SANDIA REPORT

SAND2016-10932
Unlimited Release
Printed October 2016XX

Uncertainty Quantification and Comparison of Weld Residual Stress Measurements and Predictions

J.R. Lewis, D. Brooks

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-mission laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online ordering: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2016-10932
Unlimited Release
Printed October 2016XX

Uncertainty Quantification and Comparison of Weld Residual Stress Measurements and Predictions

John R. Lewis
Statistical Sciences Department, Org. 00436
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185
jrlewi@sandia.gov

Dusty Brooks
Structural and Thermal Analysis Department, Org. 06233
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185
dbrooks@sandia.gov

Abstract

In pressurized water reactors, the prevention, detection, and repair of cracks within dissimilar metal welds is essential to ensure proper plant functionality and safety. Weld residual stresses, which are difficult to model and cannot be directly measured, contribute to the formation and growth of cracks due to primary water stress corrosion cracking. Additionally, the uncertainty in weld residual stress measurements and modeling predictions is not well understood, further complicating the prediction of crack evolution. The purpose of this document is to develop methodology to quantify the uncertainty associated with weld residual stress that can be applied to modeling predictions and experimental measurements. Ultimately the results can be used to assess the current state of uncertainty and to build confidence in both modeling and experimental procedures. The methodology consists of statistically modeling the variation in the weld residual stress profiles using functional data analysis techniques. Uncertainty is quantified using statistical bounds (e.g. confidence and tolerance bounds) constructed with a semi-parametric bootstrap procedure. Such bounds describe the range in which quantities

of interest, such as means, are expected to lie as evidenced by the data. The methodology is extended to provide direct comparisons between experimental measurements and modeling predictions by constructing statistical confidence bounds for the average difference between the two quantities. The statistical bounds on the average difference can be used to assess the level of agreement between measurements and predictions. The methodology is applied to experimental measurements of residual stress obtained using two strain relief measurement methods and predictions from seven finite element models developed by different organizations during a round robin study.

Contents

Figures

Tables

Glossary of Notation

Symbol	Description
D	a distance metric
d	the relative depths through the weld
$\mathscr{F}$	the class of absolutely continuous functions on the domain of interest
f_i	the smoothed WRS profile from the i-th participant
f_i^s	the i-th WRS profile sampled from the model
f^s	the mean of all n WRS profiles sampled from the model in one bootstrap
$\tilde{f}_i$	aligned function defined by function composition of f_i with its optimal depth-warping function
γ_i	a warping function
γ_i^*	the optimal warping function applied to depth to align f_i
K	number of depths at which to evaluate sampled WRS profile functions
n	the number of WRS data profiles being analyzed
N	denotes the normal distribution
$\tilde{q}_i$	the square root slope function (SRSF) of $\tilde{f}_i$
S	the number of bootstrap samples taken to estimate a statistic
Σ	a covariance matrix
μ_f	the Karcher mean of the population of smoothed WRS profiles being analyzed

List of Acronyms and Abbreviations

Acronym	Description
AVE	Average
DHD	Deep Hole Drilling
DM	Dissimilar Metal
EPRI	Electric Power Research Institute
FE	Finite Element
fPCA	Function Principal Component Analysis
ID	Inner Diameter
ISO	Isometric
iDHD	Incremental Deep Hole Drilling
KIN	Kinematic
MC	Monte Carlo
MPa	Mega Pascals
NRC	U.S. Nuclear Regulatory Commission
OD	Outer Diameter
PCA	Principal Component Analysis
PWR	Pressurized Water Reactor
PWSCC	Primary Water Stress Corrosion Cracking
SCC	Stress Corrosion Cracking
SRSF	Square Root Slope Function
UQ	Uncertainty Quantification
WRS	Weld Residual Stress

Executive Summary

Weld residual stress (WRS) contributes to the formation and growth of cracks within welds. In the context of nuclear safety, understanding WRS and uncertainty in WRS is essential for prediction and prevention of stress-induced cracks in nuclear power plants. This study demonstrates methodology that could be used to quantify uncertainty and compare measurements and predictions to inform analyses within a regulatory space.

All of the data used in this analysis was collected as part of a double-blind round robin study conducted jointly by the U.S. Nuclear Regulatory Commission (NRC) and the Electric Power Research Institute (EPRI). A mockup of a pressurizer surge nozzle for a pressurized water reactor with a dissimilar metal (DM) weld was manufactured. Two measurement vendors independently measured the WRS in this DM weld; one vendor used the deep hole drilling (DHD) method and the other used the contour method. The DHD method measures WRS by drilling holes, measuring displacements, and applying the measurements to a model that ultimately outputs an estimate of WRS. The contour method also applies measurements to a model to estimate WRS. However, the method requires cutting through the weld, destroying the mockup, and measuring the resulting contours on the cut surface.

Independently of the measurement effort, ten organizations were provided guidance and tasked with modeling the mockup and predicting WRS. Of the predictions, seven were deemed acceptable for this analysis based upon adequate adherence to the modeling guidance. The predictions were provided for axial and hoop WRS under two different material hardening assumptions: kinematic and isotropic.

This analysis is separated into three components: quantification of uncertainty in measurements, quantification of uncertainty in predictions, and comparison of measurements and predictions. Since data are provided as measurements of WRS with respect to the depth through the weld width, a functional analysis approach was applied. The method summarizes both horizontal (through the depth of the weld) and vertical (magnitude of stresses) variability in the WRS functions using functional principal components (fPCA) techniques and applies statistical bootstrap methods to these summaries to quantify uncertainty. This methodology is first used separately to quantify the uncertainty in the predictions and most of the measurements. The measurements for which this method is not applied are the contour axial and hoop WRS measurements. There are a large number contour axial measurements and sophisticated modelling is not needed to quantify uncertainty. On the other hand, there is no method available with the current resources to quantify uncertainty in the single hoop WRS measurement.

After the separate analyses, direct comparisons of measurements and predictions are made by combining the uncertainty quantification in both to estimate the mean difference. The statistical confidence bound for the mean difference measures the extent to which mean measurements and mean predictions can be expected to differ.

Results of the analysis indicate significant uncertainty in both measurements and predictions. Overall, the measurements tend to have wider confidence and tolerance bounds than the predictions

(due to the smaller sample size), though this is not always the case. For example, axial WRS near the inner diameter is expected to cover a range approximately 200 MPa wide in the measurement data. In the prediction data, axial WRS predictions near the ID cover a larger range about 400 to 650 MPa wide. Similarly, hoop WRS predictions near the inner diameter are expected to cover a range 450 to 1000 MPa in width, while the hoop measurements near the ID are expected to cover a range less than 100 MPa wide. In each case, the level of uncertainty has implications for modeling practices. The tolerance intervals can inform the range of WRS values that should be considered when simulating crack formation and growth; simulations that do not include uncertainty in WRS or that assume a narrow range of possible WRS values at the inner diameter may lead to incorrect predictions of cracks.

The results of the comparison of measurements to predictions are consistent with previous observations on predicting WRS; on average, neither the assumption of kinematic hardening nor the assumption of isotropic hardening are sufficient. Similarly, though the average of the kinematic and isotropic predictions generally performs better than the mean predictions using either hardening assumption alone, the difference in performance is not significant enough to definitively support a preference for use of the average in making predictions. In particular, the average may perform better solely as an artifact of being the average of an over-prediction and an under-prediction; there is no physical motivation for using the average. While poor performance of the prediction models cannot be attributed solely to the hardening law assumption, the comparisons suggest that a more robust model for material hardening could improve prediction performance. Further study to identify the greatest sources of uncertainty and the effect of improving the material hardening model is necessary to reduce uncertainty in measurements and predictions.

As a consequence of the high levels of uncertainty in measurements and predictions, the role that WRS predictions should play in regulatory decisions is unclear. For axial WRS, the DHD and contour methods disagree with each other and show comparable levels of uncertainty. In addition, the single contour hoop WRS measurement is insufficient to characterize uncertainty; comparisons of predictions to this measurement lack consideration of this, likely significant, uncertainty. Because both the DHD and contour methods show similar levels of agreement with neutron diffraction measurements, it may be preferable to use DHD measurements so that measurement uncertainty can be included in both axial and hoop stress comparisons [1] [2]. This determination is, however, subject to both to expert opinion and the goal of the analysis being performed. Thus, it may not be appropriate to use either set of measurement data for validation of predictions unless further experimental study can be performed to explore the accuracy of the DHD and contour methods.

If the data set to be used for validation of WRS prediction models can be determined, validation should not be limited to the measurement data set from this study. A modeling process may be valid for the specific nozzle under consideration here and yet fail when applied to another piece of equipment. While it may not be feasible to obtain WRS measurements for a wide variety of equipment to use for validation, a model should only be considered valid if it is able to match the available data within a certain margin and includes a thorough justification for its applicability to all of the components for which it is used.

1 Introduction

1.1 Background

The weld residual stress (WRS) that exists in welds as a result of the welding process contributes to the formation and growth of cracks within welds. In environments like nuclear power plants, the prevention, detection, and repair of cracks within welds is essential to ensure proper plant functionality and safety. Thus, a greater understanding of WRS and the uncertainty in WRS can lead to improvements in maintenance and licensing procedures. Though uncertainty in WRS and the contribution of WRS towards crack formation and growth is relevant to many applications, this work is motivated by needs specific to the nuclear energy industry. It is a continuation of efforts by the U.S. Nuclear Regulatory Commission (NRC) to understand and minimize uncertainty in WRS predictions [3].

The data used for this work was collected as part of a double-blind round robin study conducted jointly by the NRC and the Electric Power Research Institute (EPRI). The study was designed to determine if increased modeling guidance could reduce the uncertainty in predicting WRS. A mockup of a pressurizer surge nozzle for a pressurized water reactor (PWR) was manufactured and two measurement vendors were tasked with separately measuring the WRS through the depth of the dissimilar metal (DM) weld that joins the carbon steel side to the stainless steel safe-end. One vendor used a combination of incremental deep hole drilling (iDHD) and deep hole drilling (DHD), which involves drilling holes to release strain, and the other vendor used the contour method, measuring contours over the cross section after cutting through the weld to release strain. Neither technique measures WRS directly. Rather, these methods both involve taking intermediate experimental measurements of various quantities and applying these measurements to models that output WRS measurements. See Section 1.2 and references [2] and [1] for more details. Ten additional participants were tasked with modeling the mockup and predicting the WRS profiles through the depth of the weld. Their predictions were completed without access to the measurement data (and the measurement vendors did not have access to the prediction data). For further details on modeling guidance and methods see [4].

The goal of this study is to quantify the uncertainty in both experimental measurements and predictions of WRS and to use this as a basis for comparing predictions to measurements. Since experimental measurements of WRS are not made directly, uncertainty in these measurements is not solely due to variation/uncertainty in the physical measurements. In particular, sources of uncertainty include: the strain relief techniques used to measure WRS, the precision and accuracy limitations of the tools used for the intermediate measurements, the density of the intermediate measurements, and the modeling steps used to calculate stresses from strains. In this sense, the experimental measurements of WRS are part measurement and part model, creating an unavoidable limitation to the problem. The process of comparing prediction to experimental measurement becomes complicated; because the measurements are not purely measurements, using them to validate predictions should be done with caution. Despite the inherent uncertainty in these measurements, they are the best representations of the actual residual stress fields available and are used in this study as a basis of comparison for the WRS predictions.

This report is organized as follows. Section 1.2 concludes the introduction by providing a description of the data. Section 2 introduces the functional statistical analysis techniques used to characterize uncertainty in both experimental measurements and predictions. In particular, the two types of variation (i.e. vertical and horizontal) associated with functional data are described along with a high-level description of how they are separated and modeled to quantify uncertainty. Sections 2.1, 2.2, and 2.3 provide more mathematical details on how each type of variation is separated and modeled. Section 2.4 describes how the separate models for vertical and horizontal variation are combined into one unified model that characterizes the population (and the uncertainty therein) of the WRS profiles. The statistical bootstrapping techniques described in Section 2.5 are then applied to construct confidence bounds as described in Section 2.6. The remaining sections present the results of applying these methodologies to quantify uncertainty in the experimental measurements (Section 3) and the predictions (Section 4) and finally to compare the predictions to the measurements and compare the measurements to each other (Section 5). High level results are summarized in Section 6. Some details of the ancillary methods have been reserved for Section 7.

1.2 Data Description

The data used for this study is separated into two groups, measurement data and prediction data. The modeling participants were provided with material and thermal input decks and a description of the geometry to be modeled for the DM weld in a mockup PWR pressurizer surge nozzle. They were tasked with modeling the described weld and predicting the WRS through the depth of the weld at the center line at 26 evenly spaced points. The measurement vendors were provided with the mockup and tasked with measuring the WRS present in the mockup through the depth of the weld. The DHD measurements were taken at the weld centerline and the contour measurements were taken over the entire cross section [5].

WRS Prediction Data

Each participant who contributed WRS prediction data provided the depth from the inner diameter (ID) in millimeters. Two sets of predictions, using the same model, were provided for both hoop and axial stresses by each participant; one set assumed totally isotropic material hardening (ISO) and the other assumed totally kinematic material hardening (KIN). Both sets of predictions are studied in this analysis in addition to the average (AVE) of the ISO and KIN predictions. Because the models could not be applied with mixed hardening law, the average is considered here to estimate a combination of both types of hardening. Maximum depths in reported predictions ranged from 37.42 mm to 37.865 mm.

Three sets of predictions were excluded from the analysis. One such set provided a maximum depth of 47.309 mm. However, this indicates that the model used a different geometry than described in the modeling guidance. The two other sets of predictions excluded also deviated from modeling guidance; one for the use of a different heat input, and the other for the use of different material properties [5]. These predictions are excluded because the primary interest in assessing

uncertainty in the prediction data is to determine how much uncertainty exists when separate modelers make predictions of the same quantity using the same modeling guidance. Deviations from modeling guidance on the geometry, heat input, and material properties were all deemed to result in predictions of fundamentally different quantities. Thus, for each component of stress (axial and hoop) and each material hardening assumption (ISO, KIN, and AVE) there are predictions of WRS from seven modeling participants.

WRS Measurement Data

Two independent experimental methods were used to obtain WRS measurements from the mockup that the participants were asked to model. Both methods measure WRS by releasing strain. The first method is referred to as the deep hole drilling method (DHD) and the second as the contour method. A brief descriptions of both is provided. For detailed descriptions of the DHD method and the contour method, see [2] and [1] respectively.

First Method - DHD: The first method is the deep hole drilling (DHD) method which involves the drilling of a reference hole through the weld, measuring the reference hole, trepanning, and then measuring the reference hole again. The DHD method employed was a combination of DHD and incremental deep hole drilling (iDHD), in which the reference hole is trepanned and measured in increments of depth, rather than being trepanned in one step and then measured at each depth. This incremental method decreases the effects of plasticity. The measurements were then used to calculate WRS through the depth of the weld. Though a combination of DHD and iDHD was used, only the DHD data is used in this report. The iDHD method was applied only to the last 1 mm of depth through the weld (i.e. 1 mm from the outer diameter). This data was not consistent with the DHD data so it was excluded from the analysis and only the DHD data was used.

Four sets of experimental measurements were obtained from the mockup using the DHD method with holes drilled in 90° increments around the weld through the center line. Each set of data corresponds to one of the drilled holes. Axial, hoop, and shear stresses were provided, though only axial and hoop stresses are considered here. Unlike the contour and prediction data that was provided with respect to the ID, depths corresponding to the stresses were provided in millimeters from the outer diameter (OD) of the weld; stresses were given in megapascals (MPa). The reported depths were converted to relative depth through the weld from the ID by subtracting each reported depth from the actual thickness of the weld (37.8 mm) and normalizing by the thickness of the weld.

Second Method - Contour Method: The second method employed on the same specimen was the contour method, in which cuts were performed through the specimen, the contours resulting from the release of strain were measured, and then a finite element model was applied to calculate WRS. The stresses were provided at points defined by position within the weld. This data is sufficiently dense that it can be used to determine the WRS through the weld depth for any path defined through the cross section. With respect to axial WRS, the measurement vendor provided five extracted paths as an example and more paths were extracted for this study. Though hoop WRS could also be extracted at many locations, this study focuses on hoop stress only along the

weld centerline, so only the path at the centerline was used. The thickness of the weld, 37.8 mm, was used to normalize all of the extracted WRS profiles. This normalization is applied to facilitate comparisons of the measurement and prediction data as the thickness of the weld in the prediction models varies. The measurement vendor did not extract WRS measurements near the surface; their measurements began 1 mm from the OD and ended 1 mm from the ID. It is assumed that this was done intentionally to avoid conflating surfaces stresses with WRS. All contour WRS measurements extracted for this study follow this precedent, beginning and ending 1 mm from either surface.

2 Methods for Uncertainty Quantification and Comparisons of Functional Data

In this section, the statistical methods used to analyze the functional WRS data from both predictions and measurements are presented. The techniques allow us to quantify uncertainty through the construction of statistical bounds on the data. To construct these bounds, a semi-parametric bootstrap extension of the modeling methods described in [6] is developed which appropriately accounts for the two types of variability in functional data, known as *amplitude* and *phase*. These types of variability are also known as *vertical* and *horizontal* variability and are described below. The bootstrap technique, which is applied to models described in Sections 2.2, 2.3, and 2.4, is general enough to construct statistical bounds on essentially any quantity of interest. For the purposes of this work however, both statistical confidence bounds on the mean function as well as tolerance bounds are provided. The confidence bounds on the mean quantify uncertainty as to where the true mean WRS profile lies. The tolerance bounds quantify uncertainty regarding the range in which a certain proportion of the population of unseen WRS profiles will fall. Finally, the separate uncertainty quantification (UQ) of the measurements and the UQ of the predictions are combined to construct confidence bounds on the mean difference between experimental measurements and predictions. The estimate of the mean difference, along with its associated uncertainty, provides a way to quantitatively compare measurements to predictions.

To facilitate the description of the functional data analysis techniques, first consider the raw predictions of WRS from the seven participants assuming isotropic hardening. The predictions of axial stresses assuming isotropic hardening appear in Figure 1. Stress (MPa) is predicted at a discrete set of depths (mm) from the inner diameter (ID) of the weld. These values have been normalized to the interval $[0, 1]$ by dividing by the maximum modeled depth, which varies by participant (see Section 1.2). Though predictions at discrete depths (which vary by participant) are available, WRS profiles are expected to be a continuous function in reality. For this reason the raw data is smoothed using the methods described in Section 7.2 with cubic splines to obtain continuous functions of WRS on which to conduct the analysis. The results of the smoothing appear in Figure 2.

In general, the variability in functional data of this sort can be attributed to two sources: *amplitude* variability and *phase* variability.

- *Amplitude* variability refers to the variability in the vertical direction. This is the most well understood type of variability in functional data. In Figures 1 and 2, one can easily see that at any given depth value the WRS functions have considerable variation in the vertical direction.

- *Phase* variability refers to the variability in the horizontal direction. For example, consider the data in Figures 1 and 2 between the (normalized) ID depth values 0.4 and 0.6. In this interval, the horizontal position (relative depth) where the local minimum occurs varies. This type of variability is also important to consider when attempting to characterize uncertainty in functional data.

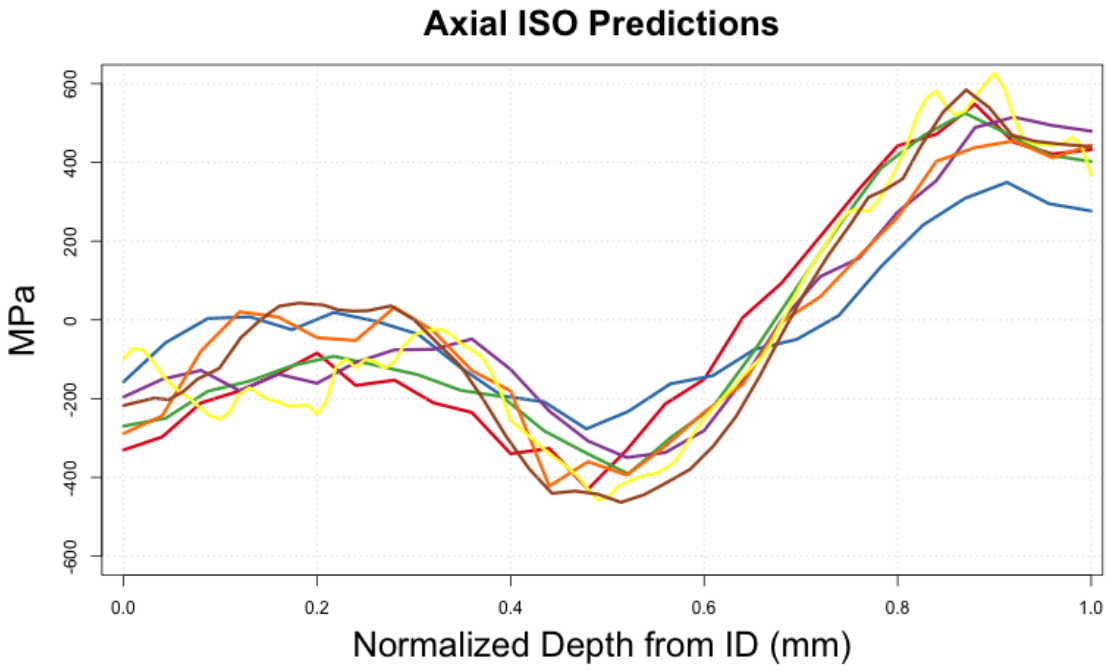


Figure 1: WRS predictions in the axial direction from the seven organizations assuming isotropic hardening.

To properly characterize uncertainty in the predictions and measurements and to make comparisons between them, both types of variability are accounted for using the methodology described in [6]. In particular, a probability model accounting for both amplitude and phase variability is applied to the space of WRS functions. For the predictions pictured in Figure 1, the modeled probability distribution of WRS functions represents the variability expected within a population of predictions. From this distribution, uncertainty can be quantified in any statistic of interest, such as the cross-sectional mean or percentiles. Likewise, on the experimental measurement side, the distribution quantifies the uncertainty in a population of measurements and, for instance, the uncertainty on the mean measurement. The models on both the experimental measurements and predictions can be extended to provide a basis for comparing the two. For example, a comparison of the difference between the mean measurement and mean prediction, while properly accounting for uncertainties in both, can be provided. The basic steps of the method are:

1. Register (i.e. align horizontally as described in Section 2.3) the smoothed WRS profiles.
2. Model the registered profiles with an adapted functional principal components (fPCA) technique.
3. Model the *warping functions* which map the original functions to the registered functions with an adapted fPCA technique.
4. Combine the two models to construct a model on the original scale of the functions.

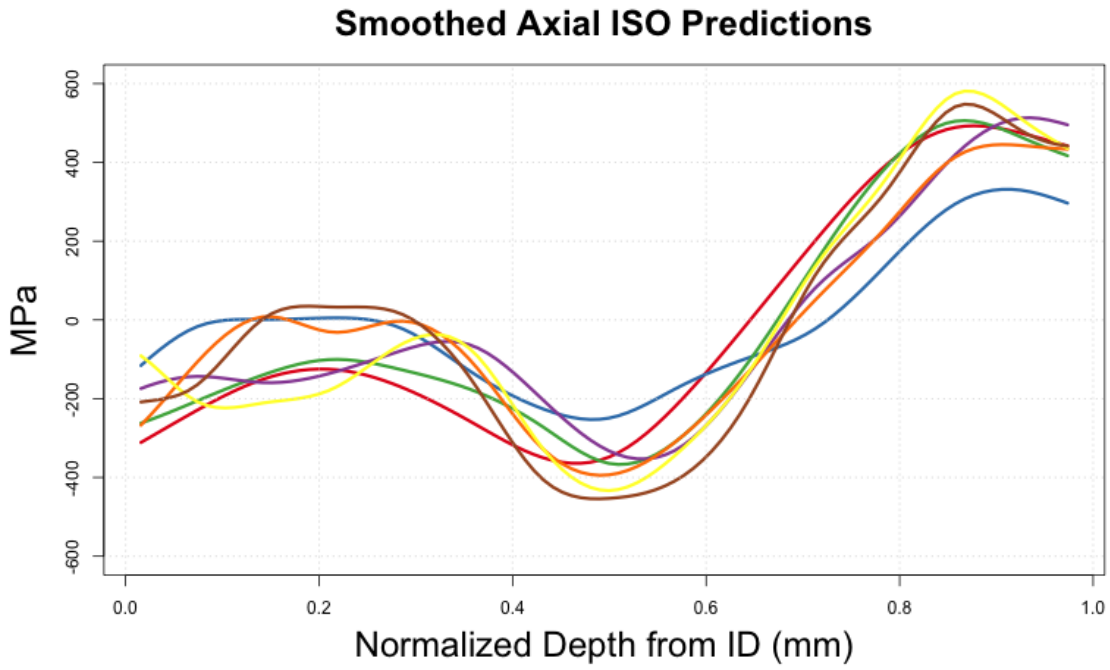


Figure 2: Smoothed WRS predictions in the axial direction from the seven organizations assuming isotropic hardening. Ten equally spaced knots across the range of normalized depths were used.

The registered WRS functions can be thought of as describing the amount of amplitude variability as all the phase variability has been removed by the warping functions. Likewise, the warping functions describe the amount of phase variability. Thus, steps 2 and 3 above separate these two types of variability and step 4 combines them to form a model on the original scale.

Steps 1-4 are first used to model both experimental measurement data and predictions separately. These steps are described in more detail in Sections 2.1-2.4. The measurement data is separated by method (DHD or contour) and stress component (axial or hoop) and the prediction data is separated by material hardening law assumption (ISO, KIN, or AVE) and stress component (axial or hoop). Within each grouping, steps 1-4 are applied to model the uncertainty within the group. For the purposes of uncertainty quantification, the model within each group is used to construct semi-parametric bootstrap confidence bounds for the mean function as well as to construct tolerance bounds intended to bound 95% of the population of WRS functions with 95% confidence. Construction of these bounds is described in Sections 2.5-2.6.

After quantifying the uncertainty within the measurements and predictions separately, the next step proceeds to pair measurements and predictions together for comparison to each other. They are paired by stress component (axial or hoop), measurement method (DHD or contour) and prediction hardening law assumption (ISO, KIN, or AVE). The quantity of interest in each comparison group is how much, on average, the measurements and predictions differ. The semi-parametric bootstrap technique used to construct confidence bounds for the means of each group separately is utilized again to construct confidence bounds on the mean difference.

Each of the above steps are described in greater detail in the following sections.

2.1 Registration

Registration (also referred to as alignment) is a process involving transformations of the depth, d [7]. The goal in this case is to transform the WRS functions via non-linear transformation of the depths to remove any phase variability. As a result, certain features of the transformed functions (in this case local extrema) become aligned horizontally. The amount of phase variability removed in the alignment process is quantified by the warping functions. The remaining variability in the aligned functions is the vertical variability. In this section, a brief description of the registration process is given and references are provided for the reader who desires more mathematical detail. The registration process is then demonstrated using the axial predictions of WRS assuming isotropic hardening.

Notationally, let $f_i(d)$ represent the i^{th} WRS profile as a function of the depth d where $i = 1,2,\ldots,n$. This can in general refer to either an experimental measurement or a model prediction. Assume without loss of generality that the depths have been normalized to the interval $[0,1]$ and that each f_i is absolutely continuous. To align the collection of functions horizontally, a *warping function* is required for each i, denoted by $\gamma_i(d)$. The transformed functions are defined as $f_i(\gamma_i(d))$. The warping function transforms the WRS profiles by warping (i.e. transforming) the depths. These functions are constrained to be:

- Boundary preserving: $\gamma(0) = 0$, $\gamma(1) = 1$
- Diffeomorphisms - This essentially means that γ is differentiable and its inverse function, γ^{-1}, exists and is also differentiable. Note that the inverse function has the property that $\gamma(\gamma^{-1}(d)) = d$

The first property ensures that the endpoints of the transformed functions remain unchanged and the second ensures the transformed function remains smooth and can be mapped back to its original function. Other than these constraints, no other assumptions are placed on the warping functions and in practice they are discretized on a fine grid. Most methods developed to define optimal $\gamma_i's$ (denoted by γ_i^*) involve minimizing a distance of the transformed functions to a reference function. Several examples are described in [7]. For the present application the method described in [6] is adopted. The main advantage of this method is that it is automatic in the sense that the reference function is defined within the alignment process using a proper distance metric. This reference function is known as the *Karcher mean* and is a generalization of the one-dimensional mean to the space of functions. It accounts for both phase and amplitude variability and is typically a better representation of the 'average' function than the cross-sectional mean when significant phase variability exists. As an example of this for demonstration only, consider Figure 3. Plotted are ten sinusoidal curves (gray) that have the same starting value but different frequencies chosen at random. The result is misalignment in the horizontal direction that becomes more noticeable as the input moves from left to right. Also plotted are the Karcher mean (red) and the cross-sectional mean (blue). For small input values between 0 and 0.2, these two means more or less

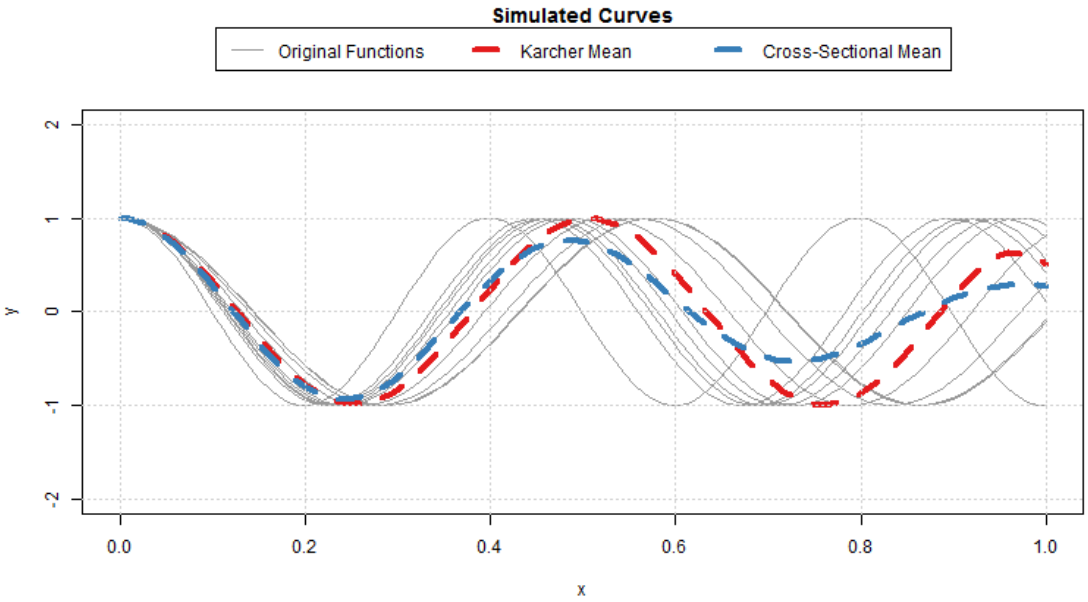


Figure 3: Example of the Karcher mean compared to the cross-sectional mean. When significant horizontal (phase) variability exists, the cross-sectional mean can be a poor representation of the average shape of the functions. The Karcher mean takes into account horizontal variability and as such is a better representation of the average shape.

agree because there is not much misalignment. However, as the misalignment becomes more substantial, the cross-sectional mean begins to deviate from the overall shape of the curves. In particular, the misalignment causes the cross-sectional mean to flatten out since the extrema do not occur at the same input values. However, the Karcher mean retains the overall shape of the curves because it considers the horizontal misalignment in its definition (given below). The method adopted here from [6] removes horizontal variability by aligning the functions to the Karcher mean, which captures well the overall shape of the curves. The removed horizontal variability can be modeled separately from the vertical variability that remains in the aligned functions. Finally the two models can be combined to create a single unified model accounting for both horizontal and vertical variability.

The Karcher mean, μ_f, is defined by:

$$\mu_f = \underset{f \in \mathscr{F}}{\arg\min} \sum_{i=1}^{n} D(f, f_i \circ \gamma_i) \tag{1}$$

where $f_i \circ \gamma_i$ is shorthand for the composition $f_i(\gamma_i(d))$, $D(\cdot,\cdot)$ is a proper distance metric defining each γ_i to be closest to f, and $\mathscr{F}$ is the class of absolutely continuous functions on the domain of interest. Essentially, this means it is assumed that the functions f_i (i.e. the WRS profiles for both measurements and predictions) are smooth and that the fundamental theorem of (Lebesgue integral) calculus applies. That is, each function can be written as an integral of its derivative. These

are technical details that are necessary to define $D(\cdot,\cdot)$ as a proper distance metric so that all of the mathematics applied works properly. In practice, the method can be applied to functional data which can be assumed to be continuous and differentiable, as is the case in the WRS application in this document. Further details are given in [6], but it should be pointed out that since $D(\cdot,\cdot)$ is a distance metric it is a minimization function in and of itself. Hence, equation (1) is really a dual minimization problem that must be solved iteratively. Notionally, the set of γ_i's that minimizes the sum in equation (1) is calculated for each $f \in \mathscr{F}$. The f at which this sum is smallest is the Karcher mean, μ_f. In practice, the sum is not calculated for every $f \in \mathscr{F}$ as this would be computationally prohibitive. Instead, efficient iterative procedures are available (see [6] and references therein). To summarize, the registration process provides three quantities:

1. The Karcher mean μ_f.
2. The set of optimal warping functions $\{\gamma_i^*\}$.
3. The set of aligned functions $\{\tilde{f}_i = f_i \circ \gamma_i^*\}$.

Computationally, the process has been implemented in **R** using the **time_warping** function in the package **fdasrvf** [8]. Implementations are also available in the Matlab, Python, and Julia languages.

Example results of the optimization in Equation 1 applied to the smoothed predictions of axial WRS assuming isotropic hardening (Figure 2) are displayed in Figures 4 - 5. Figure 4 displays the aligned functions $\{\tilde{f}_i\}$ where the black dashed line is the Karcher mean. It can be seen that compared to the unaligned functions (Figure 2) the local extrema are well aligned and the phase variability has been removed. The remaining variability in the aligned functions is just amplitude variability.

Figure 5 displays the optimal warping functions $\{\gamma_i^*\}$ where the black dashed line is the identity (i.e. the 45° line). Note that all of the warping functions intersect $(0,0)$ and $(1,1)$. This property preserves the endpoints of the f_i after warping ($f_i(x) = \tilde{f}_i(x), x = 0,1$) as mentioned above. A warping function below the diagonal pushes functions to the right and one above the diagonal pushes functions to the left. Note that the warping functions do not just shift the f_i, but also stretch or compress them. If no phase alignment was needed then the warping function would lie directly on the diagonal and $\tilde{f}_i = f_i$. It is in this sense that the variation in the warping functions represents phase variability.

2.2 Modeling Amplitude Variability

Amplitude variability is represented in the set of aligned WRS functions $\{\tilde{f}_i\}$ and there are many ways to model the variability in such functions. Perhaps the oldest and most well known method is functional principal components analysis (fPCA) which is described more thoroughly in Section 7.1. It should be noted that for technical reasons, fPCA is actually performed on a transformation

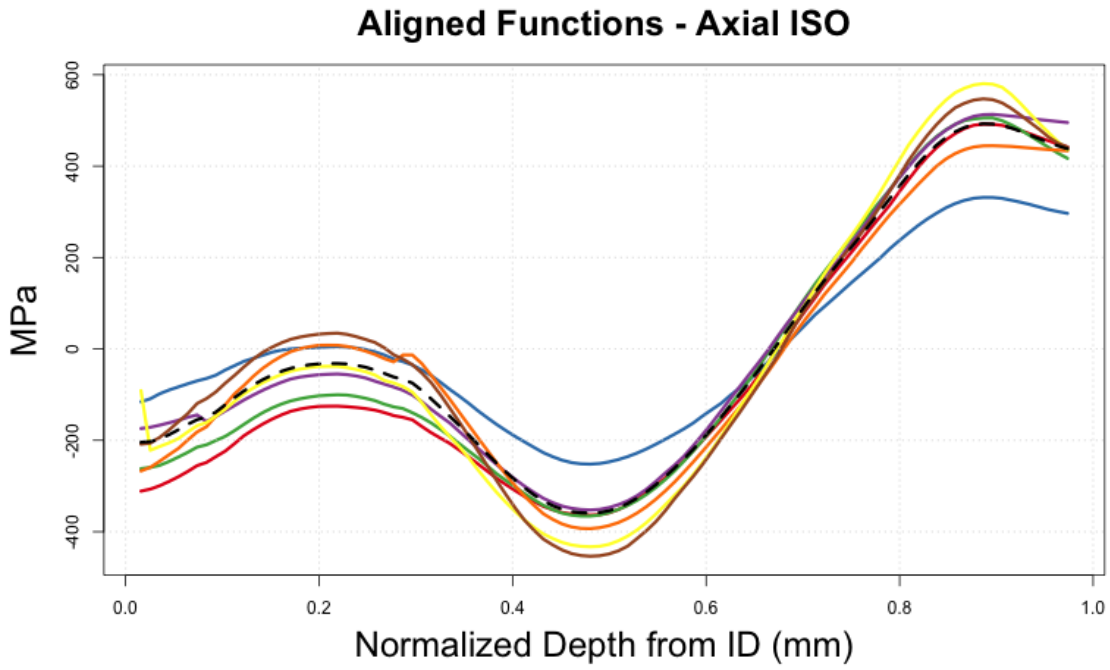


Figure 4: Aligned $\tilde{f}_i$ of the axial predictions assuming isotropic hardening. The black dashed line is the Karcher mean and the seven colored lines are the aligned predictions.

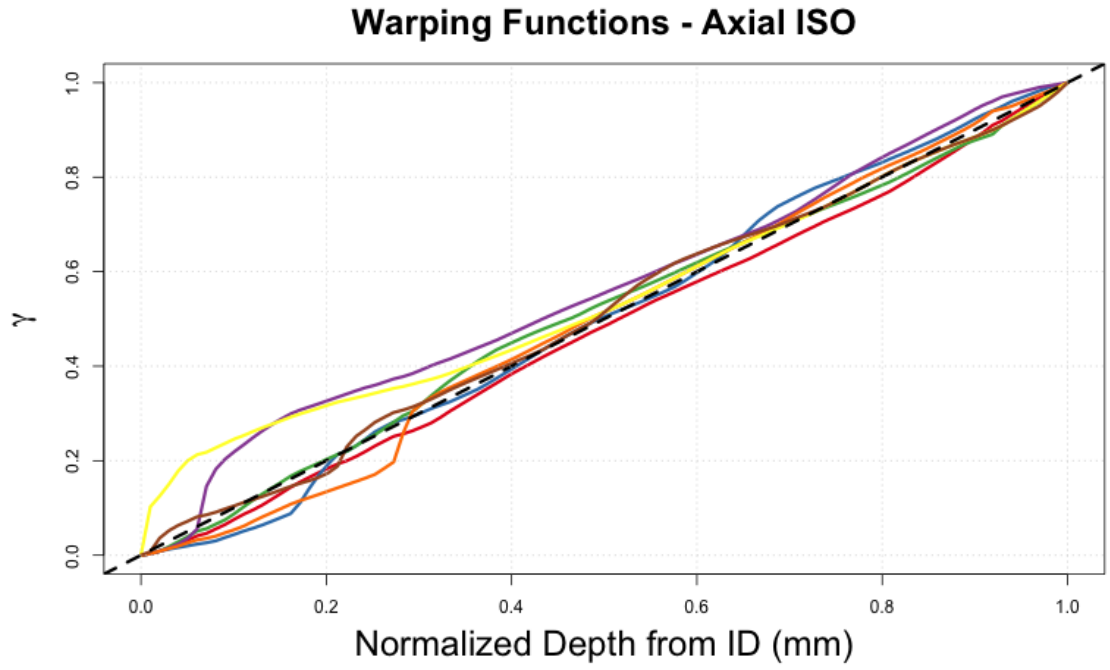


Figure 5: The warping functions γ_i of the axial predictions assuming isotropic hardening.

of the aligned functions referred to as the *square root slope function* (SRSF) defined by

$$\tilde{q}_i(d) = \text{sign}(\tilde{f}'(d))\sqrt{|\tilde{f}'(d)|} \tag{2}$$

See [6] for more details on the SRSF and reasons for its use. Note that $\tilde{f}_i$ can be reconstructed from $\tilde{q}_i$ by inverting Equation 2 due to the absolute continuity properties imposed on $\tilde{f}_i$:

$$\tilde{f}_i(d) = f_i(0) + \int_0^d \tilde{q}_i(s)|\tilde{q}_i(s)|ds \tag{3}$$

The above is mentioned only for completeness in describing the modeling process at a high level. The goal is still to model the amplitude variability in the WRS functions using aligned functions in which the phase variability has been removed. The aligned functions used to model the amplitude variability are the aligned SRSFs, yet these have a one-to-one correspondence with the aligned WRS functions through equation (3).

fPCA is described more thoroughly in Section 7.1. For now it is pointed out that this method is used to model the variation in $\{(f_i(0), \tilde{q}_i)\}$. The $\{f_i(0)\}$ is included in this part of the analysis in order to model the vertical variation at the starting point as well. At a high level, fPCA is a dimension reduction technique that associates each $(f_i(0), \tilde{q}_i)$ with a set of principal component scores (also called coefficients) $c_i = (c_{i1}, c_{i2}, \ldots, c_{ik_c})^\top$. Using just k_c scores, the infinite dimensional functional space is effectively reduced to k_c dimensions. The value of k_c is chosen to capture the majority of the variability in the $\{(f_i(0), \tilde{q}_i)\}$. The vectors c_i, $i = 1, \ldots, n$ are modeled using a multivariate Gaussian probability distribution which is completely characterized by its mean and covariance. By construction, the mean vector is zero and the covariance is a $k_c \times k_c$ diagonal. The diagonal elements of the covariance are estimated directly from the principal component scores using the sample variance of $\{c_{ij}, i = 1, 2, \ldots, n\}$ for $j = 1, \ldots, k_c$. It should be noted that from any of the c_i, the corresponding $f_i(0)$ and $\tilde{q}_i$ can be reconstructed (see Section 7.1). Hence the corresponding $\tilde{f}_i$ can be reconstructed using Equation 3.

Notationally the model is

$$c_1, c_2, \ldots, c_n \sim N_{k_c}(0, \Sigma_c) \tag{4}$$

where $N_{k_c}(0, \Sigma_c)$ denotes the multivariate Gaussian distribution of dimension k_c with mean 0 and covariance matrix Σ_c. One high level justification for the use of the Gaussian model is that the scores are linear combinations of random quantities and such linear combinations are often modeled well using Gaussian models. However, a better justification is the reasonableness of the results of the model. This is discussed further when considering samples of WRS functions from the final model in Section 2.4. For now it is noted that the amplitude variability model (4) implies a probability distribution for the aligned WRS functions $\tilde{f}$ which is used to quantify their uncertainty. To sample an aligned function from this probability distribution, sample a new c from the above model and reconstruct its corresponding SRSF $\tilde{q}$. Then Equation 3 is used to find the corresponding $\tilde{f}$.

2.3 Modeling Phase Variability

Phase variability is represented by the warping function $\{\gamma_i\}$ (the model is applied to the optimal γ_i^* but the * notation is dropped for convenience). Modeling the warping functions is more difficult due to the properties that warping functions must satisfy. As mentioned in Section 2.1, the functions must be *diffeomorphisms* that preserve the boundary points $(0,0)$ and $(1,1)$ ($\gamma_i(x) = x, x = 0,1$). This is a property that essentially says the $\gamma's$ are smooth invertible functions that preserve the endpoints of f and whose inverses are also smooth. This restriction makes modeling more difficult because the probability model used must be a probability model with respect to this restricted space of functions (i.e. its sample space must only include functions satisfying these properties). To handle these restrictions the method developed in [6] is implemented. This method applies strategically chosen transformations of the $\gamma's$ so that fPCA can be utilized. In essence, fPCA is defined only for linear spaces and, due to their constraints, the $\gamma's$ do not belong to such a space. The transformed functions do belong to a linear space and fPCA is applied to these functions. Similar to the fPCA model described for the amplitude variation in Section 2.2, this method results in a set of principal component scores $z_i = (z_{i1}, z_{i2}, \ldots, z_{ik_z})$ for each $i = 1,2,\cdots,n$ where the dimension k_z is chosen to explain most of the variability in the observed (transformed) $\{\gamma_i\}$. The scores are again modeled with a mean zero multivariate Gaussian distribution with a $k_z \times k_z$ diagonal covariance matrix Σ_z estimated directly from the observed scores. Samples of vectors z from this Gaussian distribution can be mapped to a sampled warping function γ by inversion. Thus, the Gaussian model on the scores induces a probability distribution on the warping functions.

2.4 Combining the Models

The probability models for the aligned functions (Section 2.2) and the warping functions (Section 2.3) provide a means to sample a SRSF of WRS ($\tilde{q}$), a WRS warping function (γ), and WRS values at the inner diameter ($f(0)$). The properties imposed on these functions are necessary so that these samples can be mapped back to the corresponding WRS function in the original function space. In particular, $\tilde{f}$ is a function of $\tilde{q}$ (see Equation 3 and [6]) and $f = \tilde{f} \circ \gamma^{-1}$ where γ^{-1} is the inverse function of γ (i.e. $\gamma(\gamma^{-1}(d)) = d$). Thus, by sampling from the probability models, a population of WRS functions that captures the uncertainty in WRS can be generated.

As an example, the above model is applied to the smoothed axial predictions assuming isotropic hardening displayed in Figure 2 to generate a set of 100 sampled WRS profiles. These samples are displayed in Figure 6 and have similar amplitude and phase variability characteristics to the original sample – *an indication that the combined model properly characterizes the variability in the observed data and in particular, that the Gaussian models assumed on the principal component scores are reasonable*. Samples of functions, such as those displayed in Figure 6, form the basis for constructing the statistical intervals described in Section 2.6. These intervals provide a means to quantify uncertainty as well as to compare predictions of WRS to their corresponding measurements.

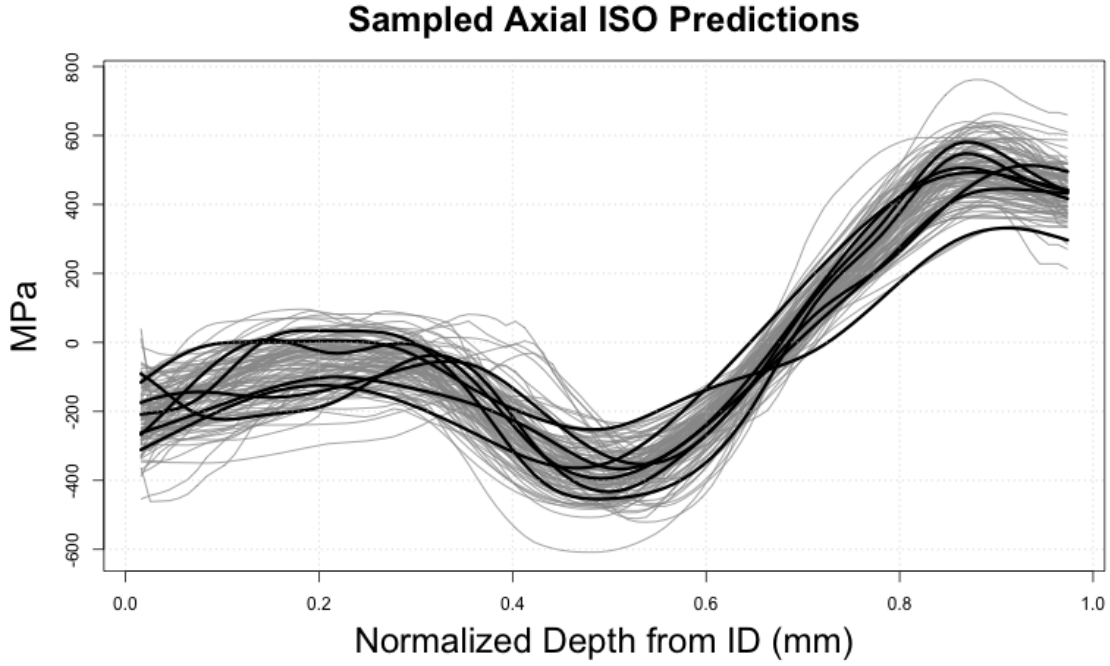


Figure 6: A sample of 100 (gray) functions from the probability model fit to the original data (black). The data are predictions of WRS in the axial direction assuming isotropic hardening.

Balancing Condition for Axial Stress

Physically, axial stresses must satisfy the balancing relation:

$$\int_0^1 f(x)\, dx = 0 \tag{5}$$

It is noted that the statistical model described does not account for this constraint. Further work is needed to develop a statistical model that explicitly constrains each sample to satisfy (5). However, the samples in Figure 6 visually have similar characteristics as the prediction data and it is believed that the uncertainty in the data is captured relatively well, despite lacking the balancing condition. The integral values of each curve generated under the model are displayed in the histogram in Figure 7. The red points represent the estimated integrals for the actual model predictions and differ from zero due to discretization. A good portion of the sampled curves differ from zero by around the same magnitude. However, others differ by significantly larger magnitudes and efforts to refine the statistical model to account for the balancing constraint have merit.

2.5 A Brief Introduction to Statistical Bootstrap Techniques

Statistical bootstrap techniques are a flexible set of re-sampling techniques that can be used to quantify uncertainty in essentially any parameter of a probability distribution, such as a mean or

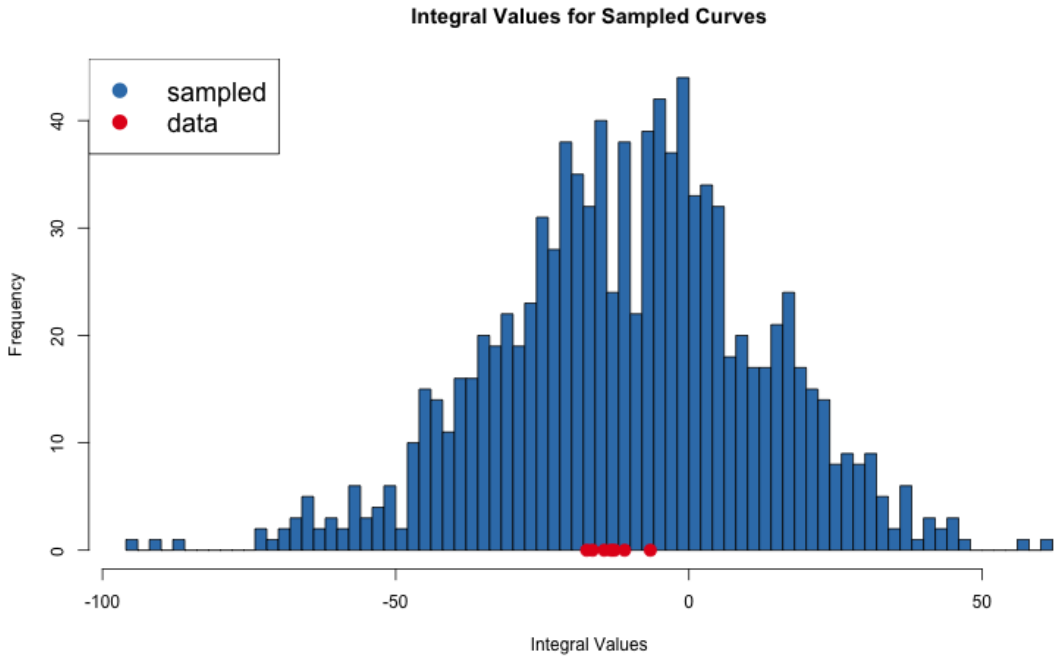
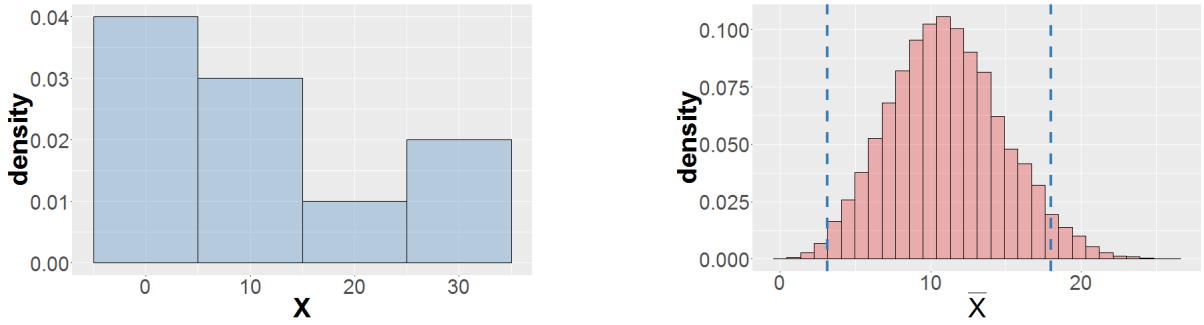


Figure 7: Integral values of the 1000 sampled curves from the model using the isotropic predictions of axial stress. The red points indicate the values for the given data.

set of quantiles. For a detailed overview of statistical bootstrap techniques see [9] and [10]. A brief overview of these techniques is given in this section.

Bootstrap techniques quantify uncertainty in parameters by estimating the sampling distribution of statistics that are used to estimate them. The following basic example illustrates the method. Consider a sample of size $n = 10$ of scalar values of some quantity of interest (i.e. random variable) X. A histogram of these values is displayed in Figure 8a. Suppose also that interest lies in estimating the mean of the distribution of X. The natural point estimate for the mean is the sample mean. To construct a confidence interval once could assume the data are normally distributed or that the central limit theorem applies and use the standard $(1-\alpha)$ confidence interval $\bar{x} \pm t_{n-1,\alpha/2} s/\sqrt{n}$ where $\bar{x}$, s, and $t_{n-1,\alpha/2}$ are the sample mean, standard deviation, and $\alpha/2$ quantile of the t-distribution with $n-1$ degrees of freedom, respectively. However, there may be reasons to be suspicious of the normality assumption and/or that the central limit theorem applies with a limited sample size of $n = 10$. This would mean that the sampling distribution of $\bar{x}$ may not be a t-distribution as assumed in the construction of the confidence interval above. In such a case, the bootstrap method offers a flexible alternative to estimate the sampling distribution. When applying the bootstrap, the population distribution of X is estimated with a distribution function F. Subsequently, F is used to estimate the sampling distribution $\bar{x}$ by repeatedly obtaining samples of size n from F and recording the value of $\bar{x}$. The collection of values of $\bar{x}$ estimates the sampling distribution of the sample mean. An example histogram of samples of $\bar{x}$ using the data in Figure 8a appears in Figure 8b. The vertical dashed lines in the figure are the 0.025 and 0.975 quantiles and represent a 95% bootstrap confidence interval because 95% of the samples lie between these



(a) (b)

Figure 8: (a) Histogram of 10 samples of a random variable X. (b) Histogram bootstrap sample of $\bar{x}$ representing an estimate of the sampling density. The vertical dashed lines are the 0.025 and 0.975 quantiles and represent one way to construct a 95% bootstrap confidence interval for the mean. Other methods for constructing bootstrap confidence intervals are discussed in 2.6.

lines. For this example, F was taken to be the *empirical distribution function* which means the distribution of X is estimated with the original sample of 10 values. A single bootstrap sample (resulting in a single $\bar{x}$ in the histogram of Figure 8b) consists of obtaining a random sample (with replacement) of size $n = 10$ from the original observed values. It should be noted that F can also be estimated in other ways. A parametric bootstrap approach would fit a parametric distribution to the data (such as a Gaussian) and re-sample from this distribution instead. In the current application of WRS uncertainty, there is limited data and using the empirical distribution function is unreliable. Instead, the model described in Sections 2.2, 2.3, and 2.4 is used for the re-sampling. This model is partly parametric (due to the Gaussian models on the scores) and partly non-parametric (due to the fPCA) and hence it is referred to as a 'semi-parametric' model.

2.6 Constructing Bootstrap Confidence Bounds on WRS Functional Data

The simple example given in the previous section is only intended to give a conceptual understanding of bootstrap techniques. Such techniques can be applied to estimate the sampling distribution of essentially any statistic of interest, including those computed with functional data. Since a methodology is available to sample functions using the model described in Sections 2.2, 2.3, and 2.4, the sampling distribution of functional statistics can be estimated by using such bootstrapping techniques. Confidence bounds can be constructed from these estimated sampling distributions.

In this section, the semi-parametric bootstrap technique used to construct such confidence bounds is described. The term 'semi-parametric' is used to describe the technique because the fPCA method is non-parametric yet the principal component scores are modeled with a parametric distribution. Hence there are both non-parametric and parametric aspects of the technique. In the next two sections, the steps taken to construct confidence bounds for the mean parameter as well as population quantiles are described. Such steps could be adjusted to construct confidence bounds for essentially any other parameter of interest. In particular, the section concludes with a description on how one would construct a confidence bound on the difference in means between

two populations. These types of confidence bounds are used to assess the level agreement between WRS predictions and experimental measurements.

Constructing Confidence Bounds on the Mean: Assessing confidence on the mean WRS profile

The steps to construct confidence bounds on the mean are as follows:

1. Sample n WRS functions using the model described in Sections 2.2, 2.3, and 2.4 on a fine grid of K values of d: $\{d_k : k = 1,2,\ldots,K\}$. This results in samples:

$$f_i^s(d_k),\ k = 1,2,\ldots,K \text{ and } i = 1,2,\ldots,n \tag{6}$$

   where the superscript s implies f is a randomly sampled WRS profile from the model.

2. Compute the mean WRS profile denoted by $\bar{f}^s = (\bar{f}^s(d_1), \bar{f}^s(d_2), \ldots, \bar{f}^s(d_K))$ where

$$\bar{f}^s(d_k) = \frac{1}{n}\sum_{i=1}^{n} f_i^s(d_k). \tag{7}$$

3. Repeat steps 1 and 2 many times. Denote the number of repetitions by S (e.g. $S = 1000$).

4. For each k, compute the $(1-\alpha/2)$ and $\alpha/2$ quantiles of the sample mean WRS values $\bar{f}^s(d_k)$ over the S samples. These quantiles form a pointwise $100(1-\alpha)\%$ bootstrap confidence bound for the population mean function.

As an example, the above steps are applied to construct a confidence bound for the mean of the axial predictions assuming isotropic hardening in Figure 2. Here and throughout the rest of the report, $K = 100$ and $S = 1000$. Recall that K is the number of depths at which the sampled functions are evaluated and S is the number of bootstrap samples at each depth d_k for $k \in \{1,2,...,K\}$. Figure 9 displays 30 of the $S = 1000$ bootstrap means computed in step 2 above as blue lines. The gray lines are the original smoothed data. The pointwise 0.025 and 0.975 quantiles of the collection of all $S = 1000$ bootstrap means (as computed in step 3) are displayed as the thick dashed lines in Figure 10. These bands represent 95% pointwise confidence bounds on the mean of the predictions.

Constructing Tolerance Bounds: Assessing lower and upper bounds for the WRS profiles

Tolerance bounds are bounds that contain a certain proportion of the population with a specified level of confidence [11]. For two-sided tolerances bounds there is both an upper and a lower tolerance bound. The upper tolerance bound is an upper confidence bound on an upper population quantile. Likewise, the lower tolerance bound is a lower confidence bound on a lower population quantile. In this sense, tolerance bounds are simply confidence bounds on population quantiles.

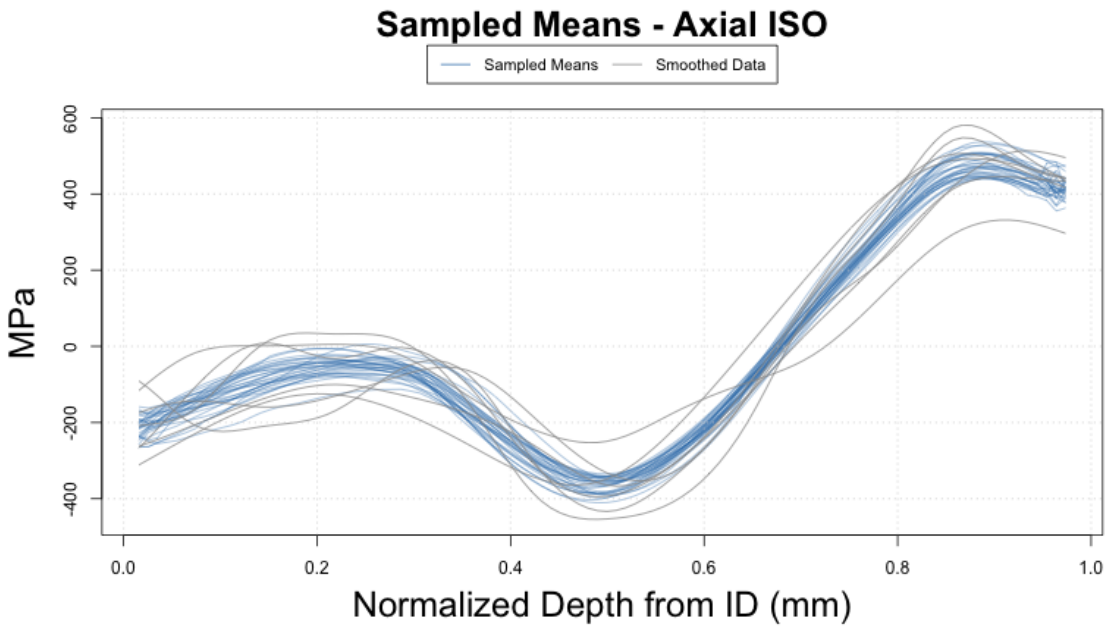


Figure 9: Thirty of the $S = 1000$ bootstrap sample means (thin blue lines). These estimate the sampling distribution of the sample mean function and form a basis for the confidence bounds in Figure 10. The gray lines are the original (smoothed) data. Only thirty bootstrap sample means are plotted for illustrative clarity.

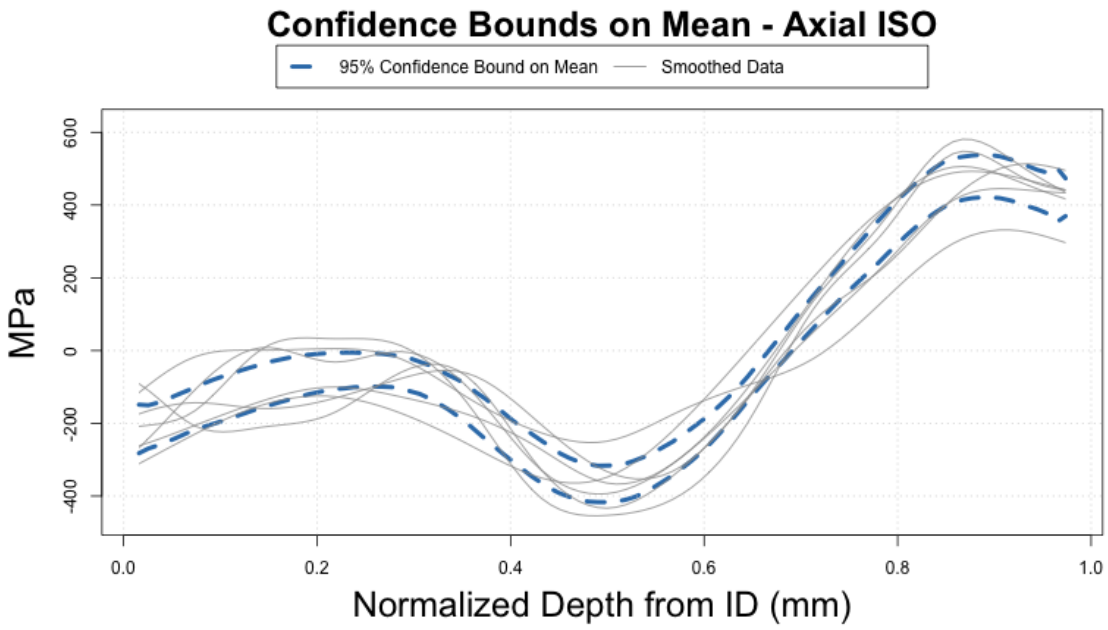


Figure 10: Pointwise 95% bootstrap confidence bounds for the mean prediction of WRS in the axial direction with isotropic hardening (thick dashed blue lines). These are computed using the pointwise 0.025 and 0.975 quantiles of the collection of all $S = 1000$ bootstrap means. The gray lines are the original (smoothed) data.

A $(1-\alpha)100\%$ confident tolerance bound with $(1-q)100\%$ coverage has the following interpretation: 'we are $(1-\alpha)100\%$ confident that $(1-q)100\%$ of the population will fall between the given bounds.'

Construction of equal-tailed tolerance bounds using the same bootstrap approach is now described. This means the proportion of the population above and below the bounds is split equally. That is a $(1-\alpha/2)100\%$ upper confidence bound on the $1-q/2$ quantile and a $(1-\alpha/2)100\%$ lower confidence bound on the $q/2$ quantile is constructed. Construction of bootstrap tolerance bounds follows essentially the same steps as for confidence bounds on the mean described earlier. The only difference is that the the desired quantiles are computed for each bootstrap sample. The steps are as follows:

1. Sample n WRS functions using the model described in Sections 2.2, 2.3, and 2.4 on a fine grid of K values of d: $\{d_k : k = 1,2,\ldots,K\}$. This results in samples:

$$f_i^s(d_k),\ k = 1,2,\ldots,K \text{ and } i = 1,2,\ldots,n \tag{8}$$

   where the superscript s implies f is a randomly sampled WRS profile from the model.

2. Compute the p^{th} quantile of the WRS profile denoted by $f_p^s = (f_p^s(d_1), f_p^s(d_2), \ldots, f_p^s(d_K))$ for $p = q/2$ and for $p = 1-q/2$ where $f_p^s(d_k)$ is the p^{th} quantile of $\{f_i^s(d_k)\}_{i=1}^n$. It should be noted that there are many ways to compute these quantiles, but with limited data (small n) a parametric assumption is made in order to get reliable quantile estimates. In particular, for each k the collection of $\{f_i^s(d_k)\}_{i=1}^n$ is assumed to follow a Gaussian distribution resulting in the quantile estimate

$$f_p^s(d_k) = \bar{f}^s(d_k) + z_p s_k/\sqrt{n}, \tag{9}$$

   where $\bar{f}^s(d_k)$ and s_k are the sample mean (see equation (7)) and sample standard deviation of $\{f_i^s(d_k)\}_{i=1}^n$, respectively, and z_p is the p^{th} quantile of the standard normal distribution.

3. Repeat steps 1 and 2 many times. Denote the number of repetitions by S (e.g. $S = 1000$).

4. Compute the pointwise $(1-\alpha/2)$ and $\alpha/2$ quantiles [1] of $f_p^s(d_k)$ over the S samples for each $k = 1,\ldots,K$. These quantiles form a pointwise $100(1-\alpha)\%$ bootstrap tolerance bound with $(1-q)100\%$ coverage.

As an example the above steps are applied to construct a tolerance bound for the population of axial predictions assuming isotropic hardening given the sample in Figure 2. Figure 11 displays 30 of the $S = 1000$ bootstrap upper (red) and lower (blue) quantiles for $q = 0.05$ computed in step 2 above. The black lines are the original smoothed predictions. The pointwise 0.025 and 0.975 quantiles of the $S = 1000$ bootstrapped quantiles (as computed in step 3) are displayed as the thick dashed lines in Figure 12. These bands represent 95% coverage 95% pointwise tolerance bounds. For shorthand this is called a $95/95$ tolerance bound where the first number refers to the confidence level ($(1-\alpha)100$) and the second to the coverage ($(1-q)100\%$).

[1] These quantiles are computed using standard nonparametric sample quantiles. Since a large bootstrap sample ($S = 1000$) is available there is no need to make any parametric assumptions.

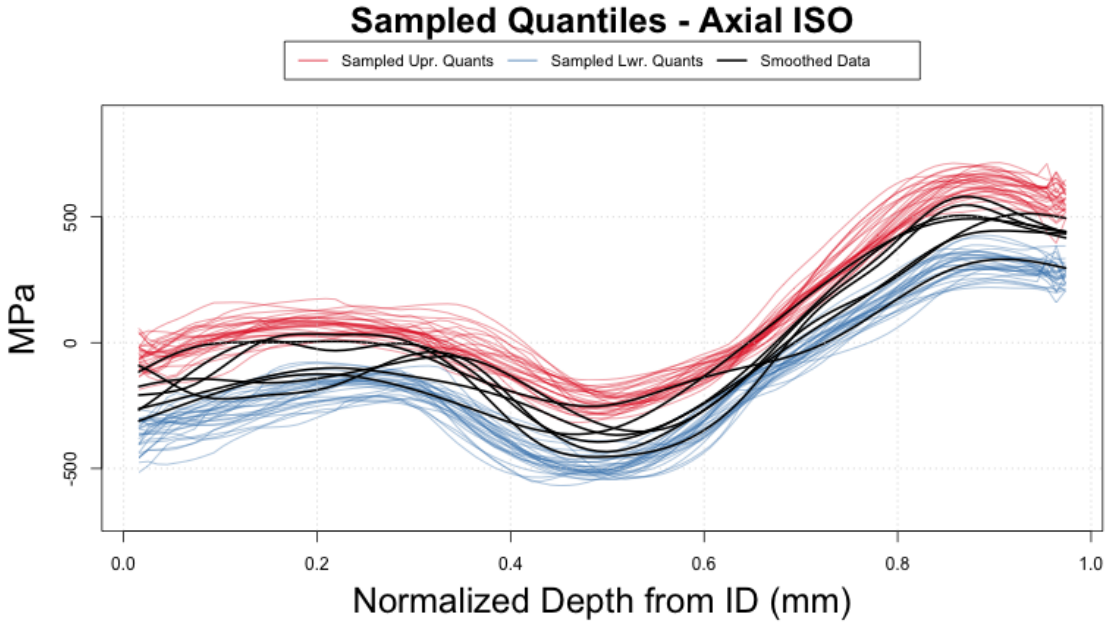


Figure 11: Thirty of the $S = 1000$ bootstrap upper (red) and lower (blue) quantiles for $q = .05$. These estimate the sampling distribution of the sample quantile functions and form a basis for the tolerance bounds in Figure 12. The black lines are the original (smoothed) data.

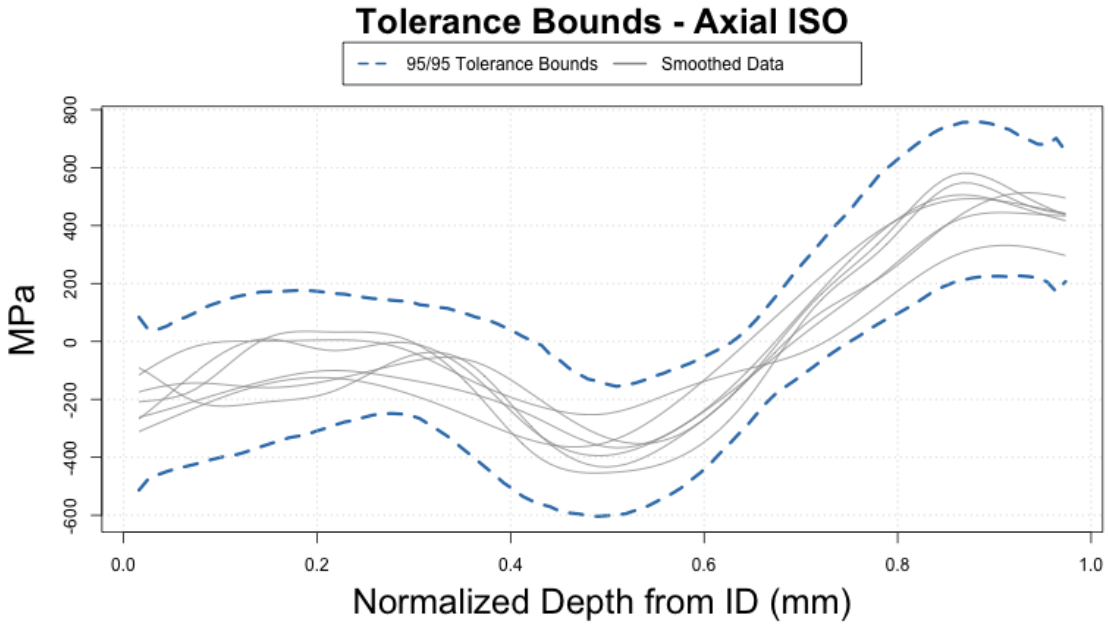


Figure 12: Pointwise 95/95 tolerance bounds for the WRS predictions in the axial direction (thick dashed blue lines). These are computed using the pointwise 0.025 and 0.975 quantiles of the collection of all $S = 1000$ bootstrap quantiles. The gray lines are the original (smoothed) data.

Constructing Confidence Bounds on the Mean Difference: Comparing WRS predictions to experimental measurements

To compare measurements and predictions, pointwise confidence bounds on the difference between mean measurements and mean predictions are constructed. The methodology for constructing these bounds is the same as that described in Section 2.6, except that the samples are differences of means. Let g denote a function from the statistical model used to quantify uncertainty in the experimental WRS measurements, with n_e denoting the number of observed experimental measurements. Similarly, let f denote the function from the statistical model used to quantify uncertainty in the WRS predictions, with n_p denoting the number of predictions available. The steps to construct confidence bounds on the mean difference are as follows:

1. Sample n_e WRS functions from the model used for UQ of the experimental measurements and n_p WRS functions from the model used for UQ of the predictions on a fine grid of K values of d: $\{d_k : k = 1,2,\ldots,K\}$. This results in samples of WRS functions from both measurement and predictions representing uncertainty in both:

$$g_i^s(d_k),\ k = 1,2,\ldots,K \text{ and } i = 1,2,\ldots,n_e$$
$$f_i^s(d_k),\ k = 1,2,\ldots,K \text{ and } i = 1,2,\ldots,n_p$$

   where a superscript s implies the function is a sample from the corresponding model.

2. Compute $h^s = (h^s(d_1), h^s(d_2), \ldots, h^s(d_K))$ where

$$h^s(d_k) = \frac{1}{n_e}\sum_{i=1}^{n_e} g_i^s(d_k) - \frac{1}{n_p}\sum_{i=1}^{n_p} f_i^s(d_k). \tag{10}$$

3. Repeat steps 1 and 2 many times. Denote the number of repetitions by S (e.g. $S = 1000$).

4. Compute the pointwise $(1-\alpha/2)$ and $\alpha/2$ quantiles of $h_p^s(d_k)$ over the S samples for each k. These quantiles form a pointwise $100(1-\alpha)\%$ bootstrap confidence bound for the population difference of means function.

An example is plotted in Figure 13, which compares the axial WRS predictions assuming kinematic hardening to the DHD axial WRS measurements. The confidence intervals, plotted as thick dashed lines, were calculated with $K = 100$ depths and $S = 1000$ bootstrap differences of means as computed in step 3. These bands represent 95% pointwise confidence bounds on the difference between mean KIN axial predictions and mean DHD axial measurements.

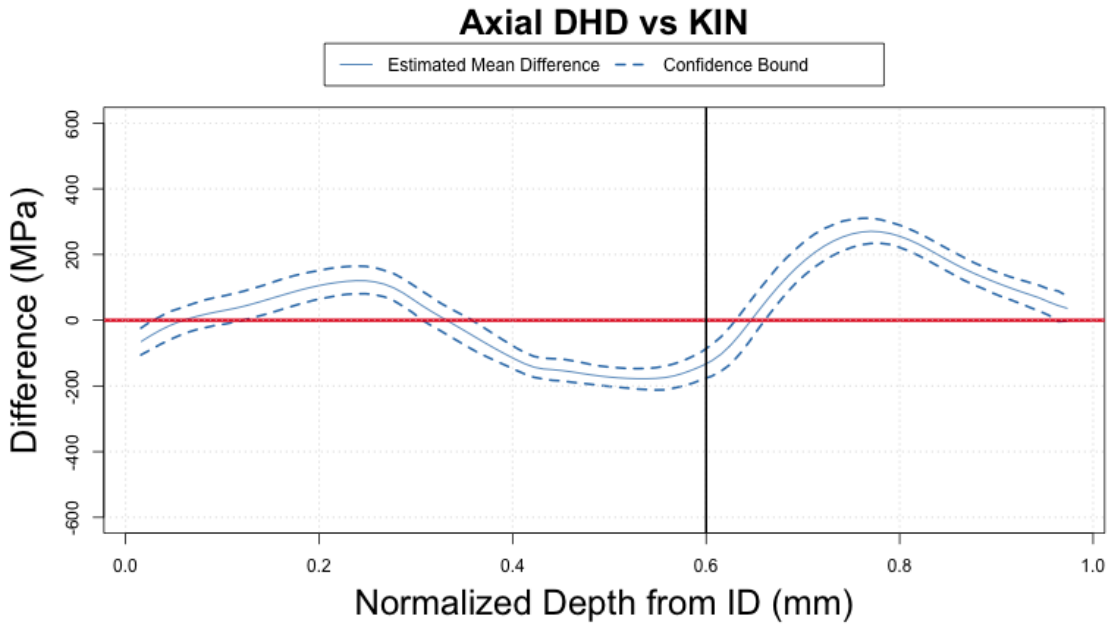


Figure 13: 95% Confidence interval on the mean difference between the axial DHD measurements and the axial predictions assuming kinematic hardening. A vertical line at ID depth $d = 0.6$ appears as a reminder that the smoothing error for the DHD measurements is large for $d > 0.6$.

3 Uncertainty Quantification Results for the Experimental Measurements

For uncertainty quantification of the experimental measurements, confidence bounds for the mean are constructed as well as tolerance bounds using the bootstrap method described in Section 2.6. A confidence level of $1-\alpha = 0.95$ was used. For the tolerance bounds, a coverage level 0.95 (i.e. $q = 0.05$) was used, resulting in the construction of $95/95$ tolerance bounds. Thus, the confidence bounds for the mean have the interpretation, 'at each value of depth we are 95% confident that the true mean lies within the specified bound.' The tolerance bounds have the interpretation 'at each value of depth we are 95% confident that 95% of all values from the population will fall between the bounds.' The tolerance bounds provide a range as to where to expect the majority (95%) of all possible measurements to fall.

A brief summary of the results in this section is as follows. There is substantial uncertainty in the DHD measurements, both hoop and axial, for normalized ID depths greater than 0.6. This uncertainty is largely due to the coarseness of the measurements taken in this region. However, since PWSCC initiates at the ID, this coarseness does not contribute much to the uncertainty of WRS in the ID regions that are of most interest in terms of crack initiations. Tolerance bounds indicate that the plausible range of the majority of DHD measurements at the ID, both axial and hoop, is around 200 MPa wide. Analysis of the axial contour measurements suggests a similar range. No uncertainty quantification on the hoop contour measurements could be performed with the current information.

The first step in modeling the variation in the group of experimental WRS measurement profiles is to apply smoothing. For this application the smoothing method described in Section 7.2 is applied. Smoothing induces a level of uncertainty due to the difference between the raw data and the smoothed data. The residuals for each smoothed data set are analyzed to qualitatively assess how much difference there is between smoothed and raw data. Since the models described in Section 2 are applied, it should be noted that the uncertainty due to smoothing is not accounted for. However, if the residuals are small compared to the overall variability between the functions, this level of uncertainty is of second order. Also, the smoothed functions are assumed to be good representations of WRS which reduce the effects of variability not caused by WRS variations. For example, the non-smoothness in the measurement data is mostly caused by measurement error of the inputs to the measurement models and is not caused by physical variations in WRS. Treating these errors as random implies the non-smoothness of the measurement results is also random and smoothing is justified.

For each of the following UQ sections for the DHD method, the smoothing is assessed and then confidence bounds for the mean and tolerance bounds for 95% of the population are provided. Together these provide a comprehensive UQ assessment for DHD measurement data. The contour data must be treated differently. There is only one hoop WRS profile at the weld centerline. This lack of information prohibits any examination of uncertainty in the hoop WRS measurements. Fortunately, five axial WRS profiles were provided in the data. Additionally, the mesh from which these profiles were obtained was also provided so that many more WRS profiles could be extracted.

This allows us to directly sample many axial WRS profiles in order to characterize the uncertainty in those measurements, thereby removing the necessity for modeling and bootstrapping.

3.1 Uncertainty Quantification of the Axial DHD Measurements

Smoothing

Raw and smoothed axial DHD measurements appear in Figures 14a and 14b, respectively. Notice that the residuals of the smoothed profiles (Figure 14c) are relatively small up to about the normalized ID depth value of $d = 0.6$. This is due to the coarseness of the measurements after this depth value and is considered a limitation in the collected data. As a result, the uncertainty induced by smoothing is not negligible after ID depth $d = 0.6$. The constructed confidence bounds, as previously noted, do not account for smoothing uncertainty and should only be trusted for depths less than 0.6. In future studies, this could be resolved by taking measurements at smaller increments near the outer diameter of the weld. However, it is noted that crack formation and propagation due to primary water stress corrosion cracking (PWSCC) begins on the inner diameter of the weld. Hence, smoothing in this context introduces negligible uncertainty with respect to the formation of cracks. However, the uncertainty is not negligible once cracks begin to propagate through the weld.

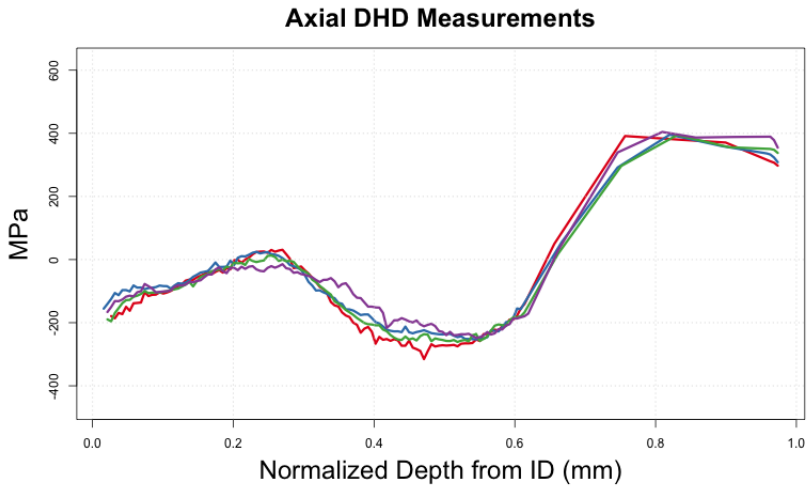
Confidence Bounds

The 95% confidence bounds on the mean DHD axial WRS measurements is plotted with the data in Figure 15. The width of the confidence bounds is less than 100 MPa through the weld depth except near the OD. This indicates 95% confidence that the true mean axial WRS measurement lies within that range. The 95/95 tolerance bounds and the data are plotted in Figure 16. These bounds describe the region at each depth in which it is expected that, with 95% confidence, 95% of axial WRS measurements will fall. Near the ID, the tolerance bounds suggest a range of about 200 MPa would be expected to contain 95% of all ID axial WRS measurements. As expected, due to the low resolution of measurements, the tolerance bounds at the OD are somewhat wider with a width of about 300 MPa.

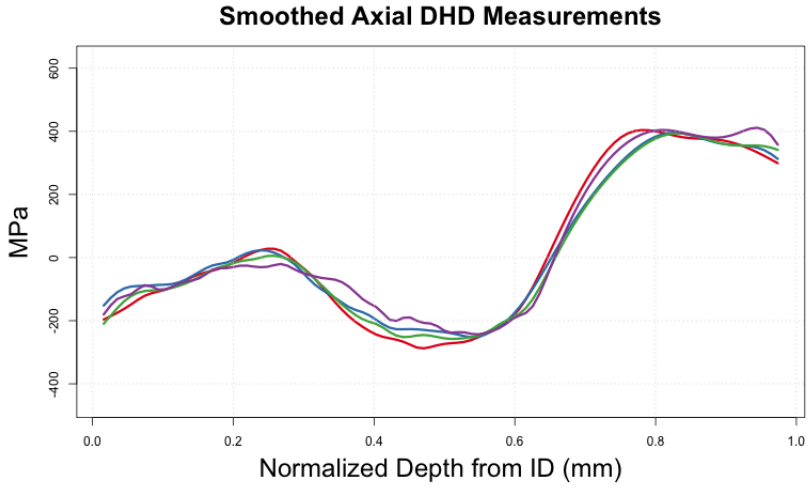
3.2 Uncertainty Quantification of the Hoop DHD Measurements

Smoothing

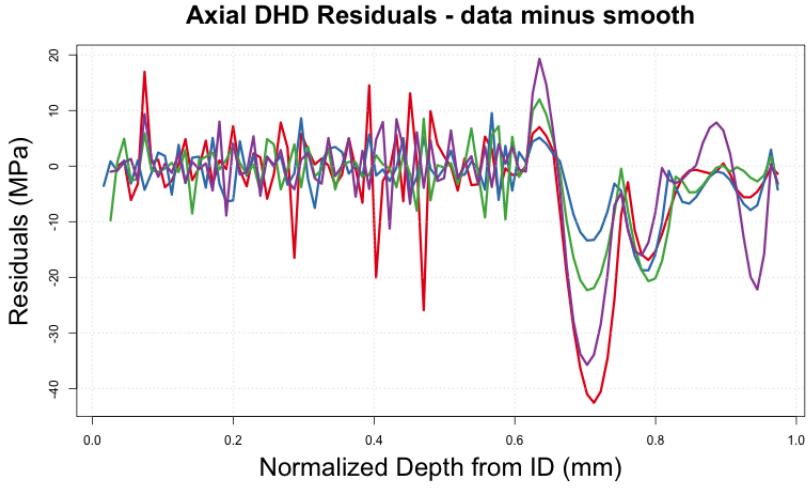
Raw and smoothed DHD hoop measurements appear in Figures 17a and 17b, respectively. As with the DHD axial measurements, the uncertainty induced by smoothing is not negligible after normalized depth $d = 0.6$ from the ID (see Figure 17c). The confidence bounds should only be



(a) Raw axial DHD measurements



(b) Smoothed axial DHD measurements.



(c) Residuals of the smoothed axial DHD measurements.

Figure 14: Smoothing results for the axial DHD measurements.

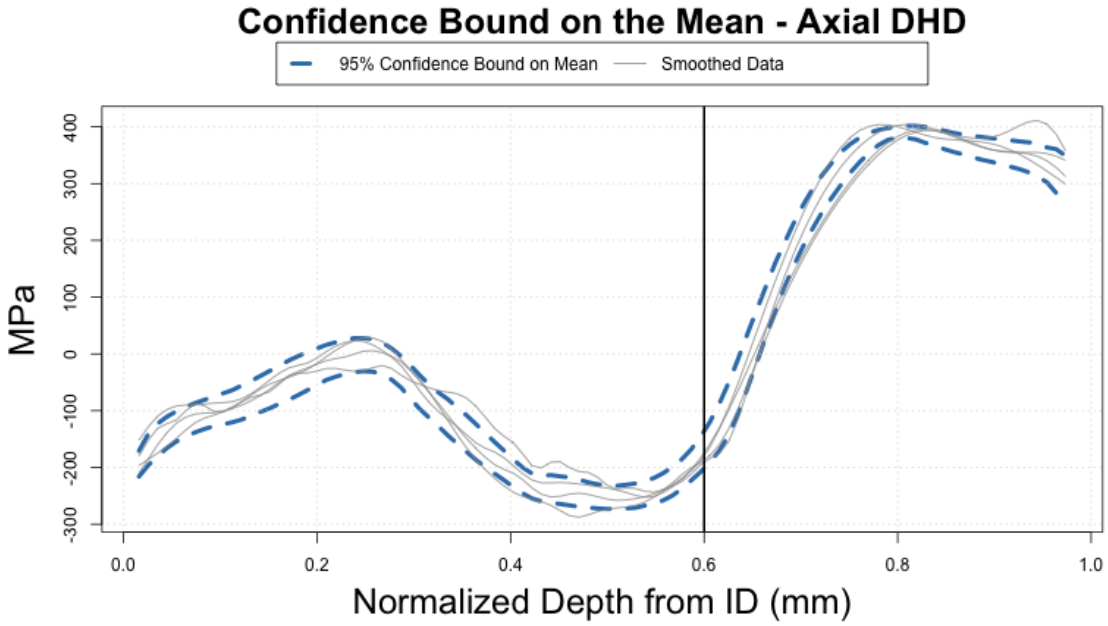


Figure 15: Confidence bound for the mean of DHD axial measurements.

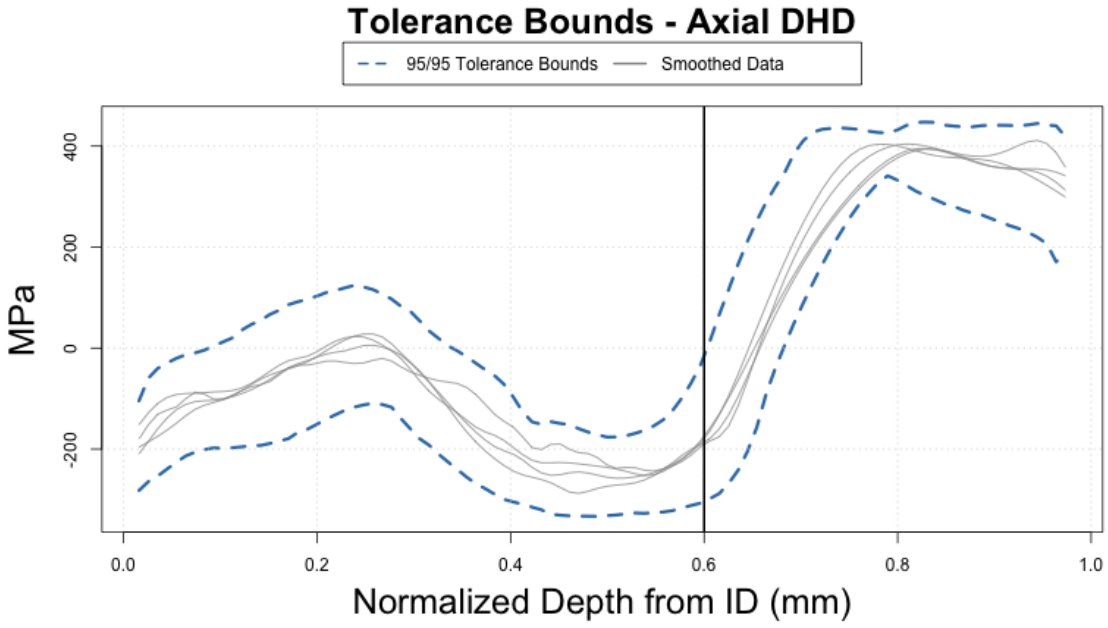


Figure 16: 95/95 tolerance bound on the DHD axial measurements.

considered reliable at relative depths from the ID less than $d = 0.6$ where the uncertainty induced by smoothing is negligible.

Confidence Bounds

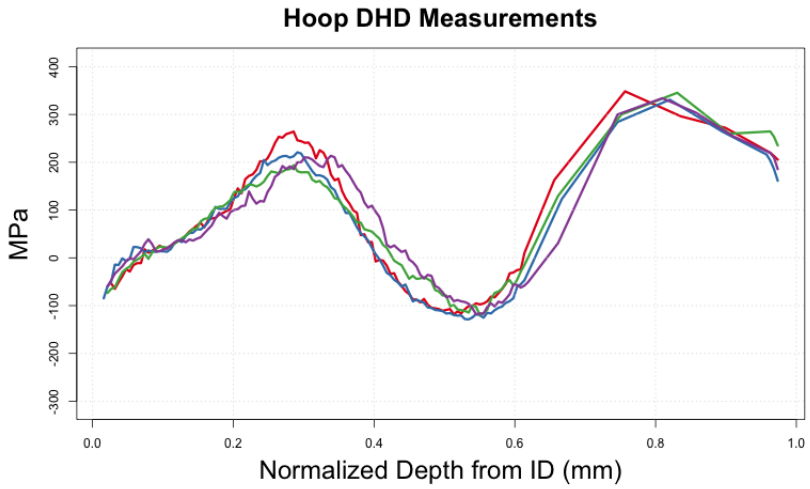
Confidence bounds for the mean DHD hoop measurements are plotted in Figure 18. The width of the confidence bounds is less than 100 MPa through the depth of the weld. It is expected that the mean of the measured hoop WRS near the ID will fall within a range less than 50 MPa wide with 95% confidence. Confidence bounds are significantly wider in the region of the weld near the OD; the mean of the measured hoop WRS is expected to fall within a larger band approximately 100 MPa in width.

The 95/95 tolerance bounds on DHD hoop measurements of WRS are plotted with the data in Figure 19. Except near the OD, this plot indicates 95% confidence at each depth that 95% of WRS measurements will fall between the bounds. Near the ID, this means that there is 95% confidence that 95% of WRS measurements will fall within a range approximately 100 MPa wide. As with the confidence bounds, notice that the tolerance bounds widen near the OD.

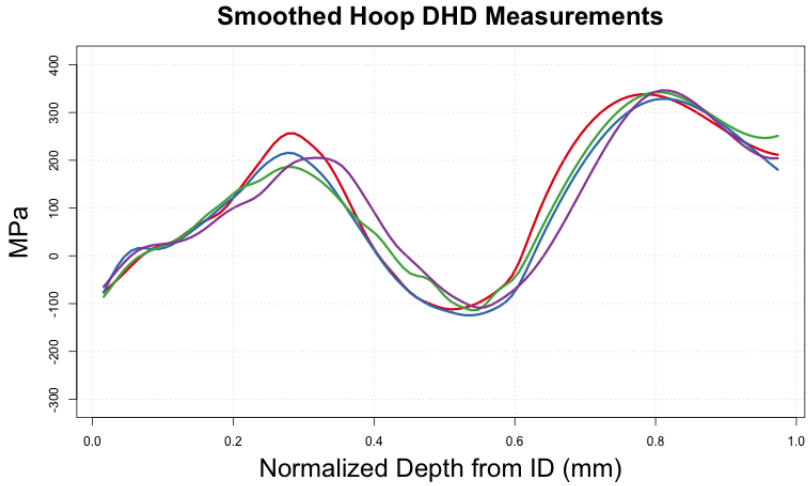
3.3 Uncertainty Quantification of the Axial Contour Measurements

The original set of axial contour measurement data consists of five axial WRS profiles that were extracted from the mid-weld cross-sectional results along five different paths. The paths for extracting the profiles were defined at the 25°, 35°, 45°, 55°, and 65°angular locations around the weld (see [3]). However, the set of contour data for axial WRS does not limit the number of locations at which stresses can be extracted. The large amount of data in this case allows for the characterization of uncertainty in the axial contour measurements without applying the modeling and bootstrapping procedure. The available data with example paths for data extraction is plotted in Figure 20. The corresponding axial WRS profiles are plotted in Figure 21 along with the original five WRS profiles provided by the measurement vendor (black). The paths and WRS profiles are colored in order of extraction; the first profile was extracted from a path on the right side of the mesh (blue) and subsequent paths were extracted sequentially moving to the left (green). This coloration is relevant because there is an apparent pattern in the WRS profiles based upon angular location around the weld. The five WRS profiles originally provided by the measurement vendor exhibit the same behavior as the profiles extracted for this study, suggesting that the pattern is present in the data and is not an artifact of the extraction method. The pattern may indicate either that the axial WRS is not axisymmetric (as would theoretically be the case) or that the machining process did not release strain evenly. Variation in the profiles represents measurement uncertainty (due to error in the FE model input data and systematic uncertainty due to uneven release of strain during machining) or stresses that are not axisymmetric.

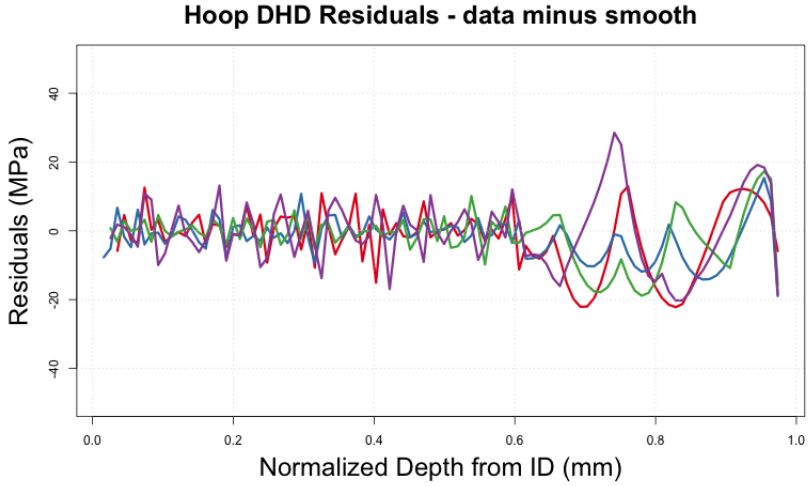
The data provided by the measurement vendor covers a wedge of the cross section that can be defined between two circles. The smaller circle defines the inner diameter of the weld and the



(a) Raw hoop DHD measurements.



(b) Smoothed hoop DHD measurements.



(c) Residuals of the smoothed hoop DHD measurements.

Figure 17: Smoothing results for the hoop DHD measurements.

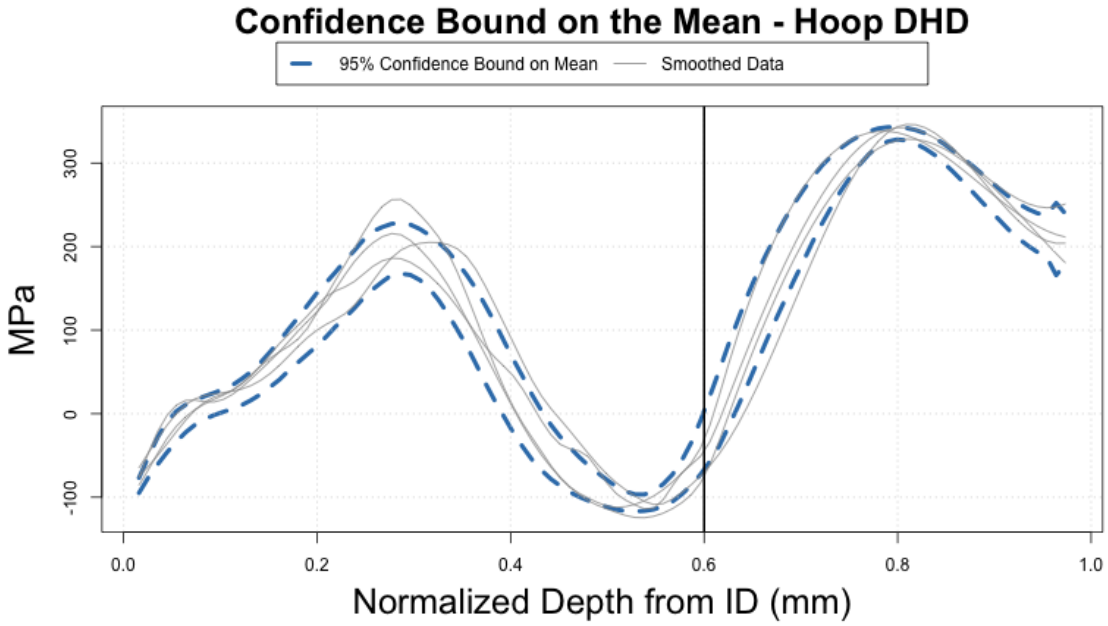


Figure 18: Confidence bound for the mean of DHD hoop measurements.

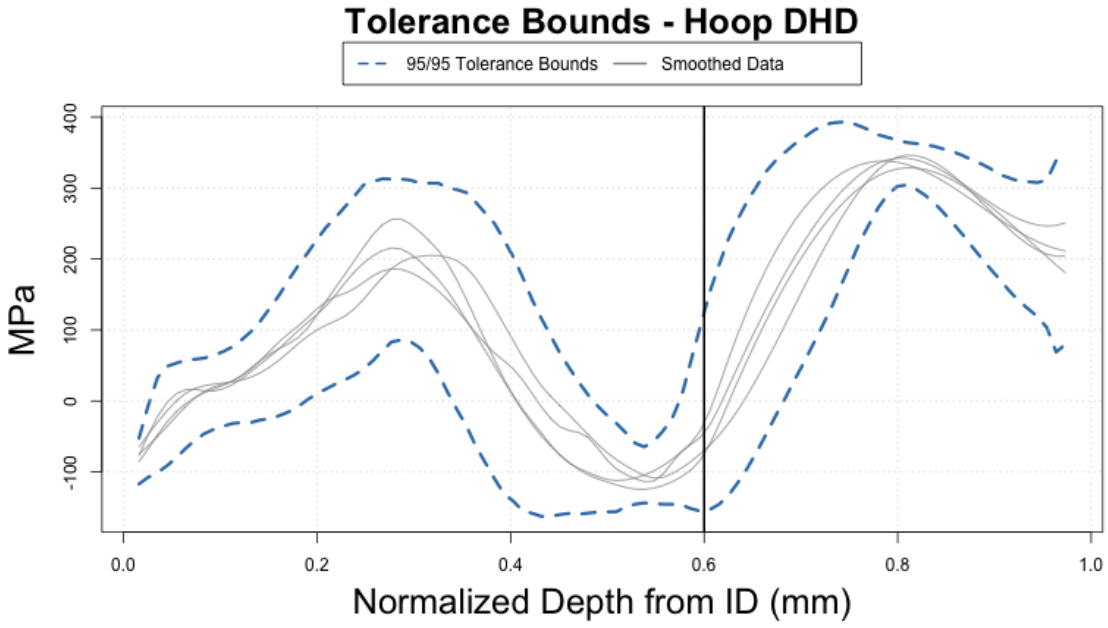


Figure 19: 95/95 tolerance bound on the DHD hoop measurements.

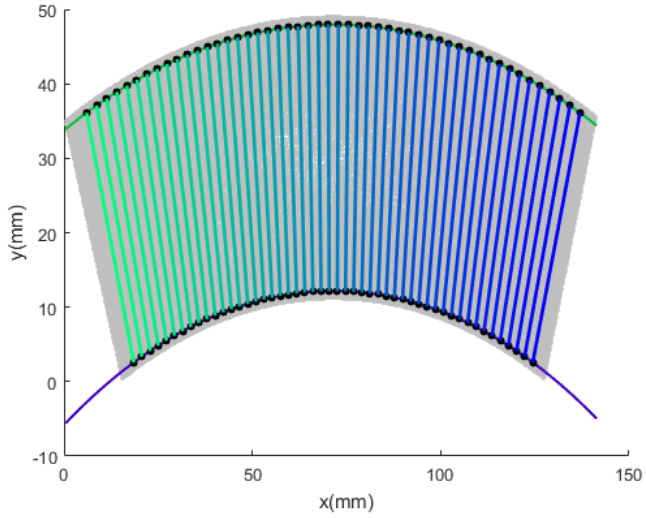


Figure 20: The FE data from the contour method with 50 through-weld paths, perpendicular to the ID boundary, on which WRS profiles can be extracted.

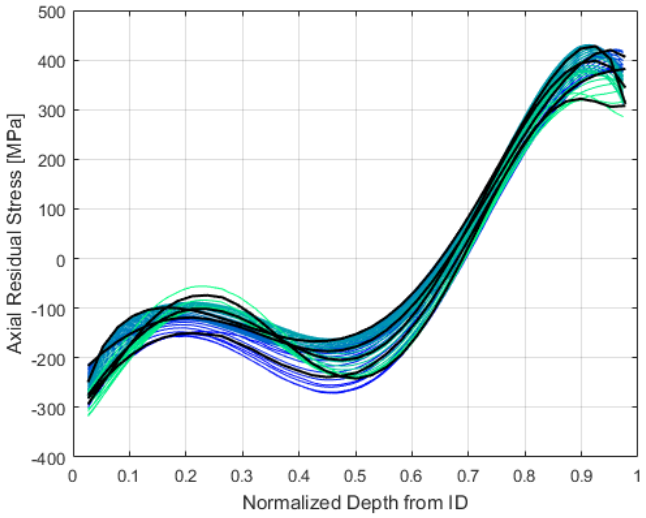


Figure 21: The axial WRS data from the contour method extracted on 50 through-weld paths at different angular locations.

larger circle defines the outer diameter of the weld. Data extraction was performed using Matlab. Three data points were chosen on the ID boundary and applied in a system of three equations to find the center and the radius of the ID boundary. Then, the circle defining the OD boundary was determined by adding the weld depth (37.8 mm) to the radius of the circle defining the ID boundary. Alternatively, one could choose three points along the OD boundary and solve a system of equations to find the center and radius of the circle defining the OD boundary as was done for the ID boundary. When this method is used, the centers of the two circles are slightly offset. This is likely due to the discrete nature of the data and does not substantively affect the results of data extraction. Once both circles were defined, the radius of the ID boundary was increased by 1 mm and the radius of the OD boundary was decreased by 1 mm. This adjustment was made so that WRS profiles would be consistent with those provided by the measurement vendor, beginning and ending within 1 mm of either surface. A sequence of evenly spaced points was chosen on the new ID boundary and line segments were defined perpendicular to the new ID boundary from each point on the new ID boundary to the points of intersection with the new OD boundary. Example line segment paths are plotted in Figure 20. Note that the segments are perpendicular to the ID boundary; they do not appear perpendicular due to scale.

Once the paths for extraction were chosen, an interpolation function was created using the **scatteredInterpolant** function in Matlab with the 'linear' method option. This function was applied to evenly spaced points along each path to calculate the WRS profile along each path. 500 paths were extracted using this methodology. The extraction procedure was also completed with 5000 paths but this did not change the results. Increasing the number of paths beyond 500 only increases the density of the profiles and does not increase the spread of profiles. The 500 paths are plotted in Figure 22 and the corresponding extracted profiles are plotted in Figure 23 with the five original profiles. Note that the original five profiles were normalized by the depth of the weld, 37.8 mm. Because the original data did not include data exactly at the ID or the OD, the domain of the normalized profiles does not include zero or one. The additional profiles were extracted on the same domain, excluding 1 mm near either surface.

The axial WRS profiles plotted in Figure 23 characterize the amount of uncertainty that can be expected in contour measurements of axial WRS. Away from the ID and the OD, the WRS measurements at any depth cover a range of depths between about 100 and 150 MPa wide. Near the ID, however, the measurements range from about -310 to -210 MPa and at the OD the measurements range from about 290 to 420 MPa.

It should be noted that the modeling and bootstrap methods described in Section 2 can be applied to either the 5 original profiles provided by the measurement vendor or to the 500 profiles extracted using the methods described above. For example, tolerance bounds in each case are displayed in Figure 24. Note that the tolerance bounds calculated using all 500 profiles are not as wide as those calculated using only 5 profiles. The 500 profiles provide more information about the distribution of profiles, providing less uncertainty on quantile estimates and hence narrower tolerance bounds.

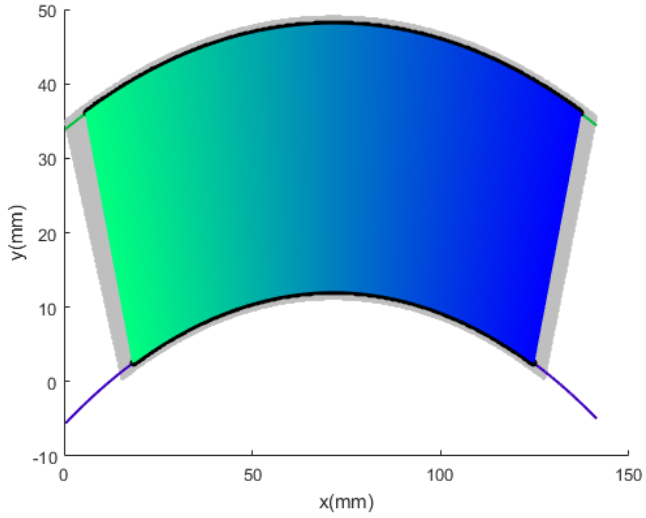


Figure 22: The finite element (FE) measurement data from the contour method with 500 through-weld paths on which WRS profiles can be extracted.

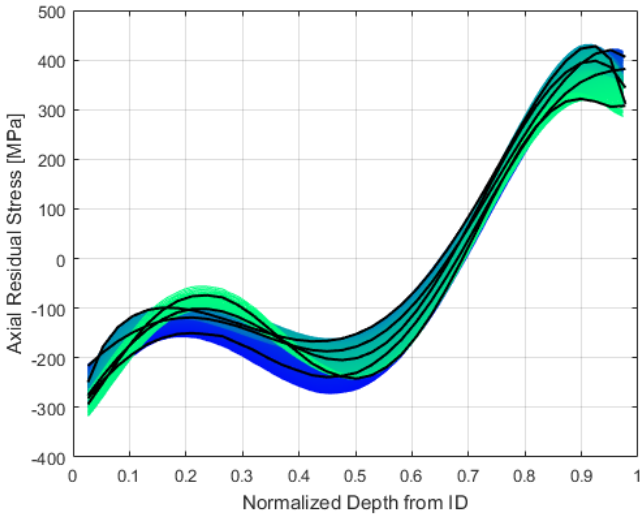


Figure 23: The axial WRS data from the contour method extracted on 500 through-weld paths at different angular locations.

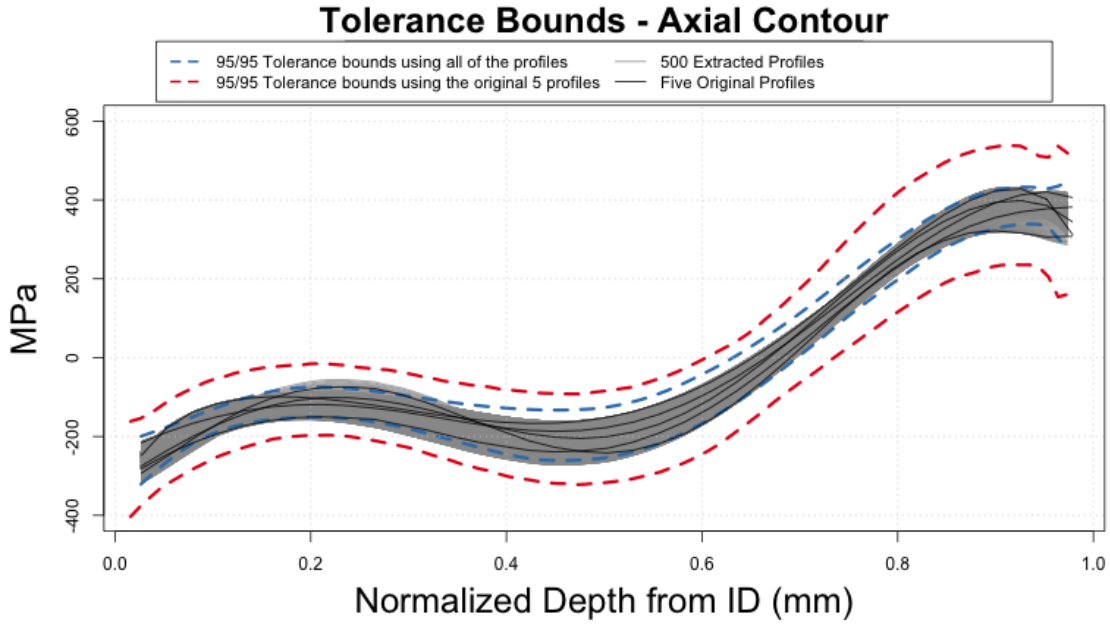


Figure 24: Comparison of tolerance bands constructed with either 500 or 5 axial contour profiles.

3.4 Limitations and Recommendations for Future Work

The DHD and contour methods for measuring WRS use machining to release strain and apply a model to changes in strain to calculate stresses. For the DHD method, this means that uncertainties in the initial measurements and those taken after releasing strain affect the overall uncertainty in WRS in addition to any model uncertainty in the equations applied to the strain to obtain WRS. The sources of uncertainty are, therefore, a combination of measurement errors introduced by the machining process, the precision and accuracy of all of the measurements tools, and the resolution of measurements through the thickness of the weld. Further sources of uncertainty include the validity of the elasticity assumption and the model form error of the equations used to calculate stresses from strains. With the information currently at hand, it is not possible to describe in greater detail which sources contribute the most uncertainty and the ability to quantify uncertainty is somewhat limited by the number of available measurements. The present analysis addresses the limited number of measurements by applying semi-parametric bootstrapping techniques, but the sources of uncertainty remain unexplained. That is, the sources of observed variation within and between the measured DHD profiles are completely confounded and rolled up into an overall total uncertainty.

Future studies of uncertainty in DHD measurements of WRS could expand this analysis to better understand how much uncertainty is due to measurement errors. This task would require the experimental measurements, any known information about the precision and accuracy of the measurement tools, and the WRS models employed. By varying the measurements reasonably based upon the magnitude of measurement errors and propagating those values through the models

using a Monte Carlo (MC) approach [12], one could determine the extent to which measurements of WRS vary solely due to errors in the experimental measurements of displacement which are then converted to strains. This variation could be compared to the between-curve variation (variation observed when DHD is performed at different locations along the weld) to help determine the magnitude of systematic WRS variation around the weld.

The same general process could similarly be applied to the contour method to better understand measurement uncertainty in WRS. Possible sources of uncertainty are essentially the same, though the use of a finite element (FE) model for the contour method introduces further computational complexity above and beyond the computational complexity of the equations used in the DHD models. The model is applied to the contour measurements to obtain strains first and then stresses from the strains. In addition to this final model, the contour WRS calculations include intermediate measurements that account for the release of strain that occurs after sequential cuts during the machining process. The current analysis combines all of these sources of uncertainty without attribution. That is, the total variation in the extracted profiles is considered and this variation is due to a combination of any of the uncertainties discussed here. Propagating uncertainty in the measurements through the finite element model and intermediate steps using a MC approach would help to determine the major source of this variation. For example, if the axisymmetric assumption holds, then the variation between the profiles can be attributed to measurement errors or lack of symmetry in the strain release during the machining process. If the measurement errors turn out to be relatively small as determined by the MC study, then attribution to the machining process is plausible. One approach to explore for conducting such an analysis is discussed by [13]. However, a possible hurdle of UQ through MC techniques in this case is the large computational time of the FE models. In such cases, pure MC approaches may not be feasible and careful consideration of the inputs to propagate through the models is necessary due to the limited number of possible runs. The use of surrogate models with associated uncertainty, such as a Gaussian process model [14], is one possible approach to take in this case. For more examples of possible surrogates see [15].

If the data and models are obtained to pursue this extension, probability distributions should be assigned for measurements based on available information about measurement error. To determine how this uncertainty affects WRS measurements, new values should be sampled from those distributions and applied to the models. The large population of resulting WRS profiles (either from direct MC or through the use of surrogates) would then characterize the uncertainty in WRS that results from uncertainty in the measurements.

4 Results for Uncertainty Quantification of the Predictions

The same methodology used for the DHD measurements is applied for uncertainty quantification of the predictions. Confidence bounds for the mean as well as tolerance bounds using the bootstrap method described in Section 2.6 are constructed. A confidence level of $1-\alpha = 0.95$ is again used. For the tolerance bounds, a coverage level of 0.95 (i.e. $q = 0.05$) is again used to construct $95/95$ tolerance bounds. Thus, the confidence bounds for the mean have the interpretation 'at each value of depth there is 95% confidence that the true mean lies within the specified bound.' The tolerance bounds have the interpretation 'at each value of depth there is 95% confidence that 95% of all values from the population will fall between the bounds.' For the predictions, the tolerance bounds provide a range where the majority (95%) of all predictions from a large group of different modelers are expected to fall.

A brief summary of the results in this section is as follows. The variation amongst the predictions is larger than amongst the repeated experimental measurements. Perhaps this is not surprising since the experimental measurements are taken using the same process (either DHD or contour) whereas the predictions were made by several different organizations using different modeling techniques. The constructed tolerance bounds indicate the plausible range of the majority of WRS axial predictions to be roughly 400-600 MPa wide near the ID. The plausible range of hoop predictions has a similar width. Further, one hoop prediction assuming isotropic hardening appears to be an outlier, contributing significantly to the variation between the predictions.

The first step in modeling the variation in the group of functions is to apply smoothing. For this application, the smoothing method described in Section 7.2 is applied. Smoothing induces a level of uncertainty due to the difference between the raw predictions and the smoothed data. The residuals for each smoothed data set are analyzed to qualitatively assess how much difference there is in smoothed and raw data. Since the models described in Section 2 are applied, the uncertainty due to the smoothing is not accounted for. However, if the residuals are small compared to the overall variability between the functions, this level of uncertainty is of second order. Also, the smoothed functions are assumed to be good representations of WRS which reduce the effects of variability not caused by WRS variations. For example, the non-smoothness in prediction data is caused by discretization error in the models. Assuming each model would produce a smooth WRS profile if observed continuously along the weld depth, smoothing the predictions is justified.

For each of the following UQ sections, the smoothing of the original data is first assessed and then confidence bounds for the mean and tolerance bounds for 95% of the population are provided. Together these provide a comprehensive UQ assessment for the prediction data.

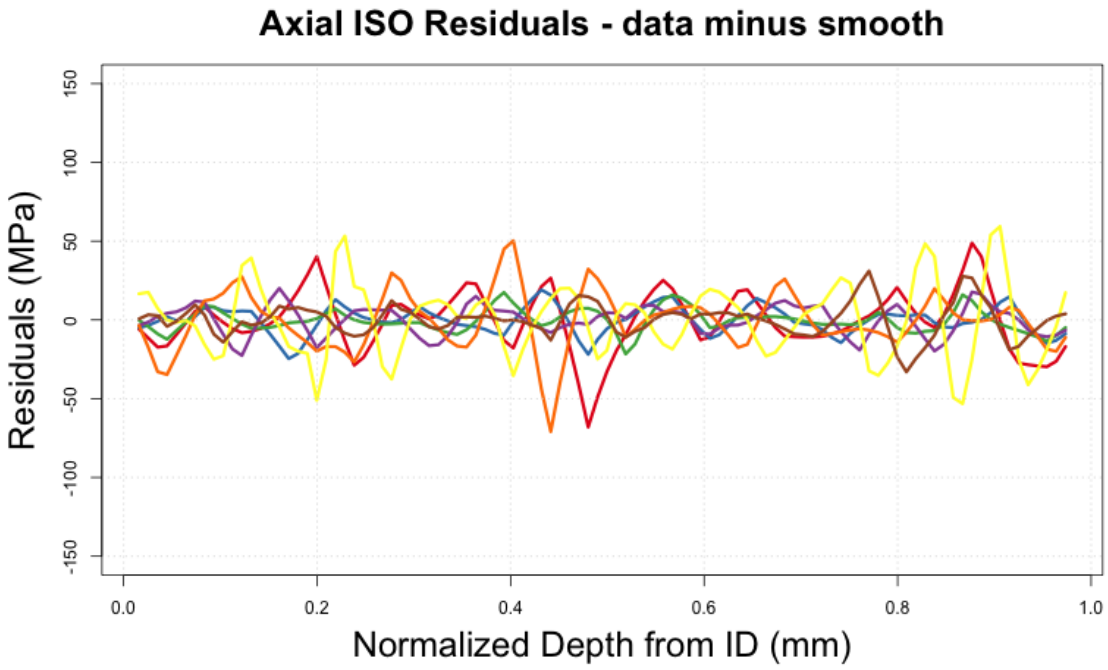


Figure 25: Residuals of the smooth profiles for the predictions in the axial direction assuming isotropic hardening.

4.1 Uncertainty Quantification of Axial Predictions with ISO Hardening

Smoothing

Raw and smoothed data appear in Figures 1 and 2. The residuals of the smooth profiles appear in Figure 25. Note that the range of these WRS predictions is over about 900 MPa and the residuals have a maximum absolute value just over 60 (roughly 7% of the range). Compared to the scale of the data, the residuals introduce nominal uncertainty.

Confidence Bounds

Confidence bounds and tolerance bounds for this group of predictions appear in Figures 10 and 12. The confidence bounds indicate approximately 100 MPa of uncertainty around the mean at the ID and through the depth of the weld. The tolerance bounds show a wider range of values bounding the predictions near the ID than at any other location in the weld. They indicate that a 600 MPa range is required to have 95% confidence that 95% of ISO axial predictions at the ID will occur within the bounds.

4.2 Uncertainty Quantification of Axial Predictions with KIN Hardening

Smoothing

Figures 26a, 26b, and 26c show the axial KIN predictions, the smoothed profiles, and the residuals of the smoothed profiles respectively. The residuals over the entire domain introduce a negligible level of uncertainty; predictions range from -200 to about 275 MPa and the maximum absolute value of the residuals is around 30 MPa.

Confidence Bounds

Confidence bounds on the mean axial KIN prediction are plotted in Figure 27. It is expected, with 95% confidence, that the true mean profile exists within the range determined by the confidence bounds. The bounds indicate a range of uncertainty for the mean of between about 50 and 100 MPa at any depth through the weld. The 95/95 tolerance bounds are plotted with the predictions in Figure 28. At the ID, there is 95% confidence that the predicted WRS will fall within a 350 MPa range 95% of the time.

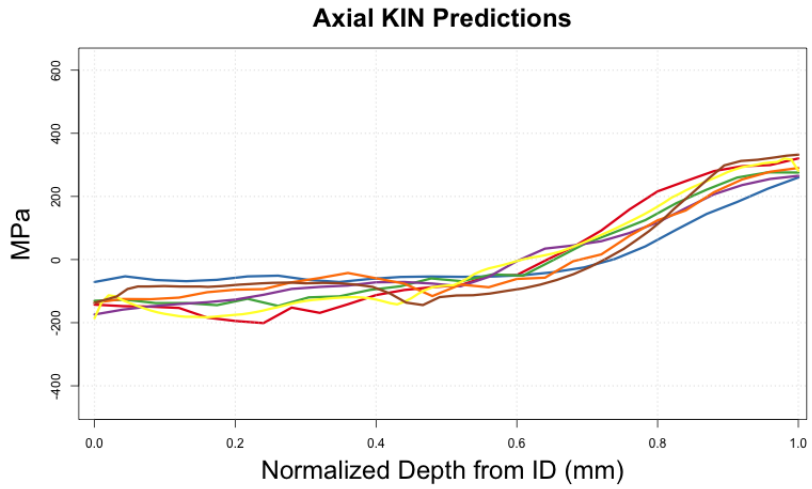
4.3 Uncertainty Quantification of Axial Predictions with AVE Hardening

Smoothing

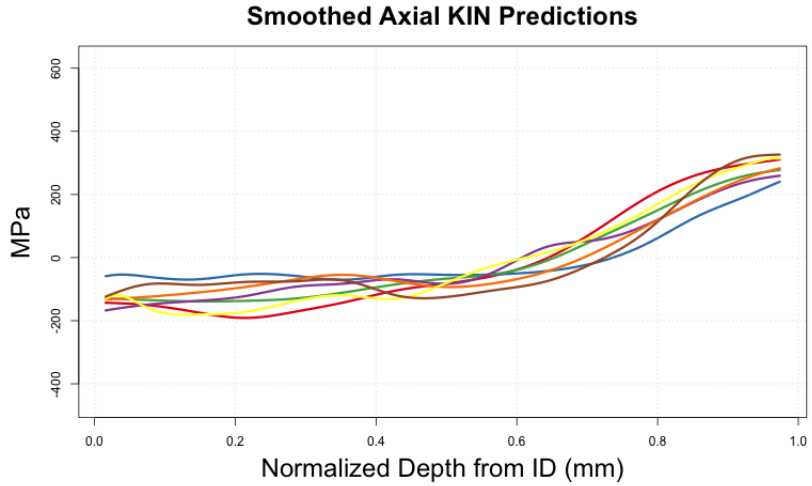
Figures 29a, 29b, and 29c show the averages of the KIN and ISO axial WRS predictions, the smoothed average predictions, and the residuals for the smoothed profiles. Predictions range from about -250 MPa and about 400 MPa and the absolute values of the residuals are bounded by about 30 MPa; smoothing introduces a negligible level of uncertainty for these predictions.

Confidence Bounds

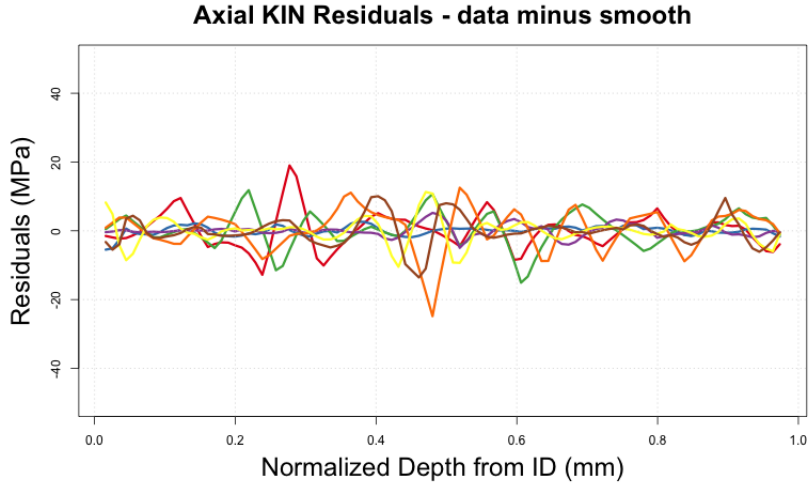
Figure 30 shows the averages of the KIN and ISO predictions and 95% confidence bounds on the mean prediction of axial WRS assuming average material hardening. The bounds are approximately 75 to 100 MPa in width through the depth of the weld. In particular, at the ID there is 95% confidence that the mean AVE axial WRS prediction will vary within a 100 MPa range. The 95/95 tolerance bounds are plotted with the smoothed data in Figure 31. Tolerance bounds indicate a range of approximately 400 MPa is required through most of the depth of the weld to contain 95% of AVE axial predictions with 95% confidence.



(a) Raw predictions in the axial direction assuming kinematic hardening.



(b) Smoothed predictions in the axial direction assuming kinematic hardening.



(c) Residuals of the smooth profiles for the predictions in the axial direction assuming kinematic hardening.

Figure 26: Smoothing results for axial predictions assuming kinematic hardening.

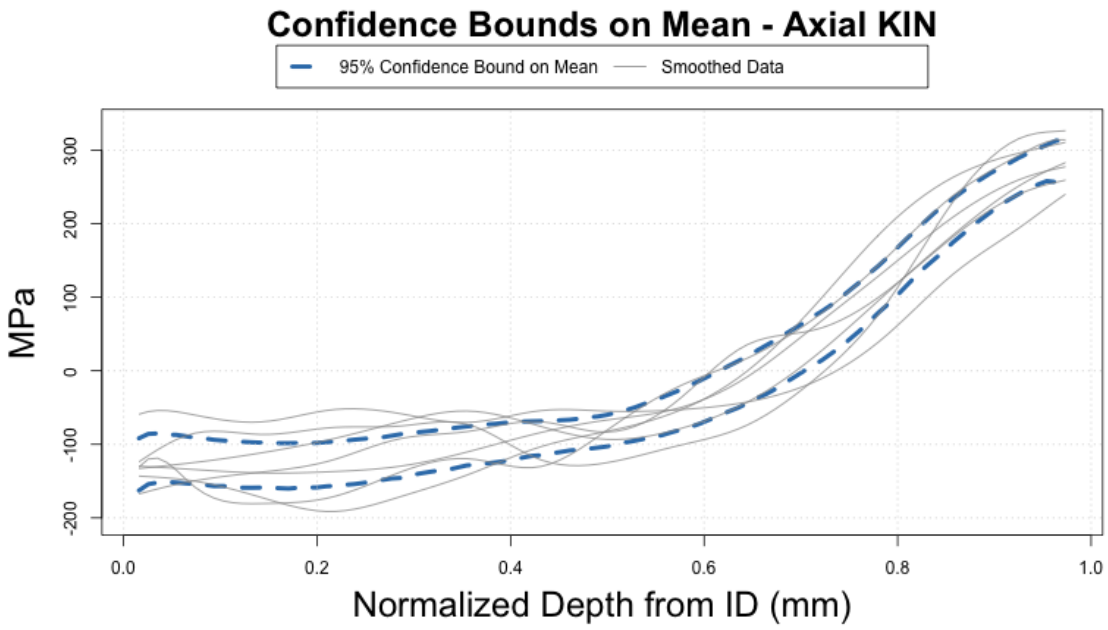


Figure 27: Pointwise 95% bootstrap confidence bounds for the mean prediction of WRS in the axial direction with kinematic hardening (thick dashed blue lines). These are computed using the pointwise 0.025 and 0.975 quantiles of the collection of all $S = 1000$ bootstrap means. The gray lines are the original (smoothed) data.

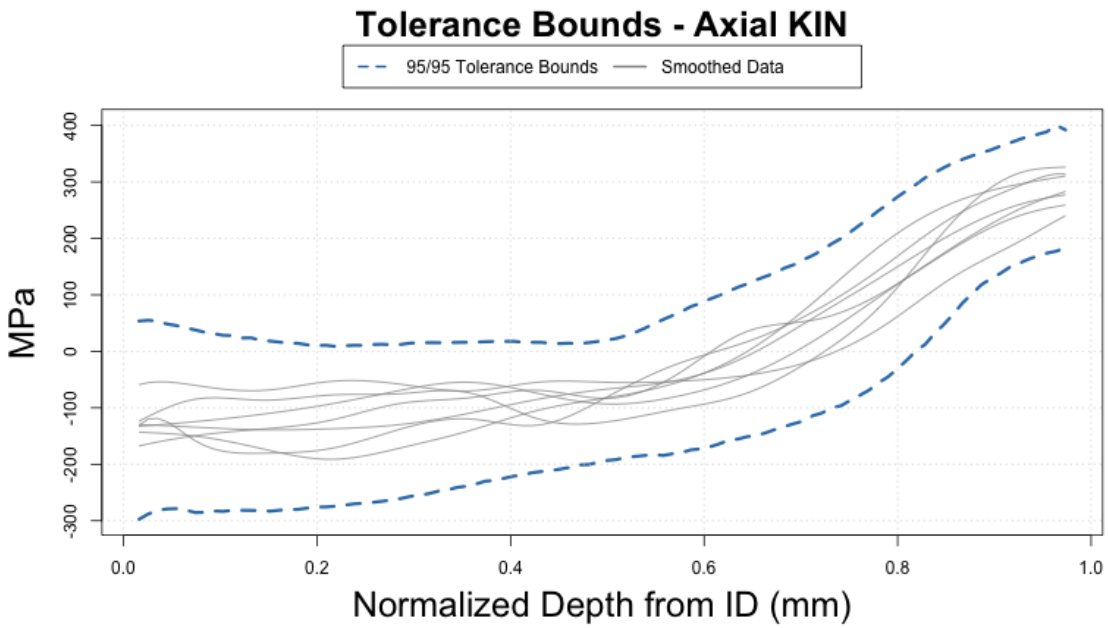
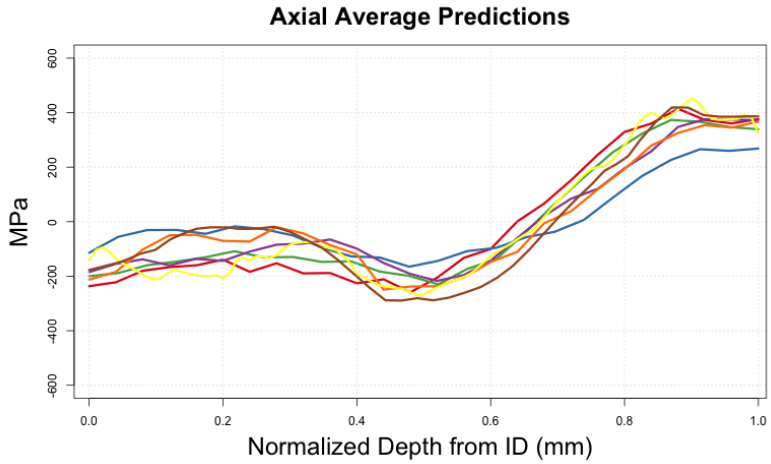
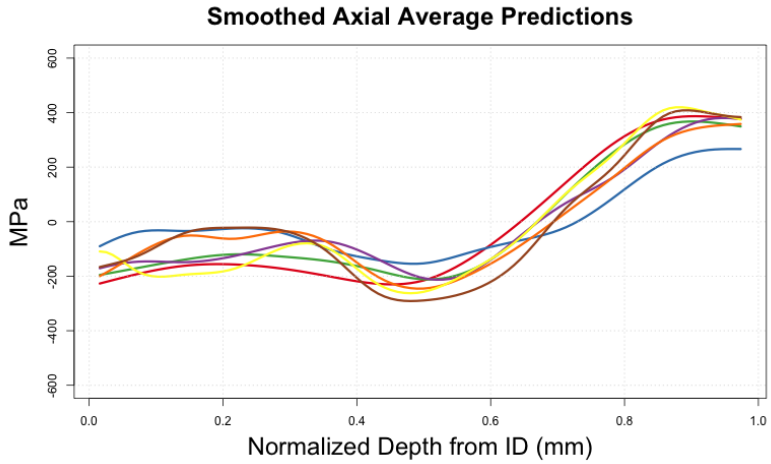


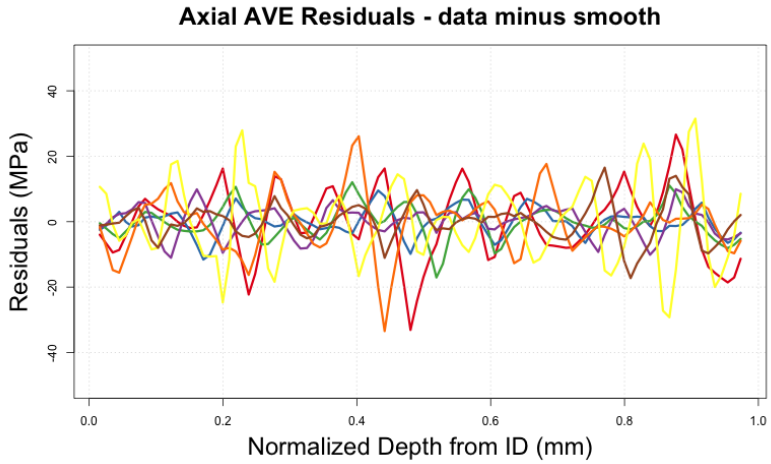
Figure 28: Pointwise $95/95$ tolerance bounds for the WRS predictions in the axial direction with kinematic hardening (thick dashed blue lines). These are computed using the pointwise 0.025 and 0.975 quantiles of the collection of all $S = 1000$ bootstrap quantiles. The gray lines are the original (smoothed) data.



(a) Raw predictions in the axial direction assuming an average of kinematic and isotropic hardening.



(b) Smoothed predictions in the axial direction assuming an average of kinematic and isotropic hardening.



(c) Residuals of the smooth profiles for the predictions in the axial direction assuming an average of kinematic and isotropic hardening.

Figure 29: Smoothing results for axial predictions assuming an average of kinematic and isotropic hardening.

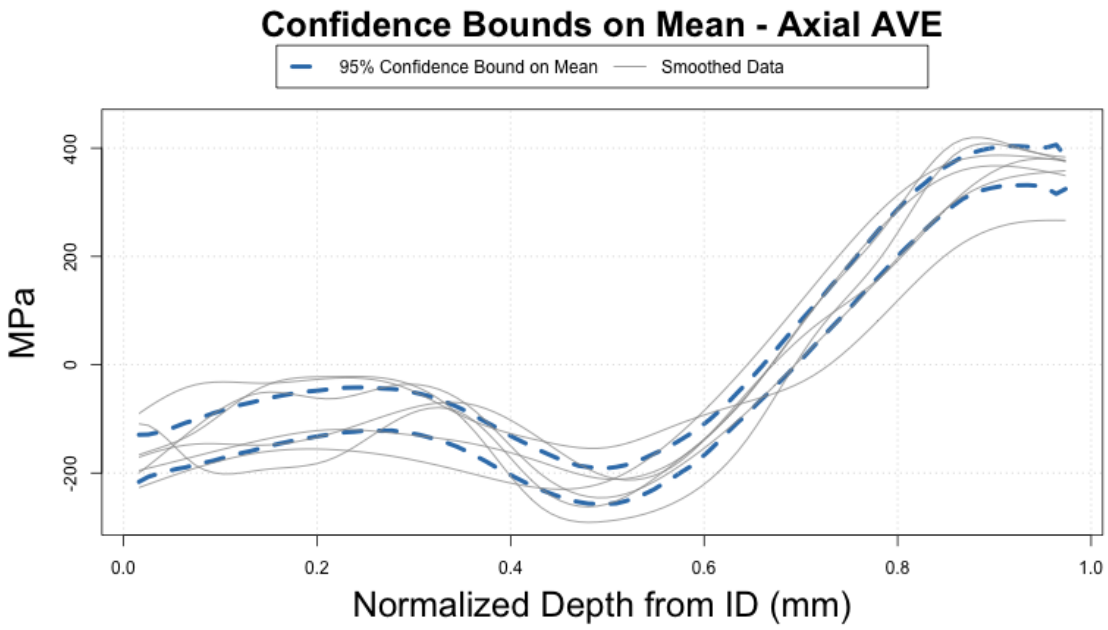


Figure 30: Pointwise 95% bootstrap confidence bounds for the mean prediction of WRS in the axial direction using an average of the two hardening laws (thick dashed blue lines). These are computed using the pointwise 0.025 and 0.975 quantiles of the collection of all $S = 1000$ bootstrap means. The gray lines are the original (smoothed) data.

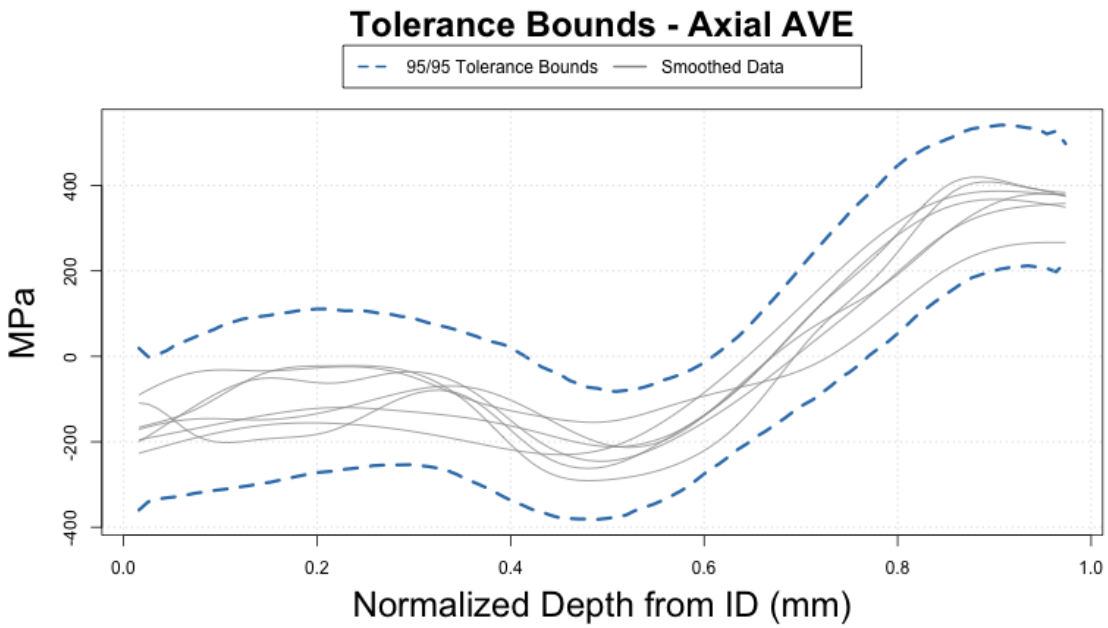


Figure 31: Pointwise $95/95$ tolerance bounds for the WRS predictions in the axial direction using an average of the two hardening laws (thick dashed blue lines). These are computed using the pointwise 0.025 and 0.975 quantiles of the collection of all $S = 1000$ bootstrap quantiles. The gray lines are the original (smoothed) data.

4.4 Uncertainty Quantification of Hoop Predictions with ISO Hardening

Smoothing

Figures 32a, 32b, and 32c show the original predictions, smoothed profiles, and residuals for predictions of hoop WRS assuming isotropic hardening. The predictions cover a range about of 700 MPa and the absolute values of the residuals are bounded by about 50 MPa through the majority of the weld with the exception of one prediction. There is one region around a relative ID depth of $d = 0.2$ where the residuals may introduce a more significant level of uncertainty. There, the absolute value of the residuals is about 16% of the range of predictions. The dramatic oscillating behavior of the yellow prediction around that depth may account for the higher residuals here. Without the yellow prediction, the absolute value of the residuals is about 6% of the range of predictions.

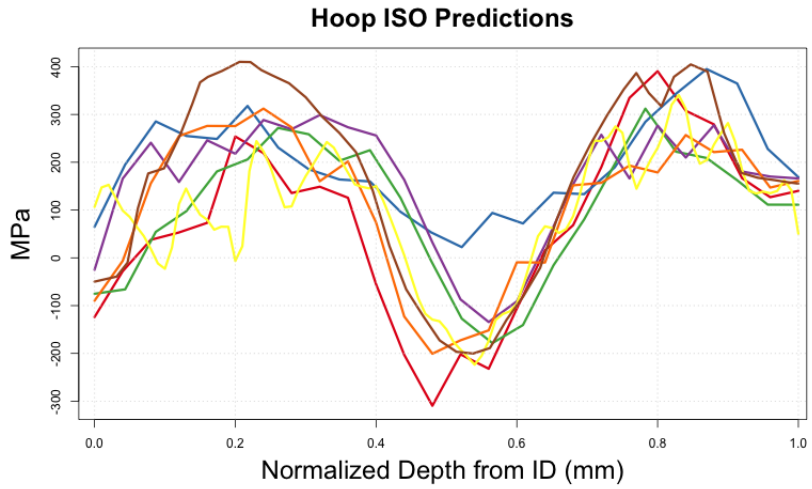
Confidence Bounds

The hoop predictions assuming isotropic hardening are plotted with 95% confidence bounds in Figure 33. The bounds suggest approximately 200 to 250 MPa of uncertainty around the mean at any given depth. In particular, the mean prediction at the ID is expected to fall within a range of about 250 MPa. Tolerance bounds are plotted with the smoothed data in Figure 34. In order to capture 95% of predictions with 95% confidence, the tolerance bounds must be greater than about 1000 MPa in width at every depth. It should be noted that a single prediction contributes significantly to this wide range. Removing the blue prediction, which appears to have outlying behavior between $d = 0.4$ and $d = 0.6$, results in the tolerance bounds in Figure 35 which have a much smaller range between roughly 200 and 600 MPa throughout the weld. This is an indication that the present approach for constructing statistical intervals is sensitive to outlying observations. This does not mean that the blue prediction should necessarily be excluded from the analysis. Its behavior may be indicative of the differences that can occur between predictions that are based on the same guidelines or there may be an aspect of the modeling that is not consistent with the other predictions. If the former is true, the blue prediction provides valuable insight about uncertainty between predictions. If the latter is true, it may be inappropriate to include the blue prediction with the others. Note also that the raw plot of the yellow prediction in 32a shows oscillating behavior that is not expected in a WRS profile. Further study of the modeling might yield insights into whether that prediction should be treated as an outlier or not.

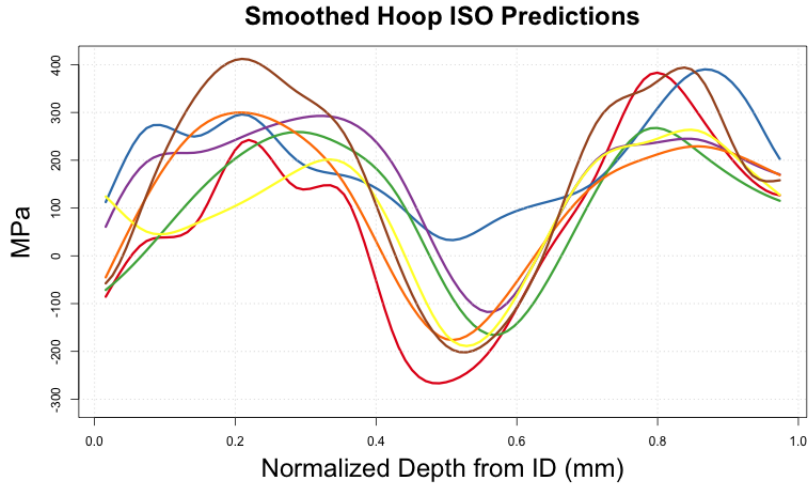
4.5 Uncertainty Quantification of Hoop Predictions with KIN Hardening

Smoothing

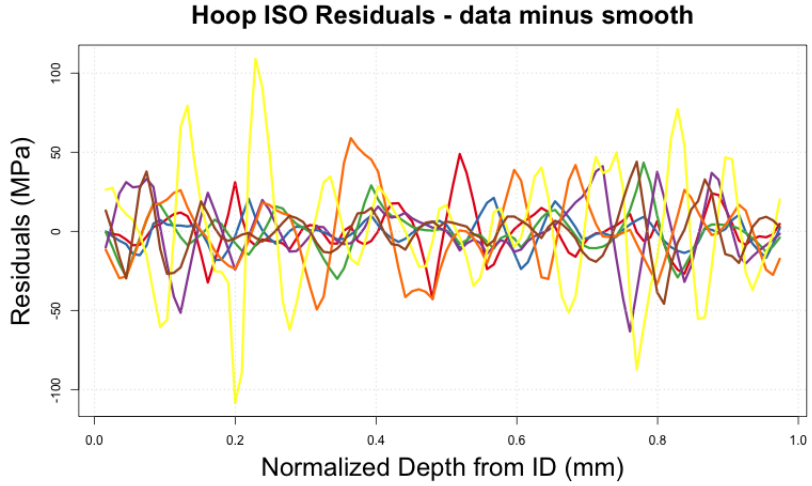
Figures 36a, 36b, and 36c show the raw predictions of hoop WRS assuming kinematic hardening, the smoothed profiles, and the residuals of the smoothed profiles respectively. Note that the pre-



(a) Raw predictions in the hoop direction assuming isotropic hardening.



(b) Smoothed predictions in the hoop direction assuming isotropic hardening.



(c) Residuals of the smooth profiles for the predictions in the hoop direction assuming isotropic hardening.

Figure 32: Smoothing results for hoop predictions assuming isotropic hardening.

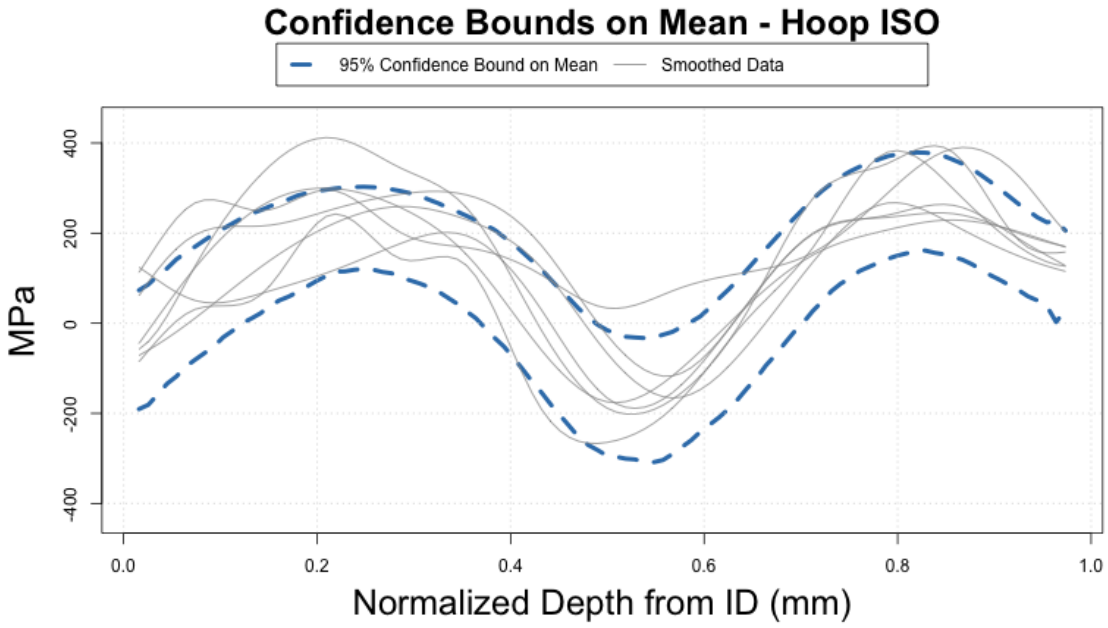


Figure 33: Pointwise 95% bootstrap confidence bounds for the mean prediction of WRS in the hoop direction using isotropic hardening (thick dashed blue lines). These are computed using the pointwise 0.025 and 0.975 quantiles of the collection of all $S = 1000$ bootstrap means. The gray lines are the original (smoothed) data.

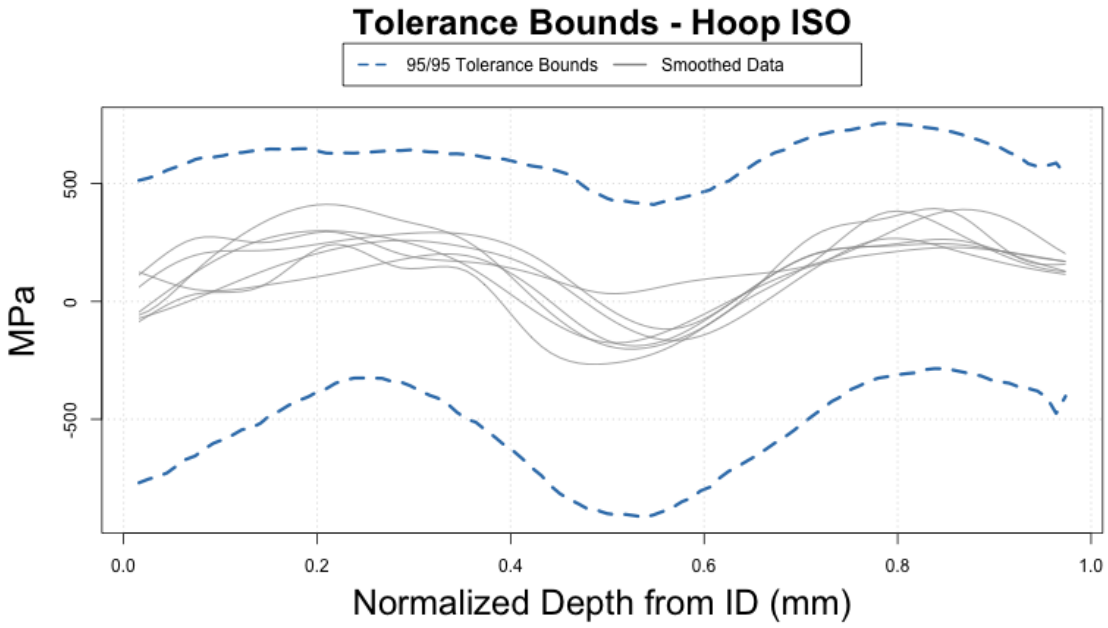


Figure 34: Pointwise 95/95 tolerance bounds for the WRS predictions in the hoop direction using using isotropic hardening (thick dashed blue lines). These are computed using the pointwise 0.025 and 0.975 quantiles of the collection of all $S = 1000$ bootstrap quantiles. The gray lines are the original (smoothed) data.

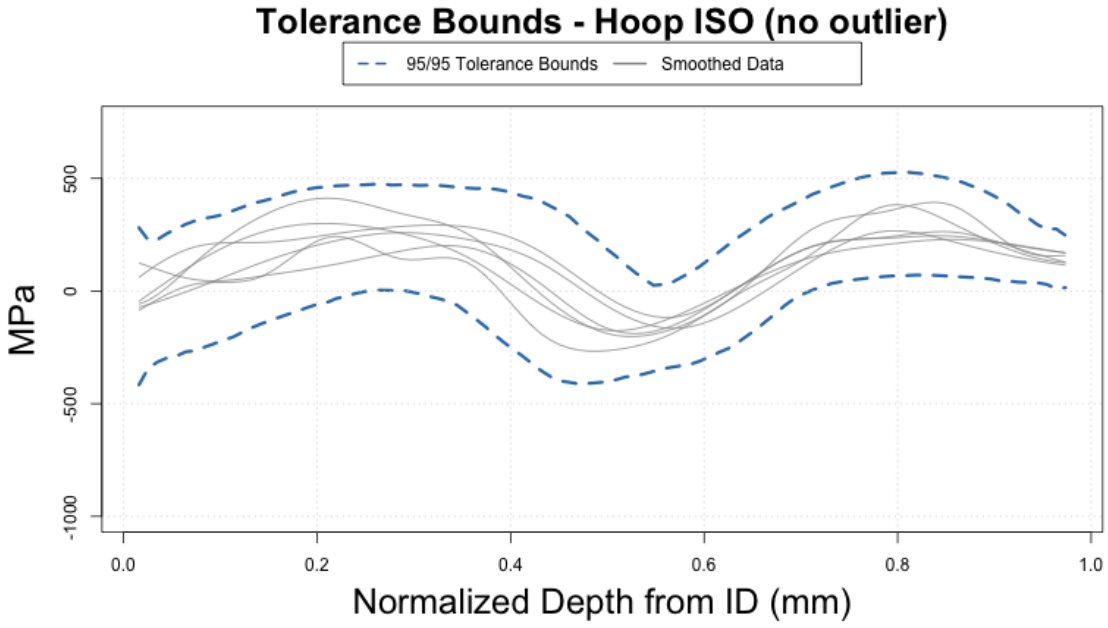


Figure 35: Pointwise 95/95 tolerance bounds for the WRS predictions in the hoop direction using using isotropic hardening (thick dashed blue lines) after removing the single outlier.

dictions cover a range of about 350 MPa. With the exception of a few points where the absolute values of the residuals are bounded by 15 MPa, the absolute values of the residuals are bounded by about 10 MPa through the depth of the weld. Smoothing introduced a negligible amount of uncertainty.

Confidence Bounds

The confidence bounds on the mean prediction of hoop WRS assuming kinematic hardening are plotted with the smoothed data in Figure 37. Notice that the bounds indicate 50-100 MPa of uncertainty in the mean prediction. In particular, the mean prediction at the ID is expected to fall within a range of 100 MPa with 95% confidence. The 95/95 tolerance bounds are plotted with the smoothed data in Figure 38. Note that the tolerance bounds are significantly wider at the ID than at any other location through the weld. At the ID, a range wider than 450 MPa is required to have 95% confidence than 95% of predictions will fall within the range. In contrast, the tolerance bounds are about 250 MPa wide at a relative depth of about $d = 0.6$. Predictions appear to be more variable near the ID.

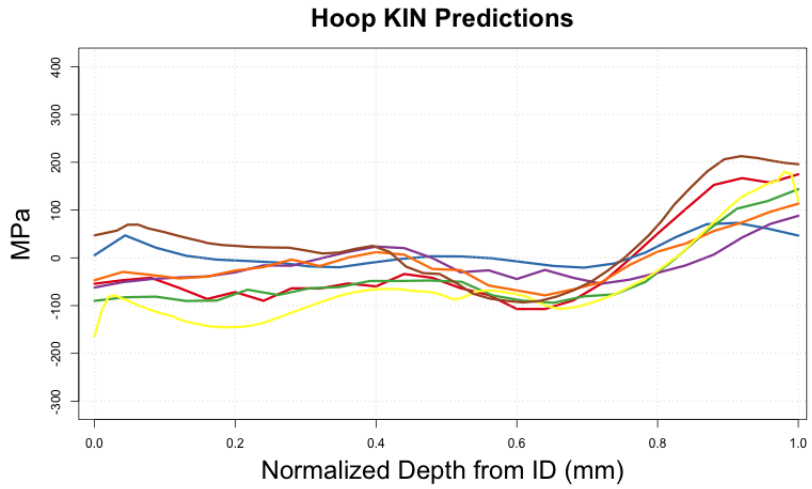
4.6 Uncertainty Quantification of Hoop Predictions with AVE Hardening

Smoothing

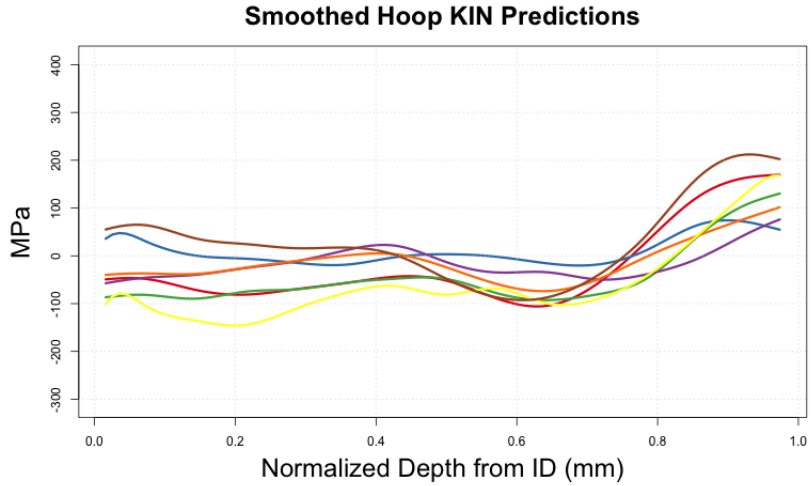
Figures 39a, 39b, and 39c show the raw hoop predictions assuming an average of kinematic and isotropic hardening, the smoothed profiles, and the residuals of the smoothed profiles respectively. Note that the WRS predictions cover a range of about 450 MPa and the absolute values of the residuals are bounded by about 35 MPa (except for one of the predictions). This indicates that smoothing introduces a negligible level of uncertainty. As with the ISO hoop predictions, the yellow prediction exhibits unexpected oscillating behavior. Study of the model that generated this prediction is suggested to determine whether its inclusion makes sense with the other models.

Confidence Bounds

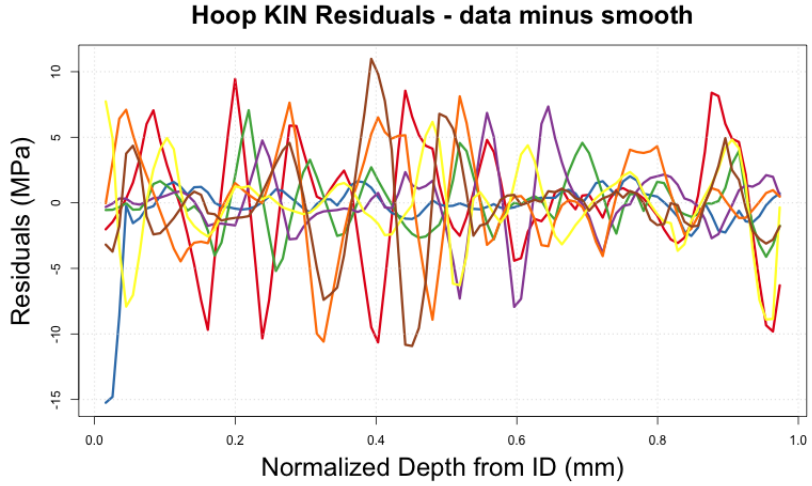
Confidence bounds on the mean prediction assuming an average of kinematic and isotropic hardening are plotted with the smoothed data in Figure 40. The 95% confidence bounds on the mean are less than approximately 200 MPa wide throughout the depth of the weld. At the ID, the mean is expected to fall within a range about 150 MPa wide. Figure 41 shows the smoothed data with 95/95 tolerance bounds. These bounds are quite wide and are again affected mostly by the one outlying prediction as was the case for the predictions assuming just isotropic hardening. Removing this prediction and redoing the analysis results in the tolerance bounds pictured in Figure 42. These have a much smaller range of less than 400 MPa through the weld. Whether it is appropriate to remove this prediction or not should be determined by reviewing the model



(a) Raw predictions in the hoop direction assuming kinematic hardening.



(b) Smoothed predictions in the hoop direction assuming kinematic hardening.



(c) Residuals of the smooth profiles for the predictions in the hoop direction assuming kinematic hardening.

Figure 36: Smoothing results for hoop predictions assuming kinematic hardening.

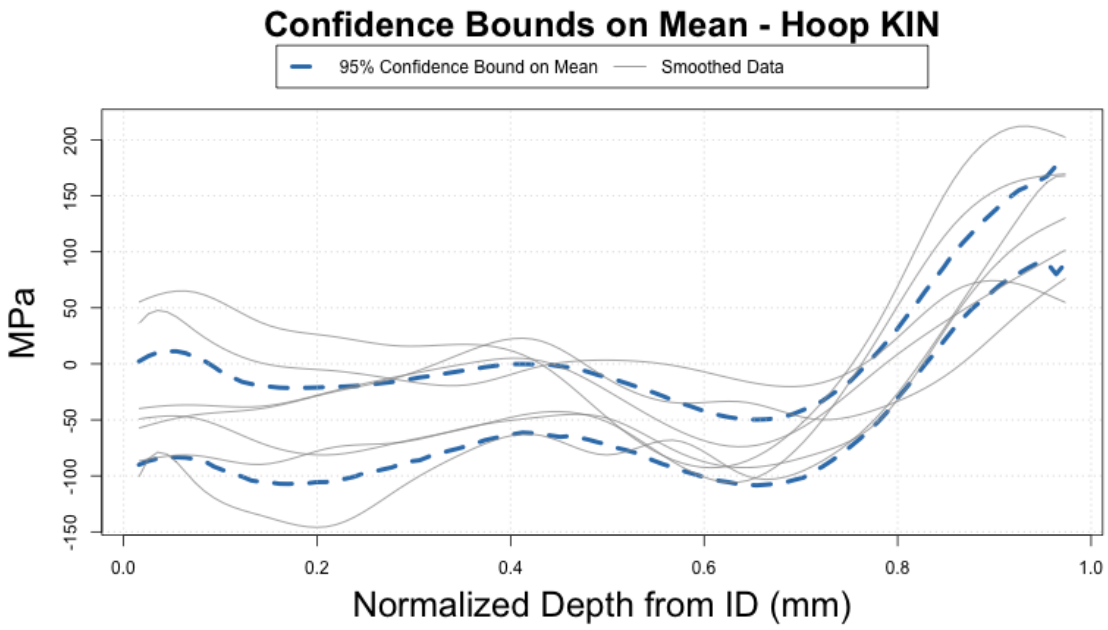


Figure 37: Pointwise 95% bootstrap confidence bounds for the mean prediction of WRS in the hoop direction using kinematic hardening (thick dashed blue lines). These are computed using the pointwise 0.025 and 0.975 quantiles of the collection of all $S = 1000$ bootstrap means. The gray lines are the original (smoothed) data.

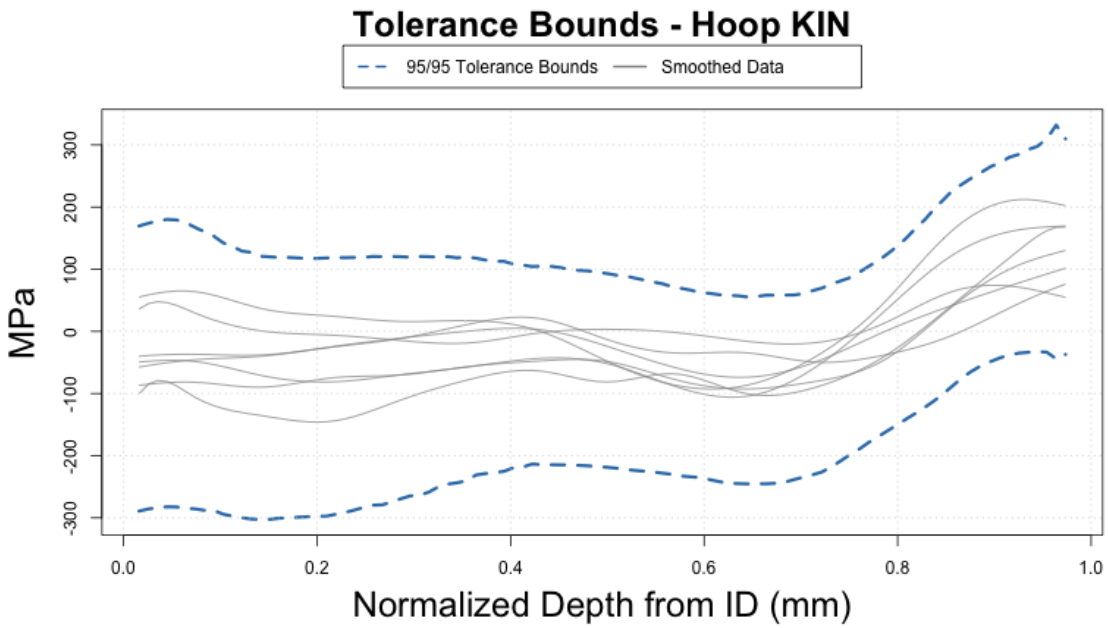
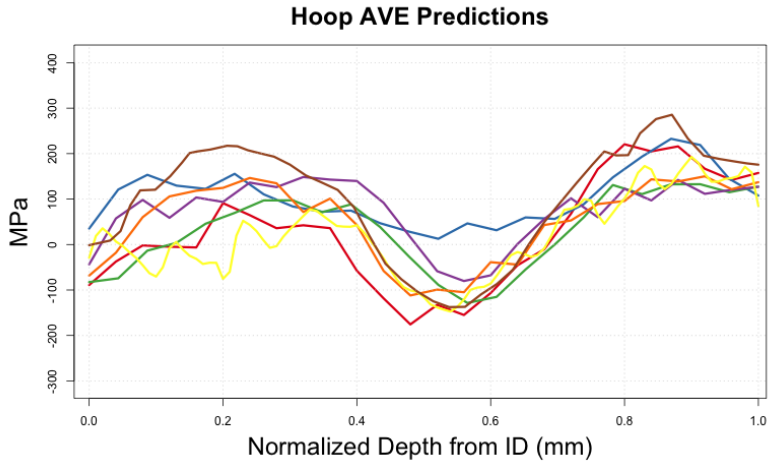
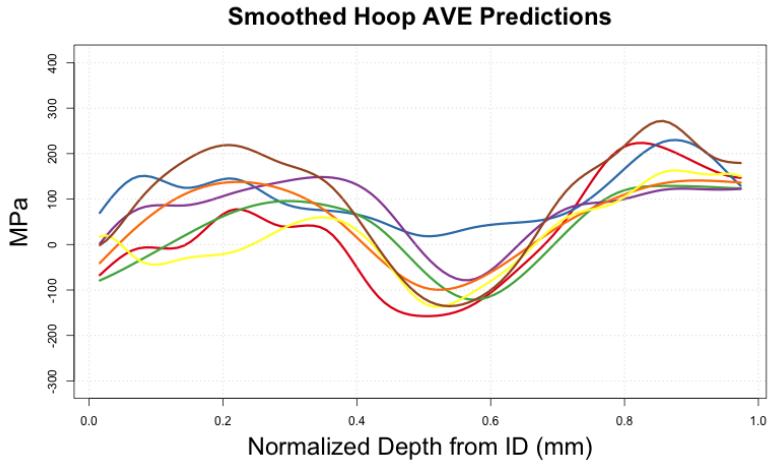


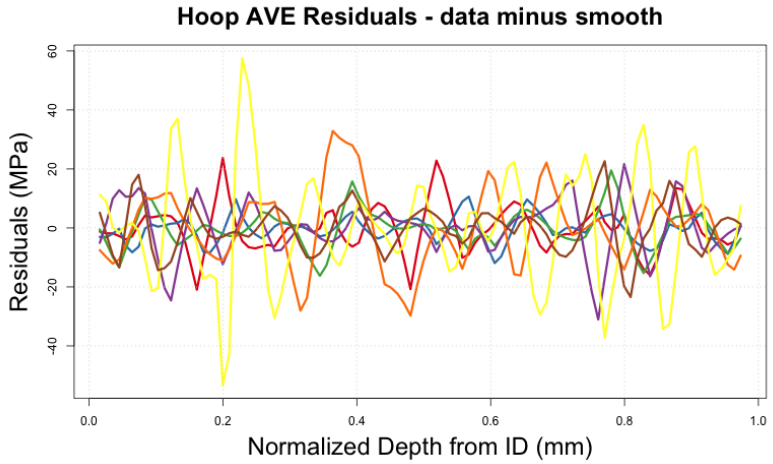
Figure 38: Pointwise 95/95 tolerance bounds for the WRS predictions in the hoop direction using using kinematic hardening (thick dashed blue lines). These are computed using the pointwise 0.025 and 0.975 quantiles of the collection of all $S = 1000$ bootstrap quantiles. The gray lines are the original (smoothed) data.



(a) Raw predictions in the hoop direction assuming an average of kinematic and isotropic hardening.



(b) Smoothed predictions in the hoop direction assuming an average of kinematic and isotropic hardening.



(c) Residuals of the smooth profiles for the predictions in the hoop direction assuming an average of kinematic and isotropic hardening.

Figure 39: Smoothing results for hoop predictions assuming an average of kinematic and isotropic hardening.

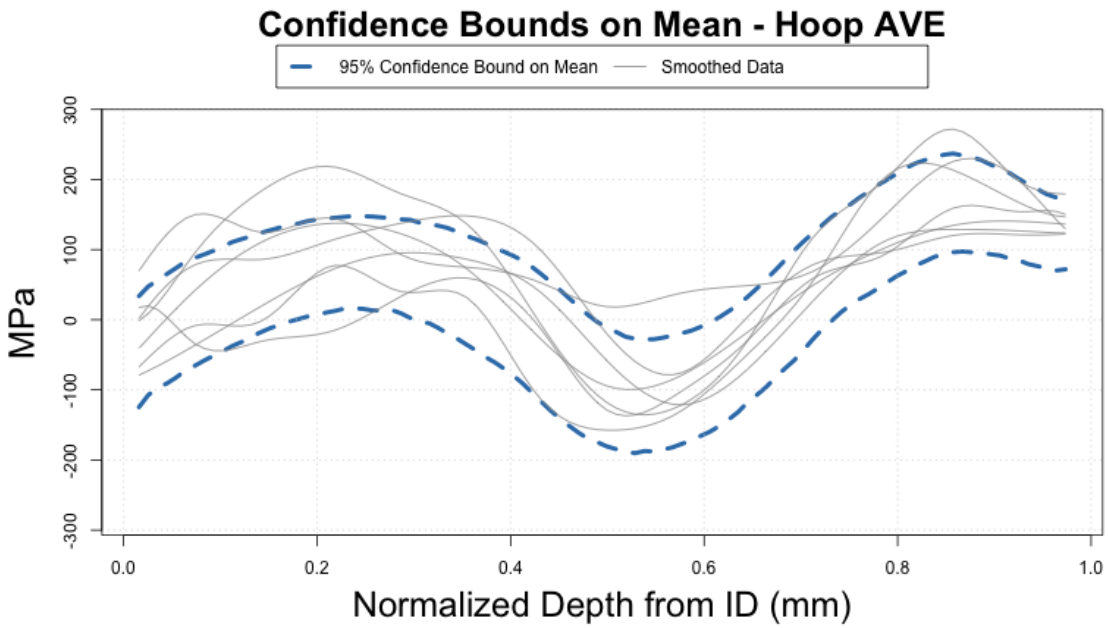


Figure 40: Pointwise 95% bootstrap confidence bounds for the mean prediction of WRS in the hoop direction using an average of the two hardening laws (thick dashed blue lines). These are computed using the pointwise 0.025 and 0.975 quantiles of the collection of all $S = 1000$ bootstrap means. The gray lines are the original (smoothed) data.

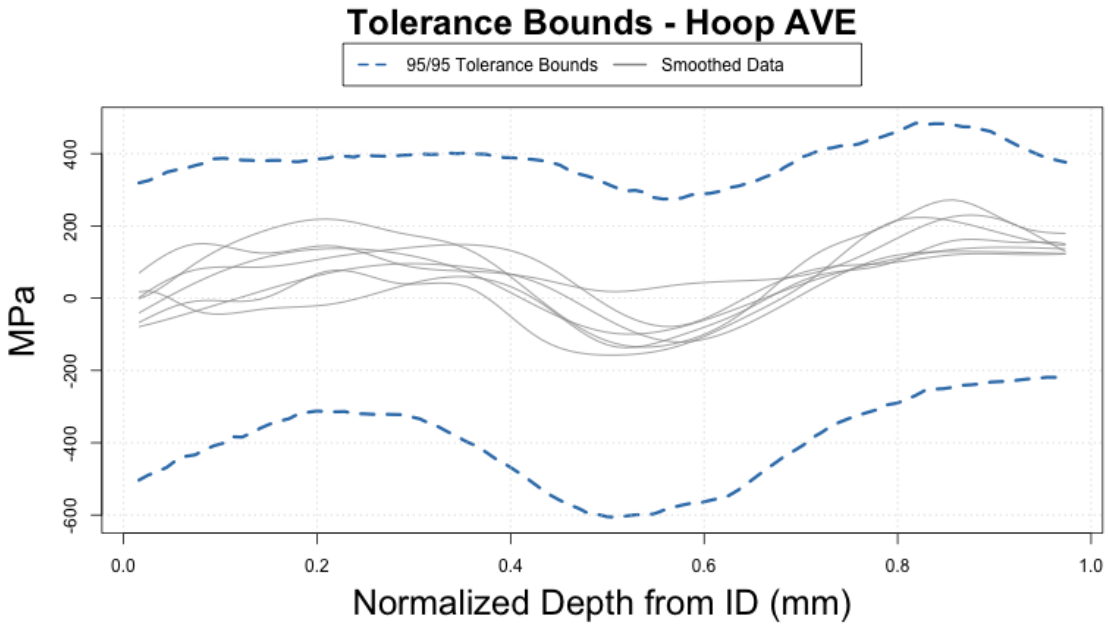


Figure 41: Pointwise 95/95 tolerance bounds for the WRS predictions in the hoop direction using an average of the two hardening laws (thick dashed blue lines). These are computed using the pointwise 0.025 and 0.975 quantiles of the collection of all $S = 1000$ bootstrap quantiles. The gray lines are the original (smoothed) data.

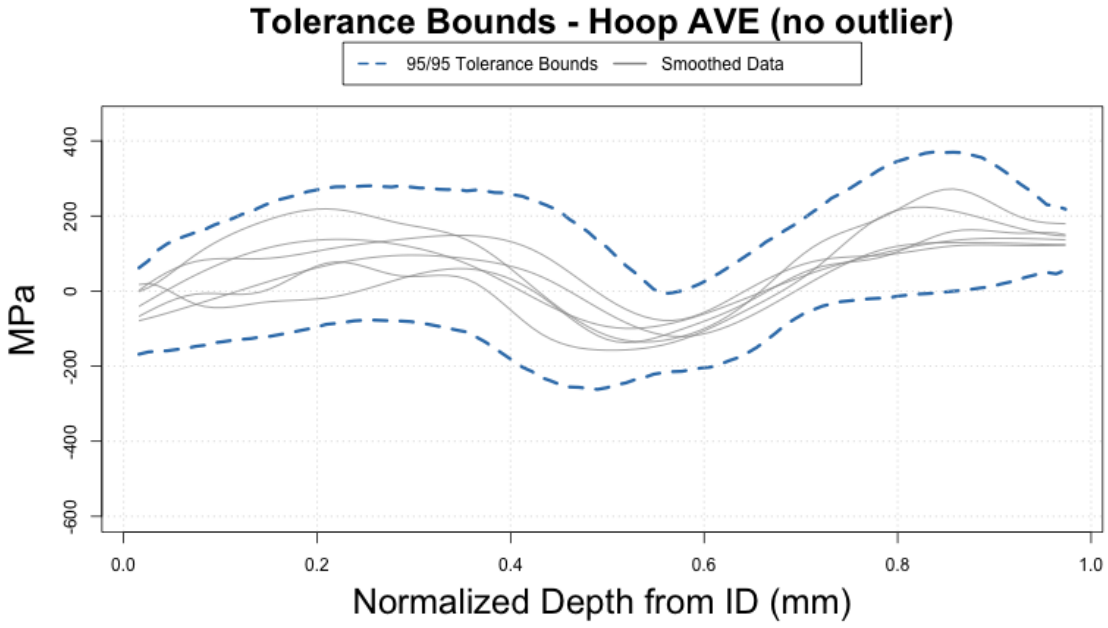


Figure 42: Pointwise 95/95 tolerance bounds for the WRS predictions in the hoop direction using an average of the two hardening laws (thick dashed blue lines) when removing the single outlier.

4.7 Limitations and Recommendations for Future Work

The modeling uncertainty described in this section only accounts for variation between individual modelers. Ideally, bounds should also account for uncertainty within each model. All of the participants constructed an FE model to predict WRS in the mockup and provided one WRS profile (for each direction) as their prediction. However, participants could provide their prediction with bounds according to common guidance. Alternatively, they could provide a population of predictions which represent uncertainty in that individual prediction due to various sources determined by the modeler. The bounds or population of predictions would represent the level of uncertainty in the prediction much like the collection of predictions from different modelers represents the uncertainty (only due to variation between modelers) in the population of predictions.

5 Comparison Results

The methods used to quantify uncertainty within predictions and within measurements set the foundation for comparing the two. The analyses in this section use the models already constructed to quantify uncertainty in Sections 3 and 4 to construct confidence bounds on the mean difference described in Section 2.6. The mean difference curves plotted in this section estimate the extent to which the mean of the predictions over- or under-estimates the mean of the measurements. Comparisons are performed by grouping based upon hardening law (KIN, ISO, or AVE), measurement method (DHD or contour) and direction of stress (axial or hoop). The confidence bounds in each analysis have the interpretation 'at each value of depth we are 95% confident that the true difference between the mean prediction and the mean measurement lies within the specified bound.' By using means, this analysis takes into account both measurement and prediction uncertainty in comparing measurements to predictions.

The methodology for comparing the contour hoop measurement to the predictions differs slightly in that there is not enough information to quantify uncertainty in the contour hoop measurement. Since there is only one contour hoop measurement, the predictions can be compared only to this single observation. Hence, this comparison includes modeling uncertainty but not experimental measurement uncertainty. For the comparison of contour axial predictions, all 500 extracted paths were used to form the generative model and the analysis of difference in means was performed as done with the DHD measurements.

A brief summary of the findings in this section is as follows. Near the ID, the average of the axial DHD measurements differs from the average prediction by roughly 100 MPa in absolute value depending on the hardening assumption chosen. The average of the hoop DHD measurements generally agrees with the hoop predictions near the ID. For the ISO and AVE assumptions, there is no statistical difference in the means at the ID, though given the uncertainty in the data, the true mean difference could be on the order of 150 MPa in absolute magnitude. For the KIN assumption, there is a statistical difference, though the KIN models perform better near the ID than at almost every other depth. The ISO predictions are not statistically different from the contour axial measurement near the ID, but they are statistically different from both the KIN and AVE predictions. The ISO, KIN, and AVE predictions are all statistically different from the contour hoop measurement near the ID; the mean predictions over-predict the contour hoop measurement by over 100 MPa in all three cases. There is a tendency for the average of the predictions assuming isotropic and kinematic hardening to differ from the mean experimental measurements less than the predictions using either one of the hardening laws separately.

The section concludes with a comparison of the DHD measurements and the contour measurements to each other using the same techniques applied to compare measurements to predictions.

5.1 Comparison of DHD Measurements and Predictions

Comparison of DHD Axial Measurements and Axial Predictions

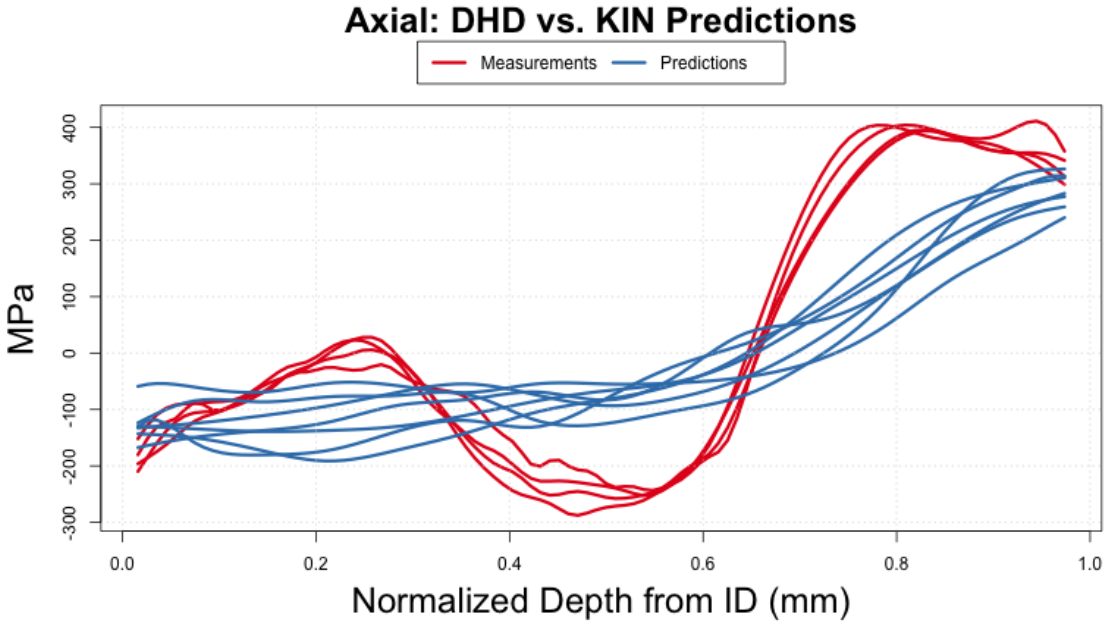
The DHD axial measurements (red) are plotted with predictions assuming kinematic hardening (blue) in Figure 43a. 95% confidence bounds on the difference between measurement and prediction means are plotted in Figure 43b. Note that the variation in the confidence bound indicates systematic differences between measurement and prediction, alternating between over and under-predicting, that can reach as high as about 200 MPa for normalized ID depth $d < 0.6$ (recall the smoothing of the DHD measurements has increased uncertainty for $d > 0.6$). However, the mean KIN axial predictions and mean DHD measurements show better agreement near the ID, with an average difference between about 40 and 120 MPa.

The axial predictions assuming isotropic hardening (blue) are plotted with the axial DHD measurements (red) in Figure 44a. The 95% confidence bounds on the difference between mean ISO axial predictions and the mean axial DHD measurements are plotted in Figure 44b. These bounds suggest that the isotropic hardening assumption tends to lead to under-predictions of the mean measurement through most of the weld depth. Except near the OD, where the smoothing introduces further uncertainty, the 95% confidence bounds suggest the difference between mean DHD measurements and mean ISO axial predictions is between approximately 0 and 200 MPa. For the first two-fifths of the weld from the ID, the expected difference in means is even smaller, bounded by about 100 MPa.

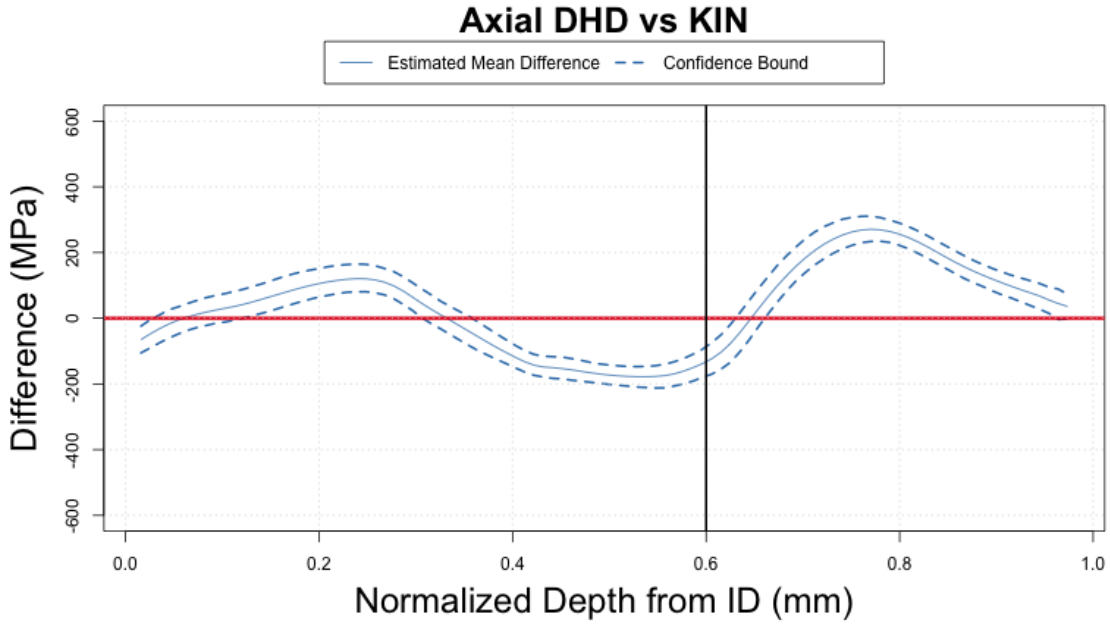
The axial predictions assuming an average of isotropic and kinematic hardening (blue) are plotted with the DHD axial measurements (red) in Figure 45a. The AVE predictions of axial WRS appear to agree with the DHD axial measurements fairly well, particularly at relative ID depths through the weld less than $d = 0.6$. This is supported by the confidence bounds in Figure 45b which show an average difference in the means being bounded by about 100 MPa with 95% confidence. The average prediction somewhat under-predicts the mean for the ID depths between about $d = 0.05$ to $d = 0.35$ where both the KIN and ISO predictions also under-predict. The over-prediction from the KIN assumption predictions from ID depths $d = 0.35$ to $d = 0.6$ dominates the under-prediction from the ISO assumption predictions, leading to a slight over-prediction on this region under the average hardening assumption. Again, caution should be taken in analyzing the mean differences beyond ID depths of $d = 0.6$ due to the uncertainty induced by smoothing sparser data.

Comparison of DHD Hoop Measurements and Hoop Predictions

The predictions of WRS in the hoop direction assuming kinematic hardening (blue) are plotted with the DHD measurements of hoop WRS (red) in Figure 46a. The estimated difference in mean KIN hoop predictions and DHD hoop measurements is plotted in Figure 46b with 95% confidence bounds. The bounds suggest systematic differences between measurement and prediction that can be as large as 300 MPa for ID depths less than 0.6. However, near the ID the difference

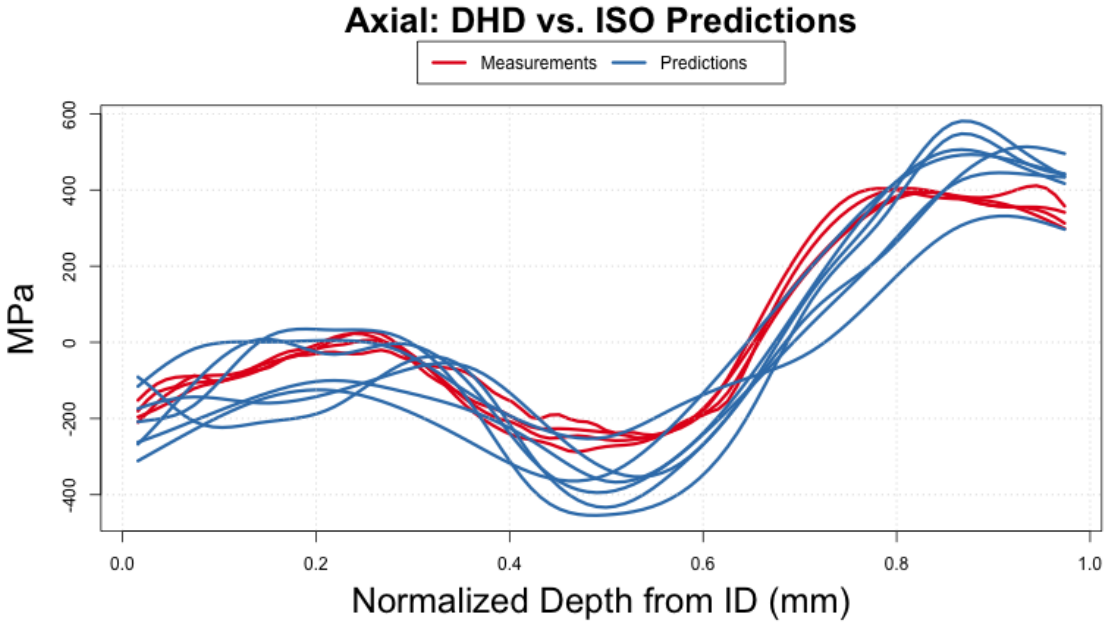


(a) Axial DHD measurements plotted with the axial predictions assuming kinematic hardening.

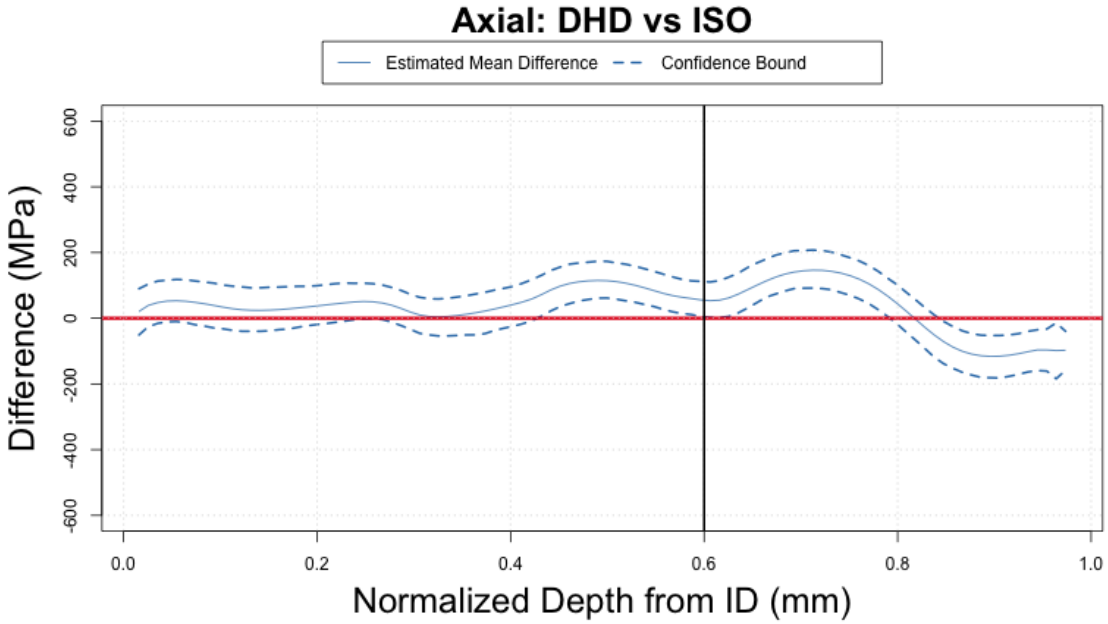


(b) 95% Confidence interval on the mean difference between the axial DHD measurements and the axial predictions assuming kinematic hardening. A vertical line at normalized ID depth $d = 0.6$ appears as a reminder that the smoothing error for the DHD measurements is large for $d > 0.6$

Figure 43: Comparison of axial DHD measurements to axial predictions assuming kinematic hardening.

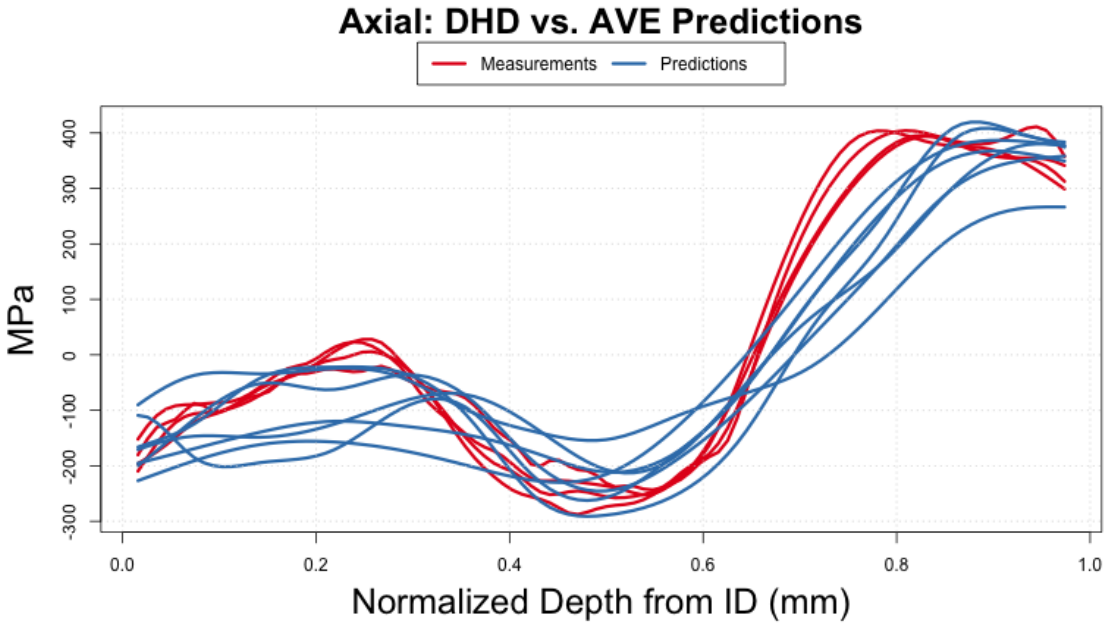


(a) Axial DHD measurements plotted with the axial predictions assuming isotropic hardening.

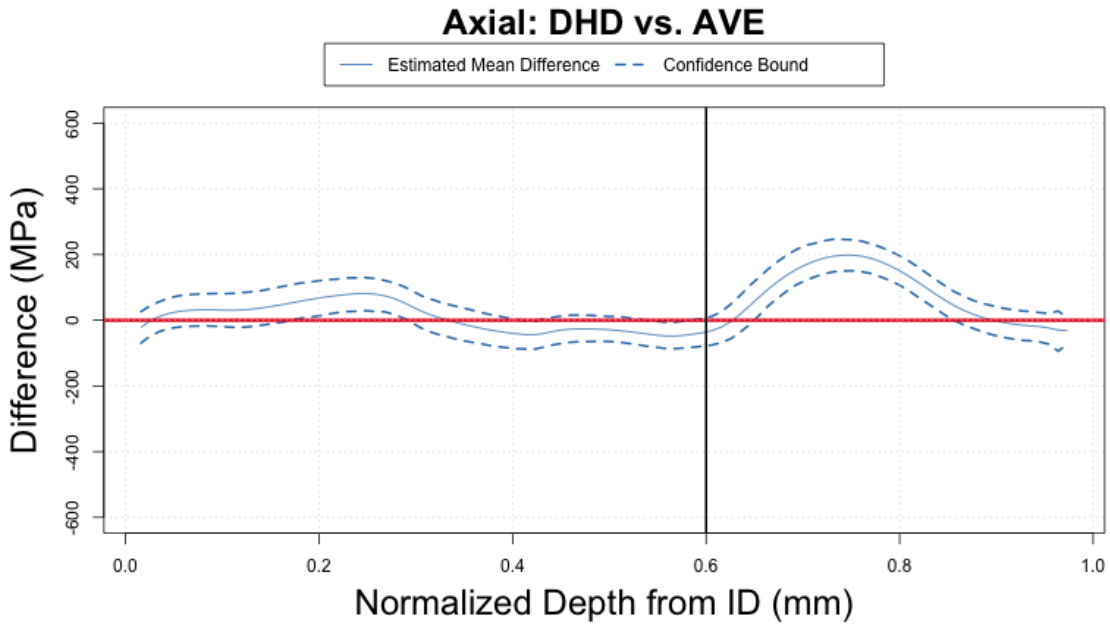


(b) 95% Confidence interval on the mean difference between the axial DHD measurements and the axial predictions assuming isotropic hardening. A vertical line at normalized ID depth $d = 0.6$ appears as a reminder that the smoothing error for the DHD measurements is large for $d > 0.6$.

Figure 44: Comparison of axial DHD measurements to axial predictions assuming isotropic hardening.



(a) Axial DHD measurements plotted with the axial predictions assuming an average of the two hardening laws.



(b) 95% Confidence interval on the mean difference between the axial DHD measurements and the axial predictions assuming an average of the two hardening laws. A vertical line at normalized ID depth $d = 0.6$ appears as a reminder that the smoothing error for the DHD measurements is large for $d > 0.6$.

Figure 45: Comparison of axial DHD measurements to axial predictions assuming an average of the two hardening laws.

between mean DHD measurement and mean KIN prediction has a 95% confidence interval from approximately -100 to 0 MPa. On average, the predictions are over-estimating the measurement by this much.

The hoop WRS predictions assuming isotropic hardening (blue) are plotted with the DHD hoop measurements (red) in Figure 47a. The estimated difference in prediction and measurement means is plotted with 95% confidence bounds in Figure 47b. Note that the confidence bounds contain zero for most of the depth of the weld up to ID depth $d = 0.6$. This suggests there is no strong bias towards under- or over- prediction of the mean on this interval, even though the mean difference is consistently negative up to about an ID depth $d = 0.45$; the confidence bounds include the possibility that the mean prediction at a specific depth correctly predicts the mean measurement. The confidence bounds span over 250 MPa, indicating that one should expect the difference between mean ISO hoop predictions and mean DHD measurements to vary within a 250 MPa range.

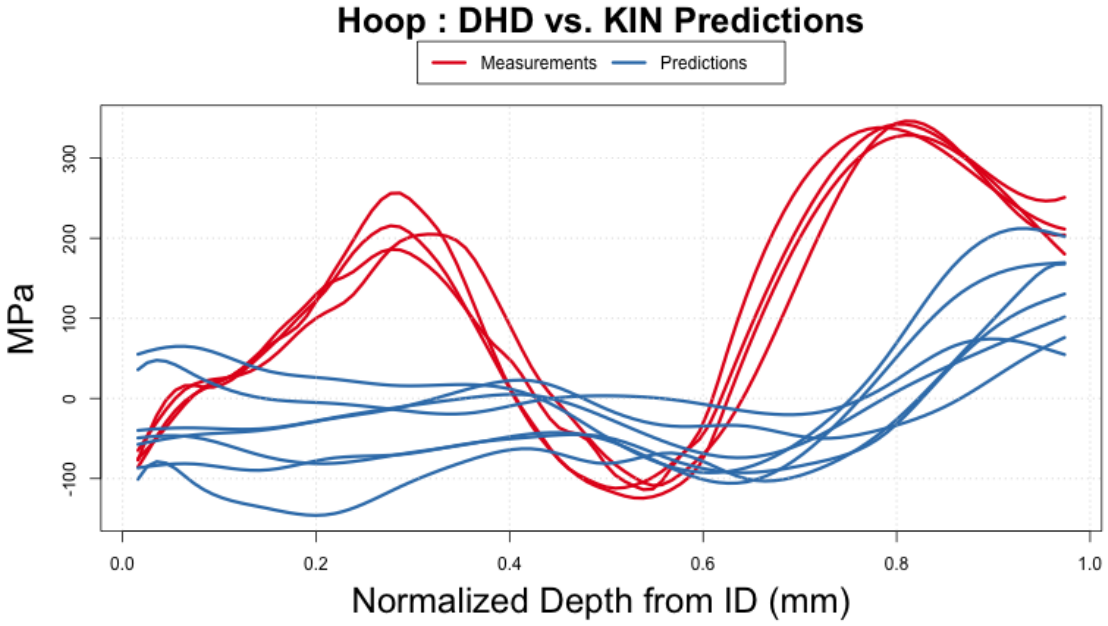
The averages of the hoop predictions (blue) assuming kinematic hardening and assuming isotropic hardening are plotted with the DHD hoop measurements (red) in Figure 48a. Figure 48b shows the estimated difference in the AVE hoop prediction mean and the DHD hoop measurement mean with 95% confidence bounds. The confidence bounds are between 125 and 200 MPa wide up through the ID depth $d = 0.6$. On this interval, the AVE of the predictions does not consistently over- or under- predict the mean hoop WRS. Near the ID, the 95% confidence interval for the difference between the measurement mean and the prediction mean ranges from about -120 and 30 MPa.

Root Mean Squared Errors of the Predictions Compared to DHD Measurements

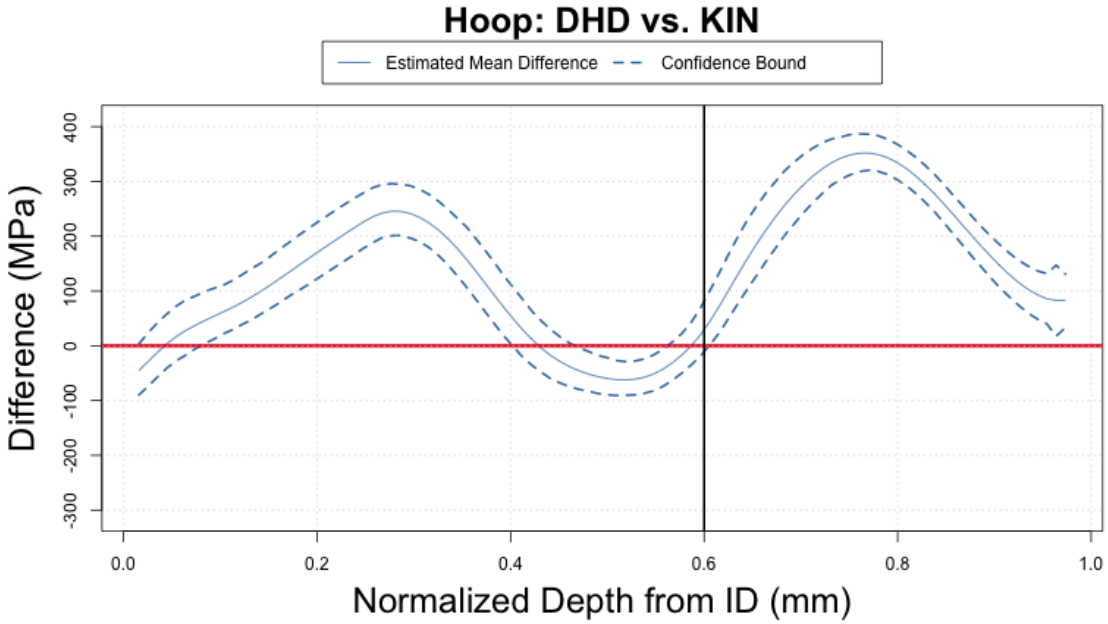
Visual inspection of the confidence bounds can be informative but makes it difficult to compare how the predictions that assume different hardening assumptions (KIN, ISO, AVE) perform. One metric that can compare the hardening laws is the root mean squared error (RMSE) of the difference between mean prediction and mean measurement. The difference in means, $h(d)$, is observed on a grid of depth values $d = d_1, \ldots, d_K$ and the RMSE is defined as

$$\sqrt{\frac{1}{K}\sum_{k=1}^{K} h(d_k)^2} \tag{11}$$

Note the RMSE is calculated for each of the $S = 1000$ bootstrap differences. The mean of these is the estimated RMSE and the 2.5 and 97.5 percentiles form a 95% bootstrap confidence interval. The estimated RMSE (along with their 95% confidence intervals) comparing the DHD measurements (of both hoop and axial WRS) to their corresponding predictions appear in Table 1. Though the confidence intervals overlap, there is a trend that the predictions using the average of the individual isotropic and kinematic hardening assumption predictions result in better predictions, on average, than individual predictions. There is also a suggestion that the isotropic assumption performs better than the kinematic assumption.

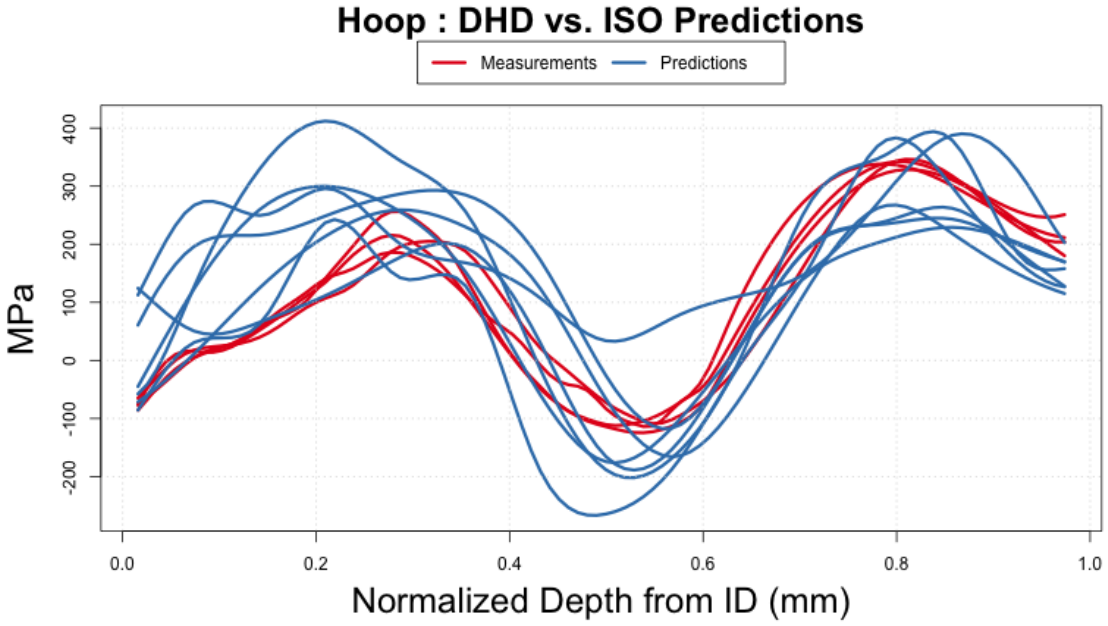


(a) Hoop DHD measurements plotted with the hoop predictions assuming kinematic hardening.

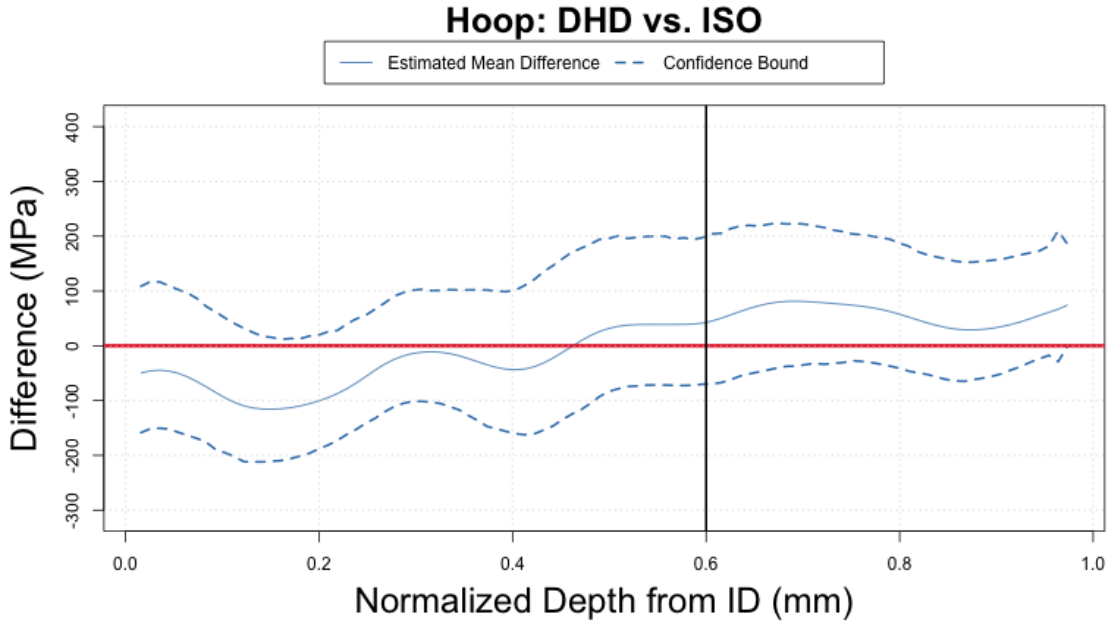


(b) 95% Confidence interval on the mean difference between the hoop DHD measurements and the hoop predictions assuming kinematic hardening. A vertical line at normalized ID depth $d = 0.6$ appears as a reminder that the smoothing error for the DHD measurements is large for $d > 0.6$

Figure 46: Comparison of the hoop DHD measurements to the hoop predictions assuming kinematic hardening.

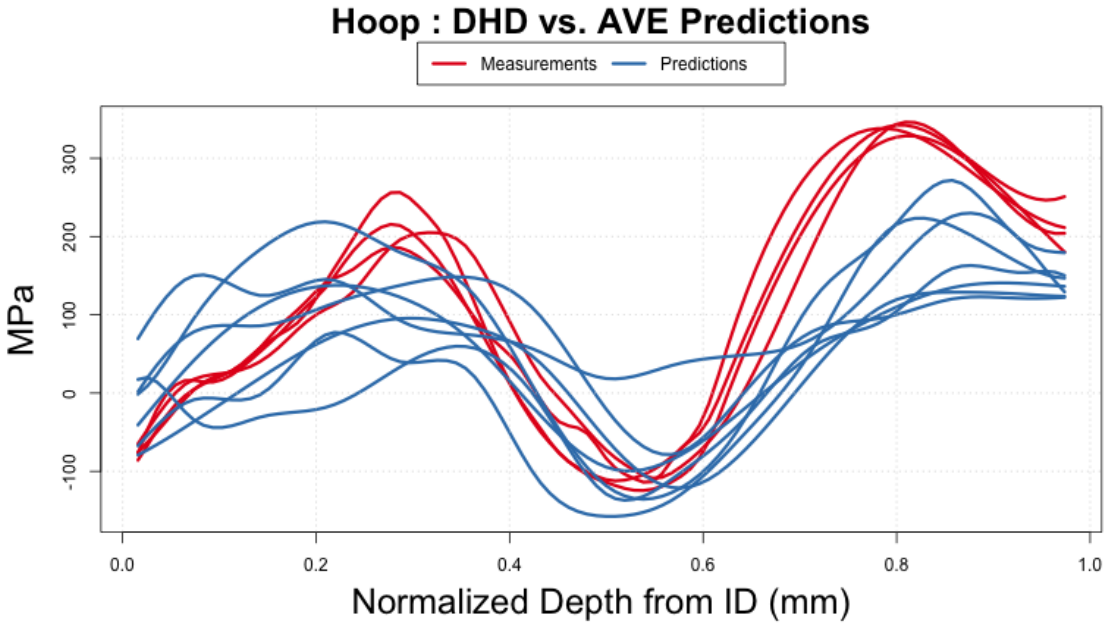


(a) Hoop DHD measurements plotted with the hoop predictions assuming isotropic hardening.

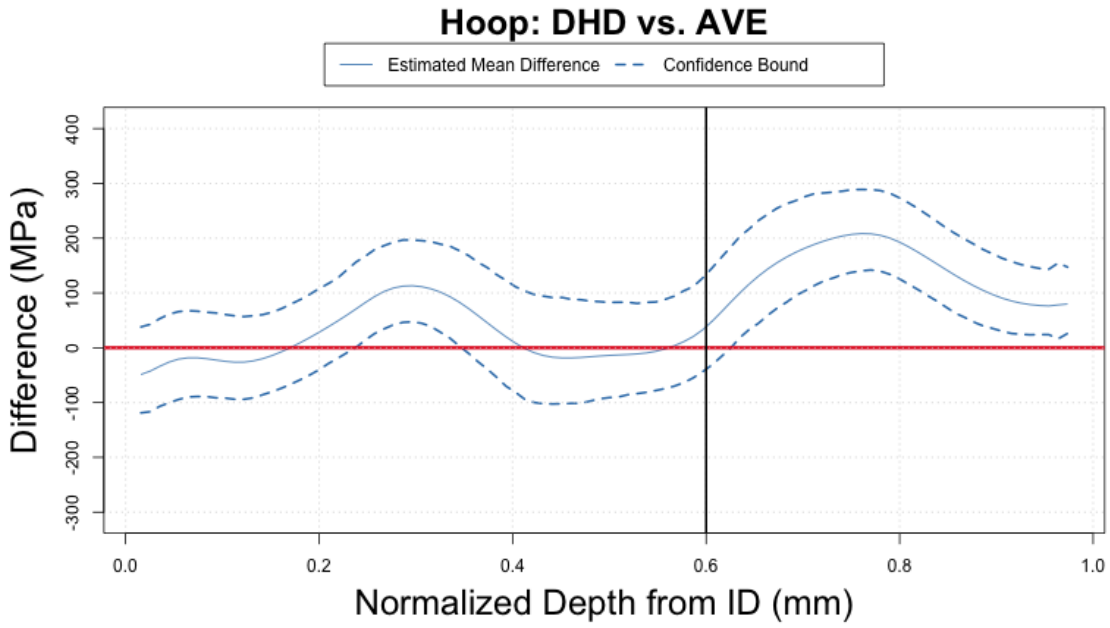


(b) 95% Confidence interval on the mean difference between the hoop DHD measurements and the hoop predictions assuming isotropic hardening. A vertical line at normalized ID depth $d = 0.6$ appears as a reminder that the smoothing error for the DHD measurements is large for $d > 0.6$.

Figure 47: Comparison of hoop DHD measurements to the hoop predictions assuming isotropic hardening.



(a) Hoop DHD measurements plotted with the hoop predictions assuming an average of the two hardening laws.



(b) 95% Confidence interval on the mean difference between the hoop DHD measurements and the hoop predictions assuming an average of the two hardening laws. A vertical line at normalized ID depth $d = 0.6$ appears as a reminder that the smoothing error for the DHD measurements is large for $d > 0.6$.

Figure 48: Comparison of the hoop DHD measurements to the hoop predictions assuming an average of the two hardening laws.

Table 1: RMSE (95% CI) of the difference in means between DHD measurements and predictions. Due to the increased uncertainty of the DHD measurements after normalized ID depth of 0.6, this value is the upper bound for computing the MSE

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)

5.2 Comparison of Contour Measurements and Predictions

Comparison of Contour Axial Measurements and Predictions

The 500 extracted axial contour measurements (red) are plotted with the kinematic prediction (blue) in Figure 49a. The estimated mean difference with 95% confidence bound is plotted in Figure 49. This indicates systematic differences between predictions and measurements with a 95% confidence interval near the ID from approximately -150 to -90 MPa.

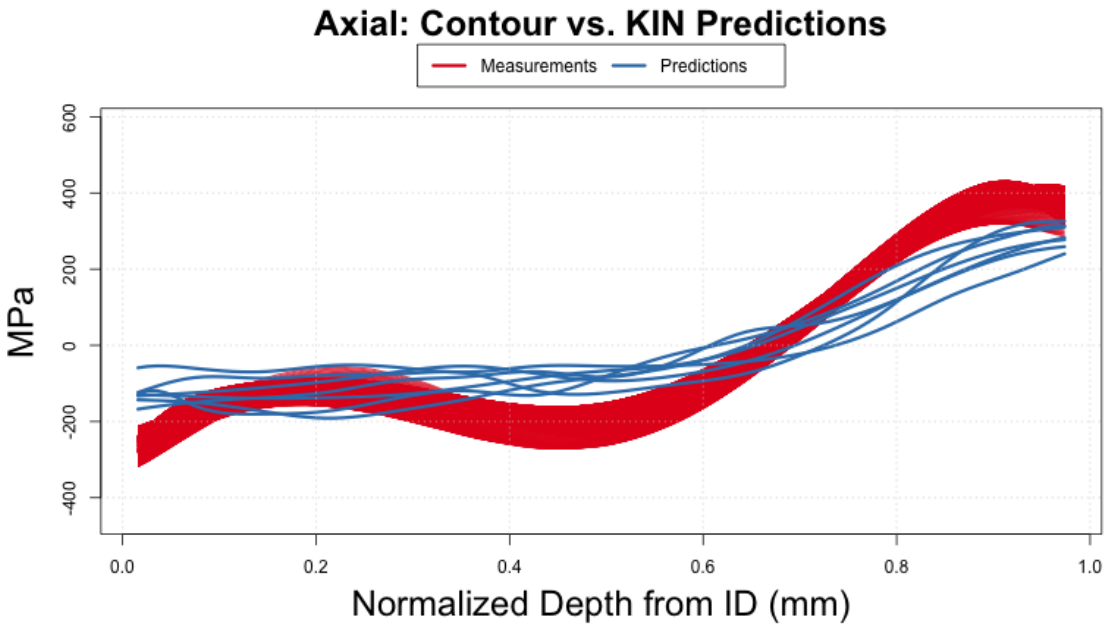
Axial isotropic predictions (blue) appear with the contour measurements in Figure 50a. The estimated mean difference with 95% confidence bound is plotted in Figure 50. This indicates some agreement between predictions and measurements with a 95% confidence interval at the ID from approximately -100 to 100 MPa.

The averaged predictions (blue) along with the contour measurements are plotted in Figure 51a. The estimated mean difference with 95% confidence bound is plotted in Figure 51. Smaller systematic differences throughout the weld appear here. The 95% confidence interval near the ID is from approximately -130 to -30 MPa.

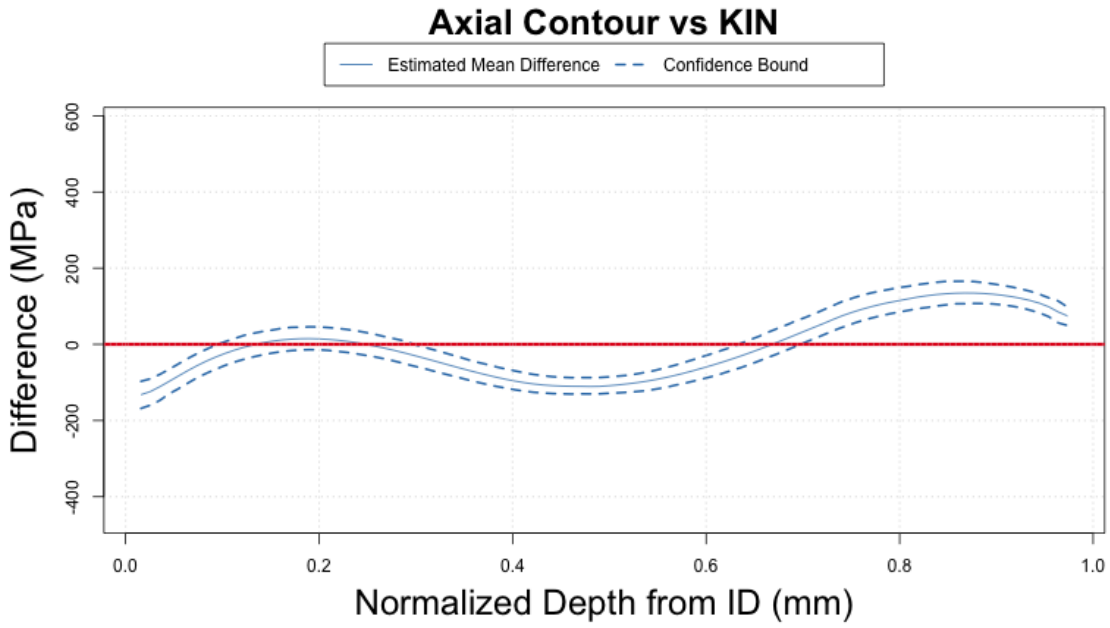
Comparison of Contour Hoop Measurement and Predictions

Figure 52a shows the hoop WRS predictions assuming kinematic hardening (blue) with the contour measurement of hoop WRS (red). Recall that there is only one contour measurement of hoop WRS. The estimate of mean difference only takes into account the uncertainty in the predictions as the information needed to characterize uncertainty in the measurements is not available. The 95% confidence bounds on the estimate of the mean difference are plotted in Figure 52b. Without accounting for any measurement uncertainty, the mean difference is expected to vary by over 200 MPa. The mean predictions are expected to over-predict the measurement near the ID.

The contour measurement of hoop WRS (red) is plotted with the predictions assuming isotropic hardening (blue) in Figure 53a. The 95% confidence bounds plotted in 53b indicate that the magnitude of the difference between the mean prediction and the contour measurement could be larger

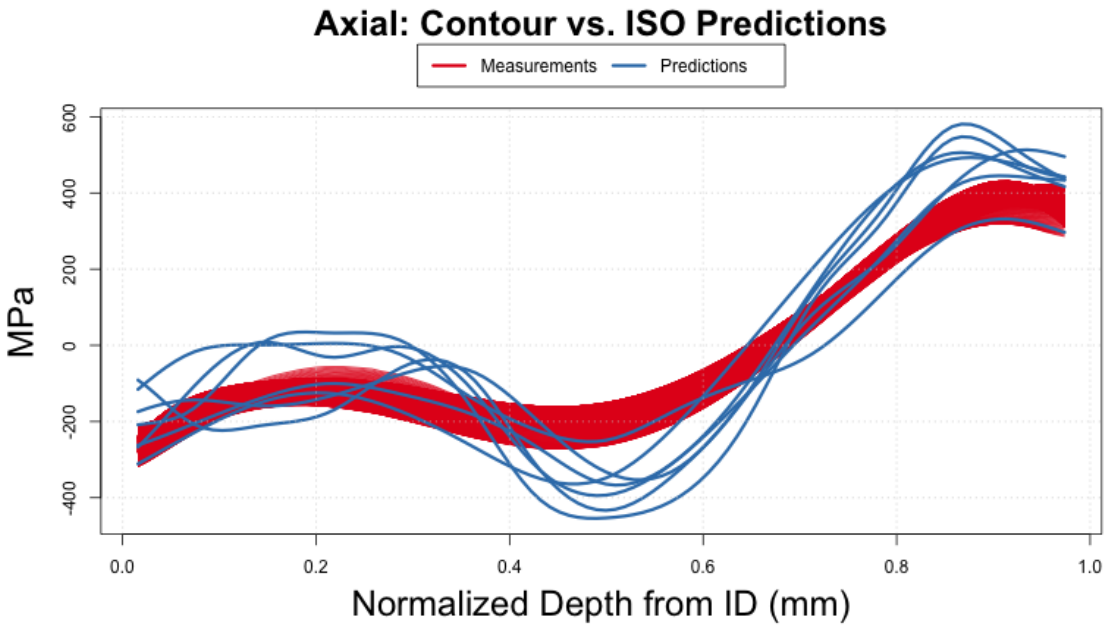


(a) Extracted axial contour measurements plotted with the Axial predictions assuming kinematic hardening.

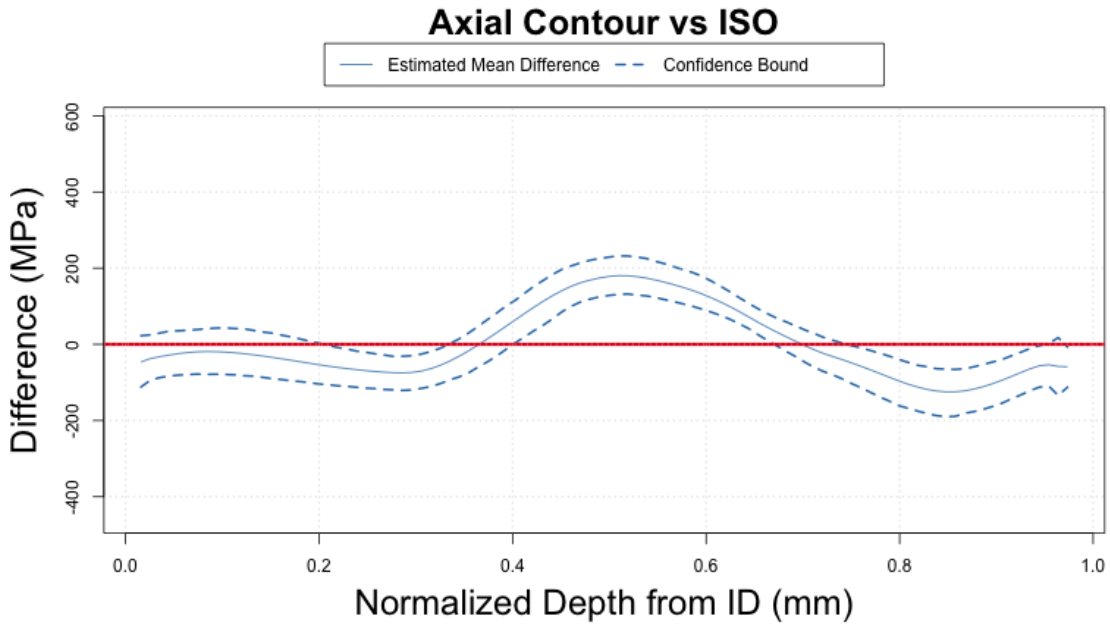


(b) 95% Confidence interval on the mean difference between the axial contour measurements and the axial predictions assuming kinematic hardening.

Figure 49: Comparison of axial contour measurements with the axial predictions assuming kinematic hardening.

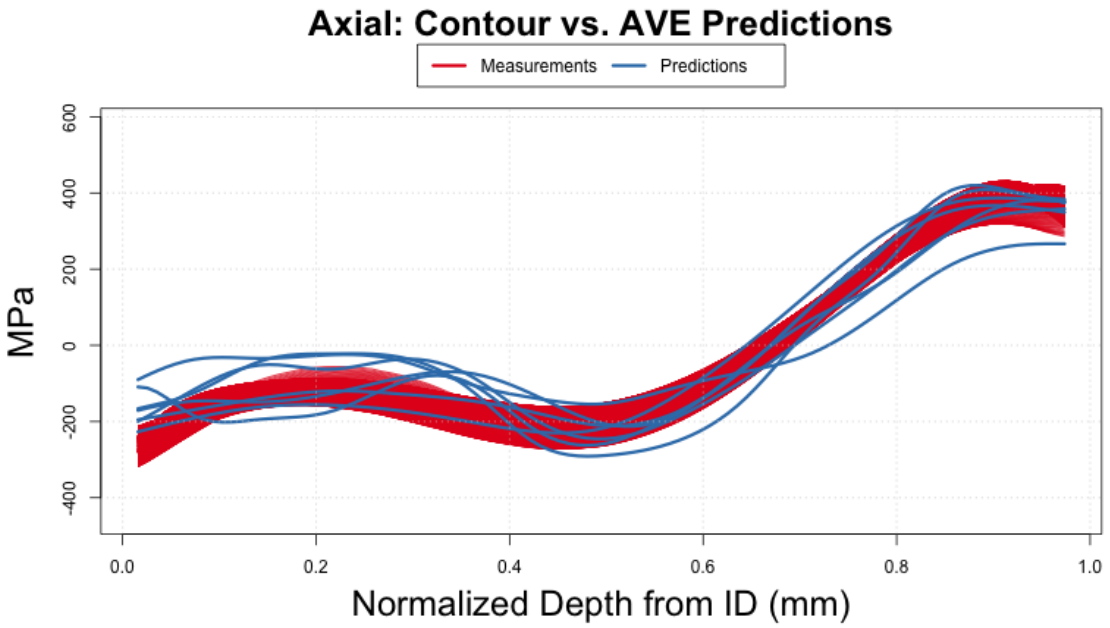


(a) Extracted axial contour measurements plotted with the axial predictions assuming isotropic hardening.

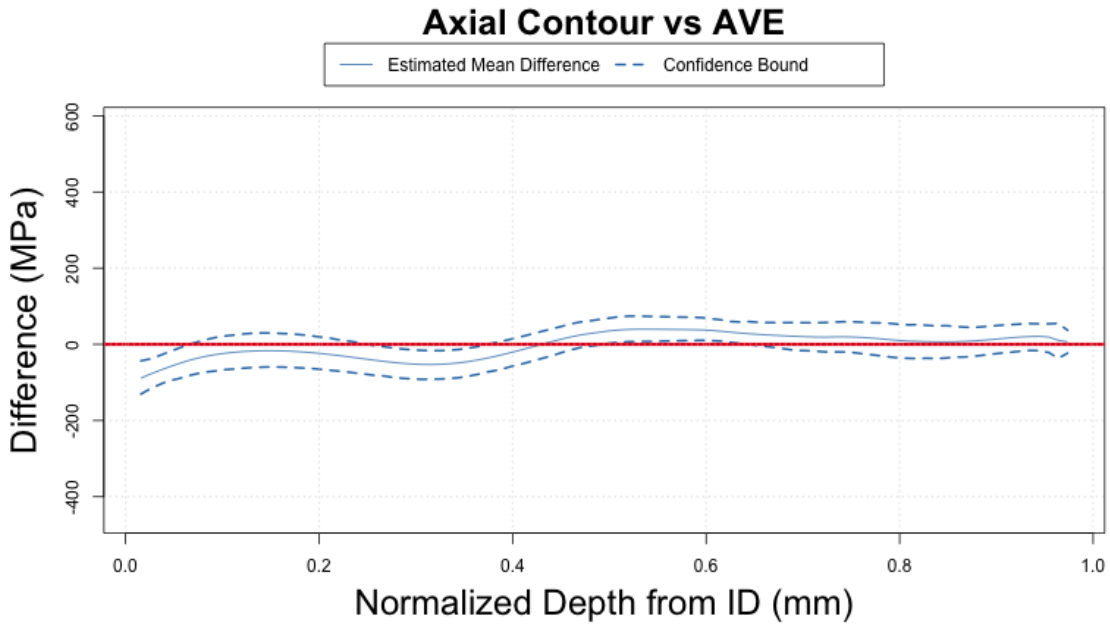


(b) 95% Confidence interval on the mean difference between the axial contour measurements and the axial predictions assuming isotropic hardening.

Figure 50: Comparison of axial contour measurements with the axial predictions assuming isotropic hardening.

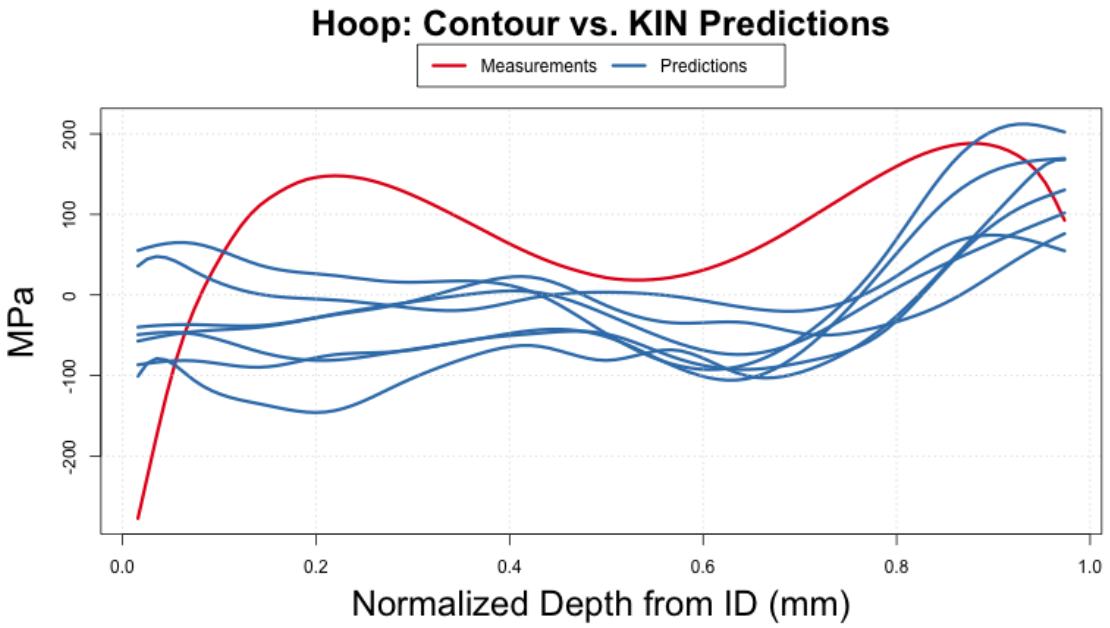


(a) Extracted axial contour measurements plotted with the average of the axial predictions.

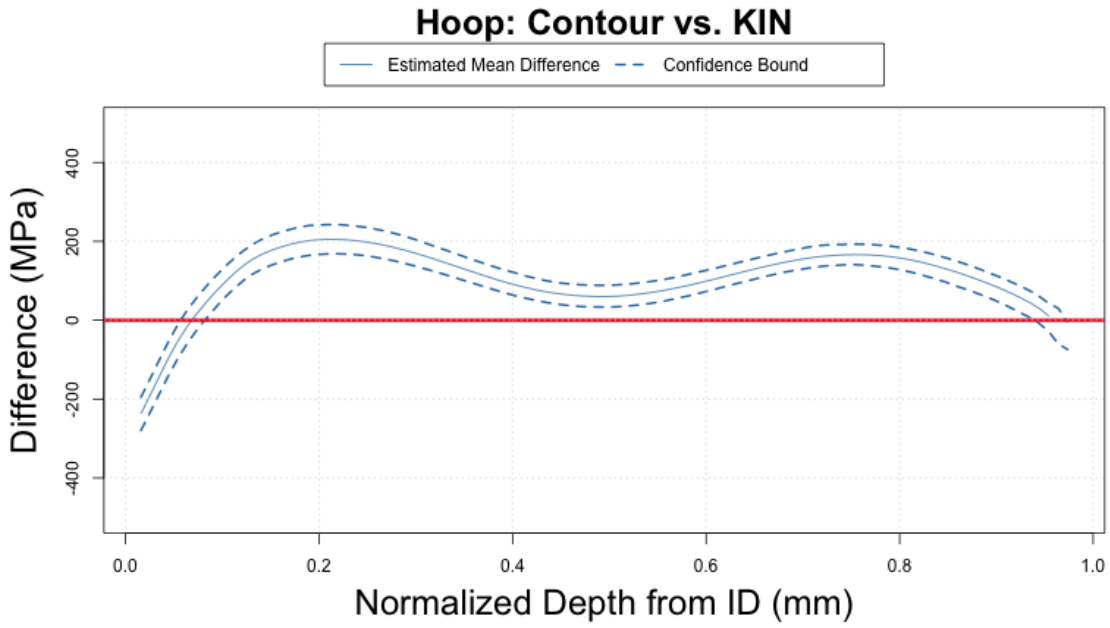


(b) 95% Confidence interval on the mean difference between the axial contour measurements and the average of the axial predictions.

Figure 51: Comparison of the axial contour measurements and the axial predictions assuming an average of the hardening laws.



(a) Hoop contour measurement plotted with the hoop predictions assuming kinematic hardening.



(b) 95% Confidence interval on the mean difference between the hoop contour measurements and the hoop predictions assuming kinematic hardening.

Figure 52: Comparison of the hoop contour measurement with the hoop predictions assuming kinematic hardening.

than about 375 MPa. Near the ID, the mean of the predictions is expected to differ from the measurement by over 100 MPa.

The averages of the hoop predictions assuming kinematic hardening and assuming isotropic hardening (blue) are plotted with the hoop contour measurement in 54a. Figure 54b shows the estimated difference between the mean AVE hoop prediction and the hoop contour measurement with 95% confidence bounds. Because the mean ISO and mean KIN predictions over-predict the contour measurement near the ID, so does the mean of the AVE predictions; it over-predicts by up to 300 MPa. Away from the ID, however, the 95% confidence bounds suggest that the magnitude of this difference is expected to be less than 200 MPa. At some depths the AVE means outperform the KIN and ISO means. For example, at ID depth $d = 0.8$, the KIN means tend to under-predict the measurement whereas the ISO means tend to over-predict the measurement. This over-prediction in the ISO means balances the under-prediction from the KIN means in the AVE predictions so that the AVE predictions are expected to fall less than 100 MPa from the measurement.

Root Mean Squared Errors of the Predictions Compared to Contour Measurements

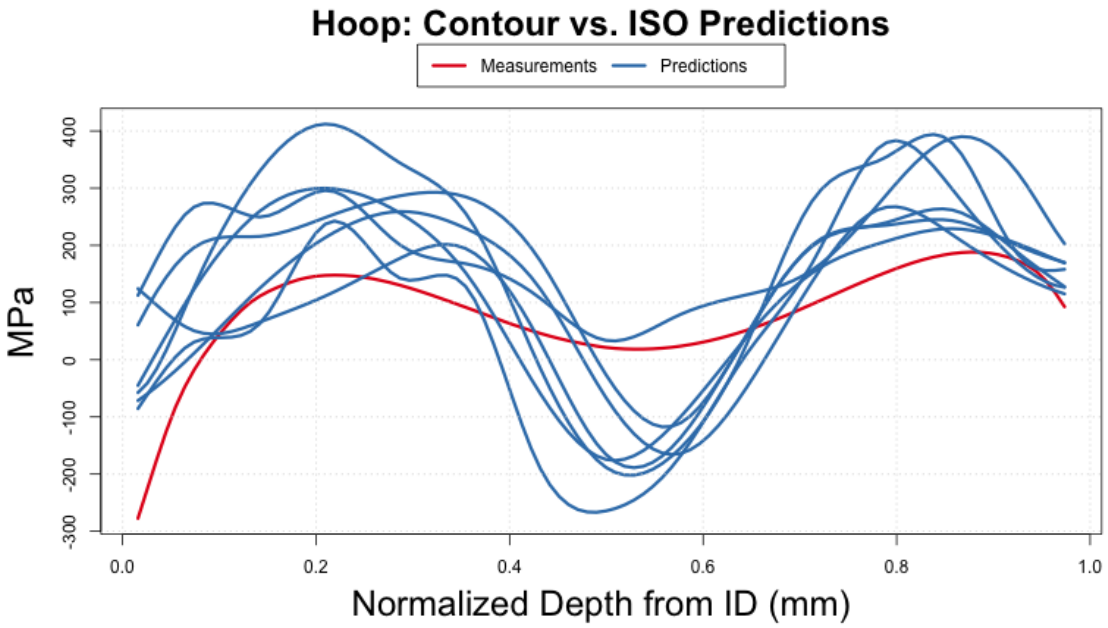
As performed for the DHD measurements in Section 5.1, the contour measurements and predictions are compared using the RMSE metric. The RMSE along with their 95% bootstrap confidence bounds appear in Table 2. Recall, only one measurement of hoop WRS at the mid-weld was provided and the hoop predictions were compared directly to this single measurement. For the axial measurements, 500 profiles along the mid-weld plane were extracted and the mean of these was compared to the mean prediction. It should be noted that, as measured by RMSE, there is a tendency for the average of the isotropic and kinematic predictions to perform better than predictions assuming just isotropic or just kinematic hardening, though there is overlap between some of the confidence intervals.

Table 2: RMSE (95% CI) of the difference in means between contour measurements and predictions.

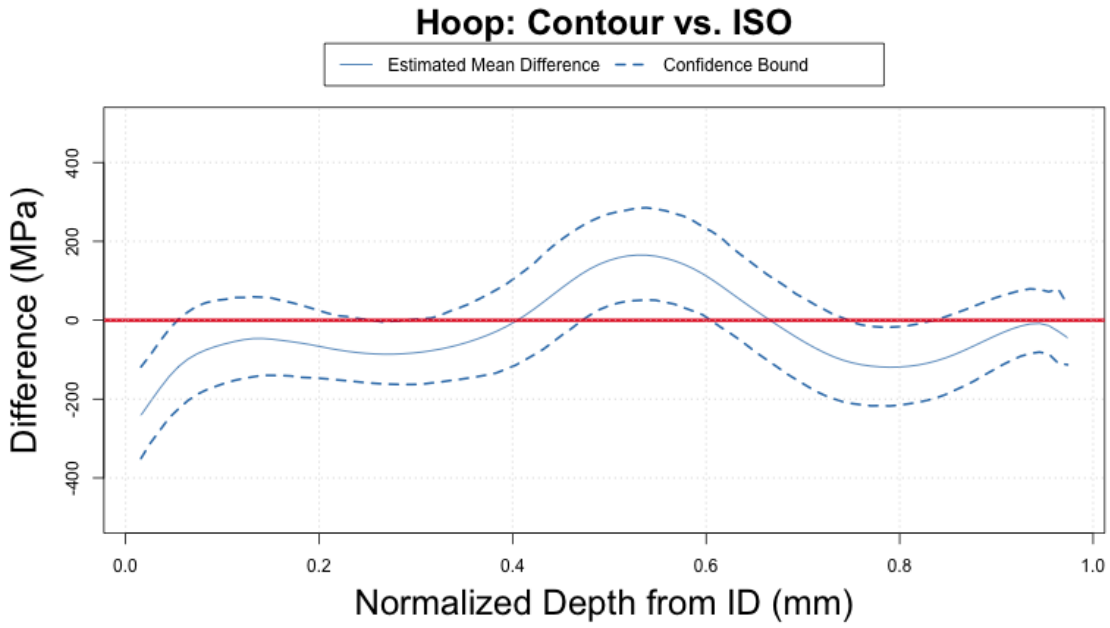
RMSE	Axial(MPa)	Hoop(MPa)
KIN	83.43 (69.61,95.11)	133.22 (116.83,154.31)
ISO	93.99 (69.42,117.84)	109.62 (86.22,142.57)
AVE	36.77 (23.29,51.76)	78.33 (56.01,120.01)

5.3 Comparison of DHD and Contour Measurements

The ultimate goal when building prediction models is to validate models with measurements. To do this, however, accurate and repeatable measurements must be obtainable. In separate experiments, both the contour measurements and the DHD measurements showed similar agreement (and disagreement) with neutron diffraction measurements of WRS [1, 2]. However, neither method lead

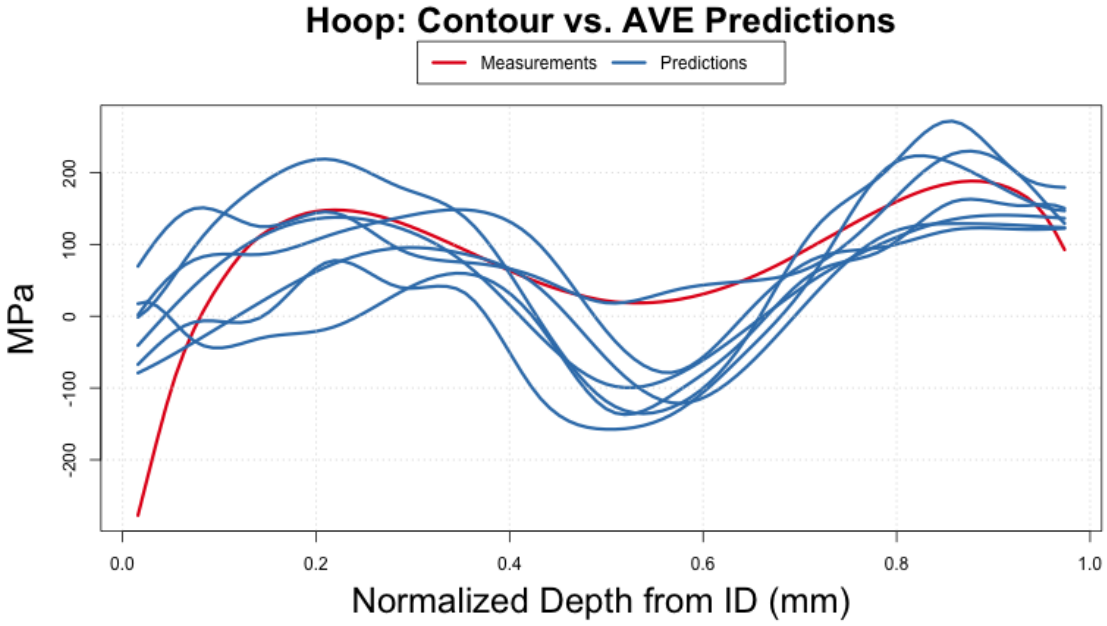


(a) Hoop contour measurement plotted with the hoop predictions assuming isotropic hardening.

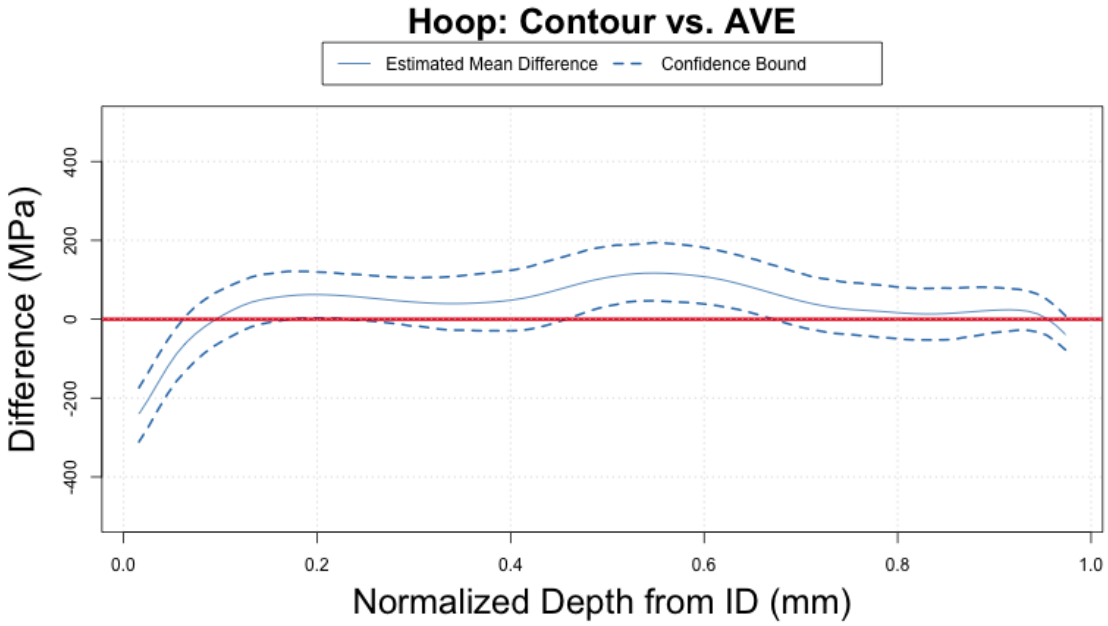


(b) 95% Confidence interval on the mean difference between the hoop contour measurements and the hoop predictions assuming isotropic hardening.

Figure 53: Comparison of the hoop contour measurement and the hoop predictions assuming isotropic hardening.



(a) Hoop contour measurement plotted with the hoop predictions assuming an average of kinematic and isotropic hardening.



(b) 95% Confidence interval on the mean difference between the hoop contour measurements and the hoop predictions assuming an average of kinematic and isotropic hardening.

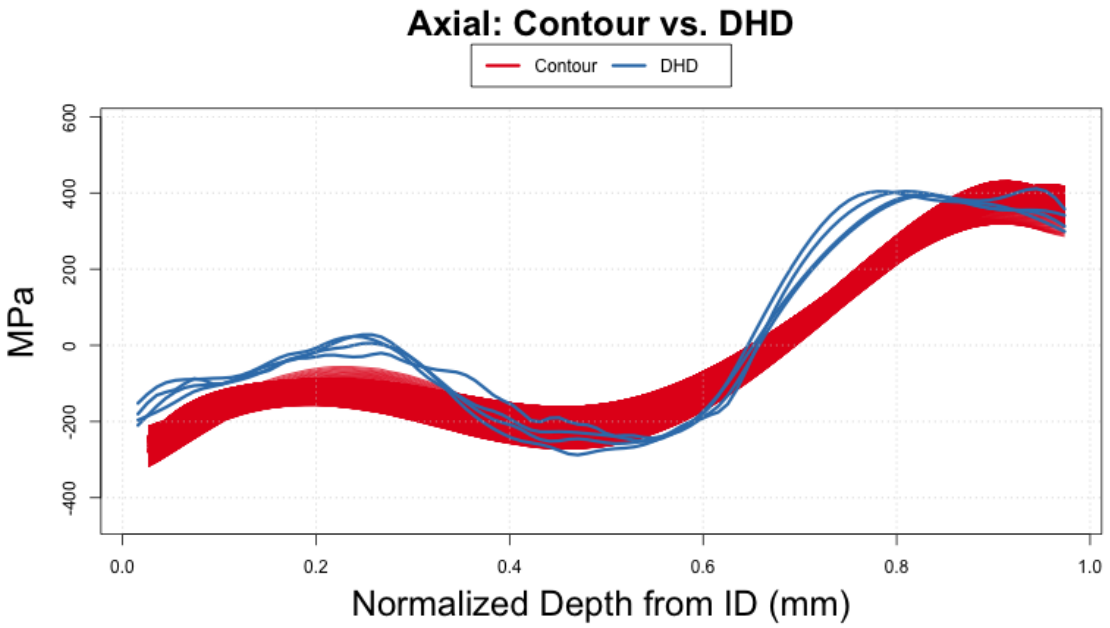
Figure 54: Comparison of the hoop contour measurement and the hoop predictions assuming an average of the hardening laws.

to measurements that always agreed with neutron diffraction measurements. The inherent challenge of measuring weld residual stresses complicates the validation process. In particular, while the method used here to compare measurements to predictions could be used as part of a model validation effort, it is not clear which measurements should be used; comparing the DHD measurements to the contour measurements suggests that the two measurements agree to about the same extent that the predictions agree with measurements.

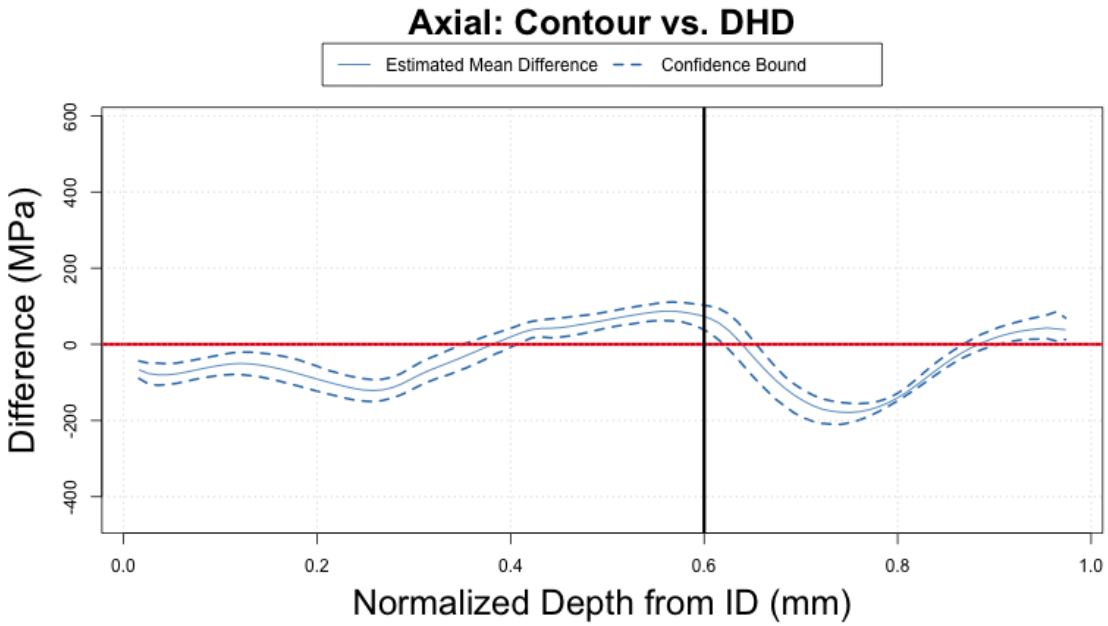
Consider the axial measurements. The initial five contour axial measurements provided by the measurement vendor are plotted with the DHD axial measurements in Figure 55a. The behavior of the measurements is similar, though there is significant disagreement on the magnitude of WRS. Figure 55b shows the estimated difference between the mean contour measurements and the mean DHD measurements with 95% bootstrap confidence bounds. For the majority of weld depths, the confidence bound does not include zero; this indicates a statistical difference between the mean contour and mean DHD measurements. Near the ID, the mean measurements are expected to differ between about 40 and 90 Mpa. Away from the ID they may differ by about 175 MPa. As with previous comparisons with the DHD measurements, the comparison beyond $d = 0.6$ includes non-negligible smoothing uncertainty due to the coarseness of the DHD measurements there.

Figure 56a shows the contour hoop measurement with the DHD hoop measurements. The data sets differ substantially; for example, the DHD measurements indicate tensile stresses in the middle of the weld, whereas the contour measurement indicates compressive stresses in the same region. This is also reflected in the plot of the estimated mean difference with 95% confidence bounds in Figure 56. The confidence bounds include zero, which indicates no statistical difference, on a small region somewhat near the ID and two slightly larger intervals away from the ID. Everywhere else, the mean measurements are expected to differ by up to about 200 MPa.

In order to use these measurements to validate prediction models, further study is recommended to determine the sources of disagreement between measurement techniques. For both hoop and axial WRS, the mean measurements using the DHD method and the mean measurements using the contour method are expected to differ near the ID. This may have significant implications for crack growth after formation near the ID.

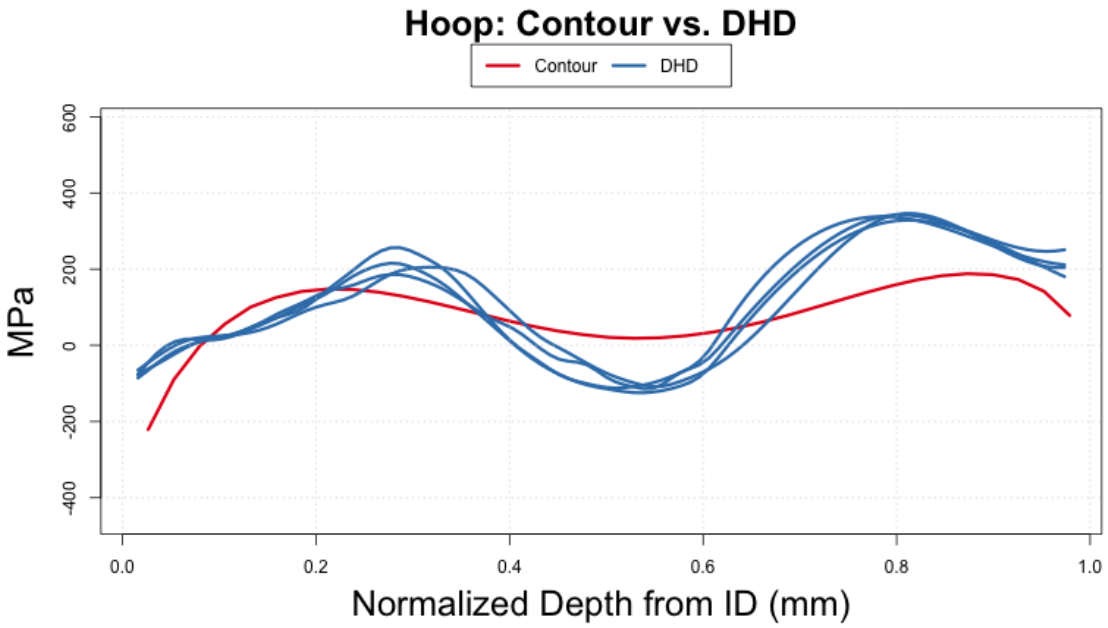


(a) 500 contour axial measurements plotted with the DHD axial measurements.

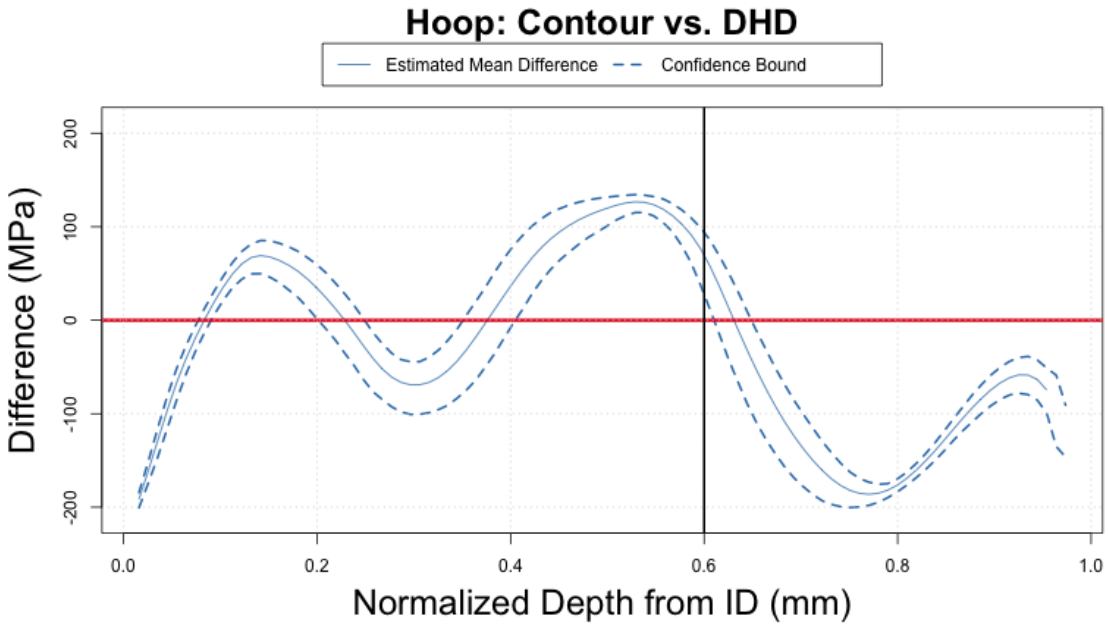


(b) 95% Confidence interval on the mean difference between the contour axial measurements and the DHD axial measurements.

Figure 55: Comparison of the contour axial measurements and the DHD axial measurements.



(a) The contour hoop measurements plotted with the DHD hoop measurements.



(b) 95% Confidence interval on the mean difference between the contour hoop measurements and the DHD hoop measurements.

Figure 56: Comparison of the contour hoop measurements and the DHD hoop measurements.

6 Discussion

The figures in Sections 3 and 4 illustrate the level of uncertainty within experimental measurements and within predictions, respectively. The confidence bounds given represent the range in which it is expected that the true mean WRS profile (of either the measurements or the model predictions) is expected to fall with 95% confidence. For the experimental measurements, the tolerance bounds given represent a range in which a large proportion (95%) of new WRS measurements of the same profile in the same weld are expected to fall (with 95% confidence). For the predictions, the tolerance bounds represent a range in which a large proportion (95%) of new predictions (from different modeling groups following the same modeling guidelines) would be expected to fall (with 95% confidence), based on the observed modeler-to-modeler variation in the double-blind round robin study. Both the statistical confidence level (in this case 95%) and the content for the tolerance bounds (in this case 95%) can be adjusted depending on the needs of a specific application.

It should be noted that, while the confidence and tolerance bounds on the measurements are narrower than those on the predictions at many depths, this is not always the case. There is enough uncertainty present in many of the measurements and predictions to affect the prediction of crack formation and growth in welds. For both contour and DHD measurements, the tolerance interval on axial WRS near the ID covers a range of about 200 MPa. For DHD measurements, the tolerance interval on hoop WRS near the ID is also about 200 MPa wide. Among the three hardening assumption predictions the tolerance intervals on hoop WRS near the ID range from about 450 MPa to over 1000 MPa in width. The tolerance intervals on axial WRS among the predictions range in width from about 400 to 650 MPa. WRS values differing by 200+ MPa may have a very different effect on predictions of crack formation and growth, particularly if the difference crosses from positive (tensile) to negative (compressive) stresses.

For both the measurements and model predictions, the resulting tolerance intervals can inform best practices for crack simulation. For example, if a single model predicts a large compressive stress at the ID of a weld, a simulation using this value may be unlikely to result in the formation of cracks. However, a simulation that considers values within the range of the tolerance interval estimating modeler-to-modeler uncertainty may predict the formation of cracks, particularly if the tolerance interval around a compressive estimate contains tensile values.

Tables 1 and 2 summarize the results of comparing predictions to measurements across the entire depth of the weld using the RMSE values. Based on this metric, the mean predictions of axial WRS tend to agree better with their corresponding measurements (both DHD and contour) than do the mean predictions of hoop WRS. Additionally, averaging the kinematic and isotropic predictions appears to match the mean measurements (both DHD and contour) better in both the hoop and axial directions than the individual predictions. However, it is not clear whether this is an artifact of averaging out an over- and under- prediction, rather than an indication that the average is more representative of the physical reality. The ordering of predictive ability of the individual isotropic and kinematic assumptions is not clear. The kinematic assumption appears to predict the contour data better in the axial direction but does worse overall when comparing to the DHD data. However, in many cases the 95% confidence intervals for RMSE overlap considerably between the two assumptions (and the average of the two assumptions), indicating the differences are not

necessarily statistically significant. Thus, one cannot be certain if a single type of prediction will consistently outperform others. These comparisons suggest that a more robust model for material hardening could improve prediction performance.

The results of this study suggest that uncertainty in measurements or predictions of WRS should be considered when predicting crack formation and growth due to PWSCC. In order to improve the accuracy of crack predictions, further study should focus on identifying (and potentially reducing) the greatest sources of uncertainty. Suggestions for pursuing such an extension based on this work are discussed in Sections 3.4 and 4.7. Additionally, the types of comparisons between models and measurements made in the report cannot necessarily be used to validate predictions. Both the DHD and contour measurements of WRS exhibit uncertainty and do not necessarily agree with each other, and hence it is unclear if either of these measurements, which require models themselves, represents a reliable baseline from which to compare model predictions. Since the hoop contour measurement could not be used to consider uncertainty, it may be preferable to use the DHD data for validation so measurement uncertainty can be included. It may not be appropriate to use either set of measurement data for validation of predictions unless further experimental study can be performed to explore the accuracy of each.

A complication of the comparison of measurements to predictions is that it is unknown which measurement method is best to use as a baseline. For example, the average method appears to be significantly the best predictor of the axial contour measurements. For the axial DHD measurements however, the RMSE confidence intervals for the isotropic and average assumption predictions overlap significantly and the best method according the RMSE is somewhat ambiguous. When comparing the measurements to each other with respect to axial stresses, the contour and DHD measurements show agreement on par with that seen between measurements and predictions; the measurements do not match each other better than they match the predictions. Without more experimental data, any determination of a baseline is ultimately a question of expert opinion. Experimental efforts to better understand these measurements could provide valuable insight for validation of prediction models so they could be used to support regulatory decisions.

If a reliable baseline measurement method can be determined, validation of a particular model could be made by using the methods for constructing confidence bounds of the mean difference between measurements and prediction. Ideally, the comparison would take into account uncertainty in the model prediction. Types of model uncertainty include modeler-to-modeler uncertainty and model input uncertainty – uncertainty arising from uncertain model inputs. In the current examples, only the former was considered, but with more information on each of the models it would be possible to adapt this work to address input uncertainty. Ultimately, the acceptance criterion would depend on the magnitude of the estimated mean difference between prediction and measurement, with associated uncertainty. The acceptable magnitude must be determined by subject matter experts, perhaps by appealing to sensitivity studies on crack initiation, and could depend on the relative depths in the weld considered the most important. Metrics similar to the RMSE could be constructed to address specific objectives of the comparison. Further, the scope of the validation must be evaluated by subject matter experts. A modeling process may be adequate for the specific nozzle under current consideration and yet fail when applied to another piece of equipment. Ideally, WRS measurements would be obtained from a wide variety of equipment to use

for validation. As this may not be feasible, a model should only be considered valid if it is able to match the available data within a certain margin and includes a thorough justification for its applicability to all of the components for which it will be used.

Finally, some of the uncertainties that were not accounted for in this study are re-iterated below.

- Multiple experimental measurements from only a single pipe mock-up were provided. As a consequence there is no ability to understand any unit-to-unit variation that would occur in a population of relevant pipe welds.
- The multiple experimental measurements on the single mock-up were used to represent measurement variability. It is important to understand that this variability also includes (and is completely confounded with) any spatial variability of WRS throughout the weld.
- Both measurements and predictions used models in different ways to produce an output WRS from inputs. Input uncertainty was not accounted for in either case because the information or resources to account for input uncertainty were not available.
- Finally, the smoothed data was treated as the best estimate of the WRS profiles throughout the weld depth. For the most part, residual variation was no more than roughly 5% of the range of the WRS profiles. This uncertainty was not accounted for but this amount of residual variation is not expected to practically affect inferences on means and differences in means.

7 Appendix

7.1 Functional Principal Components

In this section, functional principal components (fPCA) is briefly described. fPCA is used as a dimension reduction technique to model the dominant modes of variation in both the aligned SRSFs coupled with initial WRS values, $\{f_i(0), \tilde{q}_i\}$, and a transformation of the optimal warping functions $\{\gamma_i^*\}$ (see Sections 2.2 and 2.3 as well as [6] for explanations regarding the choice of modeling these two sets of functions). Further detail on fPCA can be found in many references (e.g. [7]) but it is first noted that fPCA is an extension of multivariate PCA for functional data. Functions are infinite dimensional objects, but discretization of the functions on a fine grid is a way to approximate the function in finite dimensional space. Once this is done, standard multivariate PCA is applied as is now described.

Since fPCA is performed on two different sets of functions (the predicted and measured WRS functions), let $x_i(t), i = 1, \ldots, n$ denote n generic functional observations on some domain $t \in \mathscr{T}$. Assume that the observations have been mean-centered so that $\sum_{i=1}^{N} x_i(t) = 0$ for each a. This assumption is without loss of generality as the observed mean can always be subtracted before performing fPCA.

Finding the dominant modes of variation involves approximating the x_i using basis function expansions and determining which elements of these expansions are dominant. In particular for a fixed K, the goal is to find a set of exactly K orthonormal basis functions $\xi_k(a)$ so that the basis function expansion

$$\widehat{x_i}(a) = \sum_{k=1}^{K} c_{ik}\xi_k(a), \text{ where } c_{ik} = \int x_i(t)\xi_k(t)\, dt \tag{12}$$

minimizes the global approximation criterion

$$\sum_{i=1}^{n} ||x_i - \hat{x}_i||^2, \tag{13}$$

where $||x_i - \hat{x}_i||^2 = \int (x_i(t) - \hat{x}_i(t))^2\, dt$ is the *integrated squared error*. Under this minimization criterion the ξ_k are exactly the first K eigenfunctions corresponding the K largest eigenvalues of the covariance function

$$v(s,t) = n^{-1} \sum_{i=1}^{n} x_i(s)x_i(t). \tag{14}$$

That is, the ξ_k in the basis expansion satisfy

$$\int v(s,t)\xi(t)\, dt = \rho\xi(s) \tag{15}$$

for an appropriate eigenvalue ρ.

The value of K is chosen to explain all of the dominate modes of variation so that $\hat{x}_i$ closely approximate to x_i. To explain further note the following. The c_{ik} $i = 1,2,\ldots,n$, $k = 1,2,\ldots K$ are the principal component scores and have mean zero ($\sum_i c_{ik} = 0$). Let $v_k = n^{-1}\sum_{i=1}^{n} c_{ik}^2$ and call $\sum_k v_k$ the *total variation*. The first eigenfunction ξ_1 maximizes v_1, subject to $||\xi_1||^2 = \int \xi_1^2 = 1$. Iteratively, ξ_k maximizes v_k subject to $||\xi_1||^2 = 1$ and $\int \xi_m \xi_k = 0$ for $m = 1,2,\ldots,k-1$ (i.e. the eigenfunctions are orthogonal to each other). Further, $v_1 \geq v_2 \geq \cdots \geq v_K$. It is the c_{ik} which are modeled using multivariate Gaussian distributions as described in Sections 2.2 and 2.3. Their variation is ordered from largest to smallest and only the first K components that account for the majority (e.g. 90%) of the total variation will substantially contribute. This is precisely what is meant by 'explaining all the dominant modes of variation' to choose K.

To estimate the eigenfunctions, eigenvalues, and scores, multivariate PCA is applied to a discretized version of the problem. The discretization changes all of the above integrals to sums. To proceed, each x_i is approximated on a common grid $t_1, t_2, ..., t_p$. The $N \times p$ matrix X with i^{th} row $(x_i(t_1), x_i(t_2), ..., x_i(t_p))^T$ has covariance function $\Sigma_c = n^{-1}X^T X$ which is $v(\cdot,\cdot)$ evaluated on the grid. The k^{th} eigenvector of Σ_c is $\tilde{\xi}_k = (\xi_k(t_1), \xi_k(t_2), ..., \xi_k(t_p))^T$, a discretization of the eigenfunction ξ_k. The score c_{ik} is the i,k value of the matrix $C = XE$ where E is the $p \times K$ matrix whose columns are $\tilde{\xi}_k$.

7.2 Smoothing

The WRS measurements and predictions exist on a discrete set of points on the relative depth (from the ID) interval $[0,1]$. WRS, however, exists through the depth of the weld in what is expected to be a continuous manner. The discrete measurements and predictions are approximations of the smooth WRS profiles of interest and smoothing techniques are used to approximate the true WRS profiles.

In general, functions are observed as discrete noisy data $y_j = x(t_j) + \varepsilon_j$. Smoothing the noisy data involves using a basis function expansion for $x(t)$:

$$x(t) = \sum_{k}^{K} c_k \phi_k(t). \tag{16}$$

The basis functions used here are *spline* functions (see [7]). A spline basis is specified by its *nodes* and *degree*. The nodes (a.k.a. knots) are a sequence of points that divide the interval on which the function is to be approximated. On each interval defined by the knots, a spline is a polynomial of chosen degree. For an example, see Figure 57 where linear (degree 1) and cubic (degree 3) splines are used to smooth noisy data using the standard least squares estimates of the c_k in Equation (16) (i.e. the $c's$ minimize $\sum_j (y_j - x(t_j))^2$). The linear spline uses each data point as a node and clearly over-fits the data. The cubic splines using $5, 10$ and 15 evenly spaced nodes to fit the example data are plotted in the left, middle and right plots, respectively. Each of these fits the data fairly well, but the fit using 15 nodes appears to be a slight over-fit of the data. Increasing the number of nodes will result in more over-fitting and closer interpolation of the data. Note that

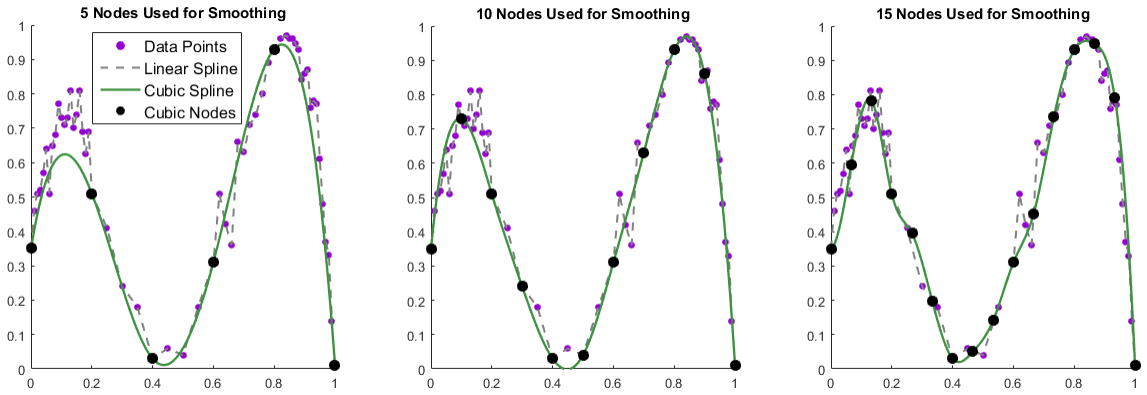


Figure 57: Example data with cubic splines using a different number of knots for smoothing.

a cubic spline fit with nodes at each data point (as was done for the linear spline) would have also severely over-fit the data. Thus, choosing too many nodes can be problematic.

A popular way of resolving the dilemma of choosing an appropriate number of nodes and their locations is to use a penalized smoothing technique. This is often referred to as *spline smoothing*. The idea is to augment the least squares fitting criterion with a term that penalizes the 'roughness' of the resulting fit. The result is that $x(t)$ cannot be too rough, even if a large number of nodes is chosen. While there are reasons to use other measures of 'roughness', the most common is the integrated squared second derivative:

$$P_2(x) = \int [x''(t)]^2 \, dt. \tag{17}$$

The penalized criterion function to minimize is then

$$\sum_j (y_j - x(t_j))^2 + \lambda P_2(x), \tag{18}$$

where λ is called the *smoothing parameter* because it controls the trade-off between the amount of roughness and the amount of smoothing. If $\lambda = 0$, there is no penalization and the standard least squares criterion is used. Therefore, choosing too many nodes will result in over-fitting (under-smoothing) the data. As λ becomes larger, the fitted function becomes less rough (more smooth). Interestingly, the curve that minimizes (18) is a cubic spline with knots at each of the points t_j [16, see]. Hence we use cubic splines, which effectively reduces the choice of nodes to a choice of λ. To make this choice, the λ that minimizes the *generalized cross-validation* measure [17] is used. We note that, in some cases, using knots at each t_j and the cross-validation measure still results in overly rough functions. In these cases we reduce the number of knots while still using cubic splines and solving (18) via cross-validation. The spline smoothing in this document is implemented in **R** using the **fda** package [16]. Data analytic choices on the number of knots is documented in the accompanying code.

References

[1] Michael B. Prime, Robert J. Sebring, John M. Edwards, Darren J. Hughes, and Peter J. Webster. Laser surface-contouring and spline data-smoothing for residual-stress measurement. *Experimental Mechanics*, 44(2):176–184, 2004.

[2] A.h. Mahmoudi, S. Hossain, C.E. Truman, D.J. Smith, and M.J. Pavier. A new procedure to measure near yield residual stresses using the deep hole drilling technique. *Experimental Mechanics*, 49:595–604, 2008.

[3] Michael L. Benson, Minh N. Tran, and Michael R Hill. Phase 2b weld residual stress round robin: Mockup design and comparisons of measurement and simulation results. In *Proceedings of the ASME 2015 Pressure Vessels and Piping Conference*, Boston, MA, July 2015.

[4] Michael Benson, David Rudland, and Aladar Csontos. Weld residual stress finite element analysis validation: Part 1 - data development effort. NUREG- 2162, U. S. Nuclear Regulatory Commission, 2014.

[5] Benson, Michael. Description of WRS Mockup and Data, 2015.

[6] J Derek Tucker, Wei Wu, and Anuj Srivastava. Generative models for functional data using phase and amplitude separation. *Computational Statistics & Data Analysis*, 61:50–66, 2013.

[7] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer, New York, 2005.

[8] J. Derek Tucker. *fdasrvf: Elastic Functional Data Analysis*, 2016. R package version 1.6.1.

[9] Anthony Christopher Davison and David Victor Hinkley. *Bootstrap methods and their application*, volume 1. Cambridge university press, 1997.

[10] Anthony C Davison, David V Hinkley, and G Alastair Young. Recent developments in bootstrap methodology. *Statistical Science*, pages 141–157, 2003.

[11] Gerald J Hahn and William Q Meeker. *Statistical intervals: a guide for practitioners*, volume 328. John Wiley & Sons, 2011.

[12] Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

[13] Mitchell D Olson, Adrian T DeWald, Michael B Prime, and Michael R Hill. Estimation of uncertainty for contour method residual stress measurements. *Experimental Mechanics*, 55(3):577–585, 2015.

[14] Marc C. Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society, Series B, Methodological*, 63:425–464, 2001.

[15] Lorenz Biegler, George Biros, Omar Ghattas, Matthias Heinkenschloss, David Keyes, Bani Mallick, Luis Tenorio, Bart van Bloemen Waanders, Karen Willcox, and Youssef Marzouk. *Large-scale inverse problems and quantification of uncertainty*, volume 712. John Wiley & Sons, 2011.

[16] J. O. Ramsay, Hadley Wickham, Spencer Graves, and Giles Hooker. *fda: Functional Data Analysis*, 2014. R package version 2.4.4.

[17] Peter Craven and Grace Wahba. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31(4):377–403, 1979.

DISTRIBUTION:

1 Michael L. Benson
U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research
Division of Engineering
Mail Stop: TWFN 10-A36

1 Shah N. Malik
U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research
Division of Engineering
Mail Stop: TWFN 10-A36

1 Matthew J. Homiack
U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research
Division of Engineering
Mail Stop: TWFN 10-A36

1 MS 0748 Dusty M. Brooks, 6233
1 MS 0748 Rémi Dingreville, 6233
1 MS 0748 Aubrey C. Eckert-Gallup, 6233
1 MS 0748 Scott E. Sanborn, 6233
1 MS 0830 John R. Lewis, 0436
1 MS 0899 Technical Library, 9536 (electronic copy)

Sandia National Laboratories