines in Section 5.3 and reference [7]. The model should be run twice: once with isotropic hardening properties and once for nonlinear kinematic hardening properties (see Appendix B).
2. The analyst extracts a number of WRS profiles from the FE results.
   a. axial and hoop WRS for isotropic hardening
   b. axial and hoop WRS for nonlinear kinematic hardening

   The profiles should be extracted from the centerline location. The analyst may smooth the FE WRS profiles (see [13] for one approach), provided the smoothed profiles are reasonable representations of the FE results. The final WRS profiles should consist of 100 uniformly-spaced data points.
3. The analyst should average the axial nonlinear kinematic and isotropic WRS profiles to obtain an axial average hardening profile. The analyst should repeat this step for hoop WRS.
4. The analyst calculates the first derivative of the WRS profiles. If an analytical form of the WRS profiles is available (e.g., as obtained from smoothing), then the profiles may be differentiated analytically. Otherwise, Equation 5-4 may be used as an approximation to the first derivative.
5. The means of the average hardening Phase 2b predictions, along with their first derivatives, are given in tabular form in Appendix C. These values are needed for step 6.
6. Using the data gathered in steps 2-5, the analyst should calculate the three quality metrics; $RMSE_{WRS}$, $RMSE_{D1}$, and $diff_{avg}$; according to the equations given in Section 5.4.7. The analyst should calculate $diff_{avg}$ according to the initial flaw depth of interest. In the absence of other information, the analyst may assume a/t=0.1 initial flaw depth.
7. The analyst compares the calculated values with the acceptance measures summarized in Table 5-10 and Table 5-11.
8. The analyst plots the first derivative of the extracted profiles and the derivatives of the mean profiles, similar to the plot in Figure 5-18. The two curves should be of the same algebraic sign, especially up to the initial flaw depth.
9. If all acceptance measures are met, then the analyst has passed the validation exercise. The analyst proceeds to step 11.

10. If one or more acceptance measures are not met, then the analyst and an independent individual in the organization should carefully review the model.
    a. They should verify that (i) all modeling guidelines in Section 5.3 were followed, (ii) all dimensions were correct, and (iii) all material property inputs were correct.
    b. They should discuss whether any improvements can be made to the model (e.g., refined mesh in a specific geometric region).
    c. If the model can be improved, then the analyst should rerun the model and repeat the validation exercise with the new results.
11. The analyst should document the validation exercise, including any lessons-learned during the process. An independent individual in the organization should review the validation exercise.

Table 5-10: Acceptance Measures for Axial Stresses

Quality Metric	Acceptance Criteria
$RMSE_{WRS}$	≤ 55 MPa
$RMSE_{D1}$	≤ 520 MPa
$diff_{avg}$	≥ -15 MPa ≤ 15 MPa

Table 5-11: Acceptance Measures for Hoop Stresses

Quality Metric	Acceptance Criteria
$RMSE_{WRS}$	≤ 70 MPa
$RMSE_{D1}$	≤ 550 MPa
$diff_{avg}$	≥ 0 ≤ 65 MPa

5.6 Modeling a Nuclear Plant Application

The approach outlined in Section 5.5 provides a means for an analyst to demonstrate capability in FE modeling of WRS. The approach uses an existing dataset related to a prototypical mockup to draw conclusions about validation criteria. This section provides recommendations for advancing from the validation stage to modeling real components.

5.6.1 Applicability of Validation Scheme and Acceptance Measures

The proposed validation scheme is most directly applicable to a dissimilar metal butt weld. Some details such as constraint condition, repair geometry and location, safe end length, and radius-to-thickness ratio need not be identical to the Phase 2b mockup to be considered covered by the validation process proposed in this document. However, there are potential scenarios encountered in the U.S. nuclear fleet where a validation approach based on the Phase 2b dataset is not appropriate.

One example of additional benchmark data is the mockup dataset from the excavate and weld repair (EWR) investigation [30]-[33]. EWR was proposed as a partial arc repair process, as illustrated in Figure 5-22 [33]. Therefore, three-dimensional analysis methods were required to adequately assess residual stresses for this process. The mockup shown in Figure 5-23 was

designed such that a 3D analysis was required. The measurement data from this mockup can therefore be used to benchmark a 3D analytical procedure applied to EWR, but the Phase 2b data would not be appropriate. In a similar vein, modeling residual stress in a control rod drive mechanism nozzle (i.e., j-groove weld) would require a validation dataset more representative of that application.

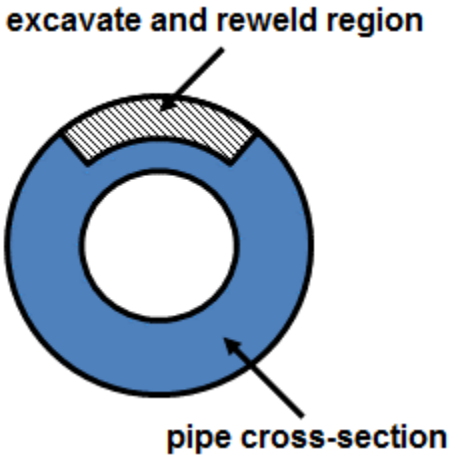


Figure 5-22: A Partial Arc Weld Repair

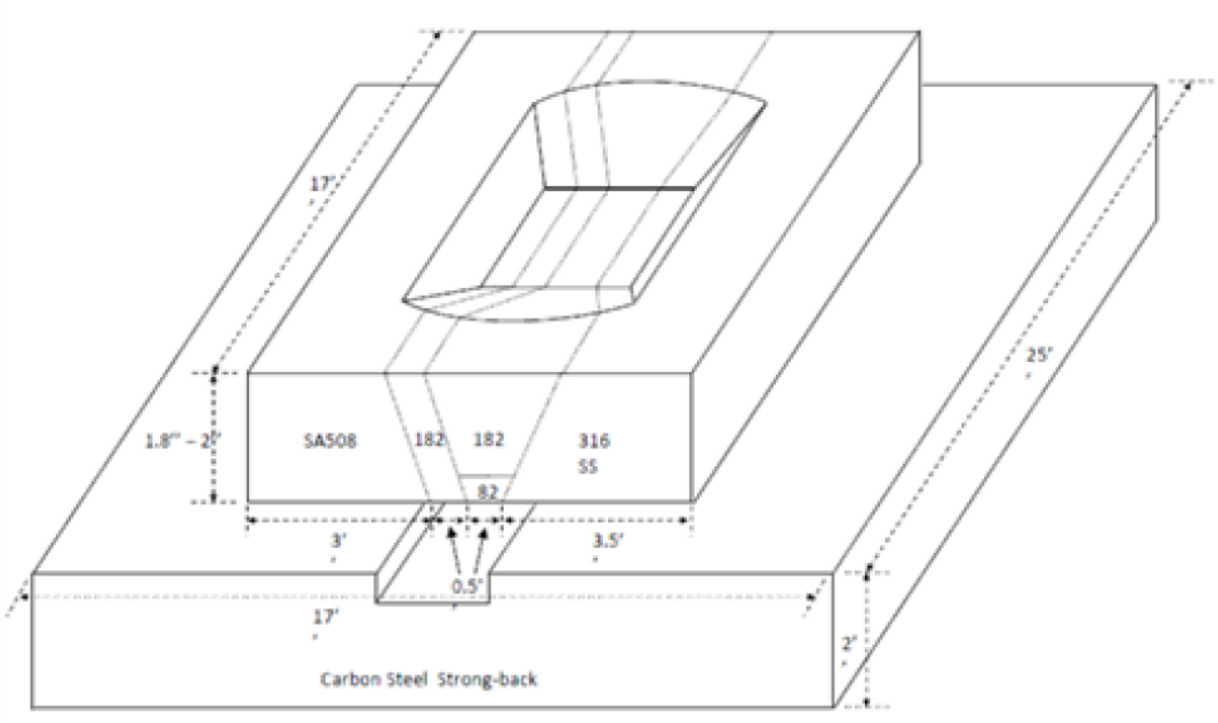


Figure 5-23: EWR Mockup

Furthermore, the validation metrics and associated acceptance measures in Section 5.4 are most directly applicable to a dissimilar metal butt weld case. The metrics and acceptance measures would likely have to be redeveloped for other applications. When developing new metrics and

acceptance measures, the user should carefully consider the end use of the model results. The work presented here was focused on flaw growth calculations, so the metrics and acceptance measures were developed with a flaw growth argument in mind. The same philosophy should be applied to broader applications of this approach.

5.6.2 Welding Process

The Phase 2b mockup was fabricated with a combination of gas tungsten arc welding and shielded metal arc welding [7]. The typical weld bead cross-sectional area was around 35 mm^2, with some variation. Submerged arc welding is often characterized by higher deposition rate, higher heat input, and larger weld beads than the gas tungsten arc and shielded metal arc processes. Large weld beads and high heat inputs can lead to large nodal displacements in an FE model. For these reasons, the proposed validation scheme may not be applicable to submerged arc welding.

5.6.3 Hardening Law

Hardening law has large influence on the results of WRS predictions using FE. Since hardening law is a matter of material behavior, the recommendations on hardening law should not change from application to application. However, if the requisite material data are available, future investigators should continue to study hardening law, including consideration of the Lemaitre-Chaboche approach.

5.6.4 Best Practices for a Plant Application

When modeling a plant application, the analyst should generally apply the modeling guidelines of Section 5.3. Any lessons-learned through the validation process should also be applied. In general, the model should represent the physical situation as closely as possible, including (but not limited to) mechanical constraint condition, groove geometry, material properties, number and size of weld passes and layers, and thermal boundary conditions. Inclusion of a repair weld in the model according to MRP-287 [36] may be advisable, unless plant records demonstrate that such a repair does not exist.

5.7 Conclusion

The procedure recommended in this chapter consists of two aspects: (1) modeling guidelines developed over the course of the EPRI/NRC WRS Validation Program and (2) a validation procedure developed from results of the Phase 2b round robin study. These recommendations drew heavily on the uncertainty quantification methodology presented in Chapter 3. Validation against the Phase 2b dataset is most directly applicable to axisymmetric FE models of dissimilar metal butt welds. This process may be adapted to other applications as discussed in Section 5.6.1, but the analyst should first evaluate whether additional data are needed. For instance, additional WRS measurement data may be needed for unique geometries or welding processes.

The modeling guidelines presented in Section 5.3 were principally drawn from the recommendations of MRP-316 [2] . Many of these guidelines were also provided to the Phase 2b participants, as in reference [7]. The same bootstrapping methods discussed in Chapter 3 were used in Section 5.2.1 to develop a method to compare uncertain FE predictions to uncertain measurements. This led to RMSE and associated confidence bounds, as shown in Figure 5-4 and Figure 5-5. The recommendation to use average hardening was based on qualitative (Table

5-2) and quantitative assessments (Figure 5-4 and Figure 5-5). The Phase 2b study did not include the mixed hardening law. Future work should address mixed hardening law.

Development of a potential validation process involved proposing a benchmark, a set of metrics, and acceptance measures. Choice of a benchmark proved to be challenging due to difficulties in discriminating between measurements that do not entirely agree (see Section 5.4.2). The quality metrics for validation were developed in Section 5.4.7. They include the RMSE on WRS magnitude averaged over the entire wall thickness, the RMSE on the first derivative of the WRS curve, and the average difference up to the initial crack depth. The acceptance measures (Table 5-10 and Table 5-11) were developed based on results of flaw growth calculations. The quality metrics, and their associated acceptance measures, were chosen as a way to ensure a reasonable flaw growth prediction.

6 CONCLUSIONS

This document describes a validation process for FE modeling of WRS and the associated technical background. The validation process was based on the Phase 2b round robin dataset. This dataset consisted of 2 sets of measurement results (contour and DHD) and 10 independent predictions based on axisymmetric FE modeling. Chapter 2 and reference [7] discuss the round-robin study in greater detail. The Phase 2b dataset was input into an uncertainty analysis procedure as the first step in developing the WRS guidelines.

The characterization of WRS uncertainty started with the recognition that WRS is functional data. In other words, the magnitude of the stress depends on the spatial location within the component. Using the methods of [15]-[16], a statistical model of the dataset was created. This model enabled bootstrapping techniques for the determination of appropriate confidence bounds and 95/95 tolerance bounds. Chapter 3 describes this methodology, with further details contained in reference [13].

Chapter 4 of this document explores the impacts of WRS variations on flaw growth calculation. It examines several features of the stress curve (i.e. residual plus operating stresses), including the inner surface stress, the stress at the initial flaw depth, and the through-wall positions where the curve crossed zero. Regulatory staff can review these features to gain confidence in an assumed WRS input found in an industry licensing submittal. A potential review strategy is to compare these features to a reference WRS curve. The reference WRS curve may come from previously-accepted curves or from confirmatory analyses. WRS impacts on flaw growth calculations were also used in Section 5.4 to define quality metrics and acceptance measures for the validation procedure.

The WRS modeling guidance presented here was largely based on reference [11]. However, additional work was performed here to explore three different hardening law choices: isotropic, nonlinear kinematic, and the average of isotropic and kinematic. Qualitative and quantitative comparisons of measurements to models (see Section 5.2) led to average hardening as the recommend hardening approach. Future work should consider mixed hardening as an option.

Three quality metrics were developed in Section 5.4 for the validation process: RMSE on WRS magnitude averaged over the entire wall thickness, the RMSE on the first derivative of the WRS curve, and the average difference up to the initial crack depth. These metrics and associated acceptance measures (see Table 5-10 and Table 5-11) were chosen in a way that promoted reasonable flaw growth results using the Phase 2b WRS predictions.

The validation procedure and WRS input guidelines developed in this document provide a potential method for increasing confidence in WRS predictions for nuclear power applications. Such a process may lead to more efficient NRC reviews of industry submittals and increased regulatory certainty. Actual implementation of the suggestions in this document will depend upon further interactions with industry representatives and consensus standards bodies.

7 REFERENCES

[1] U.S. Nuclear Regulatory Commission, NUREG-2162, “Weld Residual Stress Finite Element Analysis Validation: Part I – Data Development Effort,” Washington, DC, March 2014 (Agencywide Documents Access and Management System (ADAMS) Accession No. ML14087A118).

[2] Electric Power Research Institute, MRP-316 Revision 1, “Materials Reliability Program: Finite-Element Model Validation for Dissimilar Metal Butt-Welds,” Report 3002005498, Palo Alto, CA, September 2015.

[3] U.S. Nuclear Regulatory Commission/Electric Power Research Institute, NUREG-2110/EPRI 1022860, “xLPR Pilot Study Report,” Washington, DC, February 2012.

[4] D. Rudland and C. Harrington, “Initial Development of the Extremely Low Probability of Rupture (xLPR) Version 2.0 Code,” Proceedings of the ASME Pressure Vessels and Piping Conference, PVP2012-78186, Toronto, Canada, July 2012.

[5] Robert E. Kurth, Cedric J. Sallaberry, Frederick W. Brust, Elizabeth A. Kurth, Michael L. Benson, and David L. Rudland, “Uncertainty Sampling of Weld Residual Stress Fields in Probabilistic Analysis: Part I Theory,” Proceedings for the 2016 ASME Pressure Vessels and Piping Conference, PVP2016-63962, Vancouver, Canada, July 2016.

[6] Robert E. Kurth, Cedric J. Sallaberry, Frederick W. Brust, Elizabeth A. Kurth, Michael L. Benson, and David L. Rudland, “Uncertainty Sampling of Weld Residual Stress Fields in Probabilistic Analysis: Part II Example,” Proceedings for the 2016 ASME Pressure Vessels and Piping Conference, PVP2016-63963, Vancouver, Canada, July 2016.

[7] U.S. Nuclear Regulatory Commission, “Phase 2b Finite Element Round Robin Results,” Technical Letter Report, Washington, DC, December 2015 (ADAMS Accession No. ML16020A034).

[8] H. J. Rathbun, L. F. Fredette, P. M. Scott, A. A. Csontos, and D. L. Rudland, "NRC Welding Residual Stress Validation Program International Round Robin Program and Findings," Proceedings of the ASME Pressure Vessels and Piping Conference, PVP2011-57642, Baltimore, MD, July 2011.

[9] L. F. Fredette, M. Kerr, H. J. Rathbun, and J. E. Broussard, "NRC/EPRI Welding Residual Stress Validation Program: Phase III - Details and Findings," PVP2011-57645, Proceedings of the ASME Pressure Vessels and Piping Conference, Baltimore, MD, July 2011.

[10] M. Kerr, D. L. Rudland, M. B. Prime, H. Swenson, M. A. Buechler, and B. Clausen, “Characterization of a Plate Specimen from Phase I of the NRC/EPRI, Weld Residual Stress Program,” PVP2011-57687, Proceedings of the ASME Pressure Vessels and Piping Conference, Baltimore, MD, July 2011.

[11] Electric Power Research Institute, MRP-317, Revision 1, “Materials Reliability Program: Welding Residual Stress Dissimilar Metal Butt Weld Finite Element Modeling Handbook,” Report 3002005499, Palo Alto, CA, September 2015.

[12] Michael L. Benson, Frederick W. Brust, Robert E. Kurth, and John E. Broussard, “Weld Residual Stress Inputs for a Probabilistic Fracture Mechanics Code,” PVP2014-28030, Proceedings of the ASME Pressure Vessels and Piping Conference, Anaheim, CA, July 2014.

[13] J. R. Lewis and D. Brooks, “Uncertainty Quantification and Comparison of Weld Residual Stress Measurements and Predictions,” Sandia National Laboratory, SAND2016-10932, October 2016 (ADAMS Accession No. ML16301A055).

[14] J. R. Lewis, D. Brooks, and M. L. Benson, “Methods for Uncertainty Quantification and Comparison of Weld Residual Stress Measurements and Predictions,” Proceedings of the ASME Pressure Vessels and Piping Conference, PVP2017-65552, Waikoloa, HI, July 2017.

[15] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer, New York, NY, 1997.
[16] J. Derek Tucker, Wei Wu, and Anuj Srivastava, "Generative Models for Functional Data Using Phase and Amplitude Separation," Comparative Statistics and Data Analysis, 61:50–66, 2013
[17] Anthony C. Davison and David V. Hinkley, *Bootstrap Methods and Their Application*, Volume 2, Cambridge University Press, 1997.
[18] Anthony C. Davison, David V. Hinkley, and G. Alastair Young, "Recent Developments in Bootstrap Methodology," Statistical Science, 18:141–157, 2003.
[19] Minh N. Tran, Michael R. Hill, and Mitchell D. Olson, "Further Comments on Validation Approaches for Weld Residual Stress Simulation," Proceedings of the ASME Pressure Vessels and Piping Conference, PVP2015-45751, Boston, MA, July 2015.
[20] Erwin Kreyszig, *Advanced Engineering Mathematics*, 9th Edition, John Wiley & Sons, Inc. Hoboken, NJ, 2006.
[21] G. J. Hahn and W. Q. Meeker, *Statistical Intervals: A Guide for Practitioners*, Vol. 328, John Wiley & Sons, Hoboken, NJ, 2011.
[22] American Society of Mechanical Engineers, *Boiler and Pressure Vessel Code*, Section XI (2015), "Rules for Inservice Inspection of Nuclear Power Plant Components," New York, NY, July 2015.
[23] American Society of Mechanical Engineers, Code Case N-770, "Alternative Examination Requirements and Acceptance Standards for Class 1 PWR Piping and Vessel Nozzle Butt Welds Fabricated with UNS N06082 or UNS W86182 Weld Filler Material with or without Application of Listed Mitigation Activities," New York, NY, June 2011.
[24] Memorandum to G. T. Powell from Robert J. Pascarelli, "South Texas Project, Unit 2 – Request for Relief No. RR-ENG-3-20 for Extension of the Inspection Frequency of the Reactor Vessel Cold-Leg Nozzle to Safe-End Welds with Flaw Analysis," June 30, 2016 (ADAMS Accession No. ML16174A091).
[25] Electric Power Research Institute, MRP-115NP, "Crack Growth Rates for Evaluating Primary Water Stress Corrosion Cracking of Alloy 82, 182, and 132 Welds," Report 1006696, Palo Alto, CA, November 2004.
[26] American Petroleum Institute, API 579-1/ASME FFS-1, *Fitness-for-Service*, Washington, DC, 2007.
[27] S. X. Xu, D. A. Scarth, and R. C. Cipolla, "Technical Basis for Proposed Weight Function Method for Calculation of Stress Intensity Factor for Surface Flaws in ASME Section XI Appendix A," Proceedings of the 2011 ASME Pressure Vessel Piping Conference, PVP2011-57911, Baltimore, MD, July 2011.
[28] S. X. Xu, D. A. Scarth, and R. C. Cipolla, "Calculation of Stress Intensity Factor for Surface Flaws Using Universal Weight Functions with Piece-Wise Cubic Stress Interpolation," Proceedings of the ASME Pressure Vessels and Piping Conference, PVP2012-78236, Toronto, Canada, July 2012.
[29] Electric Power Research Institute, MRP-216, "Materials Reliability Program: Advance FEA Evaluation of Growth of Postulated Circumferential PWSCC Flaws in Pressurizer Nozzle Dissimilar Metal Welds," Report 1015400, Palo Alto, CA, August 2007.
[30] U.S. Nuclear Regulatory Commission, "Weld Residual Stress Analysis of Excavate and Weld Repair Mockup," Technical Letter Report, Washington, DC, September 2016 (ADAMS Accession No. ML16257A523).
[31] Mitchell D. Olson, Adrian T. DeWald, Michael R. Hill, and Steven L. McCracken, "Residual Stress Mapping for an Excavate and Weld Repair Mockup," Proceedings of the ASME Pressure Vessels and Piping Conference, PVP2016-63197, July 2016.

[32] Francis H. Ku, Pete C. Riccardella, and Steven L. McCracken, “3D Residual Stress Simulation of an Excavate and Weld Repair Mockup,” Proceedings of the ASME Pressure Vessels and Piping Conference, PVP2016-63815, July 2016.
[33] Steven McCracken, Jonathan Tatman, and Pete Riccardella, “Technical Basis for Code Case N-847 – Excavate and Weld Repair for SCC Mitigation,” Proceedings of the ASME Pressure Vessels and Piping Conference, PVP2016-63769, July 2016.
[34] J. Lemaitre and J.-L. Chaboche, *Mechanics of Solid Materials*, Cambridge University Press, Cambridge, United Kingdom, 1990.
[35] Richard L. Burden and J. Douglas Faires, *Numerical Analysis*, 6th Edition, Brooks/Cole Publishing Company, Albany, NY, 1997.
[36] Electric Power Research Institute, MRP-287, “Materials Reliability Program: Primary Water Stress Corrosion Cracking (PWSCC) Flaw Evaluation Guidance,” Report 1021023, Palo Alto, CA, December 2010.

APPENDIX A MODEL-MEASUREMENT COMPARISONS

This appendix contains all relevant data plots for model-measurement comparisons, including the calculated mean difference function and associated confidence bounds. It includes a plot for each case described in Section 5.2.2.

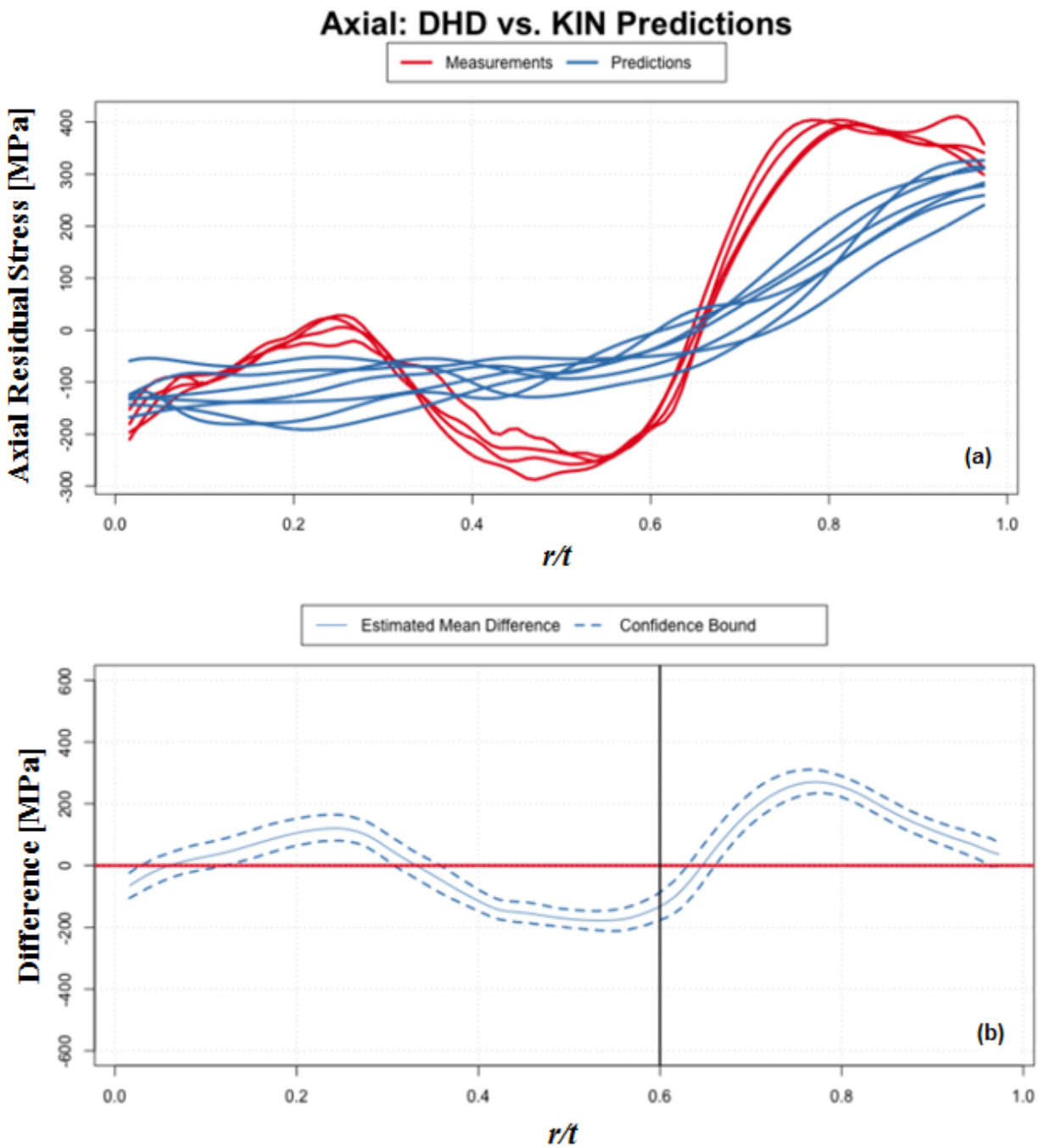


Figure A-1: Axial Stress Component with Nonlinear Kinematic Hardening and DHD Benchmark
(a) Stress Magnitude and (b) Difference in Means with Confidence Bounds

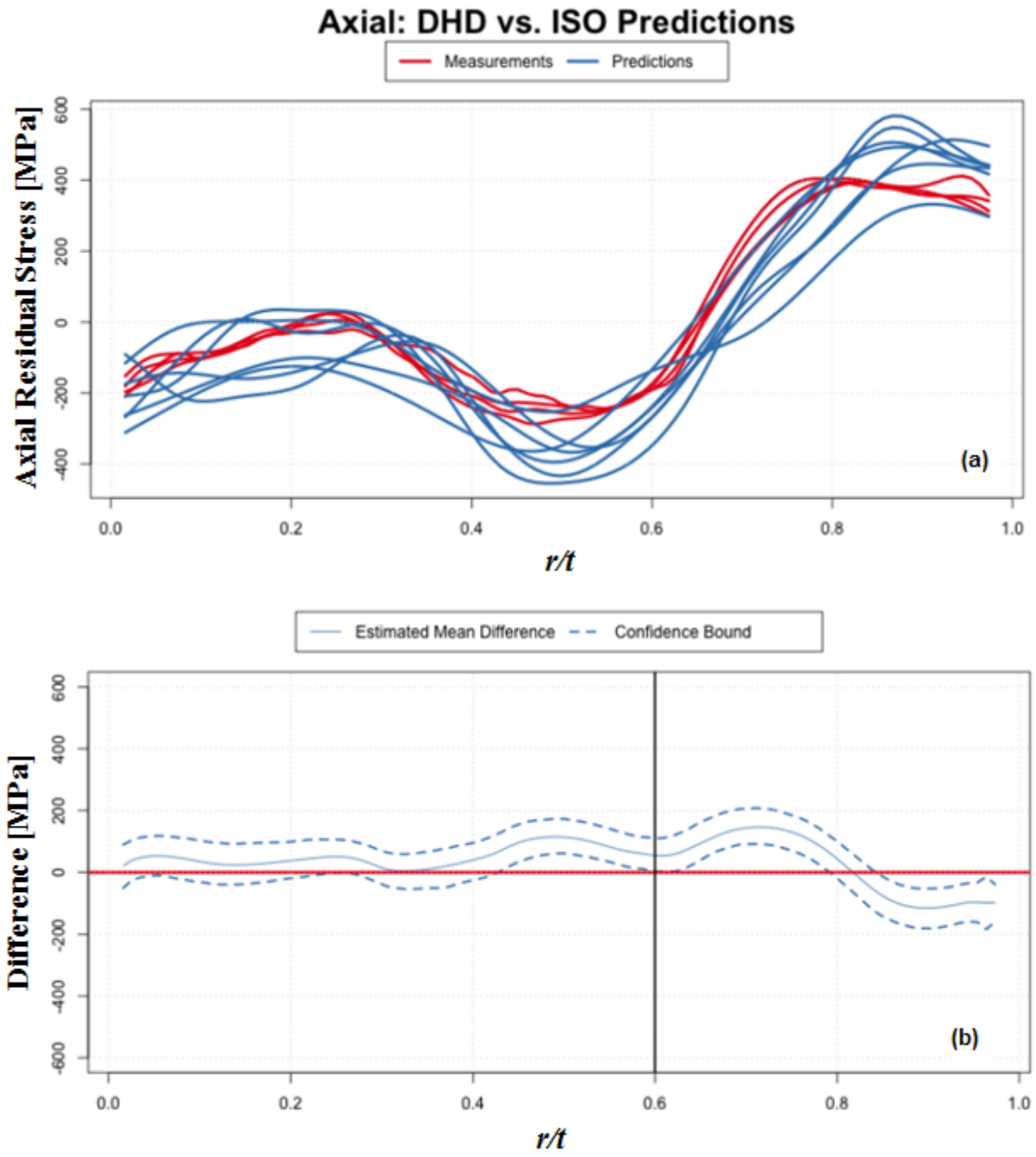


Figure A-2: Axial Stress Component with Isotropic Hardening and DHD Benchmark (a) Stress Magnitude and (b) Difference in Means with Confidence Bounds

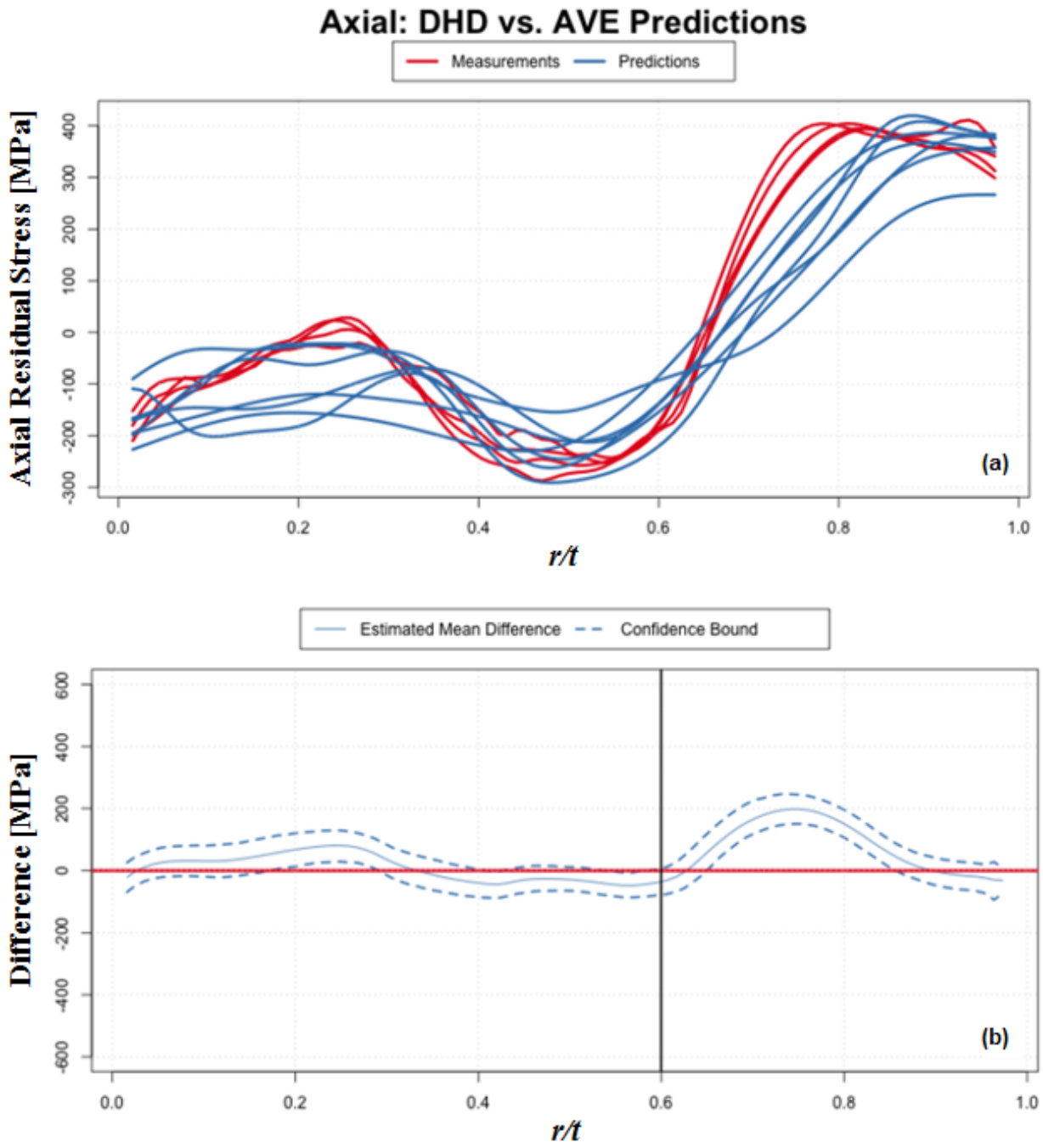


Figure A-3: Axial Stress Component with Average Hardening and DHD Benchmark (a) Stress Magnitude and (b) Difference in Means with Confidence Bounds

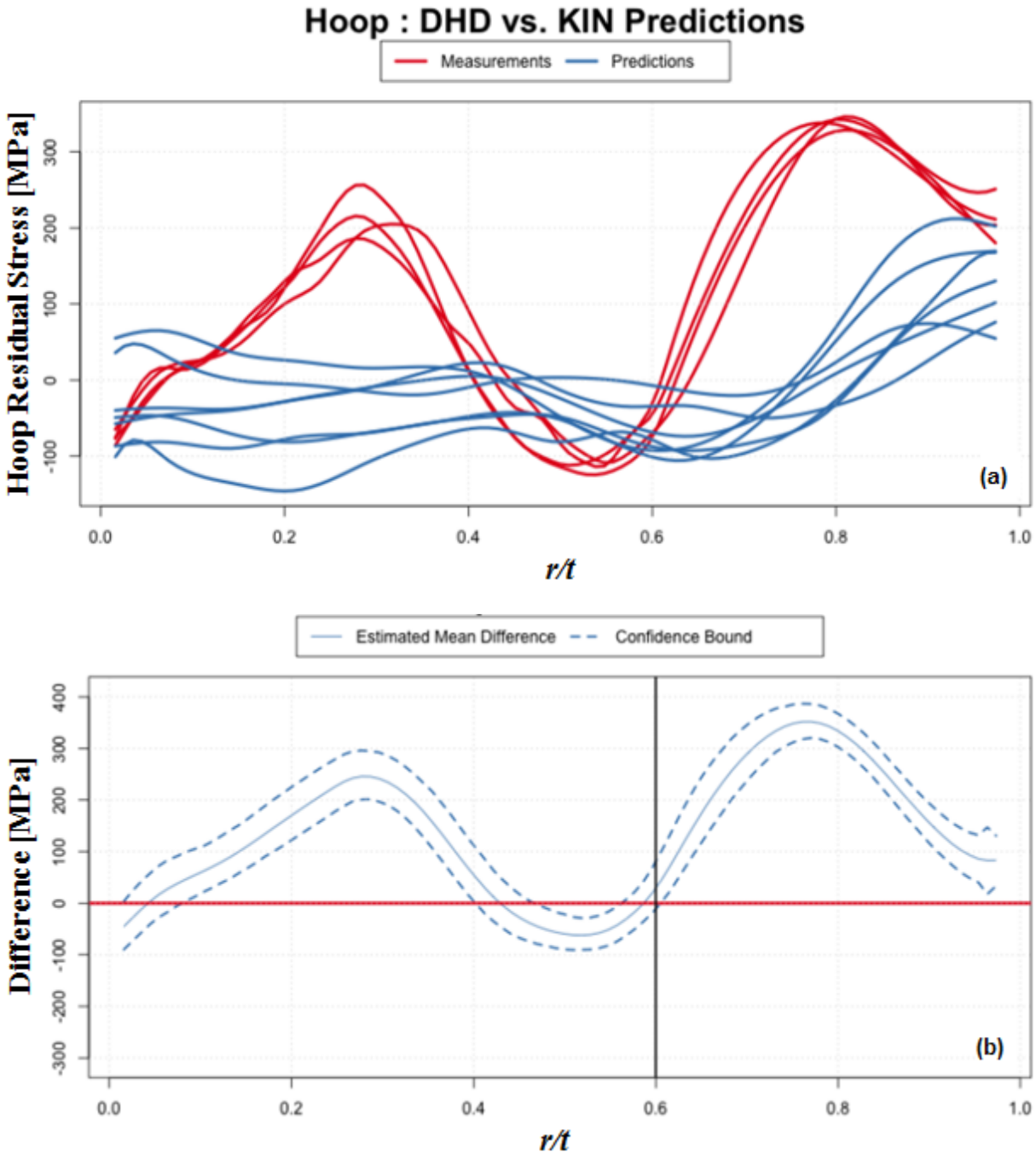


Figure A-4: Hoop Stress Component with Nonlinear Kinematic Hardening and DHD Benchmark
(a) Stress Magnitude and (b) Difference in Means with Confidence Bounds

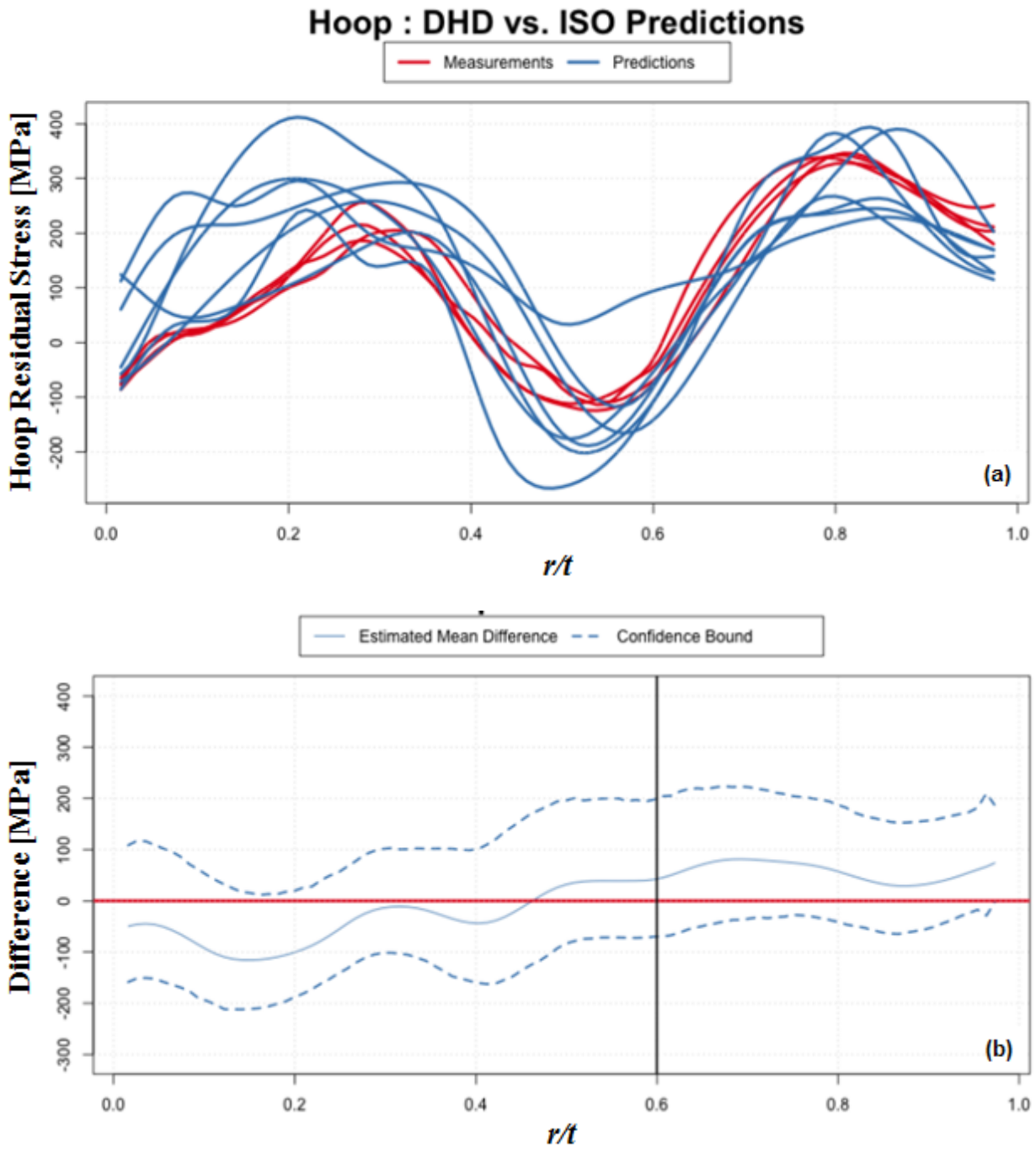


Figure A-5: Hoop Stress Component with Isotropic Hardening and DHD Benchmark (a) Stress Magnitude and (b) Difference in Means with Confidence Bounds

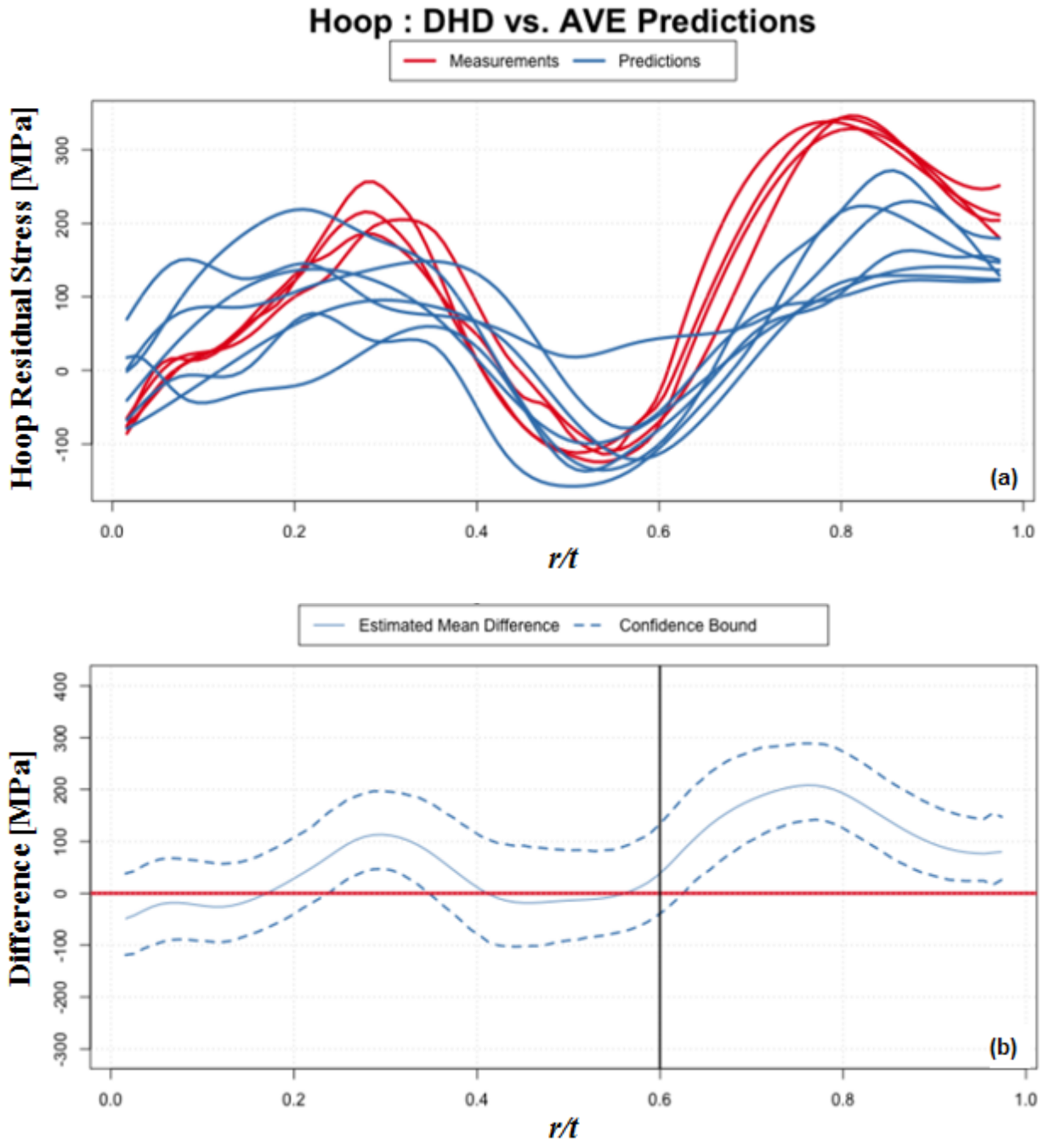


Figure A-6: Hoop Stress Component with Average Hardening and DHD Benchmark (a) Stress Magnitude and (b) Difference in Means with Confidence Bounds

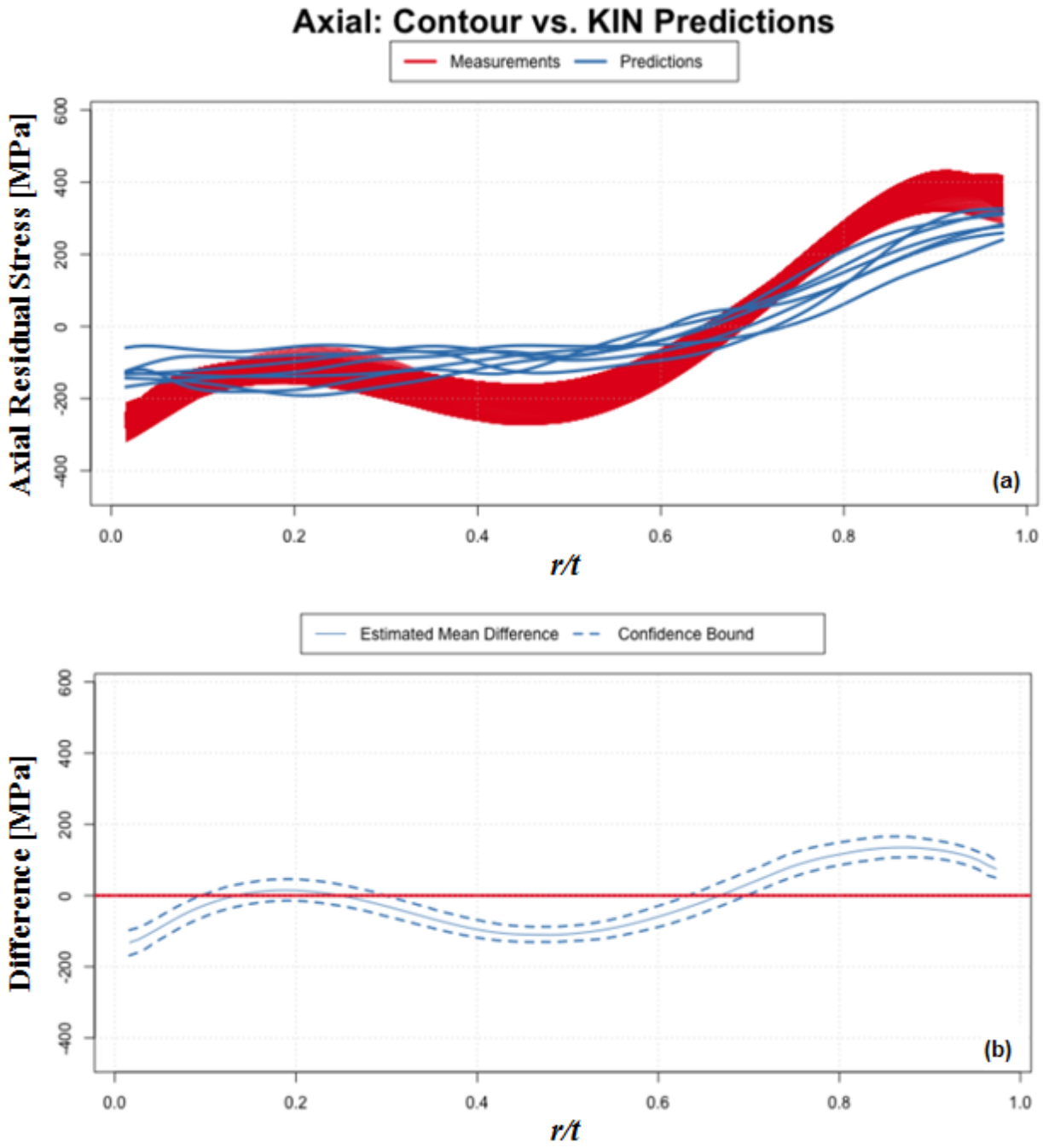


Figure A-7: Axial Stress Component with Nonlinear Kinematic Hardening and Contour Benchmark
(a) Stress Magnitude and (b) Difference in Means with Confidence Bounds

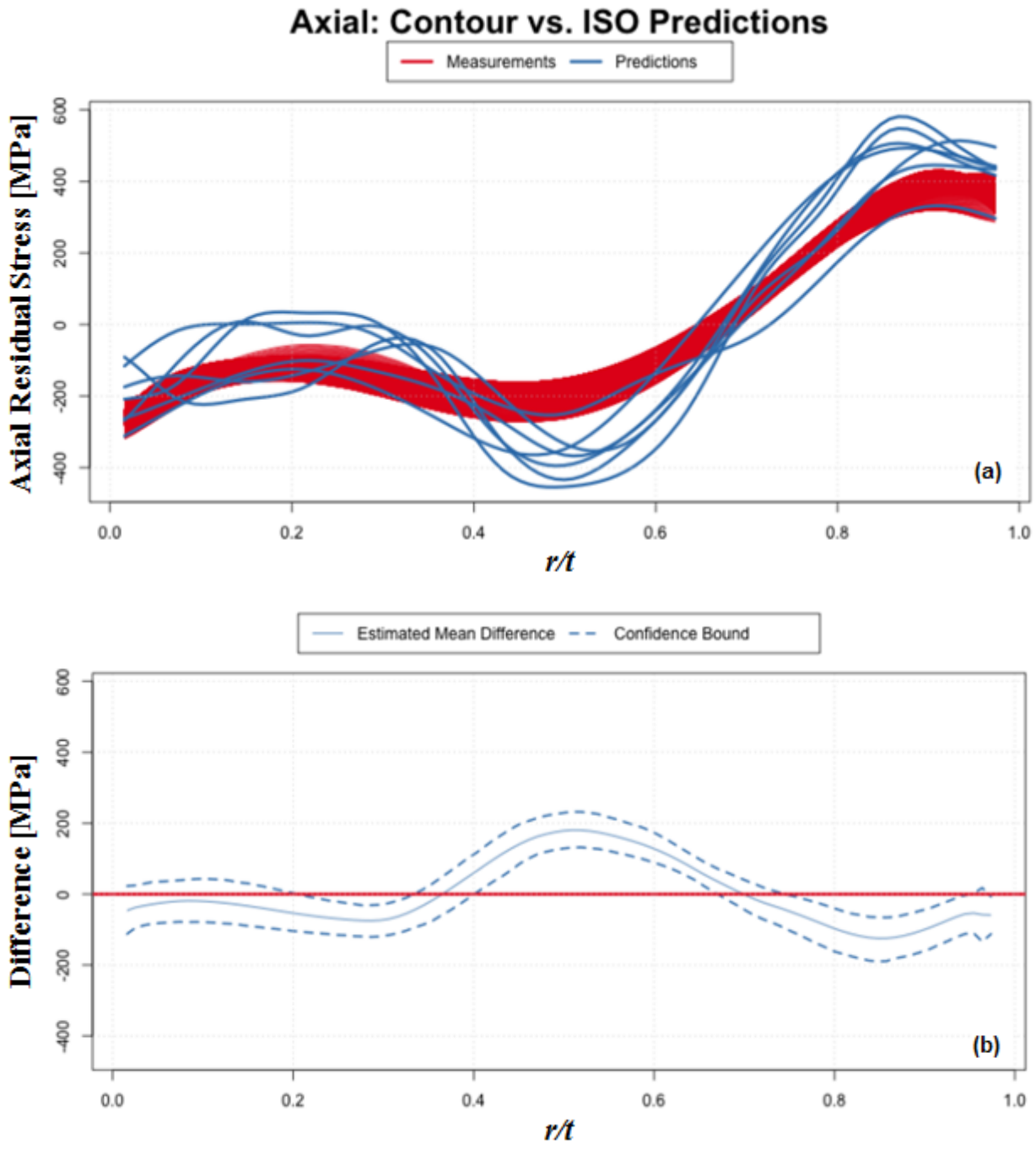


Figure A-8: Axial Stress Component with Isotropic Hardening and Contour Benchmark (a) Stress Magnitude and (b) Difference in Means with Confidence Bounds

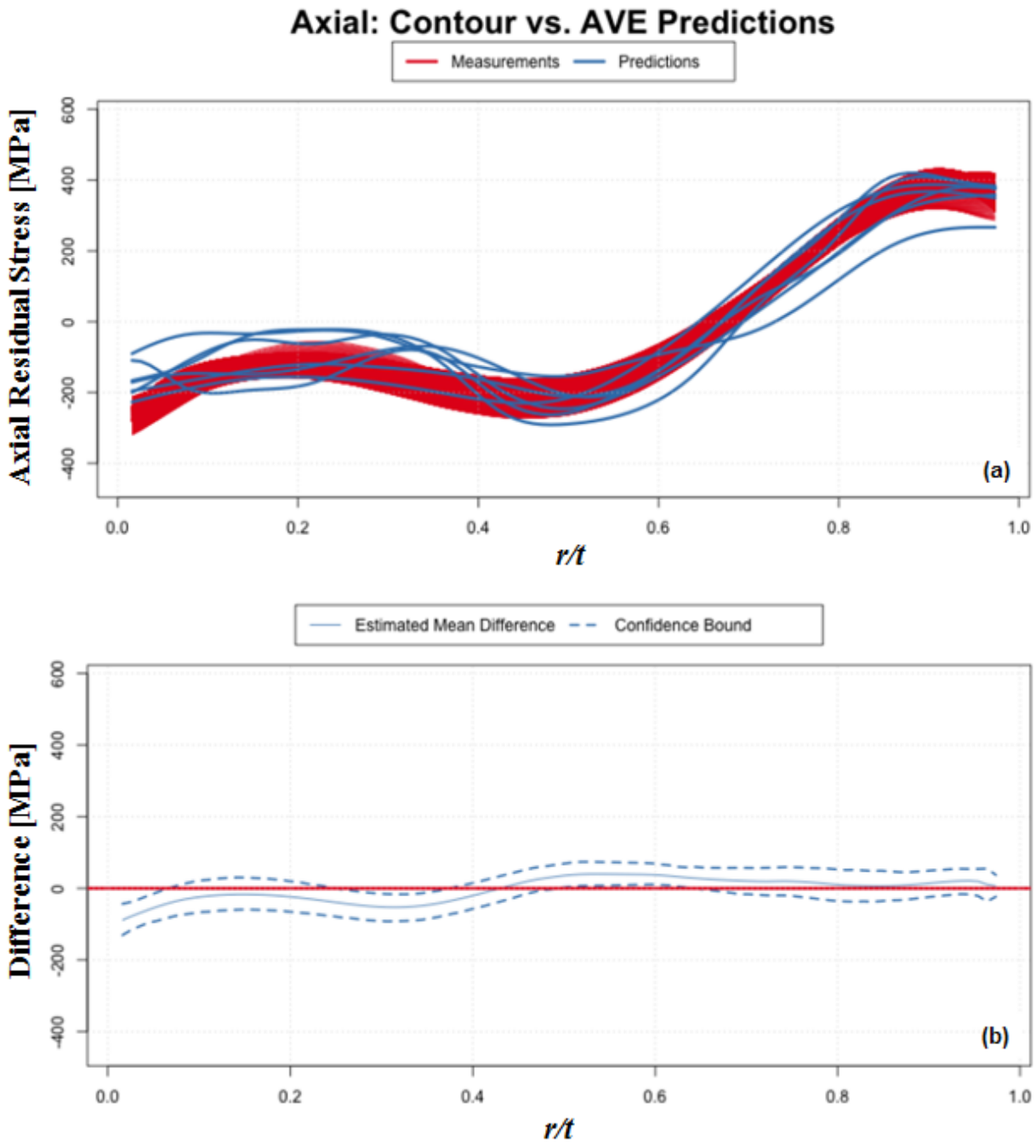


Figure A-9: Axial Stress Component with Average Hardening and Contour Benchmark (a) Stress Magnitude and (b) Difference in Means with Confidence Bounds

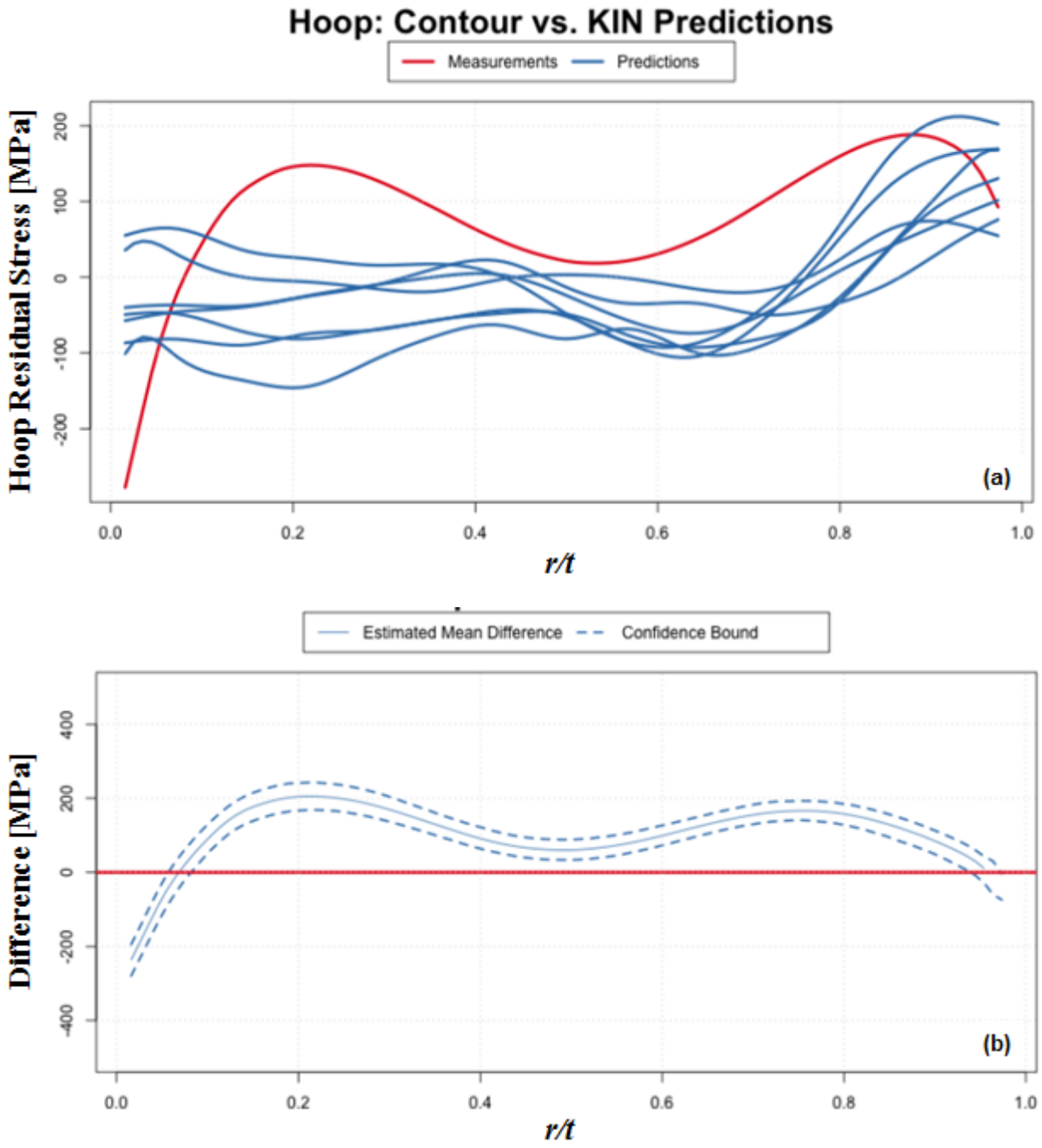


Figure A-10: Hoop Stress Component with Nonlinear Kinematic Hardening and Contour Benchmark
(a) Stress Magnitude and (b) Difference in Means with Confidence Bounds

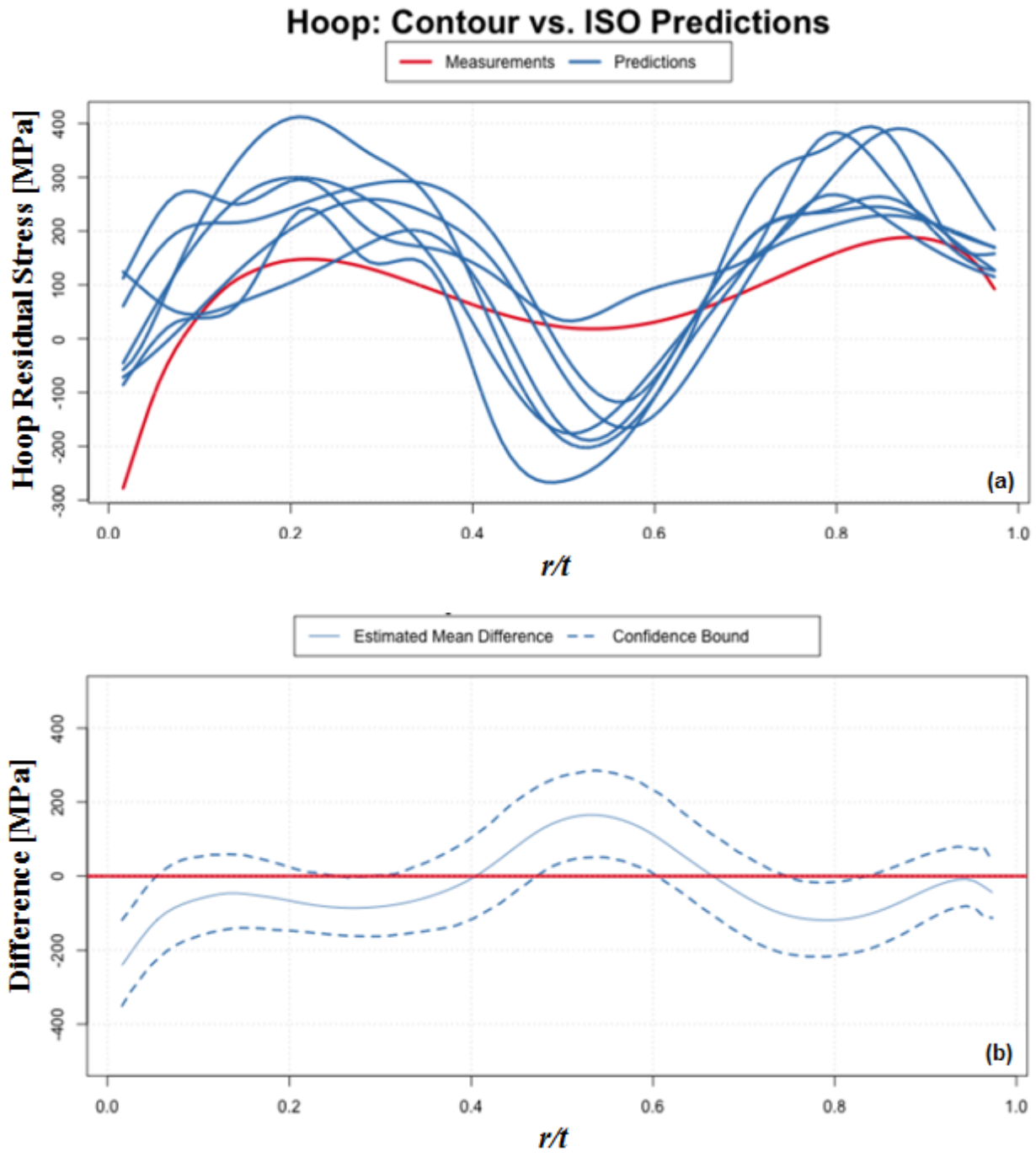


Figure A-11: Hoop Stress Component with Isotropic Hardening and Contour Benchmark (a) Stress Magnitude and (b) Difference in Means with Confidence Bounds

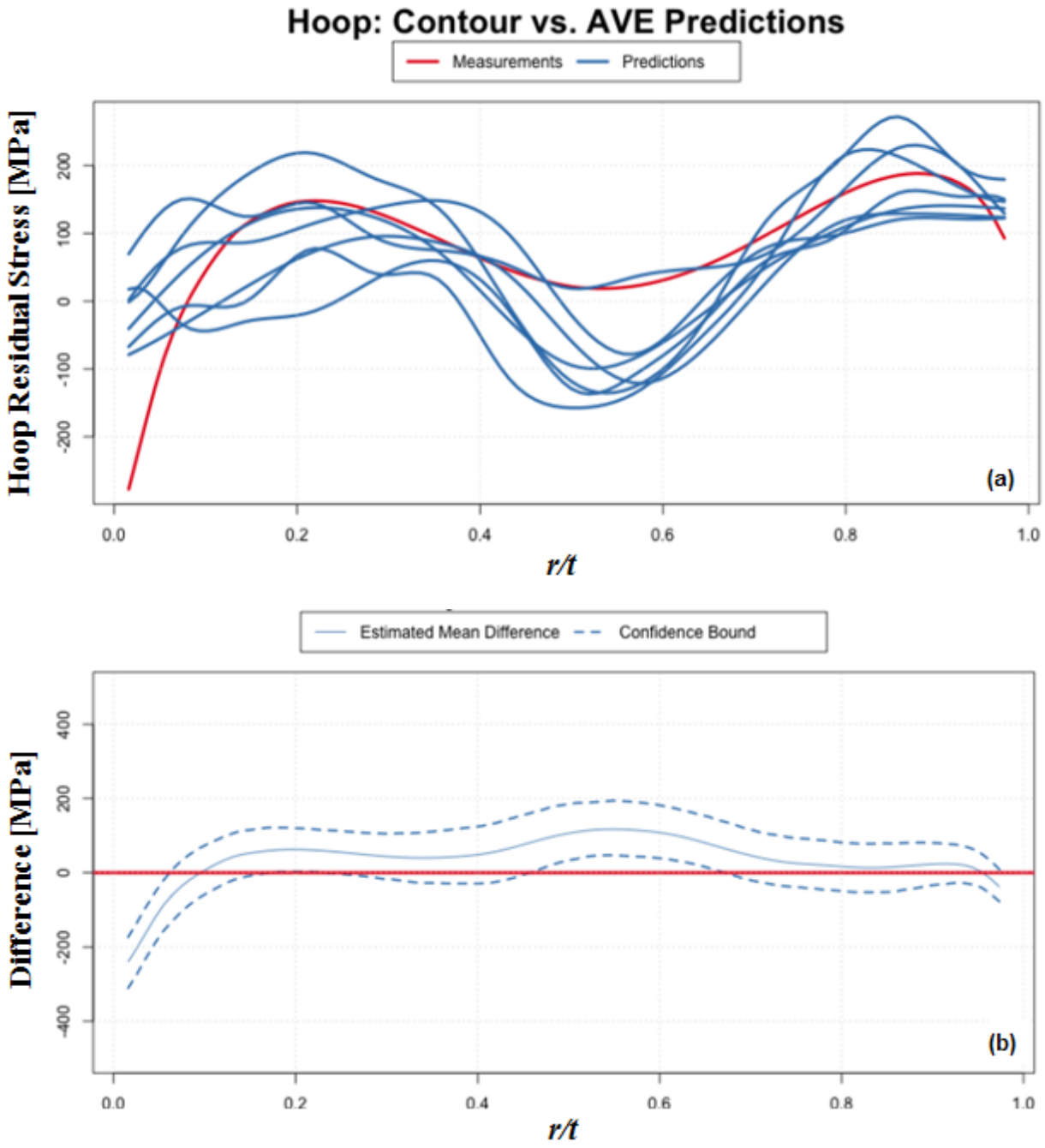


Figure A-12: Hoop Stress Component with Average Hardening and Contour Benchmark (a) Stress Magnitude and (b) Difference in Means with Confidence Bounds

APPENDIX B MATERIAL PROPERTIES

This appendix contains temperature-dependent material properties appropriate for performing the validation procedure. Tables B-1 through B-7 are the Alloy 182 weld properties. Tables B-8 through B-14 are the low alloy steel nozzle properties. Tables B-15 through B-21 are the stainless steel safe end and pipe properties. These data tables use the following conventions.

- T – temperature
- E – elastic modulus
- ν – Poisson's ratio
- CTE – coefficient of thermal expansion
- A, n, m – parameters in the creep law $\dot{\varepsilon} = Aq^{-n}\tau^{m}$, where $\dot{\varepsilon}$ is the uniaxial equivalent creep strain rate, q is the uniaxial equivalent deviatoric stress, and τ is time

Table B-1: Isotropic Hardening Law - Alloy 182

T [K]	strain [mm/mm]	stress [MPa]
295	0.00E+00	207.3
295	1.10E-04	225.0
295	1.40E-04	242.6
295	2.10E-04	259.8
295	3.10E-04	273.2
295	4.40E-04	282.7
295	6.40E-04	290.5
295	8.60E-04	295.2
295	1.04E-03	297.8
295	2.14E-03	305.9
295	3.37E-03	310.9
295	8.21E-03	322.8
295	1.32E-02	331.9
295	1.83E-02	340.3
295	2.31E-02	347.4

295	2.82E-02	355.2
295	3.32E-02	363.1
295	3.82E-02	371.2
295	4.30E-02	379.2
295	4.80E-02	387.4
295	5.30E-02	396.0
295	5.80E-02	404.4
295	6.30E-02	413.1
295	6.80E-02	421.8
295	7.30E-02	430.7
295	7.80E-02	439.6
295	8.29E-02	448.6
295	8.79E-02	457.9
295	9.29E-02	467.1
295	9.80E-02	476.2
295	1.03E-01	484.6
295	1.08E-01	494.5
295	1.13E-01	503.0
295	1.18E-01	511.5
295	1.23E-01	521.4
295	1.28E-01	529.9
295	1.33E-01	538.4
295	1.38E-01	548.3
295	1.43E-01	556.9
295	1.49E-01	567.9

295	1.54E-01	576.4
295	1.58E-01	584.8
295	1.63E-01	593.3
295	1.69E-01	604.6
295	1.78E-01	621.3
295	1.88E-01	640.6
295	1.98E-01	659.7
295	2.08E-01	678.5
589	0.00E+00	162.8
589	1.90E-04	178.7
589	2.30E-04	194.5
589	2.90E-04	204.8
589	4.00E-04	216.5
589	5.80E-04	224.1
589	8.80E-04	229.7
589	2.86E-03	242.5
589	4.72E-03	248.8
589	6.73E-03	254.1
589	8.82E-03	259.0
589	1.87E-02	279.4
589	2.85E-02	299.6
589	3.85E-02	320.7
589	4.84E-02	340.4
589	5.84E-02	356.1
589	6.84E-02	372.5

589	7.84E-02	389.9
589	8.84E-02	402.8
589	9.89E-02	424.0
589	1.09E-01	440.8
589	1.19E-01	456.2
589	1.28E-01	470.8
589	1.38E-01	484.8
589	1.48E-01	502.3
589	1.59E-01	519.4
589	1.68E-01	531.3
589	1.79E-01	552.1
589	1.89E-01	561.7
589	1.99E-01	578.9
811	0.00E+00	169.6
811	1.60E-04	178.5
811	2.10E-04	185.3
811	2.60E-04	191.7
811	3.40E-04	197.7
811	4.30E-04	201.8
811	5.00E-04	204.2
811	6.40E-04	207.4
811	7.40E-04	209.3
811	1.07E-03	211.5
811	1.53E-03	213.5
811	2.53E-03	218.7

811	3.51E-03	221.2
811	4.53E-03	225.2
811	5.53E-03	227.3
811	6.42E-03	229.5
811	7.47E-03	230.7
811	8.48E-03	233.6
811	1.83E-02	258.4
811	2.82E-02	277.4
811	3.81E-02	295.5
811	4.81E-02	307.6
811	5.80E-02	324.8
811	6.80E-02	341.0
811	7.79E-02	350.6
811	8.78E-02	369.7
811	9.79E-02	382.8
811	1.08E-01	397.6
811	1.18E-01	409.2
811	1.28E-01	425.2
811	1.38E-01	429.5
811	1.48E-01	451.5
811	1.58E-01	462.4
811	1.68E-01	478.1
811	1.78E-01	492.1
811	1.88E-01	505.3
811	1.98E-01	518.8

1033	0.00E+00	143.7
1033	1.70E-04	152.2
1033	2.10E-04	165.7
1033	2.60E-04	175.1
1033	4.40E-04	185.5
1033	7.00E-04	191.3
1033	1.00E-03	195.0
1033	1.62E-03	199.6
1033	2.57E-03	204.0
1033	3.61E-03	207.3
1033	4.58E-03	209.8
1033	5.62E-03	212.5
1033	6.59E-03	214.9
1033	7.53E-03	217.3
1033	8.48E-03	219.6
1033	1.84E-02	244.8
1033	2.82E-02	271.9
1033	3.81E-02	297.4
1033	4.80E-02	317.6
1033	5.79E-02	331.5
1033	6.79E-02	340.6
1033	7.83E-02	345.9
1033	8.80E-02	349.8
1033	9.80E-02	353.0
1033	1.08E-01	355.3

1033	1.18E-01	357.0
1033	1.28E-01	358.3
1033	1.38E-01	359.5
1033	1.48E-01	360.3
1033	1.58E-01	361.5
1033	1.68E-01	362.6
1033	1.78E-01	363.7
1033	1.88E-01	364.7
1033	1.98E-01	365.4
1255	0.00E+00	42.6
1255	1.80E-04	43.8
1255	2.90E-04	59.8
1255	4.20E-04	73.7
1255	5.80E-04	84.8
1255	7.20E-04	91.7
1255	9.10E-04	98.2
1255	1.07E-03	102.1
1255	1.25E-03	105.6
1255	1.47E-03	108.4
1255	1.96E-03	112.0
1255	2.48E-03	113.7
1255	3.41E-03	114.6
1255	4.44E-03	114.8
1573	0.00E+00	5.0

Table B-2: Elastic Properties - Alloy 182

T [K]	E [MPa]	ν
273	2.14E+05	0
477	2.10E+05	0.307
589	2.03E+05	0.316
700	1.94E+05	0.325
811	1.62E+05	0.334
922	1.32E+05	0.343
1033	1.14E+05	0.352
1144	9.65E+04	0.361
1255	7.93E+04	0.370
1573	1.00E+04	0.396
1623	1.00E+03	0.400

Table B-3: Coefficient of Thermal Expansion - Alloy 182

T [K]	CTE [K^{-1}]
295.2222	1.17E-05
588.5556	1.34E-05
810.7778	1.40E-05
1033	1.51E-05
1255.222	1.62E-05
1366.333	1.66E-05
1644.111	1.66E-05

Table B-4: Creep Law - Alloy 182

T [K]	A [MPa^n/s]	n	m
293	4.43E-50	4.00	0
840.2	4.43E-40	4.00	0
844	1.44E-21	6.1709	0
894	1.24E-20	6.6426	0
1644.1	1.24E-20	6.6426	0

Table B-5: Thermal Conductivity - Alloy 182

T [K]	Thermal Conductivity [W/mm-K]
300	0.0142
350	0.0151
400	0.016
450	0.0168
500	0.0177
600	0.0195
700	0.0213
800	0.023
900	0.0248
1000	0.0265
1100	0.0283
1200	0.0301
1300	0.0318
1400	0.0336

1500	0.0353
1600	0.0371
1665	0.0383
2000	0.0766

Table B-6: Specific Heat - Alloy 182

T [K]	Specific Heat [J/(g-K)]
73	0.272155
123	0.314025
173	0.351708
223	0.385204
273	0.414513
373	0.46057
473	0.489879
573	0.510814
673	0.527562
773	0.54431
873	0.565245
973	0.590367
1073	0.62805
1123	0.653172

Table B-7: Miscellaneous Properties - Alloy 182

Density [g/mm^3]	Latent Heat [J/g]	Solidus [K]	Liquidus [K]
0.00847	297.6	1573	1618

Table B-8: Isotropic Hardening Law - Low Alloy Steel

T [K]	strain [mm/mm]	stress [MPa]
295	0.000	402
295	0.000	418
295	0.001	415
295	0.001	416
295	0.001	416
295	0.001	416
295	0.002	414
295	0.003	416
295	0.004	419
295	0.005	419
295	0.006	423
295	0.007	427
295	0.008	431
295	0.018	473
295	0.028	513
295	0.038	545
295	0.048	572
295	0.058	593

295	0.068	612
295	0.078	625
295	0.088	635
295	0.098	643
295	0.108	649
295	0.118	652
295	0.128	651
589	0.000	269
589	0.000	289
589	0.000	308
589	0.000	332
589	0.000	344
589	0.001	354
589	0.001	361
589	0.002	374
589	0.003	383
589	0.004	390
589	0.005	397
589	0.006	404
589	0.007	410
589	0.008	415
589	0.018	462
589	0.028	498
589	0.038	525
589	0.047	547

589	0.057	563
589	0.067	576
589	0.077	587
589	0.087	595
589	0.097	600
589	0.108	600
811	0.000	187
811	0.000	211
811	0.000	226
811	0.000	244
811	0.001	263
811	0.001	277
811	0.001	288
811	0.002	300
811	0.003	310
811	0.004	317
811	0.005	322
811	0.006	327
811	0.007	331
811	0.008	335
811	0.008	338
811	0.018	359
811	0.028	372
811	0.039	379
811	0.048	384

811	0.058	387
811	0.068	389
811	0.078	391
811	0.089	392
900	0.000	115
900	0.006	155
900	0.014	168
900	0.026	177
900	0.040	182
900	0.055	182
900	0.065	185
1050	0.000	57
1050	0.007	66
1050	0.015	71
1050	0.028	74
1050	0.041	77
1050	0.058	78
1050	0.076	81
1050	0.088	83
1150	0.000	36
1150	0.005	41
1150	0.016	44
1150	0.029	46
1150	0.050	47
1150	0.079	50

1250	0.000	21
1250	0.011	25
1250	0.025	29
1250	0.053	31
1250	0.084	32
1373	0.000	10
1373	0.010	14
1373	0.030	16
1373	0.072	18
1373	0.135	19
1373	0.173	21

Table B-9: Elastic Properties - Low Alloy Steel

T [K]	E [MPa]	ν
273	2.10E+05	0.280
473	1.95E+05	0.293
673	1.75E+05	0.297
873	1.50E+05	0.310
1073	1.20E+05	0.360
1193	9.70E+04	0.370
1600	1.00E+03	0.400

Table B-10: Thermal Expansion - Low Alloy Steel

T [K]	CTE [K^{-1}]
293	1.38E-05
588.56	1.38E-05
810.78	1.50E-05
1033	1.55E-05
1255.22	1.60E-05
1773	1.60E-05

Table B-11: Creep Law - Low Alloy Steel

T [K]	A [MPa^n/s]	n	m
293	4.43E-50	4.00	0
673	4.43E-40	4.00	0
850	1.95E-17	6.0451	0
894	2.61E-11	4.8865	0
1644.1	2.61E-11	4.8865	0

Table B-12: Thermal Conductivity - Low Alloy Steel

T [K]	Thermal Conductivity [W/(mm-K)]
273	0.0595
373	0.0578
473	0.0532
573	0.0494
673	0.0456
773	0.041
873	0.0368
973	0.0331
1073	0.0285
1273	0.0276
1473	0.0297
2000	0.065

Table B-13: Specific Heat - Low Alloy Steel

T [K]	Specific Heat [J/(g-K)]
348	0.418
448	0.519
498	0.536
548	0.553
598	0.574
648	0.595
748	0.662
848	0.754
948	0.867
998	1.2
1048	0.875
1148	0.846
1185	0.605
1600	0.735
1800	0.835

Table B-14: Miscellaneous Properties - Low Alloy Steel

Density [g/mm^3]	Latent Heat [J/g]	Solidus [K]	Liquidus [K]
7.80E-03	272.0	1500	1600

Table B-15: Isotropic Hardening - Stainless Steel

T [K]	stress [MPa]	strain [mm/mm]
294	0.00E+00	241
294	1.00E-02	290
294	5.00E-02	393
294	1.00E-01	483
294	2.00E-01	621
294	4.00E-01	827
693	0.00E+00	117
693	1.00E-02	193
693	5.00E-02	276
693	1.00E-01	359
693	2.00E-01	483
693	4.00E-01	621
693	8.00E-01	724
923	0.00E+00	97
923	1.00E-02	145
923	5.00E-02	214
923	1.00E-01	283
923	2.00E-01	379
923	4.00E-01	448
923	8.00E-01	483
1089	0.00E+00	69
1089	1.00E-02	110
1089	5.00E-02	145

1089	1.00E-01	172
1089	2.00E-01	207
1089	4.00E-01	241
1089	8.00E-01	241
1366	0.00E+00	28
1366	1.00E-02	39
1366	5.00E-02	41
1366	1.00E-01	41
1366	2.00E-01	41
1366	4.00E-01	41
1366	8.00E-01	41
1550	0.00E+00	5

Table B-16: Elastic Properties - Stainless Steel

T [K]	E [MPa]	ν
300	2.07E+05	0.28
477	1.86E+05	0.29
700	1.67E+05	0.30
922	1.45E+05	0.29
1144	1.14E+05	0.25
1366	6.21E+04	0.25
1477	2.76E+04	0.25
1550	1.00E+04	0.25
1728	2.00E+03	0.25

Table B-17: Thermal Expansion - Stainless Steel

T [K]	CTE [MPa]
296.89	1.46E-05
421.89	1.58E-05
560.78	1.68E-05
644.11	1.72E-05
755.22	1.77E-05
866.33	1.82E-05
977.44	1.84E-05
1088.56	1.87E-05
1173	1.91E-05
1273	1.93E-05
1373	1.96E-05

1673	2.02E-05

Table B-18: Creep Law - Stainless Steel

T [K]	A [MPa^n/s]	n	m
294.1	4.42E-30	4.00	0
748	5.83E-33	9.78	0
773	2.05E-32	9.97	0
798	4.15E-29	9.06	0
823	5.28E-26	8.2	0
848	3.66E-25	8.2	0
873	2.27E-24	8.2	0
898	1.62E-23	8.18	0
923	1.04E-22	8.16	0
948	2.05E-20	7.42	0
973	3.08E-18	6.72	0
998	1.20E-16	6.25	0
1023	4.69E-15	5.77	0

Table B-19: Thermal Conductivity - Stainless Steel

T [K]	Thermal Conductivity [W/(mm-K)]
300	0.0133824
700	0.02091
1200	0.033456
2000	0.08

Table B-20: Specific Heat - Stainless Steel

T [K]	Specific Heat [J/(g-K)]
360	0.343318
700	0.485669
870	0.53591
925	0.544284
1300	0.552658

Table B-21: Miscellaneous Properties - Stainless Steel

Density [g/mm^3]	Latent Heat [J/g]	Solidus [K]	Liquidus [K]
7.90E-03	225.0	1560	1670

APPENDIX C TABLES FOR VALIDATION PROCESS

This appendix contains the tabular data referenced in Section 5.5. These tables are necessary to execute the recommended validation procedure. Table C-1 shows the mean weld residual stress (WRS) profiles from the average hardening predictions in the Phase 2b dataset. The mean chosen here is the cross-section mean (see [13] for a detailed description of determining means of functional data). Table C-2 contains the first derivative of the curves in Table C-1.

Table C-1: Mean of the Average Hardening WRS Predictions

r/t	Axial Mean WRS [MPa]	Hoop Mean WRS [MPa]
0.00	-165	-28
0.01	-159	-18
0.02	-154	-8
0.03	-150	1
0.04	-146	9
0.05	-142	16
0.06	-139	23
0.07	-134	29
0.08	-130	34
0.09	-126	40
0.10	-121	45
0.11	-117	50
0.12	-114	55
0.13	-110	61
0.14	-108	67
0.15	-106	73
0.16	-104	79
0.17	-103	84
0.18	-102	90

0.19	-101	95
0.20	-99	98
0.21	-97	101
0.22	-95	103
0.23	-93	103
0.24	-90	104
0.25	-88	103
0.26	-86	103
0.27	-85	102
0.28	-84	101
0.29	-85	101
0.30	-86	101
0.31	-89	101
0.32	-93	100
0.33	-98	98
0.34	-105	94
0.35	-112	89
0.36	-121	82
0.37	-132	74
0.38	-143	63
0.39	-154	51
0.40	-166	39
0.41	-177	25
0.42	-188	10
0.43	-198	-5

0.44	-206	-20
0.45	-213	-35
0.46	-219	-49
0.47	-223	-63
0.48	-225	-75
0.49	-226	-87
0.51	-226	-96
0.52	-223	-104
0.53	-219	-109
0.54	-214	-113
0.55	-207	-115
0.56	-199	-115
0.57	-190	-113
0.58	-180	-110
0.59	-169	-105
0.60	-157	-98
0.61	-144	-90
0.62	-130	-80
0.63	-115	-70
0.64	-99	-58
0.65	-83	-46
0.66	-65	-33
0.67	-47	-19
0.68	-28	-4
0.69	-9	10

0.70	11	24
0.71	30	38
0.72	49	52
0.73	68	64
0.74	87	76
0.75	105	86
0.76	123	96
0.77	140	105
0.78	158	113
0.79	176	121
0.80	194	129
0.81	213	137
0.82	233	144
0.83	252	151
0.84	271	157
0.85	289	163
0.86	306	167
0.87	321	170
0.88	334	171
0.89	343	171
0.90	351	169
0.91	356	166
0.92	360	163
0.93	362	159
0.94	363	155

0.95	363	151
0.96	362	149
0.97	361	147
0.98	359	145
0.99	357	144
1.00	355	143

Table C-2: First Derivative of the Mean WRS Profiles

r/t	Axial D1 [MPa]	Hoop D1 [MPa]
0.00	N/A	N/A
0.01	567.62	1014.01
0.02	467.53	919.31
0.03	400.60	832.83
0.04	366.84	754.55
0.05	366.18	684.50
0.06	390.84	623.52
0.07	418.42	574.13
0.08	432.59	538.14
0.09	431.70	515.73
0.10	415.75	506.91
0.11	384.73	511.67
0.12	338.66	530.02
0.13	279.24	559.99
0.14	218.38	588.09
0.15	169.93	598.60

0.16	137.59	587.31
0.17	121.37	554.23
0.18	121.26	499.33
0.19	137.27	422.64
0.20	169.04	324.56
0.21	207.75	215.78
0.22	236.19	117.09
0.23	245.22	39.53
0.24	234.45	-16.42
0.25	203.88	-50.76
0.26	153.50	-63.48
0.27	83.32	-54.61
0.28	-5.11	-30.22
0.29	-105.69	-14.29
0.30	-213.03	-27.91
0.31	-326.31	-74.32
0.32	-445.53	-153.55
0.33	-570.70	-265.57
0.34	-701.80	-410.40
0.35	-837.52	-586.25
0.36	-963.86	-774.31
0.37	-1061.07	-948.04
0.38	-1122.00	-1097.82
0.39	-1146.58	-1223.56
0.40	-1134.80	-1325.26

0.41	-1086.67	-1402.92
0.42	-1002.29	-1456.51
0.43	-887.51	-1484.82
0.44	-756.95	-1484.81
0.45	-620.19	-1454.47
0.46	-478.00	-1393.66
0.47	-330.37	-1302.36
0.48	-177.29	-1180.59
0.49	-18.78	-1028.34
0.51	144.04	-848.17
0.52	304.77	-654.39
0.53	456.65	-462.16
0.54	598.17	-274.87
0.55	729.34	-92.51
0.56	850.15	84.91
0.57	960.60	257.40
0.58	1060.84	424.82
0.59	1153.53	584.86
0.60	1243.27	733.47
0.61	1332.22	868.77
0.62	1420.45	990.71
0.63	1507.95	1099.28
0.64	1594.73	1194.49
0.65	1680.77	1276.32
0.66	1763.49	1343.43

0.67	1834.30	1391.27
0.68	1886.33	1416.23
0.69	1918.73	1417.88
0.70	1931.51	1396.21
0.71	1924.67	1351.21
0.72	1898.20	1282.89
0.73	1853.70	1192.83
0.74	1804.51	1094.27
0.75	1767.62	1004.07
0.76	1748.30	927.46
0.77	1746.58	864.46
0.78	1762.45	815.07
0.79	1795.91	779.30
0.80	1846.58	756.92
0.81	1900.38	739.79
0.82	1926.63	710.15
0.83	1907.27	657.54
0.84	1841.24	581.37
0.85	1728.53	481.63
0.86	1569.16	358.32
0.87	1363.11	211.44
0.88	1117.48	46.57
0.89	865.05	-110.50
0.90	637.31	-234.98
0.91	440.07	-322.30

0.92	273.34	-372.45
0.93	137.12	-385.44
0.94	31.39	-361.26
0.95	-44.32	-301.37
0.96	-96.64	-225.16
0.97	-135.83	-162.70
0.98	-166.12	-126.35
0.99	-187.57	-116.33
1.00	N/A	N/A

APPENDIX D ANALYSIS OF VALIDATION METRICS FOR AVERAGE HARDENING

This appendix describes analyses performed to explore potential additional or alternative metrics for weld residual stress (WRS) validation, in complement to those presented in Section 5.4.7.

D.1 Root Mean Square Error on Second Derivative

D.1.1 Definition for $RMSE_{D2}$

The first additional metric investigated was root mean square error (RMSE) on the second derivative of WRS with respect to through-wall position. This quality metric is related to through-wall WRS trends and is a true measure of concavity.

$$D2_k = \left.\frac{dD1}{dx_{norm}}\right|_k \approx \frac{1}{4 \cdot h^2}(WRS_{k+2} - 2 \cdot WRS_k + WRS_{k-2})$$

where $x_{norm} = \frac{r}{t}$ is the through-wall position normalized to the wall thickness, h is the distance between the kth and k+1th data point (h should not vary from interval to interval), WRS_k is the analyst's predicted stress magnitude at the k^{th} position through the wall thickness, and $WRS_k{}^{mean}$ is the cross-sectional mean prediction. The RMSE on the second derivative was calculated according to:

$$RMSE_{D2} = \sqrt{\frac{1}{L-4}\sum_{k=3}^{L-2}(D2_k - D2_k^{mean})}$$

where L is the number of equally-spaced points chosen to represent the WRS profile through the wall thickness (from x/t = 0 to x/t = 1), $D2_k$ is the second derivative of the analyst's predicted stress magnitude at the k^{th} position through the wall thickness, and $D2_k{}^{mean}$ is the second derivative of the cross-sectional mean prediction.

D.1.2 Results and Assessment for $RMSE_{D2}$

$RMSE_{D2}$ was calculated for both axial and hoop WRS, as shown in Figure D-1. The acceptance criterion for $RMSE_{D2}$ was set to $RMSE_{D2}$<10,050 for axial WRS, and to $RMSE_{D2}$<11,000 for hoop WRS. These values were chosen such that all the participants whose WRS prediction resulted in "acceptable" crack growth passed the criterion for the $RMSE_{D2}$ metric. In Figure D-1, the cells highlighted in green on the left-side tables meet the corresponding acceptance measures shown in the tables on the right.

For axial WRS, the acceptance criterion that had to be set for $RMSE_{D2}$ does not allow for a distinction to be made between the WRS predictions that resulted in "acceptable" crack growth (hereafter referred to as "good" predictions) and those that did not (hereafter referred to as "bad" predictions). In fact, all except one of the participants meet the acceptance criterion for the $RMSE_{D2}$ metric, because participant G, who predicted similar crack growth, had the second highest $RMSE_{D2}$ of all participants.

For hoop WRS, in all cases except for participant C, the acceptance criterion that had to be set for $RMSE_{D2}$ allows for a distinction between the WRS predictions that resulted in "acceptable" time to leakage (hereafter referred to as "good" predictions) and those that did not (hereafter referred to as "bad" predictions).

In summary, there is no added benefit to using the $RMSE_{D2}$ metric, since it is not able to distinguish between the "good" and the "bad" WRS predictions without also using the $diff_{ave}$ metric, which was already the case when using the $RMSE_{WRS}$ and $RMSE_{D1}$ metrics.

D2	Axial Stress / Circ Flaw Growth				
Participant	$RMSE_{WRS}$	$RMSE_{D1}$	$RMSE_{D2}$	$diff_{avg}$	**Crack Growth**
D	29	409	8130	-10	Acceptable
E	32	366	6458	11	Acceptable
G	52	512	10025	14	Acceptable
B	79	523	9207	92	Not Acceptable
A	59	374	5882	-55	Not Acceptable
C	28	300	6414	-31	Not Acceptable
F	46	624	12475	-20	Not Acceptable
Min	28	300	5882	-55	
25th percentile	30	370	6436	-26	
Median	46	409	8130	-10	
75th percentile	56	518	9616	12	
Max	79	624	12475	92	

Axial WRS		
Metric		Acceptance
$RMSE_{WRS}$	≤	55
$RMSE_{D1}$	≤	520
$RMSE_{D2}$	≤	10050
$diff_{avg}$	≥	15
$diff_{avg}$	≤	-15

D2	Hoop Stress / Axial Flaw Growth				
Participant	$RMSE_{WRS}$	$RMSE_{D1}$	$RMSE_{D2}$	$diff_{avg}$	**Time to leakage**
D	49	417	8569	47	Acceptable
E	26	330	5583	13	Acceptable
G	68	525	10951	63	Acceptable
B	74	717	14028	114	Not Acceptable
A	51	524	15332	-37	Not Acceptable
C	34	331	7421	-56	Not Acceptable
F	50	591	13245	-30	Not Acceptable
Min	26	330	5583	-56	
25th percentile	42	374	7995	-33	
Median	50	524	10951	13	
75th percentile	60	558	13636	55	
Max	74	717	15332	114	

Hoop WRS		
Metric		Acceptance
$RMSE_{WRS}$	≤	70
$RMSE_{D1}$	≤	550
$RMSE_{D2}$	≤	11000
$diff_{avg}$	≥	0
$diff_{avg}$	≤	65

Figure D-1: Quality metrics for Phase 2b isotropic predictions, including $RMSE_{D2}$. (left), and acceptance measures for the proposed metrics (right).

D.2 Truncated Metrics

D.2.1 Definitions for Truncated Metrics

Truncated $RMSE_{WRS}$ and $RMSE_{D1}$ metrics were investigated to determine if a truncation limit for the calculation of RMSE would improve the ability to distinguish between the good and bad WRS predictions. The following equations define $(RMSE_{WRS})_T$ and $(RMSE_{D1})_T$.

$$(RMSE_{WRS})_T = \sqrt{\frac{1}{L_T}\sum_{k=1}^{L_T}(WRS_k - WRS_k^{mean})}$$

$$(RMSE_{D1})_T = \sqrt{\frac{1}{L_T - 1}\sum_{k=2}^{L_T}(D1_k - D1_k^{mean})}$$

where L_T is the number of equally-spaced points where WRS or D1 are known, up to the truncation limit T, with 0<T<1. For example, for T=0.55, if there are 100 equally-spaced points through the wall thickness where WRS and D1 are known, L_T would be equal to 55. WRS_k and $D1_k$ are the WRS and first derivative of the WRS at the k^{th} position through the wall thickness, and WRS_k^{mean} and $D1_k^{mean}$ are the WRS and first derivative of the cross-sectional mean prediction.

$(RMSE_{WRS})_T$ and $(RMSE_{D1})_T$ were calculated for T=0.1, 0.2,…, 0.9, as shown in D.2.2. In addition, $(RMSE_{WRS})_T$ and $(RMSE_{D1})_T$ were calculated for T=a/t(10 years) and T=a/t(20 years), as shown in Section D.2.3 and Section D.2.4, respectively.

D.2.2 Assessment of Arbitrary Truncation

Figure D-2 through Figure D-10 show the results for the analysis of truncated metrics $(RMSE_{WRS})_T$ and $(RMSE_{D1})_T$ for T=0.1, 0.2,…, 0.9. As in previous figures, the cells highlighted in green on the left-side tables meet the corresponding acceptance measures shown in the tables on the right. In addition, the cells with green text and green background in the tables on the right (showing the acceptance measures) indicate the cases where the acceptance measures could be narrowed as compared to the criteria in Sections 5.4.9 and 5.4.11. These are the cases where using the truncated metrics results in a better ability to distinguish between good and bad WRS predictions.

Below each figure in this section, a case-by-case assessment is provided. Overall, the truncated metrics do not offer any clear improvement over the basic metrics proposed in this report (see Sections 5.4.8 and 5.4.10).

T=0.1	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	10	278	-10	Acceptable
E	28	890	11	Acceptable
G	17	553	14	Acceptable
B	93	397	92	Not Acceptable
A	56	141	-55	Not Acceptable
C	31	64	-31	Not Acceptable
F	49	1517	-20	Not Acceptable
Min	10	64	-55	
25th percentile	22	209	-26	
Median	31	397	-10	
75th percentile	52	722	12	
Max	93	1517	92	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	30
$(RMSE_{D1})_T$	≤	900
$diff_{avg}$	≥	15
$diff_{avg}$	≤	-15

T=0.1	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	48	412	47	Acceptable
E	19	544	13	Acceptable
G	64	899	63	Acceptable
B	117	667	114	Not Acceptable
A	36	374	-37	Not Acceptable
C	57	211	-56	Not Acceptable
F	52	1508	-30	Not Acceptable
Min	19	211	-55	
25th percentile	42	393	-26	
Median	52	544	-10	
75th percentile	61	783	12	
Max	117	1508	92	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	65
$(RMSE_{D1})_T$	≤	900
$diff_{avg}$	≥	0
$diff_{avg}$	≤	65

Figure D-2: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.1

Assessment of truncated metrics for T=0.1:

- Acceptance measures:
  - Narrower (improved) $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall no improvement in acceptance measures
- Ability to distinguish between good and bad predictions:
  - Similar to basic metrics proposed in 5.4.8 and 5.4.10

T=0.2	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	27	285	-10	Acceptable
E	42	655	11	Acceptable
G	51	518	14	Acceptable
B	85	336	92	Not Acceptable
A	54	110	-55	Not Acceptable
C	29	91	-31	Not Acceptable
F	68	1050	-20	Not Acceptable
Min	27	91	-55	
25th percentile	35	197	-26	
Median	51	336	-10	
75th percentile	61	587	12	
Max	85	1050	92	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	55
$(RMSE_{D1})_T$	≤	675
$diff_{avg}$	≥	15
$diff_{avg}$	≤	-15

T=0.2	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	38	394	47	Acceptable
E	35	411	13	Acceptable
G	97	665	63	Acceptable
B	94	648	114	Not Acceptable
A	45	461	-37	Not Acceptable
C	49	205	-56	Not Acceptable
F	80	1065	-30	Not Acceptable
Min	35	205	-55	
25th percentile	41	403	-26	
Median	49	461	-10	
75th percentile	87	657	12	
Max	97	1065	92	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	100
$(RMSE_{D1})_T$	≤	675
$diff_{avg}$	≥	0
$diff_{avg}$	≤	65

Figure D-3: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.2

Assessment of truncated metrics for T=0.2:

- Acceptance measures:
  - Identical $(RMSE_{WRS})_T$ acceptance criterion for axial WRS
  - Wider (worse) $(RMSE_{WRS})_T$ acceptance criterion for hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall widening of acceptance measures
- Ability to distinguish between good and bad predictions:
  - Similar to basic metrics proposed in 5.4.8 and 5.4.10

T=0.3	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	24	329	-10	Acceptable
E	41	541	11	Acceptable
G	56	464	14	Acceptable
B	78	369	92	Not Acceptable
A	62	226	-55	Not Acceptable
C	31	167	-31	Not Acceptable
F	63	969	-20	Not Acceptable
Min	24	167	-55	
25th percentile	36	277	-26	
Median	56	369	-10	
75th percentile	62	502	12	
Max	78	969	92	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	60
$(RMSE_{D1})_T$	≤	550
$diff_{avg}$	≥	15
$diff_{avg}$	≤	-15

T=0.3	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	35	361	47	Acceptable
E	34	371	13	Acceptable
G	98	615	63	Acceptable
B	79	651	114	Not Acceptable
A	45	461	-37	Not Acceptable
C	42	233	-56	Not Acceptable
F	85	939	-30	Not Acceptable
Min	34	233	-55	
25th percentile	38	366	-26	
Median	45	461	-10	
75th percentile	82	633	12	
Max	98	939	92	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	100
$(RMSE_{D1})_T$	≤	620
$diff_{avg}$	≥	0
$diff_{avg}$	≤	65

Figure D-4: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.3

Assessment of truncated metrics for T=0.3:

- Acceptance measures:
  - Wider (worse) $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall widening of acceptance measures
- Ability to distinguish between good and bad predictions:
  - Similar to basic metrics proposed in 5.4.8 and 5.4.10

T=0.4	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	29	390	-10	Acceptable
E	41	510	11	Acceptable
G	51	598	14	Acceptable
B	69	391	92	Not Acceptable
A	68	301	-55	Not Acceptable
C	31	288	-31	Not Acceptable
F	55	876	-20	Not Acceptable
Min	29	288	-55	
25th percentile	36	345	-26	
Median	51	391	-10	
75th percentile	61	554	12	
Max	69	876	92	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	55
$(RMSE_{D1})_T$	≤	600
$diff_{avg}$	≥	15
$diff_{avg}$	≤	-15

T=0.4	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	43	401	47	Acceptable
E	30	390	13	Acceptable
G	88	595	63	Acceptable
B	69	631	114	Not Acceptable
A	51	476	-37	Not Acceptable
C	36	281	-56	Not Acceptable
F	76	844	-30	Not Acceptable
Min	30	281	-55	
25th percentile	40	395	-26	
Median	51	476	-10	
75th percentile	72	613	12	
Max	88	844	92	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	90
$(RMSE_{D1})_T$	≤	600
$diff_{avg}$	≥	0
$diff_{avg}$	≤	65

Figure D-5: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.4

Assessment of truncated metrics for T=0.4:

- Acceptance measures:
  - Identical $(RMSE_{WRS})_T$ acceptance criterion for axial WRS
  - Wider (worse) $(RMSE_{WRS})_T$ acceptance criterion for hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall widening of acceptance measures
- Ability to distinguish between good and bad predictions:
  - Similar to basic metrics proposed in 5.4.8 and 5.4.10

T=0.5	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	34	409	-10	Acceptable
E	37	487	11	Acceptable
G	54	570	14	Acceptable
B	67	400	92	Not Acceptable
A	62	398	-55	Not Acceptable
C	29	301	-31	Not Acceptable
F	51	806	-20	Not Acceptable
Min	29	301	-55	
25th percentile	36	399	-26	
Median	51	409	-10	
75th percentile	58	528	12	
Max	67	806	92	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	55
$(RMSE_{D1})_T$	≤	575
$diff_{avg}$	≥	15
$diff_{avg}$	≤	-15

T=0.5	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	58	407	47	Acceptable
E	29	373	13	Acceptable
G	79	567	63	Acceptable
B	70	693	114	Not Acceptable
A	63	521	-37	Not Acceptable
C	38	308	-56	Not Acceptable
F	68	765	-30	Not Acceptable
Min	29	308	-55	
25th percentile	48	390	-26	
Median	63	521	-10	
75th percentile	69	630	12	
Max	79	765	92	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	80
$(RMSE_{D1})_T$	≤	575
$diff_{avg}$	≥	0
$diff_{avg}$	≤	65

Figure D-6: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.5

Assessment of truncated metrics for T=0.5:

- Acceptance measures:
  - Identical $(RMSE_{WRS})_T$ acceptance criterion for axial WRS
  - Wider (worse) $(RMSE_{WRS})_T$ acceptance criterion for hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall widening of acceptance measures
- Ability to distinguish between good and bad predictions:
  - Similar to basic metrics proposed in 5.4.8 and 5.4.10

T=0.6	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	32	407	-10	Acceptable
E	35	444	11	Acceptable
G	57	526	14	Acceptable
B	67	386	92	Not Acceptable
A	58	403	-55	Not Acceptable
C	27	286	-31	Not Acceptable
F	47	748	-20	Not Acceptable
Min	27	286	-55	
25th percentile	33	395	-26	
Median	47	407	-10	
75th percentile	58	485	12	
Max	67	748	92	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	60
$(RMSE_{D1})_T$	≤	550
$diff_{avg}$	≥	15
$diff_{avg}$	≤	-15

T=0.6	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	56	438	47	Acceptable
E	27	361	13	Acceptable
G	73	518	63	Acceptable
B	84	661	114	Not Acceptable
A	59	514	-37	Not Acceptable
C	35	385	-56	Not Acceptable
F	63	718	-30	Not Acceptable
Min	27	361	-55	
25th percentile	46	411	-26	
Median	59	514	-10	
75th percentile	68	589	12	
Max	84	718	92	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	75
$(RMSE_{D1})_T$	≤	525
$diff_{avg}$	≥	0
$diff_{avg}$	≤	65

Figure D-7: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.6

Assessment of truncated metrics for T=0.6:

- Acceptance measures:
  - Wider (worse) $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial WRS
  - Narrower (improved) $(RMSE_{D1})_T$ acceptance criterion for hoop WRS
  - Overall widening of acceptance measures
- Ability to distinguish between good and bad predictions:
  - Similar to basic metrics proposed in 5.4.8 and 5.4.10

T=0.7	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	30	389	-10	Acceptable
E	33	415	11	Acceptable
G	59	530	14	Acceptable
B	64	565	92	Not Acceptable
A	59	382	-55	Not Acceptable
C	25	282	-31	Not Acceptable
F	44	701	-20	Not Acceptable
Min	25	282	-55	
25th percentile	31	386	-26	
Median	44	415	-10	
75th percentile	59	548	12	
Max	64	701	92	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	60
$(RMSE_{D1})_T$	≤	550
$diff_{avg}$	≥	15
$diff_{avg}$	≤	-15

T=0.7	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	54	411	47	Acceptable
E	26	357	13	Acceptable
G	68	526	63	Acceptable
B	85	741	114	Not Acceptable
A	56	477	-37	Not Acceptable
C	36	365	-56	Not Acceptable
F	58	664	-30	Not Acceptable
Min	26	357	-55	
25th percentile	45	388	-26	
Median	56	477	-10	
75th percentile	63	595	12	
Max	85	741	92	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	70
$(RMSE_{D1})_T$	≤	550
$diff_{avg}$	≥	0
$diff_{avg}$	≤	65

Figure D-8: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.7

Assessment of truncated metrics for T=0.7:

- Acceptance measures:
  - Wider (worse) $(RMSE_{WRS})_T$ acceptance criterion for axial WRS
  - Identical $(RMSE_{WRS})_T$ acceptance criterion for hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial WRS
  - Identical $(RMSE_{D1})_T$ acceptance criterion for hoop WRS
  - Overall widening of acceptance measures
- Ability to distinguish between good and bad predictions:
  - Similar to basic metrics proposed in 5.4.8 and 5.4.10

T=0.8	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	29	402	-10	Acceptable
E	33	389	11	Acceptable
G	55	518	14	Acceptable
B	69	572	92	Not Acceptable
A	62	364	-55	Not Acceptable
C	28	290	-31	Not Acceptable
F	43	660	-20	Not Acceptable
Min	28	290	-55	
25th percentile	31	377	-26	
Median	43	402	-10	
75th percentile	59	545	12	
Max	69	660	92	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	60
$(RMSE_{D1})_T$	≤	525
$diff_{avg}$	≥	15
$diff_{avg}$	≤	-15

T=0.8	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	51	446	47	Acceptable
E	26	348	13	Acceptable
G	66	508	63	Acceptable
B	80	716	114	Not Acceptable
A	53	551	-37	Not Acceptable
C	35	349	-56	Not Acceptable
F	55	640	-30	Not Acceptable
Min	26	348	-55	
25th percentile	43	397	-26	
Median	53	508	-10	
75th percentile	61	596	12	
Max	80	716	92	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	70
$(RMSE_{D1})_T$	≤	510
$diff_{avg}$	≥	0
$diff_{avg}$	≤	65

Figure D-9: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.8

Assessment of truncated metrics for T=0.8:

- Acceptance measures:
  - Wider (worse) $(RMSE_{WRS})_T$ acceptance criterion for axial WRS
  - Identical $(RMSE_{WRS})_T$ acceptance criterion for hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial WRS
  - Narrower (better) $(RMSE_{D1})_T$ acceptance criterion for hoop WRS
  - Overall widening of acceptance measures
- Ability to distinguish between good and bad predictions:
  - Similar to basic metrics proposed in 5.4.8 and 5.4.10

T=0.9	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	30	400	-10	Acceptable
E	34	370	11	Acceptable
G	54	528	14	Acceptable
B	76	541	92	Not Acceptable
A	62	390	-55	Not Acceptable
C	29	310	-31	Not Acceptable
F	47	635	-20	Not Acceptable
Min	29	310	-55	
25th percentile	32	380	-26	
Median	47	400	-10	
75th percentile	58	535	12	
Max	76	635	92	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	55
$(RMSE_{D1})_T$	≤	530
$diff_{avg}$	≥	15
$diff_{avg}$	≤	-15

T=0.9	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	51	430	47	Acceptable
E	27	335	13	Acceptable
G	69	510	63	Acceptable
B	77	695	114	Not Acceptable
A	54	542	-37	Not Acceptable
C	35	339	-56	Not Acceptable
F	53	617	-30	Not Acceptable
Min	27	335	-55	
25th percentile	43	384	-26	
Median	53	510	-10	
75th percentile	61	579	12	
Max	77	695	92	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	70
$(RMSE_{D1})_T$	≤	515
$diff_{avg}$	≥	0
$diff_{avg}$	≤	65

Figure D-10: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.9

Assessment of truncated metrics for T=0.9:

- Acceptance measures:
  - Identical $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial WRS
  - Narrower (better) $(RMSE_{D1})_T$ acceptance criterion for hoop WRS
  - Overall similar acceptance measures
- Ability to distinguish between good and bad predictions:
  - Similar to basic metrics proposed in 5.4.8 and 5.4.10

D.2.3 Assessment of Truncation at 10 Years Crack Growth

Figure D-11 shows the results for the analysis of truncated metrics $(RMSE_{WRS})_T$ and $(RMSE_{D1})_T$ for T=a/t(10 years), which vary by participant. As in previous figures, the cells highlighted in green

on the left-side tables meet the corresponding acceptance measures shown in the tables on the right. In addition, the cells with green text and green background in the right-side tables (showing the acceptance measures) indicate the cases where the acceptance measures could be narrowed as compared to the criteria presented in Sections 5.4.9 and 5.4.11. These are the cases where using the truncated metrics results in a better ability to distinguish between good and bad WRS predictions.

T = a/t (10 yrs)	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	22	311	-10	Acceptable
E	38	496	11	Acceptable
G	52	594	14	Acceptable
B	79	520	92	Not Acceptable
A	56	141	-55	Not Acceptable
C	31	59	-31	Not Acceptable
F	52	1442	-20	Not Acceptable
Min	22	59	-55	
25th percentile	34	226	-26	
Median	52	496	-10	
75th percentile	54	557	12	
Max	79	1442	92	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	55
$(RMSE_{D1})_T$	≤	600
$diff_{avg}$	≥	15
$diff_{avg}$	≤	-15

T = a/t (10 yrs)	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	49	415	47	Acceptable
E	27	352	13	Acceptable
G	68	522	63	Acceptable
B	74	713	114	Not Acceptable
A	48	432	-37	Not Acceptable
C	40	225	-56	Not Acceptable
F	86	947	-30	Not Acceptable
Min	27	225	-55	
25th percentile	44	383	-26	
Median	49	432	-10	
75th percentile	71	618	12	
Max	86	947	92	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	70
$(RMSE_{D1})_T$	≤	525
$diff_{avg}$	≥	0
$diff_{avg}$	≤	65

Figure D-11: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=a/t(10 years)

Assessment of truncated metrics for T=a/t(10 years):

- Acceptance measures:
  - Identical $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial WRS
  - Narrower (better) $(RMSE_{D1})_T$ acceptance criterion for hoop WRS
  - Overall similar acceptance measures
- Ability to distinguish between good and bad predictions:
  - Similar to basic metrics proposed in 5.4.8 and 5.4.10

Overall, it can be said that truncated metrics with T=a/t(10 years) do not offer any clear improvement over the basic metrics proposed in this report (see 5.4.8 and 5.4.10).

D.2.4 Assessment of Truncation at 20 Years Crack Growth

Figure D-12 shows the results for the analysis of truncated metrics $(RMSE_{WRS})_T$ and $(RMSE_{D1})_T$ for T=a/t(20 years), which varies by participant. As in previous figures, the cells highlighted in green on the left-side tables meet the corresponding acceptance measures shown in the tables on the right. In addition, the cells with green text and green background in the right-side tables (showing the acceptance measures) indicate the cases where the acceptance measures could be narrowed as compared to the criteria presented in Sections 5.4.9 and 5.4.11. These are the cases where using the truncated metrics results in a better ability to distinguish between good and bad WRS predictions.

T = a/t (20 yrs)	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	27	286	-10	Acceptable
E	33	422	11	Acceptable
G	54	549	14	Acceptable
B	79	520	92	Not Acceptable
A	56	141	-55	Not Acceptable
C	31	63	-31	Not Acceptable
F	56	1376	-20	Not Acceptable
Min	27	63	-55	
25th percentile	32	213	-26	
Median	54	422	-10	
75th percentile	56	535	12	
Max	79	1376	92	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	55
$(RMSE_{D1})_T$	≤	550
$diff_{avg}$	≥	15
$diff_{avg}$	≤	-15

T = a/t (20 yrs)	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	49	415	47	Acceptable
E	26	329	13	Acceptable
G	68	522	63	Acceptable
B	74	713	114	Not Acceptable
A	63	529	-37	Not Acceptable
C	35	384	-56	Not Acceptable
F	66	739	-30	Not Acceptable
Min	26	329	-55	
25th percentile	42	399	-26	
Median	63	522	-10	
75th percentile	67	621	12	
Max	74	739	92	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	70
$(RMSE_{D1})_T$	≤	525
$diff_{avg}$	≥	0
$diff_{avg}$	≤	65

Figure D-12: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T= a/t(20 years)

Assessment of truncated metrics for T=a/t(20 years):

- Acceptance measures:
  - Identical $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial WRS
  - Narrower (better) $(RMSE_{D1})_T$ acceptance criterion for hoop WRS
  - Overall similar acceptance measures
- Ability to distinguish between good and bad predictions:
  - Similar to basic metrics proposed in 5.4.8 and 5.4.10

Overall, the truncated metrics with T=a/t(20 years) do not offer any clear improvement over the basic metrics proposed in this report (see Sections 5.4.8 and 5.4.10).

D.3 Weighted Metrics

D.3.1 Definitions for Weighted Metrics

Truncated $RMSE_{WRS}$ and $RMSE_{D1}$ metrics were investigated to determine if a weight function for the calculation of RMSE would improve the ability to distinguish between the good and bad WRS predictions. The following equations define $(RMSE_{WRS})_W$ and $(RMSE_{D1})_W$:

$$(RMSE_{WRS})_W = \sqrt{\frac{1}{L}\sum_{k=1}^{L}\left(1-\frac{x_k}{t}\right)^W \cdot (WRS_k - WRS_k^{mean})}$$

$$(RMSE_{D1})_W = \sqrt{\frac{1}{L-2}\sum_{k=2}^{L-1}\left(1-\frac{x_k}{t}\right)^W \cdot (D1_k - D1_k^{mean})}$$

where L is the number of equally-spaced points chosen to represent the WRS profile through the wall thickness (from x/t = 0 to x/t = 1), W is the weight exponent, x_k/t is the normalized through-thickness distance at the k^{th} position through the wall thickness, WRS_k and $D1_k$ are the WRS and first derivative of the WRS at the k^{th} position through the wall thickness, and WRS_k^{mean} and $D1_k^{mean}$ are the WRS and first derivative of the cross-sectional mean prediction.

$(RMSE_{WRS})_W$ and $(RMSE_{D1})_W$ were calculated for W=1, 2, 5, 10, as shown in Section D.3.2. The weight function gives higher weight to values of (WRS_k-WRS_k^{mean}) or ($D1_k$-$D1_k^{mean}$) closer to x/t = 0. Furthermore, as the weight exponent increases, the weight function eventually takes values smaller than 1%, essentially truncating the RMSE calculation beyond a certain point, as illustrated in Figure D-13.

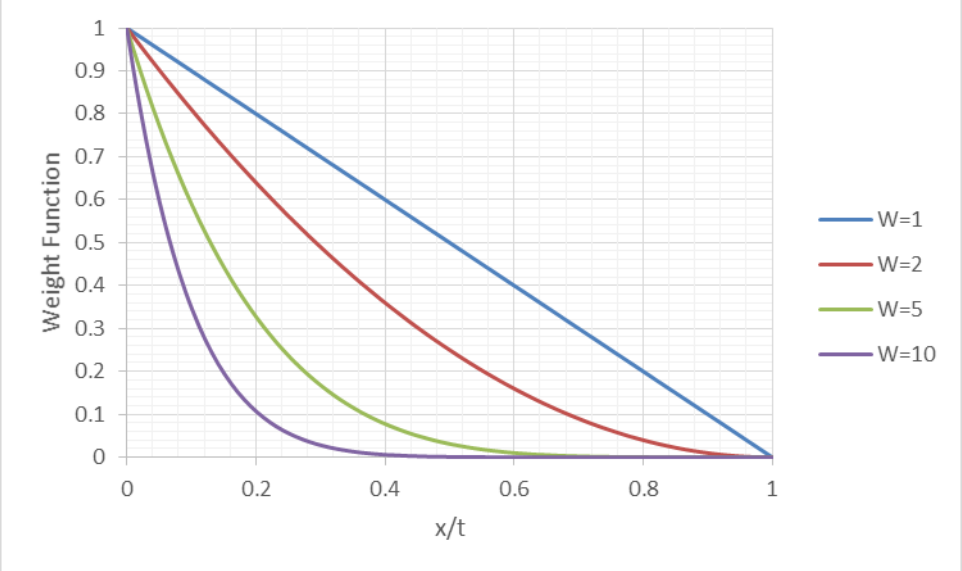


Figure D-13: Weight Function as a Function of Weight Exponent W

D.3.2 Results and Assessment for Weighted Metrics

Figure D-14 though Figure D-17 show the results for the analysis of weighted metrics $(RMSE_{WRS})_W$ and $(RMSE_{D1})_W$ for W=1, 2, 5, and 10. As in previous figures, the cells highlighted in green on the left-side tables meet the corresponding acceptance measures shown in the tables on the right. In addition, the cells with green text and green background in the right-side tables (showing the acceptance measures) indicate the cases where the acceptance measures could be narrowed as compared to the criteria presented in 5.4.9 and 5.4.11. These are the cases where using the weighted metrics results in a better ability to distinguish between good and bad WRS predictions.

Below each figure in this section, a case-by-case assessment is provided. Overall, weighted metrics allow for a significant narrowing of the acceptance measures for both axial and hoop WRS, which is an improvement over the basic metrics proposed in Sections 5.4.8 and 5.4.10. However, since the ability to distinguish between good and bad predictions is the same as for the basic metrics proposed in this report (see Sections 5.4.8 and 5.4.10), the additional analysis work required to implement weighted metrics is not considered worthwhile.

W=1	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_W$	$(RMSE_{D1})_W$	$diff_{avg}$	**Crack Growth**
D	21	277	-10	Acceptable
E	25	319	11	Acceptable
G	38	376	14	Acceptable
B	51	346	92	Not Acceptable
A	43	255	-55	Not Acceptable
C	20	198	-31	Not Acceptable
F	35	543	-20	Not Acceptable
Min	20	198	-55	
25th percentile	23	266	-26	
Median	35	319	-10	
75th percentile	41	361	12	
Max	51	543	92	

Weighted

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_W$	≤	40
$(RMSE_{D1})_W$	≤	380
$diff_{avg}$	≥	15
$diff_{avg}$	≤	-15

W=1	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_W$	$(RMSE_{D1})_W$	$diff_{avg}$	**Time to leakage**
D	36	297	47	Acceptable
E	20	256	13	Acceptable
G	53	389	63	Acceptable
B	57	489	114	Not Acceptable
A	38	359	-37	Not Acceptable
C	26	235	-56	Not Acceptable
F	44	524	-30	Not Acceptable
Min	20	235	-55	
25th percentile	31	277	-26	
Median	38	359	-10	
75th percentile	48	439	12	
Max	57	524	92	

Weighted

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_W$	≤	55
$(RMSE_{D1})_W$	≤	390
$diff_{avg}$	≥	0
$diff_{avg}$	≤	65

Figure D-14: Weighted Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for W=1

Assessment of weighted metrics for W=1:

- Acceptance measures:
  - Narrower (improved) $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Narrower (improved) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall significant improvement in acceptance measures
- Ability to distinguish between good and bad predictions:
  - Similar to basic metrics proposed in 5.4.8 and 5.4.10

W=2	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_W$	$(RMSE_{D1})_W$	$diff_{avg}$	**Crack Growth**
D	17	216	-10	Acceptable
E	21	296	11	Acceptable
G	31	308	14	Acceptable
B	42	254	92	Not Acceptable
A	35	191	-55	Not Acceptable
C	17	148	-31	Not Acceptable
F	31	504	-20	Not Acceptable
Min	17	148	-55	
25th percentile	19	204	-26	
Median	31	254	-10	
75th percentile	33	302	12	
Max	42	504	92	

Weighted

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_W$	≤	35
$(RMSE_{D1})_W$	≤	310
$diff_{avg}$	≥	15
$diff_{avg}$	≤	-15

W=2	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_W$	$(RMSE_{D1})_W$	$diff_{avg}$	**Time to leakage**
D	28	237	47	Acceptable
E	17	219	13	Acceptable
G	47	338	63	Acceptable
B	47	391	114	Not Acceptable
A	31	280	-37	Not Acceptable
C	23	179	-56	Not Acceptable
F	40	490	-30	Not Acceptable
Min	17	179	-55	
25th percentile	26	228	-26	
Median	31	280	-10	
75th percentile	43	365	12	
Max	47	490	92	

Weighted

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_W$	≤	50
$(RMSE_{D1})_W$	≤	340
$diff_{avg}$	≥	0
$diff_{avg}$	≤	65

Figure D-15: Weighted Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for W=2

Assessment of weighted metrics for W=2:

- Acceptance measures:
  - Narrower (improved) $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Narrower (improved) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall significant improvement in acceptance measures
- Ability to distinguish between good and bad predictions:
  - Similar to basic metrics proposed in 5.4.8 and 5.4.10

W=5	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_W$	$(RMSE_{D1})_W$	$diff_{avg}$	**Crack Growth**
D	10	134	-10	Acceptable
E	16	256	11	Acceptable
G	20	212	14	Acceptable
B	33	156	92	Not Acceptable
A	25	98	-55	Not Acceptable
C	12	77	-31	Not Acceptable
F	24	439	-20	Not Acceptable
Min	10	77	-55	
25th percentile	14	116	-26	
Median	20	156	-10	
75th percentile	24	234	12	
Max	33	439	92	

Weighted

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_W$	≤	20
$(RMSE_{D1})_W$	≤	260
$diff_{avg}$	≥	15
$diff_{avg}$	≤	-15

W=5	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_W$	$(RMSE_{D1})_W$	$diff_{avg}$	**Time to leakage**
D	18	164	47	Acceptable
E	12	169	13	Acceptable
G	35	270	63	Acceptable
B	37	269	114	Not Acceptable
A	19	178	-37	Not Acceptable
C	19	103	-56	Not Acceptable
F	30	434	-30	Not Acceptable
Min	12	103	-55	
25th percentile	18	167	-26	
Median	19	178	-10	
75th percentile	33	270	12	
Max	37	434	92	

Weighted

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_W$	≤	35
$(RMSE_{D1})_W$	≤	275
$diff_{avg}$	≥	0
$diff_{avg}$	≤	65

Figure D-16: Weighted Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for W=5

Assessment of weighted metrics for W=5:

- Acceptance measures:
  - Narrower (improved) $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Narrower (improved) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall significant improvement in acceptance measures
- Ability to distinguish between good and bad predictions:
  - Similar to basic metrics proposed in 5.4.8 and 5.4.10

W=10	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_W$	$(RMSE_{D1})_W$	$diff_{avg}$	**Crack Growth**
D	6	83	-10	Acceptable
E	11	221	11	Acceptable
G	12	148	14	Acceptable
B	26	117	92	Not Acceptable
A	18	49	-55	Not Acceptable
C	9	35	-31	Not Acceptable
F	18	381	-20	Not Acceptable
Min	6	35	-55	
25th percentile	10	66	-26	
Median	12	117	-10	
75th percentile	18	185	12	
Max	26	381	92	

Weighted

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_W$	≤	15
$(RMSE_{D1})_W$	≤	225
$diff_{avg}$	≥	15
$diff_{avg}$	≤	-15

W=10	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_W$	$(RMSE_{D1})_W$	$diff_{avg}$	**Time to leakage**
D	13	126	47	Acceptable
E	8	136	13	Acceptable
G	24	220	63	Acceptable
B	31	198	114	Not Acceptable
A	13	116	-37	Not Acceptable
C	16	67	-56	Not Acceptable
F	20	381	-30	Not Acceptable
Min	8	67	-55	
25th percentile	13	121	-26	
Median	16	136	-10	
75th percentile	22	209	12	
Max	31	381	92	

Weighted

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_W$	≤	25
$(RMSE_{D1})_W$	≤	225
$diff_{avg}$	≥	0
$diff_{avg}$	≤	65

Figure D-17: Weighted Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for W=10

Assessment of weighted metrics for W=10:

- Acceptance measures:
  - Narrower (improved) $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Narrower (improved) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall significant improvement in acceptance measures
- Ability to distinguish between good and bad predictions:
  - Similar to basic metrics proposed in 5.4.8 and 5.4.10

APPENDIX E ANALYSIS OF VALIDATION METRICS FOR ISOTROPIC HARDENING

This appendix describes analyses performed to explore potential additional or alternative metrics for weld residual stress (WRS) validation when using an isotropic hardening law, in complement to those presented in Section 5.4.7.

E.1 Root Mean Square Error on Second Derivative

E.1.1 Definition for $RMSE_{D2}$

The first additional metric investigated was root mean square error (RMSE) on the second derivative of WRS with respect to through-wall position. This quality metric is related to through-wall WRS trends and is a true measure of concavity.

$$D2_k = \left.\frac{dD1}{dx_{norm}}\right|_k \approx \frac{1}{4 \cdot h^2}(WRS_{k+2} - 2 \cdot WRS_k + WRS_{k-2})$$

where $x_{norm} = \frac{r}{t}$ is the through-wall position normalized to the wall thickness, h is the distance between the k^{th} and $k+1^{th}$ data point (h should not vary from interval to interval), WRS_k is the analyst's predicted stress magnitude at the k^{th} position through the wall thickness, and $WRS_k{}^{mean}$ is the cross-sectional mean prediction. The RMSE on the second derivative was calculated according to:

$$RMSE_{D2} = \sqrt{\frac{1}{L-4}\sum_{k=3}^{L-2}(D2_k - D2_k^{mean})}$$

where L is the number of equally-spaced points chosen to represent the WRS profile through the wall thickness (from x/t = 0 to x/t = 1), $D2_k$ is the second derivative of the analyst's predicted stress magnitude at the k^{th} position through the wall thickness, and $D2_k{}^{mean}$ is the second derivative of the cross-sectional mean prediction.

E.1.2 Results and Assessment for $RMSE_{D2}$

$RMSE_{D2}$ was calculated for both axial and hoop WRS, as shown in Figure E-1. The acceptance criterion for $RMSE_{D2}$ was set to $RMSE_{D2}$<22340 for axial WRS, and to $RMSE_{D2}$<23145 for hoop WRS. These values were chosen such that all the participants whose WRS prediction resulted in "acceptable" crack growth passed the criterion for the $RMSE_{D2}$ metric. In Figure E-1, the cells highlighted in green on the left-side tables meet the corresponding acceptance measures shown in the tables on the right.

For axial WRS, the acceptance criterion that had to be set for $RMSE_{D2}$ does not allow for a distinction to be made between the WRS predictions that resulted in "similar" crack growth (hereafter referred to as "good" predictions) and those that did not (hereafter referred to as "bad" predictions). In fact, all the participants meet the acceptance criterion for the $RMSE_{D2}$ metric,

because participant G, who predicted similar crack growth, had the highest $RMSE_{D2}$ of all participants.

For hoop WRS, not in all cases except for participant D, the acceptance criterion that had to be set for $RMSE_{D2}$ allows for a distinction between the WRS predictions that resulted in "acceptable" time to leakage (hereafter referred to as "good" predictions) and those that did not (hereafter referred to as "bad" predictions). Importantly, in all cases except for participant D, $RMSE_{WRS}$ and $RMSE_{D1}$ were also able to distinguish between the WRS predictions that resulted in "acceptable" time to leakage and those that did not.

In summary, there is no added benefit to using the $RMSE_{D2}$ metric, since it is not able to distinguish between the "good" and the "bad" WRS predictions any better than the $RMSE_{WRS}$ and $RMSE_{D1}$ metrics.

D2	Axial Stress / Circ Flaw Growth				
Participant	$RMSE_{WRS}$	$RMSE_{D1}$	$RMSE_{D2}$	$diff_{avg}$	**Crack Growth**
D	52	768	17734	8	Similar
E	48	660	16033	22	Similar
G	74	865	22340	-1	Similar
B	109	839	18379	120	Too Fast
A	78	605	14546	-84	Too Slow
C	43	521	15796	-55	Too Slow
F	67	1055	20787	-16	Too Slow
Min	43	521	14546	-84	
25th percentile	50	632	15915	-36	
Median	67	768	17734	-1	
75th percentile	76	852	19583	15	
Max	109	1055	22340	120	

Axial WRS		
Metric		Acceptance
$RMSE_{WRS}$	≤	75
$RMSE_{D1}$	≤	900
$RMSE_{D2}$	≤	22340
$diff_{avg}$	≥	-5

D2	Hoop Stress / Axial Flaw Growth				
Participant	$RMSE_{WRS}$	$RMSE_{D1}$	$RMSE_{D2}$	$diff_{avg}$	**Time to leakage**
D	70	698	17087	95	Too Short
E	49	713	14652	23	Acceptable
G	82	939	23145	23	Acceptable
B	114	1234	25551	156	Too Short
A	86	1079	34326	-68	Too Long
C	53	627	16645	-71	Acceptable
F	59	914	17033	1	Acceptable
Min	49	627	14652	-71	
25th percentile	56	706	16839	-34	
Median	70	914	17087	23	
75th percentile	84	1009	24348	59	
Max	114	1234	34326	156	

Hoop WRS		
Metric		Acceptance
$RMSE_{WRS}$	≤	85
$RMSE_{D1}$	≤	950
$RMSE_{D2}$	≤	23145
$diff_{avg}$	≥	-75
$diff_{avg}$	≤	75

Figure E-1: Quality metrics for Phase 2b isotropic predictions, including $RMSE_{D2}$. (left), and acceptance measures for the proposed metrics (right).

E.2 Truncated Metrics

E.2.1 Definitions for Truncated Metrics

Truncated $RMSE_{WRS}$ and $RMSE_{D1}$ metrics were investigated to determine if a truncation limit for the calculation of RMSE would improve the ability to distinguish between the good and bad WRS predictions. The following equations define $(RMSE_{WRS})_T$ and $(RMSE_{D1})_T$.

$$(RMSE_{WRS})_T = \sqrt{\frac{1}{L_T}\sum_{k=1}^{L_T}(WRS_k - WRS_k^{mean})}$$

$$(RMSE_{D1})_T = \sqrt{\frac{1}{L_T - 1}\sum_{k=2}^{L_T}(D1_k - D1_k^{mean})}$$

where L_T is the number of equally-spaced points where WRS or D1 are known, up to the truncation limit T, with 0<T<1. For example, for T=0.55, if there are 100 equally-spaced points through the wall thickness where WRS and D1 are known, L_T would be equal to 55. WRS_k and $D1_k$ are the WRS and first derivative of the WRS at the k^{th} position through the wall thickness, and WRS_k^{mean} and $D1_k^{mean}$ are the WRS and first derivative of the cross-sectional mean prediction.

$(RMSE_{WRS})_T$ and $(RMSE_{D1})_T$ were calculated for T=0.1, 0.2,…, 0.9, as shown in E.2.2. In addition, $(RMSE_{WRS})_T$ and $(RMSE_{D1})_T$ were calculated for T=a/t(10 years) and T=a/t(20 years), as shown in Section E.2.3 and Section E.2.4, respectively.

E.2.2 Assessment of Arbitrary Truncation

Figure E-2 through Figure E-10 show the results for the analysis of truncated metrics $(RMSE_{WRS})_T$ and $(RMSE_{D1})_T$ for T=0.1, 0.2,…, 0.9. As in previous figures, the cells highlighted in green on the left-side tables meet the corresponding acceptance measures shown in the tables on the right. In addition, the cells with green text and green background in the tables on the right (showing the acceptance measures) indicate the cases where the acceptance measures could be narrowed as compared to the criteria in Sections 5.4.9 and 5.4.11. These are the cases where using the truncated metrics results in a better ability to distinguish between good and bad WRS predictions. Finally, the cells highlighted in yellow/orange in the left-side tables correspond to cases where the acceptance measures for $(RMSE_{WRS})_T$, $(RMSE_{D1})_T$, and $diff_{avg}$ are met, but where the crack growth or time-to-leakage predictions are not satisfactory. These cases are considered “false positives” and correspond to cases where the truncated criteria results are unable to correctly distinguish between good and bad WRS predictions.

Below each figure in this section, a case-by-case assessment is provided. Overall, the truncated metrics do not offer any clear improvement over the basic metrics proposed in this report (see Sections 5.4.8 and 5.4.10).

T=0.1	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	22	808	8	Similar
E	51	1612	22	Similar
G	16	936	-1	Similar
B	121	721	120	Too Fast
A	86	433	-84	Too Slow
C	55	97	-55	Too Slow
F	72	2352	-16	Too Slow
Min	16	97	-84	
25th percentile	37	577	-36	
Median	55	808	-1	
75th percentile	79	1274	15	
Max	121	2352	120	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	52
$(RMSE_{D1})_T$	≤	1615
$diff_{avg}$	≥	-5

T=0.1	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	98	733	95	Too Short
E	35	1050	23	Acceptable
G	52	1755	23	Acceptable
B	160	1129	156	Too Short
A	66	748	-68	Too Long
C	72	356	-71	Acceptable
F	74	2423	1	Acceptable
Min	35	356	-71	
25th percentile	59	741	-34	
Median	72	1050	23	
75th percentile	86	1442	59	
Max	160	2423	156	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	75
$(RMSE_{D1})_T$	≤	2450
$diff_{avg}$	≥	-75
$diff_{avg}$	≤	75

Figure E-2: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.1

Assessment of truncated metrics for T=0.1:

- Acceptance measures:
  - Narrower (improved) $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall no improvement in acceptance measures
- Ability to distinguish between good and bad predictions:
  - Good for axial WRS
  - False positive for prediction A for hoop WRS
  - Less ability to distinguish than basic metrics proposed in 5.4.8 and 5.4.10

T=0.2	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	47	708	8	Similar
E	72	1221	22	Similar
G	72	1011	-1	Similar
B	109	583	120	Too Fast
A	72	357	-84	Too Slow
C	48	249	-55	Too Slow
F	95	1636	-16	Too Slow
Min	47	249	-84	
25th percentile	60	470	-36	
Median	72	708	-1	
75th percentile	84	1116	15	
Max	109	1636	120	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	75
$(RMSE_{D1})_T$	≤	1225
$diff_{avg}$	≥	-5

T=0.2	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	74	823	95	Too Short
E	58	827	23	Acceptable
G	116	1363	23	Acceptable
B	125	1094	156	Too Short
A	78	1015	-68	Too Long
C	62	326	-71	Acceptable
F	100	1721	1	Acceptable
Min	58	326	-71	
25th percentile	68	825	-34	
Median	78	1015	23	
75th percentile	108	1229	59	
Max	125	1721	156	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	120
$(RMSE_{D1})_T$	≤	1750
$diff_{avg}$	≥	-75
$diff_{avg}$	≤	75

Figure E-3: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.2

Assessment of truncated metrics for T=0.2:

- Acceptance measures:
  - Identical $(RMSE_{WRS})_T$ acceptance criterion for axial WRS
  - Wider (worse) $(RMSE_{WRS})_T$ acceptance criterion for hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall widening of acceptance measures
- Ability to distinguish between good and bad predictions:
  - Good for axial WRS
  - False positive for prediction A for hoop WRS
  - Less ability to distinguish than basic metrics proposed in 5.4.8 and 5.4.10

T=0.3	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	44	715	8	Similar
E	66	1006	22	Similar
G	81	870	-1	Similar
B	97	573	120	Too Fast
A	75	507	-84	Too Slow
C	47	336	-55	Too Slow
F	85	1544	-16	Too Slow
Min	44	336	-84	
25th percentile	57	540	-36	
Median	75	715	-1	
75th percentile	83	938	15	
Max	97	1544	120	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	82
$(RMSE_{D1})_T$	≤	1050
$diff_{avg}$	≥	-5

T=0.3	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	63	748	95	Too Short
E	51	740	23	Acceptable
G	120	1181	23	Acceptable
B	104	1064	156	Too Short
A	74	1030	-68	Too Long
C	52	456	-71	Acceptable
F	102	1498	1	Acceptable
Min	51	456	-71	
25th percentile	57	744	-34	
Median	74	1030	23	
75th percentile	103	1123	59	
Max	120	1498	156	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	120
$(RMSE_{D1})_T$	≤	1500
$diff_{avg}$	≥	-75
$diff_{avg}$	≤	75

Figure E-4: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.3

Assessment of truncated metrics for T=0.3:

- Acceptance measures:
  - Wider (worse) $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall widening of acceptance measures
- Ability to distinguish between good and bad predictions:
  - Good for axial WRS
  - False positive for prediction A for hoop WRS
  - Less ability to distinguish than basic metrics proposed in 5.4.8 and 5.4.10

T=0.4	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	53	808	8	Similar
E	62	933	22	Similar
G	73	1022	-1	Similar
B	85	606	120	Too Fast
A	86	544	-84	Too Slow
C	46	517	-55	Too Slow
F	79	1401	-16	Too Slow
Min	46	517	-84	
25th percentile	57	575	-36	
Median	73	808	-1	
75th percentile	82	977	15	
Max	86	1401	120	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	75
$(RMSE_{D1})_T$	≤	1025
$diff_{avg}$	≥	-5

T=0.4	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	68	769	95	Too Short
E	51	764	23	Acceptable
G	108	1118	23	Acceptable
B	92	1068	156	Too Short
A	82	1081	-68	Too Long
C	48	547	-71	Acceptable
F	90	1331	1	Acceptable
Min	48	547	-71	
25th percentile	60	766	-34	
Median	82	1068	23	
75th percentile	91	1099	59	
Max	108	1331	156	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	110
$(RMSE_{D1})_T$	≤	1350
$diff_{avg}$	≥	-75
$diff_{avg}$	≤	75

Figure E-5: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.4

Assessment of truncated metrics for T=0.4:

- Acceptance measures:
  - Identical $(RMSE_{WRS})_T$ acceptance criterion for axial WRS
  - Wider (worse) $(RMSE_{WRS})_T$ acceptance criterion for hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall widening of acceptance measures
- Ability to distinguish between good and bad predictions:
  - Good for axial WRS
  - False positive for prediction A for hoop WRS
  - Less ability to distinguish than basic metrics proposed in 5.4.8 and 5.4.10

T=0.5	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	61	807	8	Similar
E	57	863	22	Similar
G	78	951	-1	Similar
B	85	629	120	Too Fast
A	79	636	-84	Too Slow
C	42	522	-55	Too Slow
F	73	1338	-16	Too Slow
Min	42	522	-84	
25th percentile	59	632	-36	
Median	73	807	-1	
75th percentile	79	907	15	
Max	85	1338	120	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	78
$(RMSE_{D1})_T$	≤	955
$diff_{avg}$	≥	-5

T=0.5	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	89	765	95	Too Short
E	55	726	23	Acceptable
G	98	1037	23	Acceptable
B	97	1155	156	Too Short
A	110	1158	-68	Too Long
C	62	605	-71	Acceptable
F	80	1204	1	Acceptable
Min	55	605	-71	
25th percentile	71	746	-34	
Median	89	1037	23	
75th percentile	98	1157	59	
Max	110	1204	156	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	100
$(RMSE_{D1})_T$	≤	1205
$diff_{avg}$	≥	-75
$diff_{avg}$	≤	75

Figure E-6: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.5

Assessment of truncated metrics for T=0.5:

- Acceptance measures:
  - Wider (worse) $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall widening of acceptance measures
- Ability to distinguish between good and bad predictions:
  - Good for axial and hoop WRS
  - Same ability to distinguish as basic metrics proposed in 5.4.8 and 5.4.10

T=0.6	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	57	807	8	Similar
E	52	793	22	Similar
G	82	876	-1	Similar
B	93	601	120	Too Fast
A	77	680	-84	Too Slow
C	39	485	-55	Too Slow
F	70	1228	-16	Too Slow
Min	39	485	-84	
25th percentile	54	640	-36	
Median	70	793	-1	
75th percentile	79	841	15	
Max	93	1228	120	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	85
$(RMSE_{D1})_T$	≤	900
$diff_{avg}$	≥	-5

T=0.6	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	86	802	95	Too Short
E	52	762	23	Acceptable
G	90	964	23	Acceptable
B	124	1103	156	Too Short
A	105	1171	-68	Too Long
C	58	717	-71	Acceptable
F	74	1119	1	Acceptable
Min	52	717	-71	
25th percentile	66	782	-34	
Median	86	964	23	
75th percentile	98	1111	59	
Max	124	1171	156	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	90
$(RMSE_{D1})_T$	≤	1120
$diff_{avg}$	≥	-75
$diff_{avg}$	≤	75

Figure E-7: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.6

Assessment of truncated metrics for T=0.6:

- Acceptance measures:
  - Wider (worse) $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall widening of acceptance measures
- Ability to distinguish between good and bad predictions:
  - Good for axial and hoop WRS
  - Same ability to distinguish as basic metrics proposed in 5.4.8 and 5.4.10

T=0.7	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	53	756	8	Similar
E	49	753	22	Similar
G	83	909	-1	Similar
B	89	921	120	Too Fast
A	81	640	-84	Too Slow
C	36	472	-55	Too Slow
F	66	1157	-16	Too Slow
Min	36	472	-84	
25th percentile	51	697	-36	
Median	66	756	-1	
75th percentile	82	915	15	
Max	89	1157	120	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	85
$(RMSE_{D1})_T$	≤	910
$diff_{avg}$	≥	-5

T=0.7	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	81	743	95	Too Short
E	50	743	23	Acceptable
G	85	959	23	Acceptable
B	125	1269	156	Too Short
A	97	1094	-68	Too Long
C	57	678	-71	Acceptable
F	69	1040	1	Acceptable
Min	50	678	-71	
25th percentile	63	743	-34	
Median	81	959	23	
75th percentile	91	1067	59	
Max	125	1269	156	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	85
$(RMSE_{D1})_T$	≤	1040
$diff_{avg}$	≥	-75
$diff_{avg}$	≤	75

Figure E-8: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.7

Assessment of truncated metrics for T=0.7:

- Acceptance measures:
  - Wider (worse) $(RMSE_{WRS})_T$ acceptance criterion for axial WRS
  - Identical $(RMSE_{WRS})_T$ acceptance criterion for hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall widening of acceptance measures
- Ability to distinguish between good and bad predictions:
  - Good for axial and hoop WRS
  - Same ability to distinguish as basic metrics proposed in 5.4.8 and 5.4.10

T=0.8	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	52	734	8	Similar
E	50	710	22	Similar
G	78	893	-1	Similar
B	96	920	120	Too Fast
A	84	602	-84	Too Slow
C	42	507	-55	Too Slow
F	64	1099	-16	Too Slow
Min	42	507	-84	
25th percentile	51	656	-36	
Median	64	734	-1	
75th percentile	81	907	15	
Max	96	1099	120	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	80
$(RMSE_{D1})_T$	≤	895
$diff_{avg}$	≥	-5

T=0.8	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	76	746	95	Too Short
E	50	740	23	Acceptable
G	85	917	23	Acceptable
B	117	1233	156	Too Short
A	93	1123	-68	Too Long
C	55	656	-71	Acceptable
F	65	1007	1	Acceptable
Min	50	656	-71	
25th percentile	60	743	-34	
Median	76	917	23	
75th percentile	89	1065	59	
Max	117	1233	156	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	85
$(RMSE_{D1})_T$	≤	1010
$diff_{avg}$	≥	-75
$diff_{avg}$	≤	75

Figure E-9: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.8

Assessment of truncated metrics for T=0.8:

- Acceptance measures:
  - Wider (worse) $(RMSE_{WRS})_T$ acceptance criterion for axial WRS
  - Identical $(RMSE_{WRS})_T$ acceptance criterion for hoop WRS
  - Narrower (better) $(RMSE_{D1})_T$ acceptance criterion for axial WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for hoop WRS
  - Overall widening of acceptance measures
- Ability to distinguish between good and bad predictions:
  - Good for axial and hoop WRS
  - Same ability to distinguish as basic metrics proposed in 5.4.8 and 5.4.10

T=0.9	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	52	741	8	Similar
E	50	674	22	Similar
G	77	876	-1	Similar
B	105	871	120	Too Fast
A	82	633	-84	Too Slow
C	45	532	-55	Too Slow
F	69	1064	-16	Too Slow
Min	45	532	-84	
25th percentile	51	654	-36	
Median	69	741	-1	
75th percentile	80	874	15	
Max	105	1064	120	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	80
$(RMSE_{D1})_T$	≤	900
$diff_{avg}$	≥	-5

T=0.9	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	73	715	95	Too Short
E	51	719	23	Acceptable
G	86	910	23	Acceptable
B	114	1238	156	Too Short
A	90	1128	-68	Too Long
C	54	649	-71	Acceptable
F	62	955	1	Acceptable
Min	51	649	-71	
25th percentile	58	717	-34	
Median	73	910	23	
75th percentile	88	1042	59	
Max	114	1238	156	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	86
$(RMSE_{D1})_T$	≤	955
$diff_{avg}$	≥	-75
$diff_{avg}$	≤	75

Figure E-10: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=0.9

Assessment of truncated metrics for T=0.9:

- Acceptance measures:
  - Wider (worse) $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall widening of acceptance measures
- Ability to distinguish between good and bad predictions:
  - Good for axial and hoop WRS
  - Same ability to distinguish as basic metrics proposed in 5.4.8 and 5.4.10

E.2.3 Assessment of Truncation at 10 Years Crack Growth

Figure E-11 shows the results for the analysis of truncated metrics $(RMSE_{WRS})_T$ and $(RMSE_{D1})_T$ for T=a/t(10 years), which vary by participant. As in previous figures, the cells highlighted in green

on the left-side tables meet the corresponding acceptance measures shown in the tables on the right. In addition, the cells with green text and green background in the right-side tables (showing the acceptance measures) indicate the cases where the acceptance measures could be narrowed as compared to the criteria presented in Sections 5.4.9 and 5.4.11. These are the cases where using the truncated metrics results in a better ability to distinguish between good and bad WRS predictions. There are no such cases for truncation at T=a/t(10 years). Finally, the cells highlighted in yellow/orange in the left-side tables correspond to cases where the acceptance measures for $(RMSE_{WRS})_T$, $(RMSE_{D1})_T$, and $diff_{avg}$ are met, but where the crack growth or time-to-leakage predictions are not satisfactory. These cases are considered "false positives" and correspond to cases where the truncated criteria results are unable to correctly distinguish between good and bad WRS predictions. There are no such cases for truncation at T=a/t(10 years).

T = a/t (10 yrs)	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	34	831	8	Similar
E	58	889	22	Similar
G	75	1009	-1	Similar
B	109	835	120	Too Fast
A	86	433	-84	Too Slow
C	55	97	-55	Too Slow
F	72	2352	-16	Too Slow
Min	45	532	-84	
25th percentile	51	654	-36	
Median	69	741	-1	
75th percentile	80	874	15	
Max	105	1064	120	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	75
$(RMSE_{D1})_T$	≤	1010
$diff_{avg}$	≥	-5

T = a/t (10 yrs)	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	70	695	95	Too Short
E	49	709	23	Acceptable
G	82	935	23	Acceptable
B	114	1228	156	Too Short
A	110	1147	-68	Too Long
C	54	653	-71	Acceptable
F	69	1040	1	Acceptable
Min	51	649	-71	
25th percentile	58	717	-34	
Median	73	910	23	
75th percentile	88	1042	59	
Max	114	1238	156	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	82
$(RMSE_{D1})_T$	≤	1050
$diff_{avg}$	≥	-75
$diff_{avg}$	≤	75

Figure E-11: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T=a/t(10 years)

Assessment of truncated metrics for T=a/t(10 years):

- Acceptance measures:

    - Identical $(RMSE_{WRS})_T$ acceptance criterion for axial WRS
    - Marginally narrower (better) $(RMSE_{WRS})_T$ acceptance criterion for hoop WRS
    - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
    - Overall widening of acceptance measures
- Ability to distinguish between good and bad predictions:
    - Good for axial and hoop WRS
    - Same ability to distinguish as basic metrics proposed in 5.4.8 and 5.4.10

Overall, it can be said that truncated metrics with T=a/t(10 years) do not offer any clear improvement over the basic metrics proposed in this report (see 5.4.8 and 5.4.10).

E.2.4 Assessment of Truncation at 20 Years Crack Growth

Figure E-12 shows the results for the analysis of truncated metrics $(RMSE_{WRS})_T$ and $(RMSE_{D1})_T$ for T=a/t(20 years), which varies by participant. As in previous figures, the cells highlighted in green on the left-side tables meet the corresponding acceptance measures shown in the tables on the right. In addition, the cells with green text and green background in the right-side tables (showing the acceptance measures) indicate the cases where the acceptance measures could be narrowed as compared to the criteria presented in Sections 5.4.9 and 5.4.11. These are the cases where using the truncated metrics results in a better ability to distinguish between good and bad WRS predictions. There are no such cases for truncation at T=a/t(20 years). Finally, the cells highlighted in yellow/orange in the left-side tables correspond to cases where the acceptance measures for $(RMSE_{WRS})_T$, $(RMSE_{D1})_T$, and $diff_{avg}$ are met, but where the crack growth or time-to-leakage predictions are not satisfactory. These cases are considered "false positives" and correspond to cases where the truncated criteria results are unable to correctly distinguish between good and bad WRS predictions. There are no such cases for truncation at T=a/t(20 years).

Assessment of truncated metrics for T=a/t(20 years):

- Acceptance measures:
    - Identical $(RMSE_{WRS})_T$ acceptance criterion for axial WRS
    - Marginally narrower (better) $(RMSE_{WRS})_T$ acceptance criterion for hoop WRS
    - Wider (worse) $(RMSE_{D1})_T$ acceptance criterion for axial WRS
    - Marginally narrower (better) $(RMSE_{D1})_T$ acceptance criterion for hoop WRS
    - Overall marginal improvement on acceptance measures
- Ability to distinguish between good and bad predictions:
    - Good for axial and hoop WRS
    - Same ability to distinguish as basic metrics proposed in 5.4.8 and 5.4.10

Overall, the truncated metrics with T=a/t(20 years) do not offer any clear improvement over the basic metrics proposed in this report (see Sections 5.4.8 and 5.4.10). In fact, although these metrics provide a marginal improvement on acceptance measures ranges, they require the performance of crack growth predictions, which is a burdensome exercise that is not worth the very small gains achieved.

T = a/t (20 yrs)	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Crack Growth**
D	46	721	8	Similar
E	58	889	22	Similar
G	75	1009	-1	Similar
B	109	835	120	Too Fast
A	86	433	-84	Too Slow
C	55	97	-55	Too Slow
F	72	2352	-16	Too Slow
Min	45	532	-84	
25th percentile	51	654	-36	
Median	69	741	-1	
75th percentile	80	874	15	
Max	105	1064	120	

Truncated

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	75
$(RMSE_{D1})_T$	≤	1010
$diff_{avg}$	≥	-5

T = a/t (20 yrs)	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_T$	$(RMSE_{D1})_T$	$diff_{avg}$	**Time to leakage**
D	70	695	95	Too Short
E	49	709	23	Acceptable
G	82	935	23	Acceptable
B	114	1228	156	Too Short
A	111	1167	-68	Too Long
C	53	624	-71	Acceptable
F	59	909	1	Acceptable
Min	51	649	-71	
25th percentile	58	717	-34	
Median	73	910	23	
75th percentile	88	1042	59	
Max	114	1238	156	

Truncated

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_T$	≤	82
$(RMSE_{D1})_T$	≤	935
$diff_{avg}$	≥	-75
$diff_{avg}$	≤	75

Figure E-12: Truncated Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for T= a/t(20 years)

E.3 Weighted Metrics

E.3.1 Definitions for Weighted Metrics

Truncated $RMSE_{WRS}$ and $RMSE_{D1}$ metrics were investigated to determine if a weight function for the calculation of RMSE would improve the ability to distinguish between the good and bad WRS predictions. The following equations define $(RMSE_{WRS})_W$ and $(RMSE_{D1})_W$:

$$(RMSE_{WRS})_W = \sqrt{\frac{1}{L}\sum_{k=1}^{L}\left(1-\frac{x_k}{t}\right)^W \cdot (WRS_k - WRS_k^{mean})}$$

$$(RMSE_{D1})_W = \sqrt{\frac{1}{L-2}\sum_{k=2}^{L-1}\left(1-\frac{x_k}{t}\right)^W \cdot (D1_k - D1_k^{mean})}$$

where L is the number of equally-spaced points chosen to represent the WRS profile through the wall thickness (from x/t = 0 to x/t = 1), W is the weight exponent, x_k/t is the normalized through-thickness distance at the k^{th} position through the wall thickness, WRS_k and $D1_k$ are the WRS and first derivative of the WRS at the k^{th} position through the wall thickness, and WRS_k^{mean} and $D1_k^{mean}$ are the WRS and first derivative of the cross-sectional mean prediction.

$(RMSE_{WRS})_W$ and $(RMSE_{D1})_W$ were calculated for W=1, 2, 5, 10, as shown in Section E.3.2. The weight function gives higher weight to values of (WRS_k-WRS_k^{mean}) or ($D1_k$-$D1_k^{mean}$) closer to x/t = 0. Furthermore, as the weight exponent increases, the weight function eventually takes values smaller than 1%, essentially truncating the RMSE calculation beyond a certain point, as illustrated in Figure E-13.

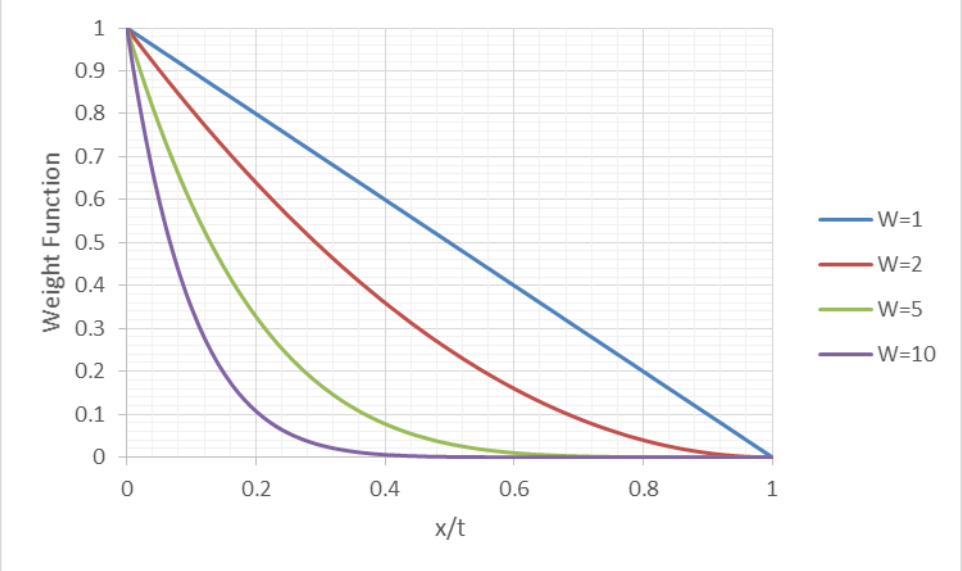


Figure E-13: Weight Function as a Function of Weight Exponent W

E.3.2 Results and Assessment for Weighted Metrics

Figure E-14 though Figure E-17 show the results for the analysis of weighted metrics $(RMSE_{WRS})_W$ and $(RMSE_{D1})_W$ for W=1, 2, 5, and 10. As in previous figures, the cells highlighted in green on the left-side tables meet the corresponding acceptance measures shown in the tables on the right. In addition, the cells with green text and green background in the right-side tables (showing the acceptance measures) indicate the cases where the acceptance measures could be narrowed as compared to the criteria presented in 5.4.9 and 5.4.11. These are the cases where using the weighted metrics results in a better ability to distinguish between good and bad WRS predictions. Finally, the cells highlighted in yellow/orange in the left-side tables correspond to cases where the acceptance measures for $(RMSE_{WRS})_W$, $(RMSE_{D1})_W$, and $diff_{avg}$ are met, but where the crack growth or time to leakage predictions are not satisfactory. These cases are considered "false positives" and correspond to cases where the weighted criteria results are unable to correctly distinguish between good and bad WRS predictions.

Below each figure in this section, a case-by-case assessment is provided. Overall, weighted metrics allow for a significant narrowing of the acceptance measures for both axial and hoop WRS, which is an improvement over the basic metrics proposed in Sections 5.4.8 and 5.4.10. However, in all cases investigated where W>1, some false positives occur for hoop WRS. Consequently, the weighted metrics do not offer any improvement over the basic metrics proposed in this report (see Sections 5.4.8 and 5.4.10).

W=1	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_W$	$(RMSE_{D1})_W$	$diff_{avg}$	**Crack Growth**
D	37	540	8	Similar
E	38	581	22	Similar
G	55	645	-1	Similar
B	68	555	120	Too Fast
A	57	429	-84	Too Slow
C	30	342	-55	Too Slow
F	51	889	-16	Too Slow
Min	30	342	-84	
25th percentile	38	485	-36	
Median	51	555	-1	
75th percentile	56	613	15	
Max	68	889	120	

Weighted

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_W$	≤	55
$(RMSE_{D1})_W$	≤	645
$diff_{avg}$	≥	-5

W=1	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_W$	$(RMSE_{D1})_W$	$diff_{avg}$	**Time to leakage**
D	55	531	95	Too Short
E	36	530	23	Acceptable
G	65	719	23	Acceptable
B	81	839	156	Too Short
A	66	778	-68	Too Long
C	40	443	-71	Acceptable
F	53	828	1	Acceptable
Min	36	443	-71	
25th percentile	46	531	-34	
Median	55	719	23	
75th percentile	66	803	59	
Max	81	839	156	

Weighted

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_W$	≤	65
$(RMSE_{D1})_W$	≤	830
$diff_{avg}$	≥	-75
$diff_{avg}$	≤	75

Figure E-14: Weighted Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for W=1

Assessment of weighted metrics for W=1:

- Acceptance measures:
  - Narrower (improved) $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Narrower (improved) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall significant improvement in acceptance measures
- Ability to distinguish between good and bad predictions:
  - Good for axial and hoop WRS
  - Same ability to distinguish as basic metrics proposed in 5.4.8 and 5.4.10

W=2	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_W$	$(RMSE_{D1})_W$	$diff_{avg}$	**Crack Growth**
D	30	442	8	Similar
E	33	541	22	Similar
G	44	538	-1	Similar
B	55	408	120	Too Fast
A	46	334	-84	Too Slow
C	25	257	-55	Too Slow
F	44	814	-16	Too Slow
Min	25	257	-84	
25th percentile	32	371	-36	
Median	44	442	-1	
75th percentile	45	539	15	
Max	55	814	120	

Weighted

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_W$	≤	45
$(RMSE_{D1})_W$	≤	550
$diff_{avg}$	≥	-5

W=2	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_W$	$(RMSE_{D1})_W$	$diff_{avg}$	**Time to leakage**
D	46	443	95	Too Short
E	30	445	23	Acceptable
G	57	638	23	Acceptable
B	66	662	156	Too Short
A	53	620	-68	Too Long
C	33	337	-71	Acceptable
F	48	782	1	Acceptable
Min	30	337	-71	
25th percentile	39	444	-34	
Median	48	620	23	
75th percentile	55	650	59	
Max	66	782	156	

Weighted

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_W$	≤	60
$(RMSE_{D1})_W$	≤	790
$diff_{avg}$	≥	-75
$diff_{avg}$	≤	75

Figure E-15: Weighted Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for W=2

Assessment of weighted metrics for W=2:

- Acceptance measures:
  - Narrower (improved) $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Narrower (improved) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall significant improvement in acceptance measures
- Ability to distinguish between good and bad predictions:
  - Good for axial WRS
  - False positive for prediction A for hoop WRS
  - Less ability to distinguish than basic metrics proposed in 5.4.8 and 5.4.10

W=5	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_W$	$(RMSE_{D1})_W$	$diff_{avg}$	**Crack Growth**
D	19	302	8	Similar
E	26	474	22	Similar
G	28	386	-1	Similar
B	42	259	120	Too Fast
A	33	200	-84	Too Slow
C	20	141	-55	Too Slow
F	34	698	-16	Too Slow
Min	19	141	-84	
25th percentile	23	230	-36	
Median	28	302	-1	
75th percentile	34	430	15	
Max	42	698	120	

Weighted

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_W$	≤	30
$(RMSE_{D1})_W$	≤	475
$diff_{avg}$	≥	-5

W=5	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_W$	$(RMSE_{D1})_W$	$diff_{avg}$	**Time to leakage**
D	32	318	95	Too Short
E	21	337	23	Acceptable
G	42	525	23	Acceptable
B	50	452	156	Too Short
A	34	395	-68	Too Long
C	25	189	-71	Acceptable
F	38	704	1	Acceptable
Min	21	189	-71	
25th percentile	28	328	-34	
Median	34	395	23	
75th percentile	40	489	59	
Max	50	704	156	

Weighted

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_W$	≤	45
$(RMSE_{D1})_W$	≤	710
$diff_{avg}$	≥	-75
$diff_{avg}$	≤	75

Figure E-16: Weighted Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for W=5

Assessment of weighted metrics for W=5:

- Acceptance measures:
  - Narrower (improved) $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Narrower (improved) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall significant improvement in acceptance measures
- Ability to distinguish between good and bad predictions:
  - Good for axial WRS
  - False positive for prediction A for hoop WRS
  - Less ability to distinguish than basic metrics proposed in 5.4.8 and 5.4.10

W=10	Axial Stress / Circ Flaw Growth			
Participant	$(RMSE_{WRS})_W$	$(RMSE_{D1})_W$	$diff_{avg}$	**Crack Growth**
D	12	208	8	Similar
E	19	411	22	Similar
G	17	276	-1	Similar
B	34	204	120	Too Fast
A	26	132	-84	Too Slow
C	16	70	-55	Too Slow
F	26	607	-16	Too Slow
Min	12	70	-84	
25th percentile	16	168	-36	
Median	19	208	-1	
75th percentile	26	344	15	
Max	34	607	120	

Weighted

Axial WRS		
Metric		Acceptance
$(RMSE_{WRS})_W$	≤	20
$(RMSE_{D1})_W$	≤	415
$diff_{avg}$	≥	-5

W=10	Hoop Stress / Axial Flaw Growth			
Participant	$(RMSE_{WRS})_W$	$(RMSE_{D1})_W$	$diff_{avg}$	**Time to leakage**
D	26	239	95	Too Short
E	14	273	23	Acceptable
G	27	428	23	Acceptable
B	42	338	156	Too Short
A	23	252	-68	Too Long
C	20	114	-71	Acceptable
F	28	628	1	Acceptable
Min	14	114	-71	
25th percentile	21	245	-34	
Median	26	273	23	
75th percentile	28	383	59	
Max	42	628	156	

Weighted

Hoop WRS		
Metric		Acceptance
$(RMSE_{WRS})_W$	≤	30
$(RMSE_{D1})_W$	≤	630
$diff_{avg}$	≥	-75
$diff_{avg}$	≤	75

Figure E-17: Weighted Quality Metrics for Phase 2b Isotropic Predictions (left), and Acceptance measures for the Proposed Metrics (right), for W=10

Assessment of weighted metrics for W=10:

- Acceptance measures:
  - Narrower (improved) $(RMSE_{WRS})_T$ acceptance criterion for axial and hoop WRS
  - Narrower (improved) $(RMSE_{D1})_T$ acceptance criterion for axial and hoop WRS
  - Overall significant improvement in acceptance measures
- Ability to distinguish between good and bad predictions:
  - Good for axial WRS
  - False positive for prediction A for hoop WRS
  - Less ability to distinguish than basic metrics proposed in 5.4.8 and 5.4.10

NRC FORM 335
(12-2010)
NRCMD 3.7

U.S. NUCLEAR REGULATORY COMMISSION

BIBLIOGRAPHIC DATA SHEET

(See instructions on the reverse)

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

NUREG-2228

2. TITLE AND SUBTITLE

Weld Residual Stress Finite Element Analysis Validation

Part II—Proposed Validation Procedure

Draft Report for Comment

3. DATE REPORT PUBLISHED

MONTH	YEAR
August	**2018**

4. FIN OR GRANT NUMBER

5. AUTHOR(S)

Michael L. Benson
Patrick A. C. Raynaud
Jay Wallace

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.)

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

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.)

Same as above

10. SUPPLEMENTARY NOTES

11. ABSTRACT (200 words or less)

Under a Memorandum of Understanding, the U.S. Nuclear Regulatory Commission and the Electric Power Research Institute conducted a research program aimed at gathering data on weld residual stress modeling. As described in NUREG-2162, "Weld Residual Stress Finite Element Analysis Validation: Part I—Data Development Effort," issued March 2014, this program consisted of round robin measurement and modeling studies on various mockups. At that time, the assessment of the data was qualitative. This report describes an additional residual stress round robin study and a methodology for capturing residual stress uncertainties. This quantitative approach informed the development of guidelines and a validation methodology for finite element prediction of weld residual stress. For example, comparisons of modeling results to measurements provided a basis for establishing guidance on a material hardening approach for residual stress models. The proposed validation procedure involves an analyst modeling a known case (the Phase 2b round robin mockup) and comparing results to three proposed quality metrics. These recommendations provide a potential method by which analysts can bolster confidence in their modeling practices for regulatory applications.

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

weld residual stress, finite element modeling, model validation, flaw evaluation, uncertainty characterization

13. AVAILABILITY STATEMENT

unlimited

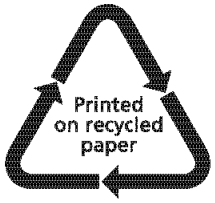
14. SECURITY CLASSIFICATION

(This Page)
unclassified

(This Report)
unclassified

15. NUMBER OF PAGES

16. PRICE



Federal Recycling Program

UNITED STATES
NUCLEAR REGULATORY COMMSSION
WASHINGTON, DC 20555-0001

OFFICIAL BUSINESS

NUREG-2228
Draft

Weld Residual Stress Finite Element Analysis Validation
Part II—Proposed Validation Procedure

August 2018