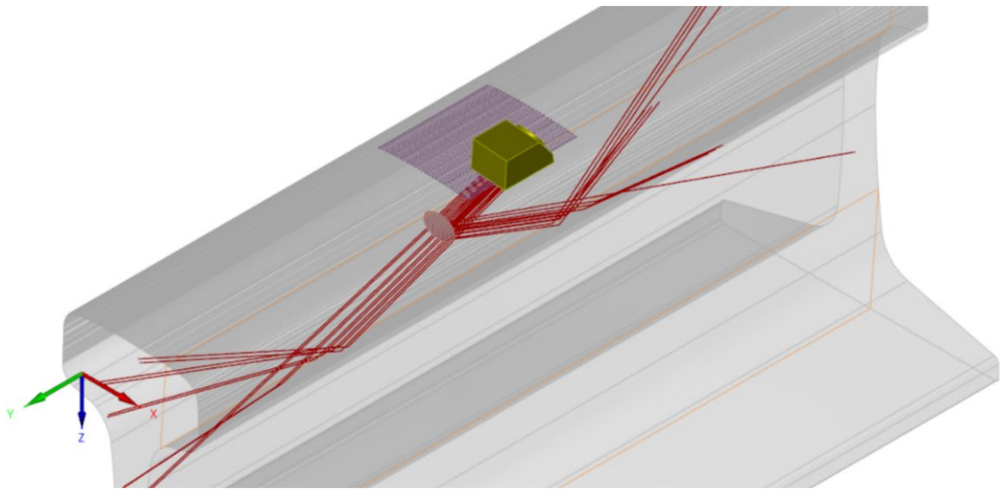
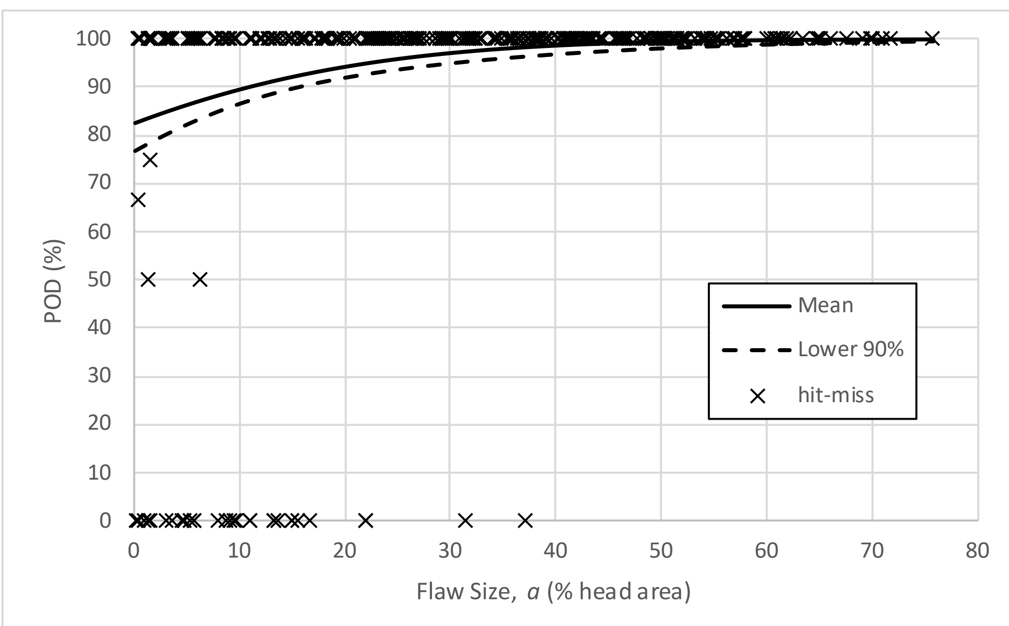
U.S. Department
of Transportation

Federal Railroad
Administration

Office of Research,
Development and Technology
Washington, DC 20590

Application of MAPOD to Handheld Ultrasonic Rail Flaw Inspection





DOT/FRA/ORD-26/07

Final Report | July 2026

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
		2019 - 2022

4. TITLE AND SUBTITLE	
Application of MAPOD to Handheld Ultrasonic Rail Flaw Inspection	**5a. CONTRACT NUMBER** 693JJ6-18-D-000005
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	
Radim Bruzek – ORCiD: 0000-0003-3069-9476 Daniel Einbinder – ORCiD: 0000-0001-7099-2381 John Tunna – ORCiD: 0000-0003-2735-0679	**5d. PROJECT NUMBER**
	5e. TASK NUMBER 693JJ6-19-F-000086
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
ENSCO, Inc. 5400 Port Royal Road Springfield, VA 22151	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	
U.S. Department of Transportation Federal Railroad Administration 1200 New Jersey Ave. SE Washington, DC 20590	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) DOT/FRA/ORD-26/07

12. DISTRIBUTION/AVAILABILITY STATEMENT
This document is available to the public through the FRA website.

13. SUPPLEMENTARY NOTES
COR: Jay Baillargeon

14. ABSTRACT
The Federal Railroad Administration contracted ENSCO to investigate the application of model-assisted probability of detection (MAPOD) to handheld ultrasonic rail flaw inspection. This research was conducted as part of a project that began in 2019 to investigate methods for evaluating the performance of track inspection technologies. The demonstration of MAPOD presented in this study points to several opportunities for further research. For example, hit-miss POD analysis could be extended to accurately show the probability of sizing flaws, and MAPOD simulations could include additional physical and human factors. Most importantly, the MAPOD approach developed in this study could also be adapted and applied to other track inspection technologies, such as track geometry measurement.

15. SUBJECT TERMS
Probability of Detection, MAPOD, Ultrasonic Rail Flaw Inspection

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**		80	**19b. TELEPHONE NUMBER** *(Include area code)*

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Acknowledgements

The authors are grateful to the human factors experts in the Federal Railroad Administration's Office of Research, Development, and Technology and the U.S. Department of Transportation's Volpe National Transportation Systems Center for reviewing and commenting on the human factors test plan used in this study.

Contents

Illustrations

Tables

Executive Summary

The Federal Railroad Administration (FRA) contracted ENSCO to investigate the application of model-assisted probability of detection (MAPOD) to handheld ultrasonic rail flaw inspection. This research was conducted as part of a project that began in 2019 to investigate methods for evaluating the performance of track inspection technologies.

Traditional probability of detection (POD) analysis uses results obtained from physical inspections of flawed samples, but confidence in the results is limited by the number of samples tested. MAPOD improves the confidence in these traditional POD estimates by simulating the inspection process with computer models.

Researchers studied the feasibility of applying MAPOD to six different types of track inspection technology. Ultimately, the team found the most feasible option for this research was handheld ultrasonic rail flaw inspection because of the availability of both commercial software for simulating ultrasonic inspection and sufficient data on rail flaw physical properties.

The POD for handheld ultrasonic rail flaw inspection depends on several physical and human factors. The research team studied the effects of the physical factors of rail profile and flaw location and conducted laboratory-based human factors testing to gather data on its effects on equipment calibration. This testing showed significant variation among operators. With a small exception, no two operators that participated produced statistically similar calibration results. Further, a single operator, using the same equipment on six occasions, produced inconsistent calibration results.

The team spent significant time setting up the computer model used to simulate ultrasonic rail flaw inspection. Simulations then required four months of continuous processing to complete, resulting in 512 predicted responses for known flaw sizes. This sample size would have been impractical to achieve by physical (i.e., laboratory or field) testing in the same timeframe. Hit-miss PODs were generated from the results showing the POD of detecting flaws occupying between 1 and 76 percent of the rail head area. The research team used the PODs to make confident statements about the performance of handheld ultrasonic inspection. For example, given the assumptions made in this research, there is a 90 percent probability that a flaw the size of 25 percent of the rail head area will be detected more than 187 times out of 200.

The demonstration of MAPOD presented in this study points to several opportunities for further research. For example, hit-miss POD analysis could be extended to accurately show the probability of sizing flaws, and MAPOD simulations could include additional physical and human factors. Most importantly, the MAPOD approach developed in this study could also be adapted and applied to other track inspection technologies, such as track geometry measurement.

1. Introduction

The Federal Railroad Administration (FRA) contracted ENSCO, Inc., to investigate the application of model-assisted probability of detection (MAPOD) to handheld ultrasonic rail flaw inspection. This research was conducted as part of a project that began in 2019 to investigate methods for evaluating the performance of track inspection technologies. This information will allow FRA to compare technologies, quantify improvements, assess the fitness for meeting regulatory requirements, and facilitate adoption of new technologies in the industry.

1.1 Background

Traditional probability of detection (POD) analysis uses results obtained from physical inspections of flawed samples, but confidence in the results is limited by the number of samples tested. MAPOD improves the confidence in these traditional POD estimates by simulating the inspection process with computer models.

FRA previously published ENSCO research in a report describing several potential evaluation methods for evaluating the effectiveness of track inspection technologies, including correlation, receiver operating characteristics, and POD (Bruzek, Maymand, Drape, Smart, & Tunna, 2020). A variation of one of these methods was model-assisted probability of detection (MAPOD). MAPOD is documented by the Department of Defense in Military Handbook 1823A (U.S. Department of Defense, 2009).

1.2 Objectives

This project was conducted in three phases with varying objectives:

- Phase I – Establish a uniform approach for the design and execution of performance evaluation testing for track inspection technologies. This phase was to include guidance on sample size requirements, repeatability and reproducibility, and acceptance criteria. A guidance document was published in 2022 (Bruzek, Drape, Einbinder, & Tunna).
- Phase II – Study the feasibility of applying MAPOD to various track inspection technologies and develop a roadmap for application. Summaries of this work were published in two Research Results reports (Bruzek, Feasibility of Applying Model-Assisted Probability of Detection to Track Inspection, 2021) (Bruzek, Model-Assisted Probability of Detection for Ultrasonic Rail Flaw Inspection, 2021).
- Phase III – Demonstrate MAPOD on a particular track inspection technology.

This report covers Phases II and III in detail.

1.3 Overall Approach

Phase I of this research established fundamental principles for evaluating inspection technologies. The research team used these principles in Phase II to study the feasibility of applying MAPOD to track inspection and select a candidate technology for demonstration. A roadmap for applying MAPOD to the selected technology also was developed in Phase II. The team followed this roadmap in Phase III for the selected inspection technology.

1.4 Scope

The scope of the research was limited to applying MAPOD to the track inspection technology that was most feasible. While the team studied several physical and human factors that affect POD, researchers did not consider all factors due to time constraints.

Although this research was focused on track inspection technologies, many of the methods described are applicable to other mechanical and electrical systems.

1.5 Organization of the Report

This report is organized as follows:

- Section 2 describes the basic POD method and discusses details of MAPOD.
- Section 3 presents the feasibility study and identifies the selected technology of handheld ultrasonic rail flaw inspection.
- Section 4 explains the roadmap for applying MAPOD to the selected technology.
- Section 5 describes the physical factors used in the MAPOD analysis and the data used to define them.
- Section 6 describes the human factors used in the MAPOD analysis and the test used to define them.
- Section 7 describes the software used to simulate handheld ultrasonic rail flaw inspection in the MAPOD analysis.
- Section 8 discusses the results from the MAPOD analysis and several sensitivity studies.
- Section 9 discusses the overall results and makes recommendations for further research.
- Section 10 provides a short conclusion to the research.

2. Model Assisted Probability of Detection

MAPOD combines experimental data and computer modeling to improve confidence in traditional estimates of POD. This section first describes traditional POD analysis then explains the MAPOD method.

2.1 Probability of Detection

The probability of detecting flaws in railroad track is a useful way of measuring inspection effectiveness. If, for example, a regulation requires flaws over a certain size to be removed from service, then it is useful to know how effective an inspection technology is at detecting that particular flaw size.

Figure 1 shows a typical POD curve in which the flaw size (in this example, the percent of rail head area) is plotted on the x-axis and the POD is plotted on the y-axis.

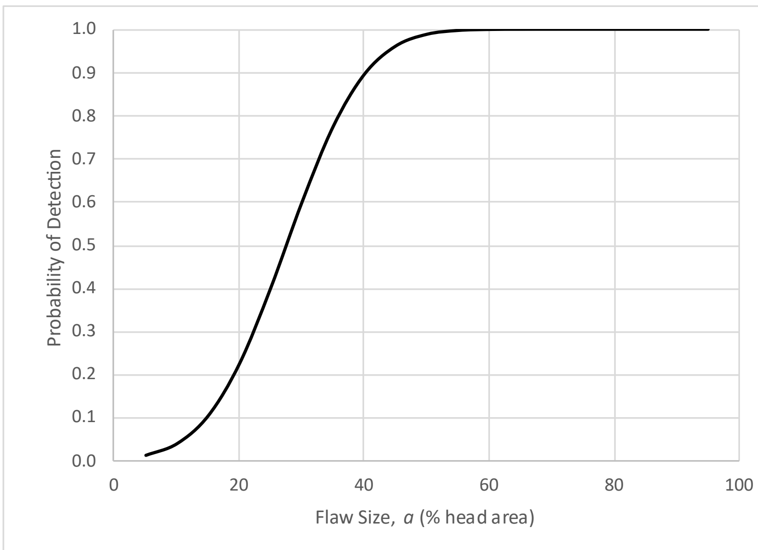


Figure 1. Typical POD curve

49 CFR Part 213 Track Safety Standards (Office of the Federal Register, 2021) require railroads to take certain actions when a transverse fissure exceeds 25 percent of the rail head. The inspection technology associated with the POD shown in Figure 1 has an approximately 0.4 probability of detecting this size of flaw. Knowing this allows a railroad to decide how often inspections should be repeated and whether a more effective technology should be sought. Therefore, it is important to generate accurate PODs.

A key step in calculating a POD curve is gathering data on the technology's response to known flaws. Figure 2 shows a typical dataset of these results (i.e., actual flaw size, or *a*) with response, (i.e., *a*-hat) plotted on the y-axis. The response is some output from the system, such as a voltage or a fraction of full-screen height.

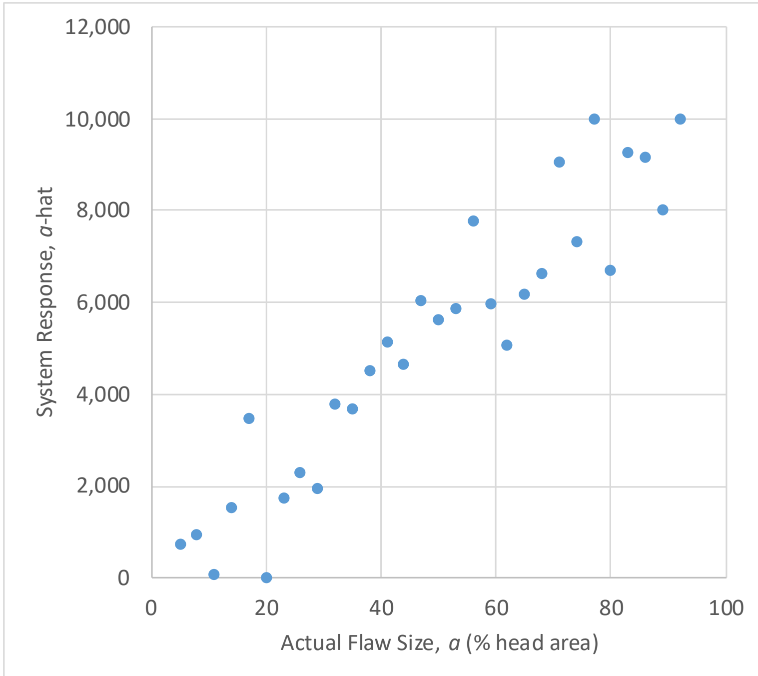


Figure 2. Typical dataset of inspection system response versus flaw size

The variation in the data in Figure 2 governs the shape of the POD curve in Figure 1. More variation shifts the POD curve to the right, indicating a lower probability of detecting a specified flaw size. Conversely, less variation shifts the POD curve to the left, indicating a more effective inspection technology. Furthermore, the quantity of data in Figure 2 governs the confidence in the POD curve, in that more available data results in higher confidence. Figure 3 shows the 90 percent confidence intervals surrounding the POD curve in Figure 1 using the 30 data points in Figure 2.

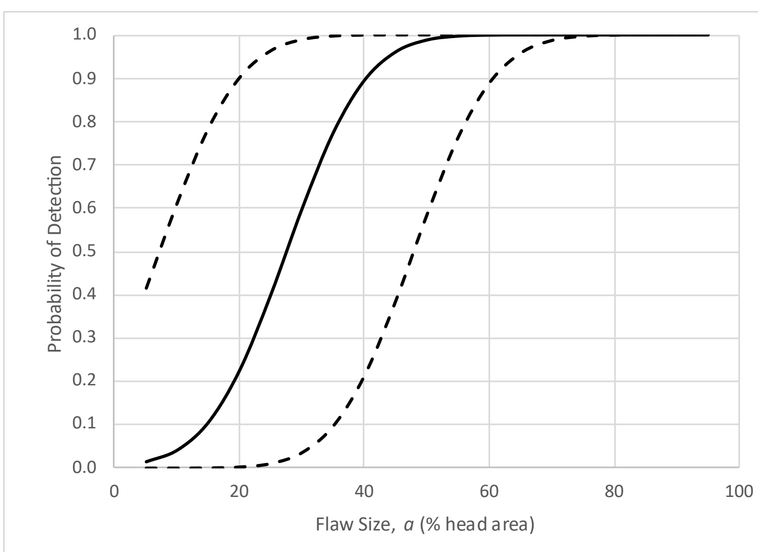


Figure 3. Typical POD with 90 percent confidence intervals

Due to limited data, the solid line in Figure 3 is only an estimate of the POD. In this example, although the estimated POD of a 25 percent flaw size is 0.4, there is only a 90 percent probability that the true POD for that flaw size lies anywhere between ~0.01 and ~0.96. This uncertainty makes it harder for the railroad to decide on inspection intervals. It also makes it difficult to compare one technology with another; the estimated POD of one technology may be higher than that for another, but the true POD may be different. It may be prudent to use a lower confidence interval estimate of the POD.

Increasing the sample size of known flaws improves confidence in POD estimates. Best practice is to have at least 30 test samples for each flaw size. However, this is impractical for most railroad-related applications involving track structure and substructure due to the cost of creating samples with known flaws, the need to avoid flaws growing during testing, and the need to control external factors, such as track support and weather. In practice, some test facilities provide the opportunity to gather limited field data. Two examples, both found at FRA's Transportation Technology Center (TTC), are the Rail Defect Test Facility (Sheehan, 2018) and the high-speed adjustable perturbation slab track (Stabler, Irani, Li, & Tajaddini, 2016).

To overcome the problem of insufficient physical samples, the U.S. Department of Defense helped develop MAPOD. MAPOD combines results from a small number of physical experiments with a large number of predictions from computer models. It was developed in collaboration with the nuclear and aeronautical industries and has been documented in a military handbook (U.S. Department of Defense, 2009).

The MAPOD method can result in thousands of *a* versus *a*-hat data points. If the computer model has been validated, this leads to a high level of confidence in the resulting POD.

2.2 MAPOD Method

Figure 4 summarizes the MAPOD method described in Military Handbook 1823A (U.S. Department of Defense, 2009).

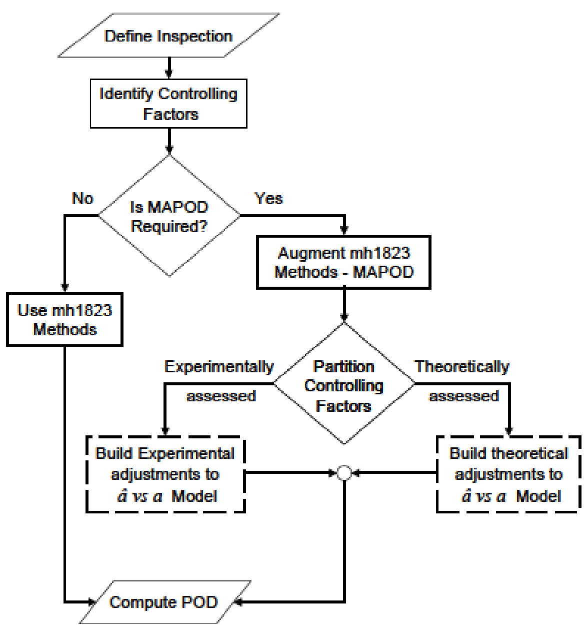


Figure 4. Process flow for MAPOD (U.S. Department of Defense, 2009)

After defining the type of inspection, the next step is to identify the controlling factors that can affect the POD. If these factors show that MAPOD is not necessary, then the traditional analysis described in Military Handbook 1823 is performed. The confidence in the resulting POD will be limited by the sample size of known flaws.

If MAPOD is applied, the first task is to examine the factors that can lead to variation in *a* versus *a*-hat results. These factors are partitioned into those that will be assessed experimentally and those that will be assessed theoretically with a computer model. It is important to note that part of the theoretical assessment is the validation of the computer model.

Bruzek, et al., described several different MAPOD methods (Bruzek, Maymand, Drape, Smart, & Tunna, 2020). The MAPOD method used in this research is non-Bayesian, Full Model Assisted, in which a computer model accounting for the physical factors is combined with knowledge of human factors to estimate the total variability in an inspection technology.

3. Feasibility of MAPOD for Track Inspection

The primary objective of this research was to demonstrate the application of the MAPOD method to a track inspection technology. This section of the report describes how the track inspection technology was chosen from six candidates:

1. Rail flaw detection by ultrasonic inspection
2. Track geometry measurement by onboard, inertial systems
3. Gage restraint measurement by track-loading vehicles
4. Joint bar condition using camera images
5. Ballast condition measurement by ground penetrating radar (GPR)
6. Grade crossing profile measurement using LiDAR images

The research team identified the controlling factors for each of the six candidate types of track inspection. They then decided whether each factor would be assessed experimentally or analytically. Details of this analysis are included in Appendix A.

To assess a controlling factor experimentally, physical tests are conducted with sample flaws. Conversely, analytical assessment involves computer modeling of the flaw detection technology. Therefore, use of the MAPOD method depends on the availability of both test facilities and computer models.

Researchers used the following criteria and ratings to rank the six candidate technologies. The ratings range from 1 to 5, where a low rating implies high feasibility.

- Model – Does computer software exist (rating of 1), or does it need to be developed? If the latter, how difficult would that be (rating of 5 for the most difficult)?
- Data – Does the model require a small, medium, or large effort to create databases of input variables (ratings of 1, 3, and 5, respectively)?
- Test Facility – Does a facility exist for physical testing (rating of 1), or does one need to be created? If the latter, how difficult would that be (rating of 5 for the most difficult)?
- Experiment – Is the experimental effort small, medium, or large (ratings of 1, 3, and 5, respectively)?

Table 1 shows the rating results for the candidate inspection technologies.

Table 1. Rating results for the feasibility of applying MAPOD

Inspection Technology	Model	Data	Test Facility	Experiment	Total
Ultrasonic Rail Flaw Detection	1	5	1	3	10
Inertial Track Geometry Measurement	5	3	1	3	12
Gage Restraint Measurement	5	1	5	3	14
Joint Bar Condition by Camera Image	5	1	5	3	14
Ballast Condition by GPR	3	5	5	5	18
Grade Crossing Profile Measurement by LiDAR	5	5	5	3	18

3.1 Ultrasonic Rail Flaw Detection

Researchers found ultrasonic rail flaw detection to be the most feasible inspection type for MAPOD. Commercial software is readily available to model the effect of several factors, such as flaw shape and size, and mechanical noise. This significantly reduces the need for physical tests, which would be necessary only to validate the model and can be done at the Rail Defect Test Facility at TTC (Sheehan, 2018). The model itself would require a large database of input variables including rail profile and material; flaw type, size, orientation, and shape; and surface condition, but some of this information is available from previous research and testing efforts.

It is important to note that human errors can arise with handheld ultrasonic rail flaw detection devices and, in turn, affect the POD. For example, calibration errors can cause excessive false positives and false negatives. These human factors must be assessed experimentally.

3.2 Inertial Track Geometry Measurement

The team found track geometry measurement to be a feasible alternative candidate technology. While a specific model does not presently exist, current vehicle-track interaction modeling software can simulate some aspects of the technology. Additional software models would need to be developed to simulate the optical and laser systems used by this technology. The resulting model could then be validated with tests on the high-speed adjustable perturbation slab track at TTC. It would be manageable to create a database of inputs for modeling the inertial track geometry measurement; 49 CFR Part 213 Track Safety Standards (Office of the Federal Register, 2021) features a well-defined set of track geometry deviations that could be used as inputs.

3.3 Gage Restraint Measurement

The team found this technology to be less feasible. A model would need to be developed for applying MAPOD to gage restraint measurements since commercial software is not readily available. Also, a full-scale test facility would be needed to validate the model and cover the wide range of possible track components, although it is possible that the High Tonnage Loop at TTC could be adapted for this purpose.

3.4 Joint Bar Condition

The team also found this technology to be less feasible. Image-based joint bar condition inspection would require model development as well as a dedicated test facility, although a database for model input could be created from existing images of flawed joint bars.

3.5 Ballast Condition by GPR and Grade Crossing Profile by LiDAR

The most challenging candidates for MAPOD analysis were ballast condition measurement by GPR and grade crossing profile measurement by LiDAR. Due to the wide range of possible ballast conditions and crossing profiles, both technologies require large databases of model inputs and test facilities capable of validating the model. The software available for ultrasonic testing could possibly be adapted for GPR, but a LiDAR model would require significant effort to develop and validate.

4. Methodology

This section discusses the main steps in the MAPOD method flow chart shown in Figure 4.

4.1 Define Inspection Technology

For the purposes of this demonstration, the research team defined handheld ultrasonic rail flaw inspection as the type of inspection to be assessed (see Section 3). This type of track inspection has been widely used in the industry for over 20 years, typically performed to confirm rail flaws detected by automated, vehicle-mounted inspection systems. Although vehicle-mounted inspection can also be used with MAPOD, the team chose handheld inspection for simplicity.

The American Society for Nondestructive Testing (ASNT) recommends ultrasonic testing practices and certifies test operators (ASNT, 2022). Operators certified to Level II have passed both a general examination and one specific to a type of testing technology. Level III operators have passed further examinations and gained significant practical experience in testing.

The American Railway Engineering and Maintenance-of-Way Association (AREMA) makes recommendations for operator certification and inspection procedures (AREMA, 2022).

4.1.1 Equipment

The team used an ultrasonic transducer fitted to a plastic wedge and connected by a cable to an ultrasonic flaw detector. The transducer and wedge together are referred to as the probe. Figure 5 shows the typical equipment setup.

Figure 5. Typical ultrasonic wedge, transducer, cable, and detector

The transducer emits an ultrasonic pulse that is transmitted through the wedge and into the rail. The pulse reflects off internal flaws and exterior rail surfaces and is transmitted back through the wedge and received by the transducer.

The detector can display the results in several different modes. The A-scan display shows the amplitude of the reflected pulse plotted against the distance it has travelled in the rail. The A-scan changes as the probe moves, which allows the operator to find flaws. B-scan and C-scan displays are discussed in Section 8 of this report.

An operator uses the equipment by sliding the probe along the top of the rail while looking for peaks on the detector's display. Peaks that do not correspond to reflections from the rail's exterior surfaces can indicate the presence of internal flaws. Once a flaw has been found it can be sized by sliding the probe backward and forward, and from side to side, repeatedly on the top of the rail. This is typically performed by finding the scan positions that result in half the maximum signal response. In these positions, half the ultrasonic beam is reflected from the flaw and the other half continues into the rail. This method is suitable for flaws larger than the area covered by the ultrasonic beam, but adjustments are necessary for smaller flaws.

It is important to note that scanning toward the gage face of the rail can be problematic. Since the face of the probe in contact with the rail is flat, it may not make good contact with a rail that has heavy gage face wear. In this situation, operators may adjust the longitudinal alignment of the probe or use probes with alternative refraction angles.

4.1.2 Calibration

The equipment should be calibrated before it can be used. Calibration is typically performed at the start of the day's inspection and after any of the equipment's components are changed. The North American railroad industry does not have a standardized calibration procedure; each organization follows its own procedure when it determines calibration is necessary.

The following four factors are typically set during calibration:

- Zero, to account for the distance the ultrasonic wave travels in the wedge
- Velocity of the ultrasonic wave in steel
- Angle of refraction of the wedge
- Gain to give a peak response at 80 percent full screen height for a known flaw

Calibrating zero and angle for a new wedge follows the manufacturer's specification. The zero will change when the wedge wears, and the angle can change if the wedge does not wear uniformly. Thus, zero and angle can change independently as the wedge wears. The velocity of ultrasonic waves in steel is typically set to a fixed value and not altered during calibration. Since its value directly affects the other calibration factors, including it in the calibration process ensures the most complete and accurate set of factors.

A fixed gain allows signals found during operation to be compared to the known flaw size used in the calibration. However, operators typically adjust the gain during operation to allow more sensitivity when looking for the edges of flaws.

Calibration is typically performed on a steel block manufactured to International Institute of Welding (IIW) specifications. Figure 6 shows an IIW Type 2 calibration block. During calibration, the probe is placed on the top surface of the block shown in Figure 6 and moved from left to right repeatedly. The block can be inverted to use the bottom surface in the same way. The 4-inch radius on the left of the block in Figure 6 is used to calibrate the zero. Behind it, on the back of the block, is a 2-inch radius that can be used in combination with the 4-inch radius to calibrate the velocity. The 1.5-inch diameter hole on the top right of the block is used to calibrate the angle. The 1/16-inch diameter hole on the bottom right of the block is typically used to calibrate the gain.

Figure 6. IIW Type 2 calibration block

Calibration requires skill and concentration from the operator, as the calibration factors are sensitive to small changes in probe position. Lubricant must be applied between the probe and the block. The operator must constantly adjust the detector's settings while monitoring the detector's display and the position of the probe on the block.

4.1.3 Rail Flaws

FRA separates rail defects into four categories: Transverse, Longitudinal, Web, and Miscellaneous (FRA, 2015). A flaw becomes a defect when its size is equal to or greater than a limit specified in FRA regulations (Office of the Federal Register, 2021).

The research reported here focuses on the following types of transverse flaws (FRA, 2015):

- Transverse Fissure – a progressive fracture originating from a nucleus inside the head of the rail, typically no closer than 1/4 inch to any surface
- Detail Fracture – a progressive fracture originating at or near the surface with no nucleus at the origin that commonly originates from a shell on the surface
- Compound Fissure – a progressive fracture originating from a horizontal split head that can grow up or down, or in both directions

These types of flaws are grouped together because they are often difficult to distinguish unless the rail is broken open. Together they account for more than half of rail-related train accidents in recent years (Poudel).

Table 2 summarizes the current limits for these types of flaws (Office of the Federal Register, 2021). The limits are shown in terms of the percentage of the existing rail head area occupied by the flaw. Three limits are shown for compound fissures to distinguish between small, medium, and large flaws of this type. Similarly, four limits are displayed to distinguish between different sized transverse fissures and detail fractures.

Once a flaw reaches the limits shown in Table 2, the rail should be removed or replaced, or some other corrective action should be taken. The corrective action depends on the severity of the defect and may include speed restrictions, applying joint bars, and visual supervision. For example, the corrective action for a transvers fissure in the range 25 to 60 percent of the rail head area is to apply joint bars and reduce the operating speed.

Table 2. Rail defect size limits (Office of the Federal Register, 2021)

Flaw Type	Less than	But not less than
Compound Fissure	70	5
	100	70
		100
Transverse Fissure or Detail Fracture	25	5
	60	25
	60	100
		100

4.2 Controlling Factors

The next step in the MAPOD method is to identify the factors that control the POD. These are separated into physical and human factors.

4.2.1 Physical Factors

Several physical factors may affect the POD.

4.2.1.1 Surface Condition

The coupling between the ultrasonic probe and the rail can be affected by the smoothness of the rail's surface. It can also be affected by the presence of rust, grease, or other deposits, as well as surface defects, including rolling contact fatigue.

4.2.1.2 Flaw Geometry

A large range of flaw sizes, locations, orientations, and shapes is possible. The face of the flaw can be rough, polished, or patterned. The flaw may break the surface of the rail or may be surrounded by rail material.

4.2.1.3 Rail Type and Shape

Many different rail cross-sections are possible. The head of the rail can be new or worn to different degrees. It can also have material flow on the gage and field faces, and the foot of the rail can be worn to different degrees. Furthermore, rail material can vary in chemical composition and hardness, and hardness can even change across the rail's cross section.

4.2.1.4 Non-flaw Features

Flaws near rail joints and welds may be difficult to detect as rail joints introduce extra surfaces for the ultrasonic signal to reflect against. Welds introduce variations in material properties.

4.2.2 Human Factors

Since the technology is manually operated, human factors can have an influence on the probability of detecting flaws. The following types of errors are possible:

- Incorrect calibration

- Insufficient lubrication of the wedge-rail interface
- Variation in vertical and horizontal alignment during operation
- Misreading A-scan indications
- Incorrect flaw sizing
- Incorrect reporting

These errors can be affected by an operator's level of experience, training, fatigue, and stress.

4.3 Data Model: *a* versus *a*-hat

The commercially available CIVA NDE 2021 modeling software used in this research can generate *a* versus *a*-hat data (EXTENDE, 2022). Specifically, CIVA has a module for simulating ultrasonic inspection which has been used extensively by non-destructive testing researchers (Foucher, Lonne, Toullelon, Mahaut, & Chatillon, 2018). FRA previously funded research to develop CIVA modeling capabilities for rail inspection (Poudel).

A CIVA model is developed by specifying the geometry of the specimen, properties of the flaw, details of the probe and wedge, and positioning of the wedge relative to the flaw. The user can specify whether these inputs are fixed or variable. Variable inputs are defined by standard mathematical distributions or by user-defined distributions.

The physical factors identified in Section 4.2.1 can readily be modeled in CIVA. Additionally, some of the human factors identified in Section 4.2.2 also can be represented in CIVA. For example, calibration factors relate directly to CIVA inputs and can be quantified experimentally.

A CIVA simulation is performed using the fixed inputs and a random selection from the variable inputs. A scan is performed along and across the specimen. By performing multiple simulations, CIVA can generate a large dataset of *a* versus *a*-hat results. This dataset can be exported for post-processing.

4.4 POD Calculation

CIVA has the functionality to calculate POD curves with confidence intervals from *a* versus *a*-hat data. The user specifies the confidence level and the noise and detection thresholds.

A useful feature of CIVA is the ability to calculate PODs from imported *a* versus *a*-hat data. This allows the user to post-process the original CIVA results and import them back to CIVA for POD analysis.

5. Physical Factors

Researchers studied the effects of rail shape and flaw geometry on POD to demonstrate how CIVA handles physical factors. Other physical factors, such as surface condition, the presence of surface defects (e.g., rolling contact fatigue), and the effects of rail joints and welds, could be included in future research.

5.1 Rail Shape

Rail shape can have a significant effect on the POD of rail flaws for three main reasons. First, the percentage of rail head area occupied by a flaw of a certain size and position depends on the amount of rail wear. For example, a flaw in a worn rail occupies a larger percentage of the head area than the same flaw in a new rail. Second, the shape of the surfaces on the sides of the rail can affect internal reflections of the ultrasonic signal. The characteristics of the reflected signal from an unworn rail with straight sides would be different to those from a rail with lateral wear resulting in a concave side. Third, the ultrasonic coupling between the probe and the rail depends on the local curvature of the rail's cross-section. It may not be possible to get a good coupling on the gage corner of rail with heavy side wear. Thus, the variation in rail profiles needs to be included when determining the POD from ultrasonic rail flaw inspection. Ignoring this effect would produce a POD that was only applicable to the chosen rail shape.

The research team identified a relatively small subset of rail shapes that could be used to represent the variation found in revenue service. Since rail on tangent track is expected to have little or no lateral wear, it seemed reasonable to analyze tangent and curved track separately. For simplicity, track with a curvature less than 1 degree was considered to be tangent.

Measurements collected by various inspection vehicles as part of the FRA Automated Track Inspection Program (ATIP) were used to determine the variation in rail profiles in revenue service. The dataset included 16,111 rail profile measurements from a random sample of surveys covering the U.S. rail network made between February 2010 and November 2015. Table 3 shows the distributions of curvature and track class in the dataset. Track classes 1 and 2 were not represented in the dataset due to the limited availability of such data.

Table 3. Curvature and track class distributions in the rail profile dataset

Curvature (deg.)	Class 3	Class 4	Class 5	All Classes	All Classes (percent)
0 to 0.2	1884	5439	2598	9921	61.6
0.2 to 1	504	748	1002	2254	14.0
1 to 2	596	878	323	1797	11.2
2 to 3	503	423	2	928	5.8
3 to 4	350	202	0	552	3.4
4 to 5	351	4	0	355	2.2
5 to 6	204	0	0	204	1.3
6 to 7	100	0	0	100	0.6
> 7	0	0	0	0	0.0
	4492	7694	3925	16111	

Each profile in the dataset was analyzed to determine the following wear parameters:

- Vertical wear – the difference between the new and worn heights at the rail centerline
- Lateral wear – the difference between the new and worn widths at a point 5/8 inch below the top, center of the worn rail at the rail centerline
- Head loss – the difference between the new and worn rail head areas

Figure 7 shows the distribution of lateral head wear for the entire rail profile dataset. The counts for bins greater than 0.25 in are too small to be seen in Figure 7.

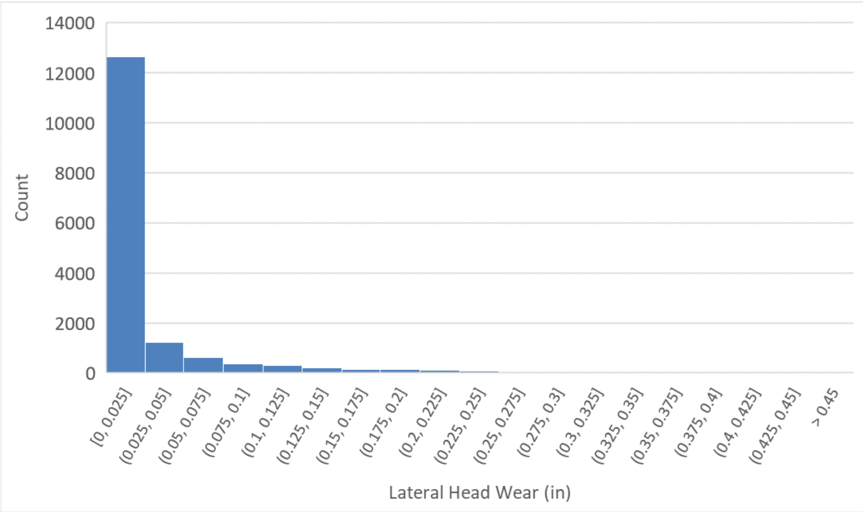


Figure 7. Distribution of lateral head wear

Figure 7 shows 78.5 percent (12,651/16,111) of profiles in the sample had lateral headwear less than 0.025 inch. This is consistent with the result from Table 3 where 75.6 percent of the dataset had a curvature less than 1 degree.

Figure 8 shows the distribution of vertical head wear for the part of the dataset with curvature less than 1 degree. The total count in Figure 8 is 12,176.

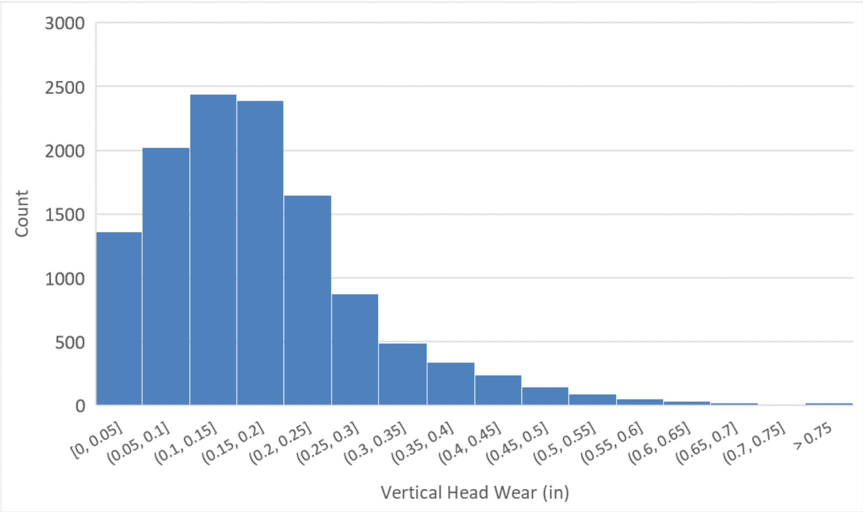


Figure 8. Distribution of vertical wear – curvatures < 1 degree

For this research, the distribution of vertical head wear on relatively straight track (i.e., curvature less than 1 degree) shown in Figure 8 was represented by the three profiles described below. It is important to note that the profiles used to represent vertical wear should not have any significant lateral wear due to the lack of notable curvature.

- Lightly worn: vertical wear < 0.05 in. – 8.4 percent (1,359 count)
- Moderately worn: vertical wear ~0.15 in. – 45.9 percent (7,393 count)
- Heavily worn: vertical wear > 0.25 in. – 21.3 percent (3,424 count)

Figure 9 shows the distribution of lateral wear on the part of the sample with curvature greater than or equal to 1 degree. The counts for bins greater than 0.45 are less than 10 and too small to be seen in Figure 9. The total count in Figure 9 is 3,935.

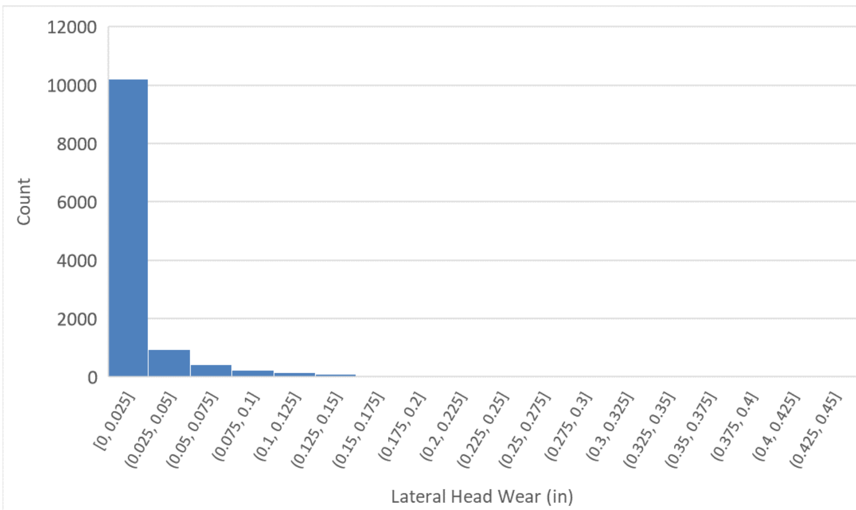


Figure 9. Distribution of lateral wear – curvatures >= 1 degree

For this research, the distribution of lateral head wear on curved track (i.e., curvature greater than or equal to 1 degree) shown in Figure 9 was represented by two profiles, described below. The profiles used to represent lateral wear are likely to also have vertical wear.

- Lightly worn: lateral wear < 0.025 in. – 15.2 percent (2,454 count)
- Moderately worn: lateral wear ~0.15 in – 9.2 percent (1,481 count)

Table 4 shows the wear measurements for the profiles available in FRA's RF-LOAD library. Some of these profiles had already been used in CIVA modeling (Poudel).

Profile S3a from Table 4 was used for the lightly worn tangent rail. It has a small amount of vertical wear and very little lateral wear or head area change.

Profile S7a in Table 4 was used for the moderately worn tangent rail. It has vertical wear that matches the central part of the distribution in Figure 8 and very little lateral wear.

None of the profiles in Table 4 were suitable for the heavily worn tangent rail. The profiles with vertical wear greater than 0.25 inch are all from rails in curves and have significant side wear. Profile E2a was artificially created by adjusting profile S7a to give 0.33 inch of vertical wear.

Profile N11a in Table 4 was used for the lightly worn rail in curves.

Profile W4a in Table 4 was used for the moderately worn rail in curves. The wear values in Table 4 are measured at 0.5 inch below the top of the rail. The lateral wear of profile W4a is approximately 0.165 inch at 5/8 inch below the top of the rail. This makes it suitable for representing the tail of the distribution in Figure 9.

Table 4. Wear measurements for library profiles

	Head Wear (in)	Side Wear[1] (in)	Area Loss (in^2)
N1a	0.039	0.056	0.097
N2a	-0.002	0.002	(0.012)
N3a	-0.001	0.034	0.020
N4a	-0.001	-0.003	(0.015)
N5a	-0.001	0.034	0.013
N6a	0.006	0.004	0.008
N7a	-0.001	0.031	0.016
N8a	-0.003	0.030	0.010
N9a	0.000	-0.001	(0.012)
N10a	0.002	0.045	0.035
N11a	0.008	0.002	0.008
N12a	0.005	0.038	0.038
N13a	0.006	-0.001	(0.001)
N14a	0.014	0.052	0.063
N15a	-0.006	0.002	(0.023)
N16a	0.009	0.000	0.009
S1a	0.041	-0.005	0.051
S2a	0.030	0.002	0.032
S3a	0.021	-0.004	(0.002)
S4a	0.019	-0.002	0.017
S5a	0.171	-0.008	0.362
S6a	0.179	0.012	0.396
S7a	0.190	-0.004	0.401
S8a	0.164	0.001	0.342
W1a	0.133	0.376	0.616
W2a	0.180	0.394	0.697
W3a	0.162	0.333	0.600
W4a	0.140	0.355	0.615
W5a	0.129	0.345	0.571
W6a	0.157	0.438	0.713
W7a	0.133	0.364	0.600
W8a	0.144	0.402	0.666

[1] Measured 0.5 inch below the top, center of the worn rail.

The selected rail profiles used for the POD analysis are shown in Appendix B.

Table 5 lists the rail profiles used in the POD analysis. The proportions of the network for each profile were calculated from the relevant parts of the distributions in Figure 8 and Figure 9.

Table 5. Rail profiles and proportions used for POD analysis

Condition	Profile	Proportion of Network
Lightly worn tangent	S3a	0.084
Moderately worn tangent	S7a	0.459
Heavily worn tangent	E2a	0.213
Lightly worn curve	N11a	0.152
Moderately worn curve	W4a	0.092

5.2 Flaw Geometry

The typical shapes of transverse rail flaws were determined from published literature and collected samples of broken rails. Images were gathered for transverse fissures, detail fractures, and compound fissures. Figure 10 shows an example of a broken rail taken from the Facility for Accelerated Service Testing (FAST) at TTC (Banerjee, Morrison, Witte, & LoPresti, 2021).

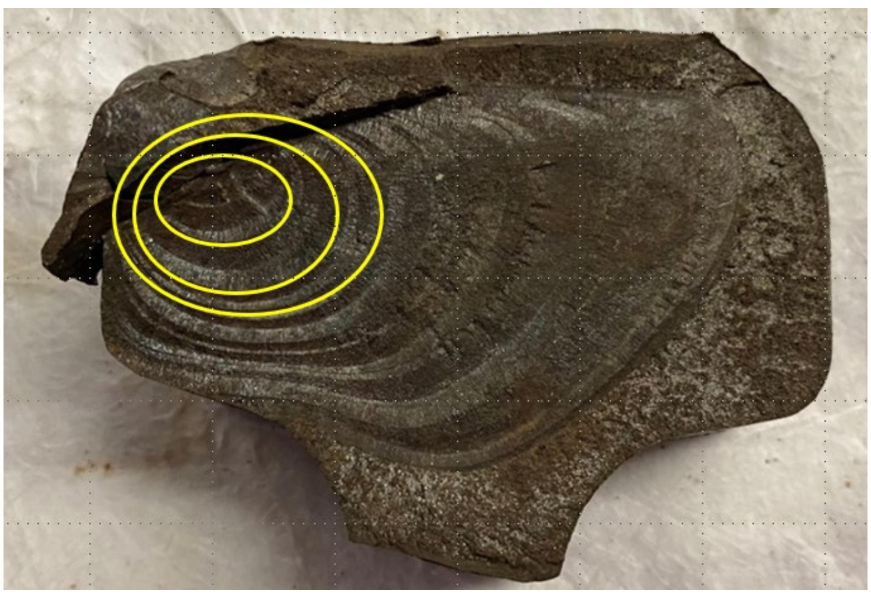


Figure 10. Example detail fracture from TTC

A train used at FAST periodically changes direction, which causes slight changes in the direction that flaws grow. As shown in Figure 10, this provides the opportunity to study the shape of flaws as they grow.

Figure 10 shows ellipses fitted to the shape of the flaw at three instances during its growth. The width and height of these ellipses were measured. Figure 11 shows the results of this analysis on 53 ellipses from 24 images.

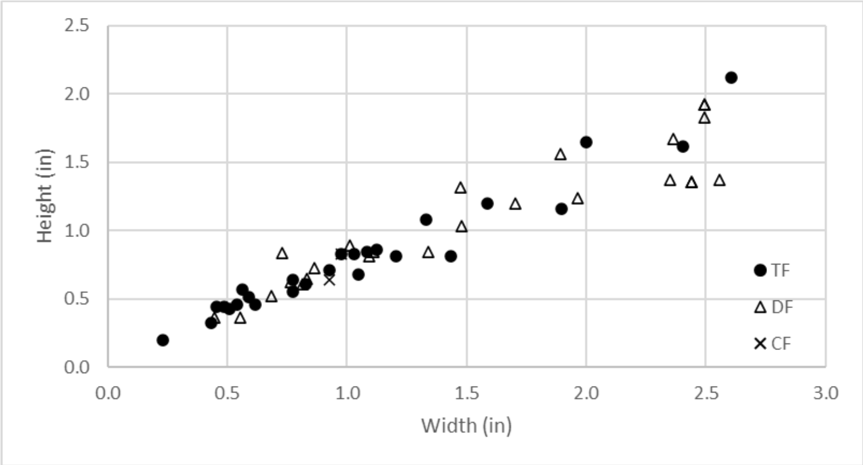


Figure 11. Flaw height vs. width

Figure 11 shows an approximately linear relationship between flaw width and height. A best fit straight line to the data had a slope of 0.71 and a R^2 value of 0.98. Thus, it was assumed that the shape of transverse rail flaws could be approximated as an ellipse with a height-to-width ratio of 0.71.

The team also recorded the coordinates of the center of each ellipse. The lateral position was measured from the rail's centerline (positive toward the gage face) and the depth was measured from the top of the rail at the rail centerline. Figure 12 shows the lateral position plotted against depth for the 53 ellipses measured.

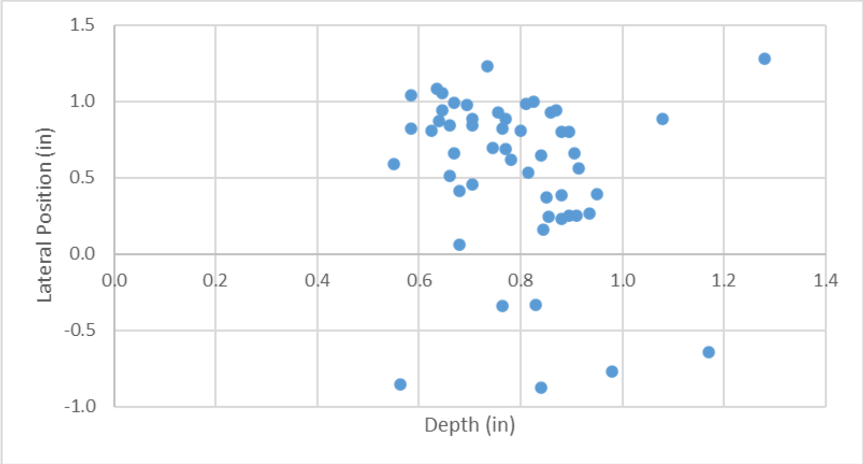


Figure 12. Flaw lateral position vs. depth

Figure 12 shows no clear relationship between flaw lateral position and depth. Further analysis also showed no relationship between the position of the flaw and its width. Thus, it was assumed that flaw lateral position and depth could be represented in CIVA by two independent distributions. Figure 13 and Figure 14 show the distributions of flaw lateral position and depth from the analysis of the 53 ellipses.

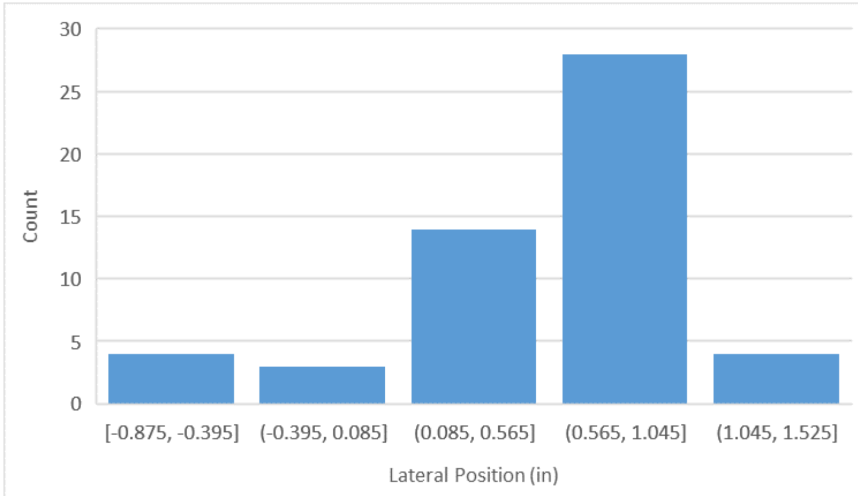


Figure 13. Distribution of flaw lateral position

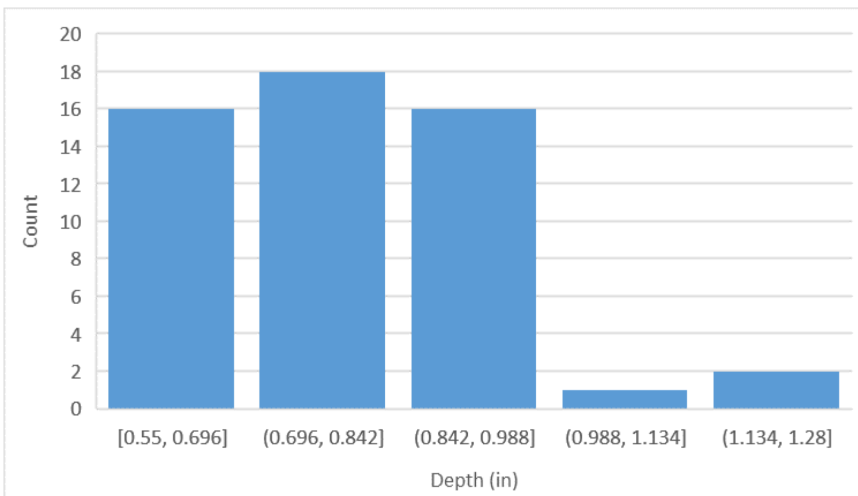


Figure 14. Distribution of flaw depth

6. Human Factors Tests

Section 4.2.2 identified a wide range of human factors that can affect the POD derived from handheld ultrasonic rail flaw inspection. The scope of this research was limited to examining the effect of equipment calibration. The team conducted a test to quantify the variation in calibration that occurs in practice.

6.1 Test Plan and Pre-Test

Researchers developed a test plan in discussion with FRA and Volpe human factors experts. Table 6 shows the risks to successful completion of the tests and the mitigation measures identified in the test plan.

Table 6. Calibration test risks and mitigations

Risk	Mitigation
Personal safety	Use experienced operators in a safe, comfortable environment
Equipment malfunctions	Test the equipment and perform dry runs before the test day
Operators behave differently because they know they are being tested	Make it clear in the briefing that operators will be anonymous
Insufficient statistical confidence in the results	Analyze variation in the results after each day's testing and add more days if necessary

The team developed a briefing note for test participants that emphasized several points, including: 1) operators' names would not be recorded (i.e., results are anonymous); 2) calibration would follow normal procedure; 3) care and time taken should be the same as normal operations; 4) there would be no correct or incorrect calibrations; 5) there would be no competition between operators; and 6) preliminary results would be shared at the end of the test day.

A dry run of the test was performed at ENSCO's facilities in Chambersburg, PA, on October 12, 2021. Three operators participated in the dry run using two sets of equipment. While this meant that one operator was free to do other tasks, the time available for other tasks was unpredictable. A recommendation from the dry run was to have equipment for each operator.

The research team developed a calibration procedure for the operators to follow in the dry run. Some operators were unfamiliar with this procedure and the equipment that was provided, which they found a little intimidating. The test plan was modified to have operators in future use equipment and follow procedures with which they are familiar.

The operators in the dry run felt that the precautions taken to ensure anonymity would work in practice. They did not report feeling under peer or time pressure.

6.2 Test Objective

The objective of the human factors test was to determine the probability distributions of three calibration factors: 1) zero, to account for the distance the ultrasonic wave travels in the wedge; 2) the angle of refraction of the wedge; and 3) the gain to give a peak response at 80 percent full screen height for a known flaw. The fourth calibration factor, velocity of sound in steel, was assumed to be constant.

The objective was not to determine the operator’s skill at detecting and sizing flaws, which would require a more comprehensive test than the one described here.

6.3 Test Scope

The equipment to be calibrated included a 70-degree wedge, 2.25 MHz transducer, and an IIW Type 2 calibration block.

The operators performed as they would in normal conditions. Since calibration in the field is normally done inside a vehicle, the test was performed indoors. The scope did not include determining the effects of weather and temperature.

The operators maintained lubrication between the wedge and the calibration block as they would in the field. They were given the amount of time they would normally take for field calibration. The scope did not include determining the effects of operator experience, fatigue, or stress.

6.4 Test Execution

The test was conducted at the TTC on December 7, 2021. The operators used the following equipment:

- Three 1/2-inch ultrasonic transducers operating at 2.25 MHz
- Three IIW Type 2 test blocks
- Two Olympus Epoch 650 ultrasonic flaw detectors
- One Olympus Epoch 600 ultrasonic flaw detector
- Ten 70°-degree wedges (see Figure 15) with the properties shown in Table 7

Table 7. Wedge properties

Wedge #	Condition	Front Height (mm)	Rear Height[2] (mm)
1	Worn	18.63	18.64
2	Worn	19.54	19.25
3	Worn	19.51	19.48
4	New	19.76	19.83
5	Worn	19.17	19.30
6	New	19.73	19.88
7	New	19.76	19.84
8	New	19.74	19.84
9	Worn	19.32	19.28
10	Worn	19.47	19.65

[2] Measured at the end of the red surface near the threaded hole in Figure 15.

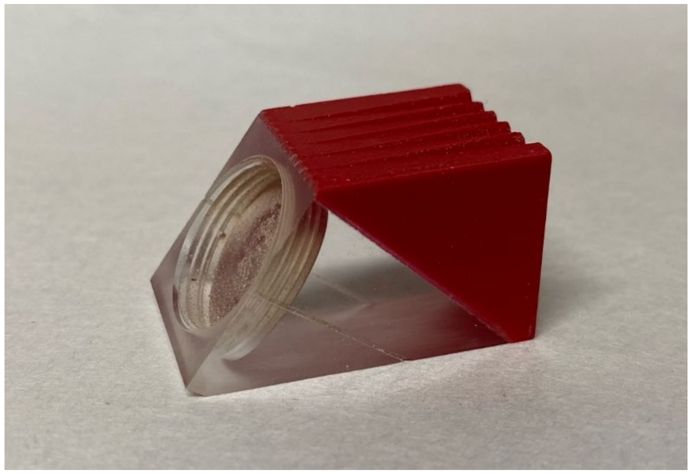


Figure 15. A 70° wedge of the type used in the calibration tests

The three operators that took part were identified by the letters A, B, and C to ensure the results were anonymous. Table 8 shows the operators' ASNT level and the ultrasonic detector each used.

Table 8. Operators' ASNT level and detector

Operator	ASNT Level	Detector
A	II	Epoch 650
B	III	Epoch 600
C	I	Epoch 650

The test was performed in a large room with calibration equipment set up on three separate tables.

Testing was conducted as follows:

1) Each operator drew a wedge from a bag at random and began a calibration.
2) Once a calibration was completed:
    a) The operator letter, wedge number, and the calibration factors were recorded.
    b) The wedge was returned to the bag of wedges.
    c) The operator drew another wedge from the bag at random and began another calibration. If the wedge was the same as the last one used, it was used again.
3) Step 2 was repeated until the end of the day's testing.

A lunch break was taken after 66 calibrations had been recorded. By the end of the day, 121 calibrations had been completed and recorded.

Each operator used the same calibration procedure with which they all were familiar. The steps in the procedure were as follows:

1) Set the velocity to 0.128 in/µs.

2) Connect the transducer to the wedge.
3) Apply lubricant to the top of the calibration block and slide the probe until the peak response from the 1.5-inch diameter hole is observed.
4) Set the wedge angle to that indicated by the wedge index point on the scale on the calibration block.
5) Apply lubricant to the top of the calibration block and slide the probe until the peak response from the 4-inch radius is observed.
6) Adjust the zero until the peak response aligns with 4 inches on the display's horizontal axis.
7) Turn the calibration block over, apply lubricant to the top, and slide the probe until the peak response from the 1/16-inch diameter hole is observed.
8) Adjust the gain until the peak response is at 80 percent of full screen height.

6.5 Test Results

The full set of test results is shown in Appendix C. The following sub-sections provide details of the analysis performed on the results.

6.5.1 Operator

The team analyzed the results by operator to determine operator consistency. If the analysis produced consistent results, then future research may only need to consider single operators. If the analysis showed significant inconsistency, then a larger cohort of operators may be necessary.

Table 9 shows the results of six calibrations performed by the same operator on the same wedge. The results are presented in the order performed.

Table 9. Example of repeat calibrations

Angle (deg.)	Zero (µs)	Gain (dB)
68.0	11.622	60.2
69.0	12.542	61.1
69.0	11.629	61.0
68.0	12.991	64.3
68.0	11.799	61.2
70.0	12.638	61.3
Min. 68.0	11.622	60.2
Max. 70.0	12.991	64.3

Table 9 shows there is inherent variation in the outcome of the calibration process. The same operator can produce different calibration factors using the same wedge, transducer, and ultrasonic device on the same day.

Figure 16 shows box plots comparing the variation in angle calibration results for the three operators and the same wedge.

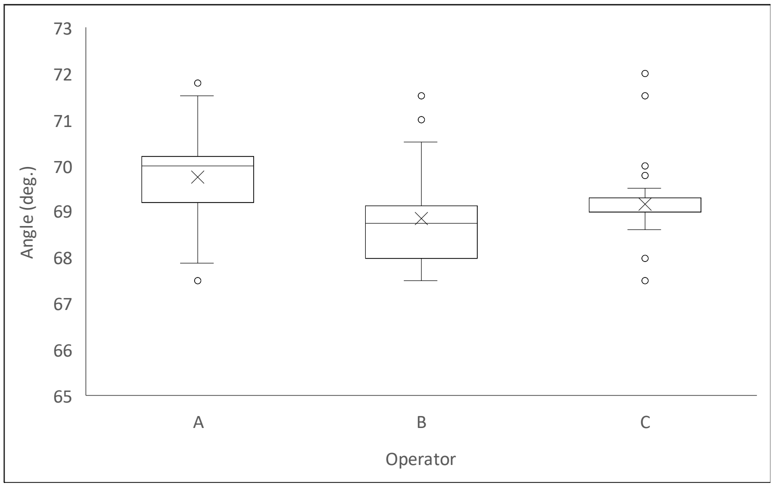


Figure 16. Variation in angle for each operator

Figure 16 indicates some differences between operators in calibrating angle, and all operators produced outliers. Table 10 shows summary statistics for the angle results from each operator.

Table 10. Angle statistics by operator

Operator	Count	Sum	Average	Variance
A	40	2791	69.76	0.6957
B	54	3718	68.84	0.8191
C	27	1868	69.17	0.8169

Table 11 shows the results of single-factor ANOVA on the angle results from the three operators with an alpha (α) level of 0.05. The null hypothesis is that all operators calibrated angle with similar results.

Table 11. ANOVA results for angle calibration between operators

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	19.54	2	9.770	12.56	1.13E-05	3.07
Within Groups	91.79	118	0.778			
Total	111.33	120				

Since the F-statistic (12.56) from Table 11 is greater than the critical value (3.07), the null hypothesis was rejected; there is a statistically significant difference between operators regarding their calibrated angle.

Table 12 shows the results from a two-sample t-test assuming equal variances for angle calibrations by operators A and B. The α value is 0.05. The null hypothesis is that operators A and B calibrated angle with similar results.

Table 12. t-test results for angle calibration between operators A and B

Operator	A	B
Mean	69.76	68.84
Variance	0.6957	0.8191
Observations	40	54
Pooled Variance	0.7668	
Hypothesized Mean Difference	0	
df	92	
t Statistic	5.036	
P(T<=t) one-tail	1.18E-06	
t Critical one-tail	1.662	
P(T<=t) two-tail	2.36E-06	
t Critical two-tail	1.986	

Since the t-statistic (5.036) from Table 12 is greater than the critical values (1.662 and 1.986), the null hypothesis was rejected; there is a statistically significant difference between operators A and B regarding their calibrated angle. Similar analysis shows there is no statistically significant difference between operators B and C, suggesting they calibrated angle with similar results.

Figure 17 compares the variation in zero calibration results for the three operators and indicates some consistency between operators A and B in calibrating zero, although operator C produced generally lower values.

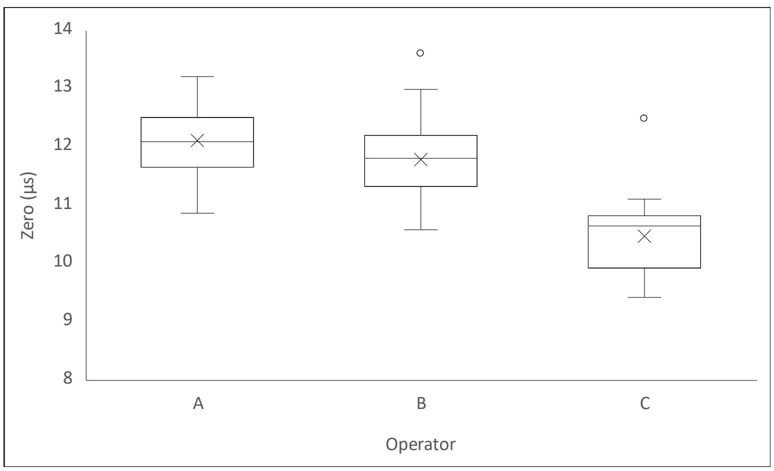


Figure 17. Variation in zero for each operator

The results of single-factor ANOVA on the zero calibration results from the three operators produced an F-statistic of 24.86. Since this is greater than the critical value of 3.07, there is statistical significance that the operators calibrated zero differently. The results of two sample t-tests assuming equal variances show there is statistical significance that no two operators calibrated zero with the same results.

Figure 18 compares the variation in gain calibration results for the three operators, and indicates some inconsistency between operators in calibrating gain, with no operators producing outliers.

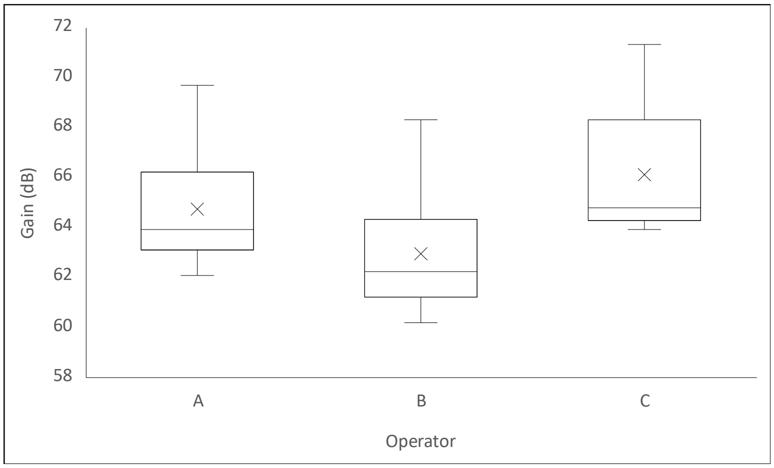


Figure 18. Variation in gain for each operator

The results of single-factor ANOVA on the gain calibration results from the three operators produced an F-statistic of 40.65. Since this is greater than the critical value of 3.07, there is a statistically significant difference between the operators' calibrated gain results. The results of two sample t-tests assuming equal variances show there is statistical significance that no two operators calibrated gain with the same results.

The comparison of results between operators demonstrates the merit in having more than one operator in the test. The variation produced by the three operators was greater than that produced by any one operator. It was common for operators to produce significantly different calibration results.

6.5.2 Wedge Condition

Analysis of the results for new and worn wedges shows if both conditions need to be represented in the calibration tests. Figure 19 compares the variation in angle calibration results for new and worn wedges.

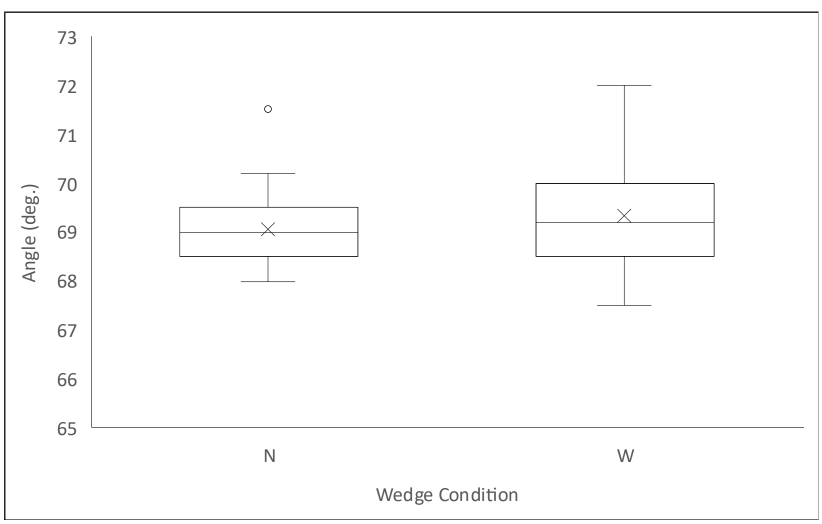


Figure 19. Variation in angle for new and worn wedges

Figure 19 shows less variation in angle for the new wedges compared to the worn wedges. This is to be expected since the new wedges are supplied to the manufacturer's tolerances and the worn wedges have different degrees of wear.

Table 13 shows summary statistics for the angle results from new and worn wedges.

Table 13. Angle statistics for new and worn wedges

Condition	Count	Sum	Average	Variance
New	50	3452	69.05	0.567
Worn	71	4923	69.34	1.158

Table 14 shows the results of single-factor ANOVA on the angle results from new and worn wedges with an alpha level of 0.05. The null hypothesis is that new and worn wedges gave similar angle calibration results.

Table 14. ANOVA results for angle calibration for new and worn wedges

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2.49	1	2.492	2.72	0.101	3.92
Within Groups	108.8	119	0.915			
Total	111.3	120				

Since the F-statistic (2.72) from Table 14 is less than the critical value (3.92), the null hypothesis is not rejected. Therefore, the differences between the angle calibration results from new and worn wedges (as shown in Figure 19) were not statistically significant.

Figure 20 compares the variation in zero calibration results for new and worn wedges.

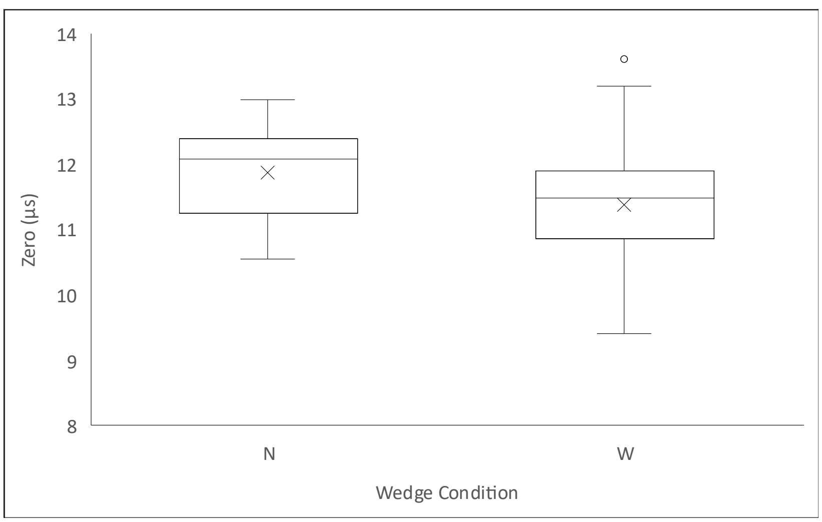


Figure 20. Variation in zero for new and worn wedges

Figure 20 shows similar variation in zero between the new and worn wedges. The worn wedges had slightly lower average zero compared to the new wedges. This is to be expected since the sound path in a worn wedge will be less than that in a new wedge.

The results of single-factor ANOVA on the zero calibration results for new and worn wedges showed an F-statistic of 10.50. Since this is greater than the critical value of 3.92, there is a statistically significant difference between the zero calibration results for new and worn wedges shown in Figure 20.

Figure 21 compares the variation in gain calibration results for new and worn wedges.

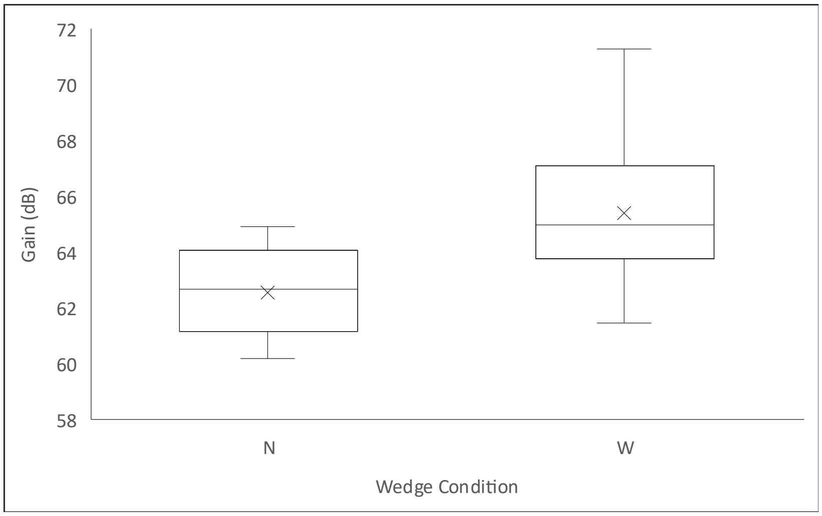


Figure 21. Variation in gain for new and worn wedges

Figure 21 shows a slightly higher gain was required for worn wedges compared to new wedges. There is no obvious physical explanation for this.

The results of single-factor ANOVA on the gain calibration results for new and worn wedges produced an F-statistic of 53.83. Since this is greater than the critical value of 3.92, there is a statistically significant difference between the gain calibration results for new and worn wedges. This confirms the observation made in Figure 21.

The variation from worn wedges was greater than that from new wedges. This demonstrates the merit of including wedges in different states of wear in the analysis.

6.5.3 Dependence Between Variables

Since angle and zero were variables in the CIVA simulation, it was necessary to determine if they were related or independent. Figure 22 shows the relationship between angle and zero for new and worn wedges.

Figure 22 shows no significant dependence between angle and zero. The overall R^2 value is 0.03. The relationship between angle and zero for new wedges did not appear to be significantly different to that for worn wedges.

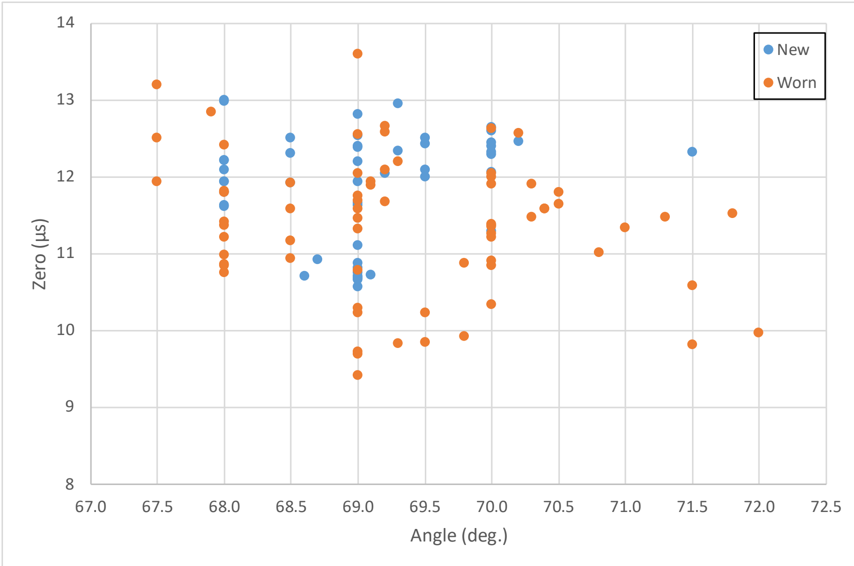


Figure 22. Angle versus zero for new and worn wedges

6.5.4 Outliers

Figure 23 shows the variation in gain calibration results for the complete dataset.

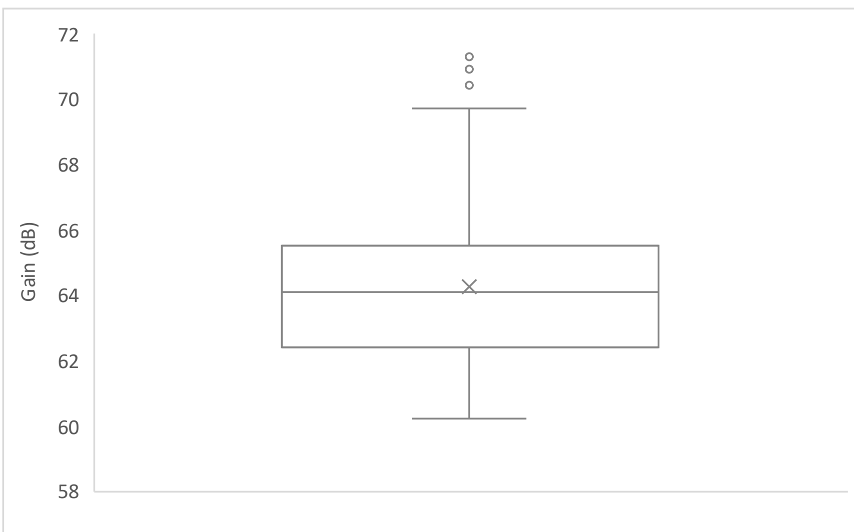


Figure 23. Overall variation in gain

Figure 23 shows three outliers in the gain calibration results. Since these records could be the result of errors and are not necessarily representative of the overall distribution, they were removed from the dataset before subsequent analysis.

No outliers were found in the angle and zero calibration results.

6.5.5 Distributions

The distributions of angle and zero are required for the CIVA analysis. Figure 24 shows the distribution of angle for the complete set of calibration results (excluding outliers).

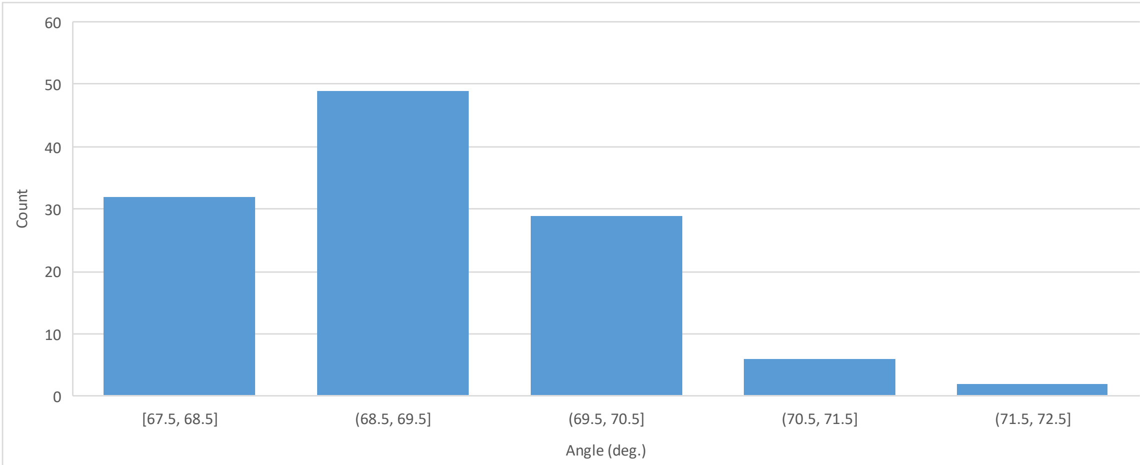


Figure 24. Distribution of refraction angle

Figure 24 shows the most common angle was centered at 69°. This differs from the manufacturer's specification of 70°. There are approximately equal counts of angles centered at 68° and 70°. The two results with angles in the bin 71.5° to 72.5° could be considered outliers, as shown in Figure 16. They could be added to the six results with angles centered at 71°.

Zero is the time taken for the signal to pass through the wedge from the probe to the surface of the calibration block and back again. Thus, it is related to the wedge size and angle. Figure 25 shows the distribution of zero for the complete set of calibration results (excluding outliers).

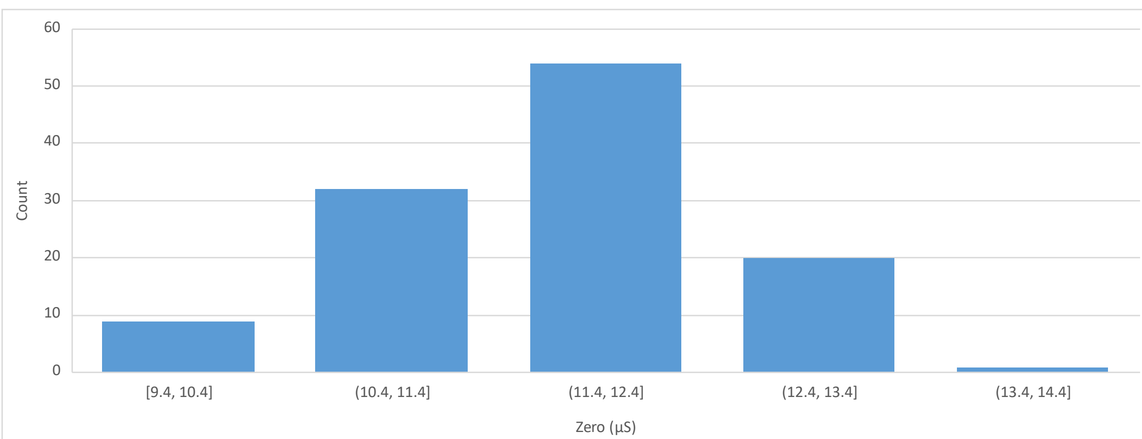


Figure 25. Distribution of zero

Figure 25 shows the most common zero was centered on 11.9 µs. The single result with zero in the bin 13.4 to 14.4 µs could be considered an outlier, as shown in Figure 17. It could be added to the 20 results with zero centered on 12.9 µs.

Figure 26 shows the distribution of gain for the complete set of calibration results. It shows a well-defined distribution with the most common gain centered on 63.2 dB. There are equal counts in the bins centered on 61.2 and 65.2 dB. The distribution has a tail extending to 70.2 dB.

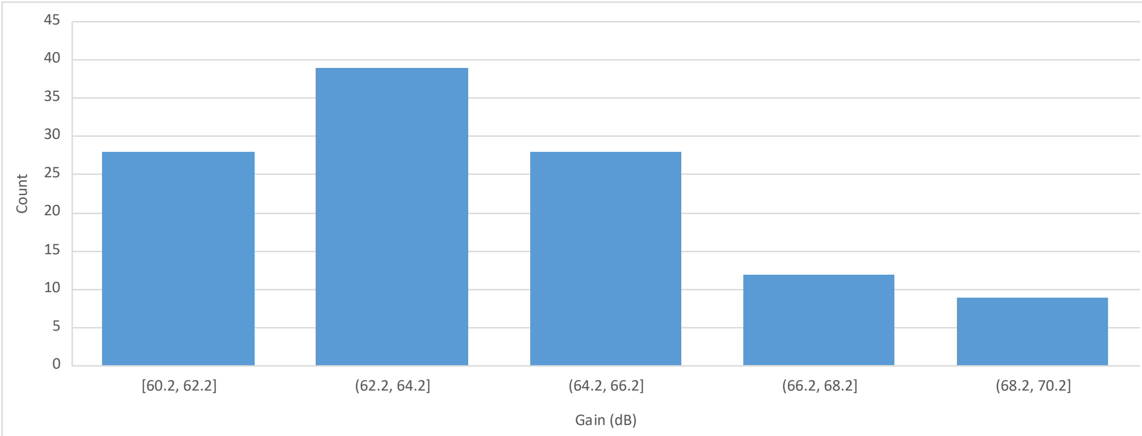


Figure 26. Distribution of gain

6.6 Test Discussion

The calibration test produced sufficient results for a meaningful statistical analysis. The test also generated well-defined distributions for angle, zero, and gain that can be used in CIVA modeling.

Although gain is not an input to the CIVA probability of detection analysis, it could be used to validate the CIVA model. This would require a separate CIVA simulation with the rail replaced by a block with a 1/16-inch diameter hole centered 0.6 inch from the surface.

Significant variation was found between operators. This suggests further tests with a different cohort of operators would provide useful data.

Significant variation in zero and gain was found between new and worn wedges. The same set of wedges should be used in any further tests to avoid introducing an additional variable.

The calibration procedure followed by the operators was relatively simple. A further test with operators employed by the railroads or their service providers might reveal differences in calibration procedure.

Repeating the calibration procedure for an entire day was tiresome and boring for the operators. Any further tests could be spread over two half days on separate days.

7. CIVA Modeling

The research team used the 2021 version of CIVA to simulate handheld ultrasonic rail flaw inspection and to calculate PODs. A dedicated computer was used that had a performance specification toward the top of the range recommended by EXTENDE, the supplier of CIVA.

Figure 27 shows a typical ultrasonic scan simulated in CIVA. The ultrasonic probe is on the top of a short section of rail. The ultrasonic rays extend into the head and web of the rail, and a small flaw under the surface is positioned ahead of the rays. The probe is scanned over the rectangular grid shown on the top of the rail. As it does so, some rays are reflected from the flaw and are received by the probe.

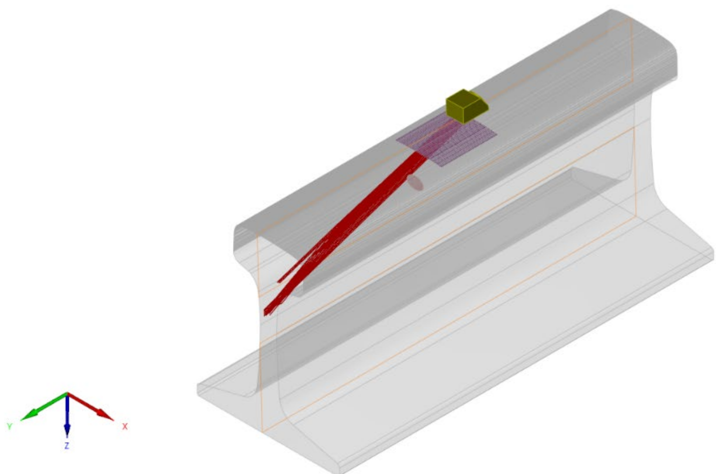


Figure 27. Simulated CIVA ultrasonic scan

The model used for this research was built on one developed for FRA in a previous project (Poudel). Full details of the parameters used in the CIVA simulations are provided in Appendix D, and the following subsections describe key parameters. By default, the CIVA software uses metric units for inputs and outputs. Thus, the inputs described below and the results in the following section are shown in metric units.

7.1 Simulation of Rail Flaw Inspection

7.1.1 Specimen Geometry

The cross-sectional geometries of the different rail profiles were defined in two-dimensional CAD (.dxf) files. A plane extension was used to give the rail a length of 600 mm.

7.1.2 Ultrasonic Probe and Wedge

The transducer used in the simulations represented the A540S model made by Olympus. This transducer has a single element, piezoelectric crystal. It is circular with a diameter of 12.7 mm. The simulations were performed using transverse waves.

The wedge represented the ABSA-5T-70 model made by Olympus. This is a 70°, flat-bottomed, acrylic wedge.

The coupling between the wedge and the rail represented the commercially available EcoGel.

7.1.3 Scanning

Transverse scanning was performed in 1 mm increments starting 25 mm from the center of the rail toward the field side and ending 25 mm toward the gage side of the rail. The wedge was kept perpendicular to the surface as it scanned across the head of the rail.

Longitudinal scanning began 175 mm along the rail from the origin and was repeated in 60 steps of 1 mm. This ensured full coverage of the largest flaws in the simulations. The wedge was aligned with the longitudinal axis of the rail as it scanned along the rail.

7.1.4 Flaw

An elliptical, planar flaw was used in the simulations. It was in the plane of the rail's cross-section, positioned 250 mm along the rail from the end of the 600-mm rail section.

7.1.5 Simulation

Figure 28 shows the parts of the rail surface designated as front wall (red), side wall (purple), and back wall (green). The CIVA simulations only considered ultrasonic reflections from the back wall, and only a single reflection was simulated.

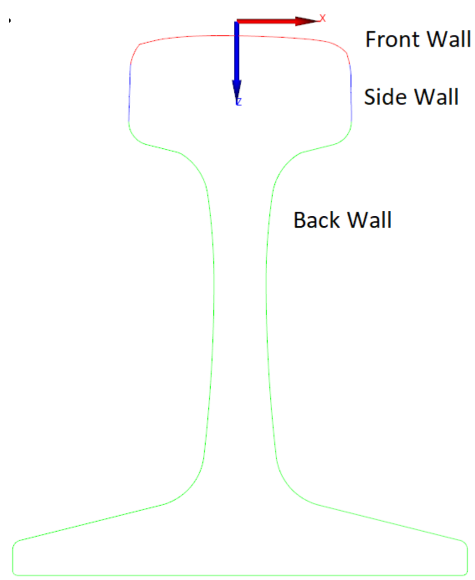


Figure 28. Specimen walls

A sensitivity zone was specified with X, Y, and Z dimensions of 100, 50, and 50 mm, respectively. CIVA does not process a scan if the flaw lies outside of this zone.

7.2 POD Variables

7.2.1 Flaw Size

The known value (*a*) in the POD analysis is the characteristic variable. In the CIVA simulation this was the flaw width. It was linearly distributed in increments of 4 mm from 5 to 65 mm to give 16 characteristic values.

From the analysis in Section 5.2, the flaw height was taken as 0.71 times the flaw width.

7.2.2 Flaw Position

Table 15 shows the distribution of flaw lateral position used in the CIVA simulations. These values are the results from the analysis described in Section 5.2. Flaw lateral position was measured from the rail centerline and was positive toward the gage face.

Table 15. Distribution of flaw lateral position

Bin	Min. (mm)	Max. (mm)	Proportion
1	-22.2	-10.2	0.075
2	-10.2	1.8	0.057
3	1.8	13.8	0.245
4	13.8	25.8	0.528
5	25.8	37.8	0.094

Table 16 shows the distribution of flaw depth used in the CIVA simulations. These values are the results from the analysis described in Section 5.2. Flaw depth was measured from a point 186 mm above the foot of the rail and was positive downward. This follows the CIVA convention that flaw position is specified in the same coordinate system as the specimen.

Table 16. Distribution of flaw depth

Bin	Min. (mm)	Max. (mm)	Proportion
1	14.0	17.7	0.302
2	17.7	21.4	0.340
3	21.4	25.1	0.302
4	25.1	28.8	0.019
5	28.8	32.5	0.038

A MATLAB® routine was created to generate random values of flaw lateral position and depths with overall distributions that matched those in Table 15 and Table 16. These random values were used for the CIVA simulations.

7.2.3 Refraction Angle

Table 17 shows the distribution of refraction angle used in the CIVA simulations. These values are the results from the calibration test described in Section 6.5.

Table 17. Distribution of refraction angle

Bin	Min. (deg.)	Max. (deg.)	Proportion
1	67.5	68.5	0.271
2	68.5	69.5	0.415
3	69.5	70.5	0.246
4	70.5	71.5	0.051
5	71.5	72.5	0.017

Table 17 shows the distribution of refraction angle was centered around 69°. This was slightly less than the nominal value of 70°.

7.2.4 Wedge Height

Table 18 shows the distribution of wedge height used in the CIVA simulations. Wedge height is dimension L4 in CIVA. The height was the zero from the calibration test results described in Section 6.5 multiplied by the velocity of longitudinal waves in the wedge (2.73 mm/µs) and then divided by 2 since zero is the time taken to travel through the wedge and back.

Table 18. Distribution of wedge height (dimension L4 in CIVA)

Bin	Min. (mm)	Max. (mm)	Proportion
1	12.8	14.2	0.076
2	14.2	15.6	0.288
3	15.6	16.9	0.458
4	16.9	18.3	0.169
5	18.3	19.7	0.008

Table 18 shows the distribution of wedge height was centered around the nominal wedge height of 15 mm, as expected. The MATLAB routine was also used to generate random values of refraction angle and wedge height with the distributions in Table 17 and Table 18 for use in the CIVA simulations.

7.3 Rail Profile

A separate CIVA model was created for each rail profile and a batch of simulations was run for each model. Within each batch, a simulation was performed for each of the 16 flaw sizes described in Section 5.2. Within each simulation, CIVA repeats its analysis several times (these repetitions are called shots). Table 19 shows the number of shots used for each rail profile. These numbers were chosen to show a distribution of rail profiles like that shown in Table 5.

Table 19. Number of shots for each rail profile

Profile	S3a	S7a	E2a	N11a	W4a
Shots	3	14	7	5	3

The total number of shots across all rail profiles was 32, meaning each flaw size was analyzed 32 times. The total number of shots across all flaw sizes was 512. Each shot used random values of flaw position, refraction angle, and wedge height. The random values were generated by a MATLAB routine that reproduced the overall distributions shown in Table 15 through Table 18.

7.4 CIVA Outputs

Figure 29 shows how input voltage from the transmitter is transformed into output voltage at the receiver. Three frequency transfer functions are involved: $t_G(\omega)$, $t_A(\omega)$, and $t_R(\omega)$. The first converts the input voltage to a force into the specimen. The last converts the force on the receiver to the output voltage. The middle function describes the relationship between input and output forces within the specimen.

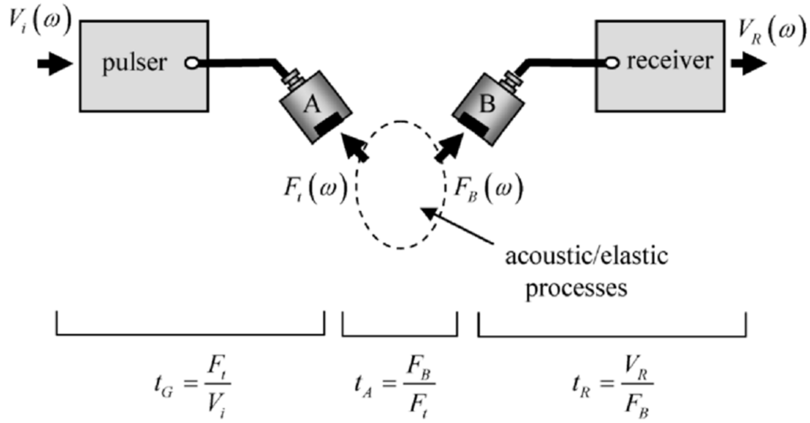


Figure 29. CIVA inputs and outputs (Schmerr & Song, 2007)

CIVA does not calculate the transfer functions between voltage and force in the transmitter and receiver. It only calculates the transfer function for force in the specimen. Output voltage can be calculated if the CIVA model is calibrated with results from a simulation with a reference flaw. For simplicity, such a calibration was not used in this study. Instead, the output from CIVA was the response to the input signal in arbitrary units referred to in CIVA as points. Since the input signal was the same for all simulations, points could be compared across simulations.

Figure 30 shows a typical A-Scan display from a CIVA simulation. It shows the size of the output signal in points plotted against time in micro-seconds. The flaw in this example has a width of 0.67 in (17 mm), height of 0.47 in (11.9 mm), transverse position measured from the rail's center line of 0.07 in (1.8 mm), and vertical position measured from the top of the rail of 0.39 in (9.9 mm).

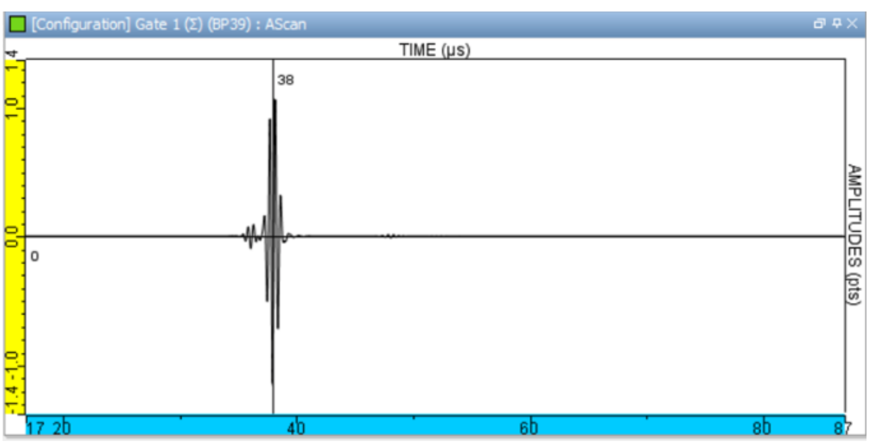


Figure 30. Example CIVA A-Scan

Figure 30 shows a reflection from the flaw after approximately 38 µs. This is the time taken for the pulse to travel from the transmitter to the flaw and back to the receiver. The maximum response in this example is 1.147 points.

Figure 31 shows a longitudinal B-Scan display from CIVA for the same example as used for Figure 30. A color scale is used to indicate the magnitude of the reflected signal. The transverse position of the probe corresponds to the scan that produced the maximum response.

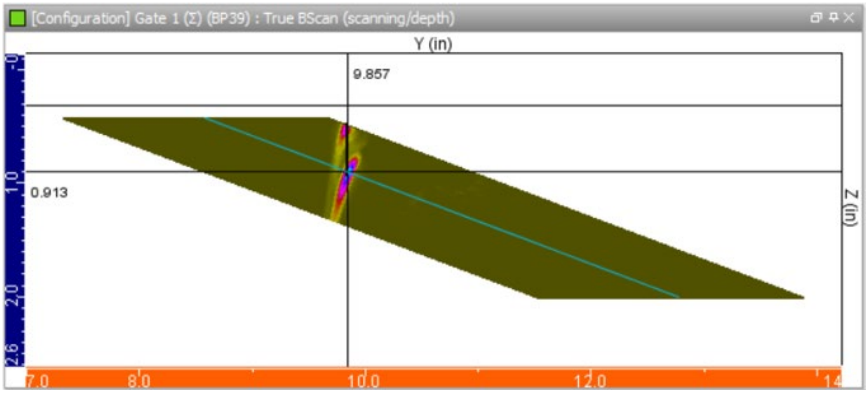


Figure 31. Example CIVA longitudinal B-Scan

The colored area in Figure 31 is that covered by the simulated scan. It starts after the dead zone below the top of the rail. The longitudinal location of the maximum response is 9.86 in (250 mm) from the start of the rail specimen. This agrees with the longitudinal location specified in the CIVA inputs. The depth of the maximum response is 0.91 in (23 mm) from the top of an unworn rail. The vertical head wear in this example is 0.33 in (8.5 mm). Thus, the depth of the maximum response is 0.58 in (15 mm) from the top of the rail. This is slightly deeper than the depth of the center of the flaw specified for this example simulation.

Figure 31 shows principal responses from the flaw and secondary responses from reflections from the undersides of the head of the rail.

Figure 32 shows a transverse B-Scan (referred to in CIVA as a D-Scan) display from CIVA for the same example as used for Figure 30 and Figure 31. The same color scale is used to indicate the magnitude of the reflected signal. The longitudinal position of the probe corresponds to the scan that produced the maximum response.

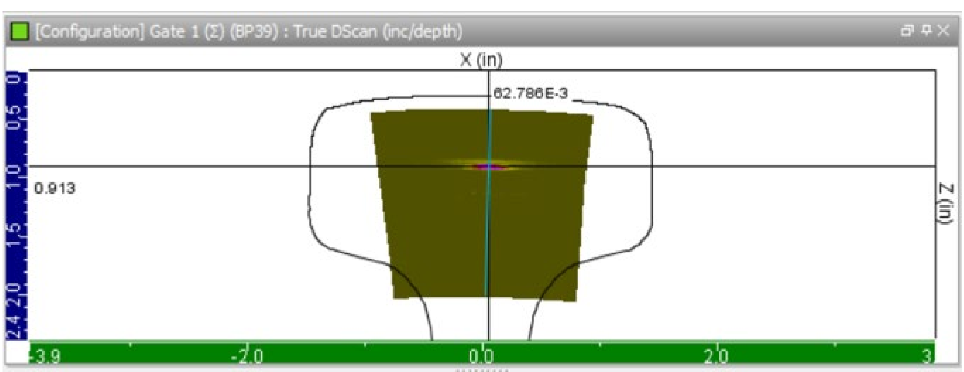


Figure 32. Example CIVA transverse B-Scan

Figure 32 shows the location of the maximum response is 0.063 in (2 mm) from the centerline of the rail. This matches the location of the center of the flaw specified in the CIVA inputs.

Figure 33 shows a C-Scan display from CIVA for the same example and color scale as used above. This display shows the magnitude of the maximum response for each scan in the simulation. In this example there were 50 increments of 0.04 in (1 mm) in the transverse direction and 60 scans separated by 0.04 in (1 mm) in the longitudinal direction.

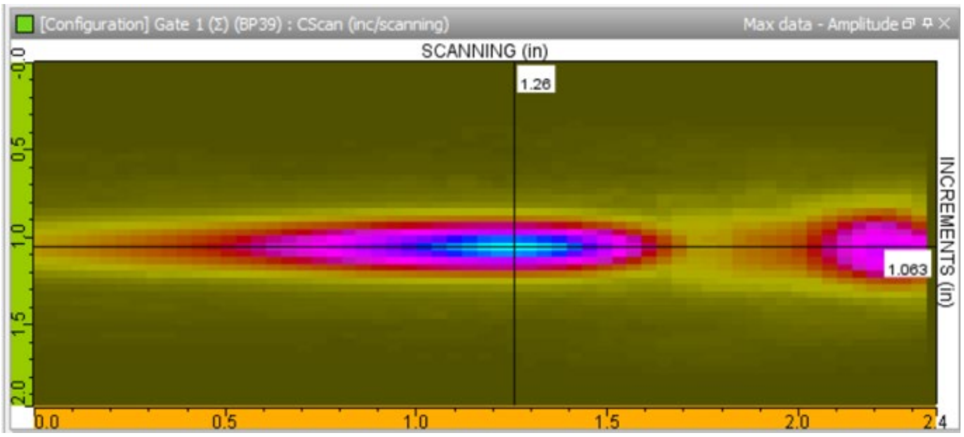


Figure 33. Example CIVA C-Scan

Figure 33 shows the maximum response occurred for the scan when the probe was located at 1.063 in (27 mm) in the transverse direction and 1.26 in (32 mm) in the longitudinal direction. In this example the transverse and longitudinal origins for the probe locations were -1 in (-25 mm) from the rail's center line and 6.7 in (170 mm) from the start of the rail specimen. Thus, the maximum response occurred when the probe was 0.063 in (2 mm) from the center line and 7.95 in (202 mm) from the start of the rail.

a_{max} is the absolute maximum value of the output signal for all scans in a simulation. This is the value that was used in the POD analysis. In this example, a_{max} is 1.147 points.

8. Results

The CIVA simulations resulted in 512 a_{max} versus flaw width data points. These results were post-processed to convert them to a_{max} versus flaw size as a percent of rail head area. These results were analyzed separately for each rail profile and then for the combined dataset.

The CIVA results were output in metric units as described below.

8.1 Post-processing

Table 20 shows a typical set of results from CIVA simulations. This example is for the S7a rail profile and a 17 mm wide flaw. The references in the last column are used in the discussion below.

Table 20. Example CIVA results file – S7a Rail Profile

Flaw Width (mm)	Flaw Height (mm)	Flaw Z (mm)	Flaw X (mm)	Wedge Height (mm)	Wedge Angle (deg.)	a_{max} (points)	Ref.
17.0	11.9	17.6	19.0	14.6	70.1	0.986	
17.0	11.9	16.5	28.8	16.6	69.4	0.191	
17.0	11.9	23.0	8.0	15.7	72.5	0.555	C
17.0	11.9	20.2	10.7	16.8	68.2	0.831	
17.0	11.9	23.6	30.0	17.1	70.0	0.086	B
17.0	11.9	17.1	25.8	15.6	71.1	0.614	
17.0	11.9	19.9	18.2	15.9	68.7	0.867	
17.0	11.9	18.3	20.8	15.0	70.6	0.892	
17.0	11.9	15.0	20.5	15.4	68.5	1.230	A
17.0	11.9	20.8	17.2	15.9	71.2	0.699	
17.0	11.9	20.0	19.4	16.8	71.4	0.749	
17.0	11.9	19.3	-16.2	14.8	69.1	0.815	
17.0	11.9	21.2	3.8	16.5	70.9	0.674	
17.0	11.9	22.3	35.4	14.0	69.8	0.017	D

Researchers developed a post-processor to convert the size and position of each flaw into a percentage of rail head area. The post-processor considered the relevant rail profile including the wear and any part of the flaw that lay outside of the rail head.

Figure 34 shows the post-processed results from Table 20. a_{max} is plotted against the known size of the flaw in terms of percent of rail head area.

Figure 34 shows the a_{max} from a simulation with 14 shots for a flaw that was 17 mm wide. Each shot had a random choice of flaw location, wedge height, and refraction angle. The scatter in the figure is a result of these randomized factors.

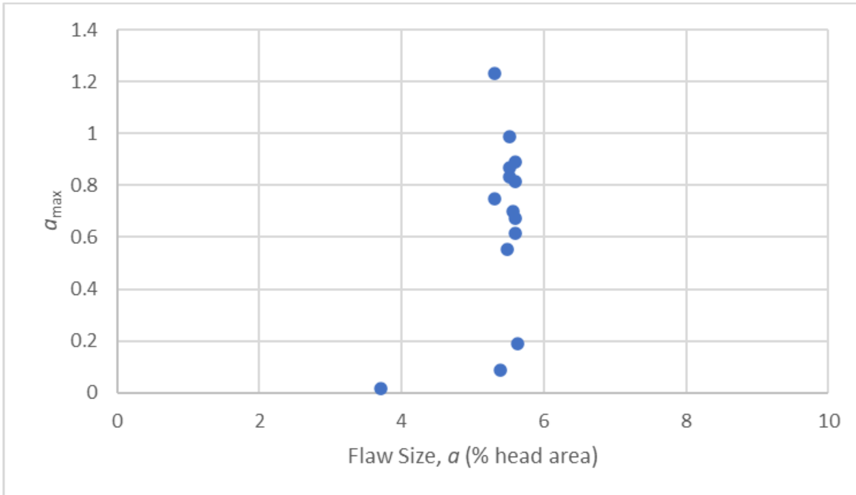


Figure 34. Example post-processed results – S7a Rail Profile

Figure 35 shows the location of four examples of rail flaws included in the results shown in Figure 34 and referenced in Table 20. The red line shows the region scanned by the probe in the simulation. The origin for the X and Z positions at the center of the flaws is also shown.

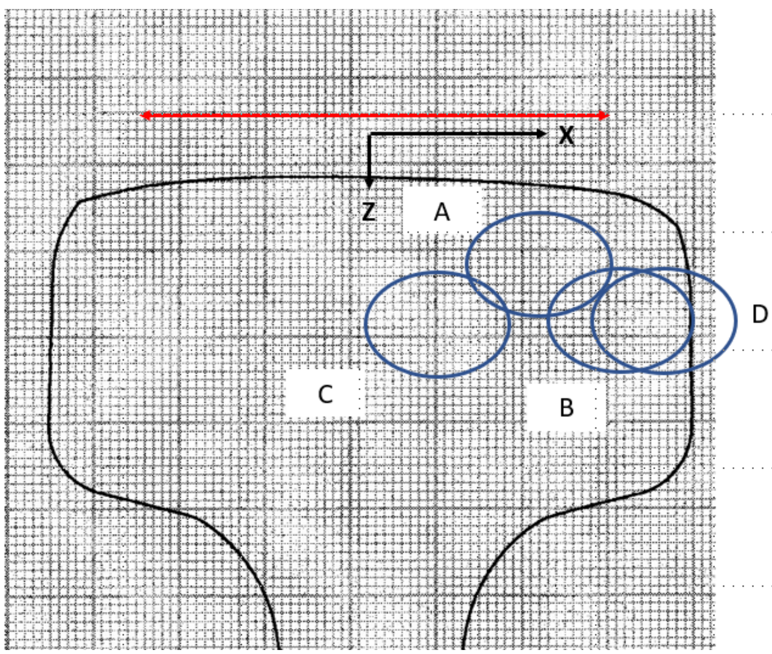


Figure 35. Example flaws – S7a Rail Profile

Flaw A in Figure 35 had the largest a_{max} of 1.230, and was also the closest flaw to the top of the rail. Flaw B had the second smallest a_{max} of 0.086 and was one of the deepest flaws. Flaw C, at the same depth as flaw B but closer to the center of the rail, had an intermediate a_{max} of 0.555.

Flaw D in Table 20 had the smallest a_{max} of 0.017. It was the same depth as Flaw B but 5 mm further toward the gage side of the rail. Since this flaw broke the surface of the rail, its percentage of rail head area was smaller than the other flaws.

[Figure 36](#) and [Figure 37](#) show the CIVA simulation for Flaws A and B, respectively. The shaded area on the top of the rail is that covered by the probe when scanning. The probe is shown in the position that produced the maximum response in both cases. The red lines show several axial rays emitted from the probe. Ultrasonic energy was dispersed in a cone surrounding these rays.

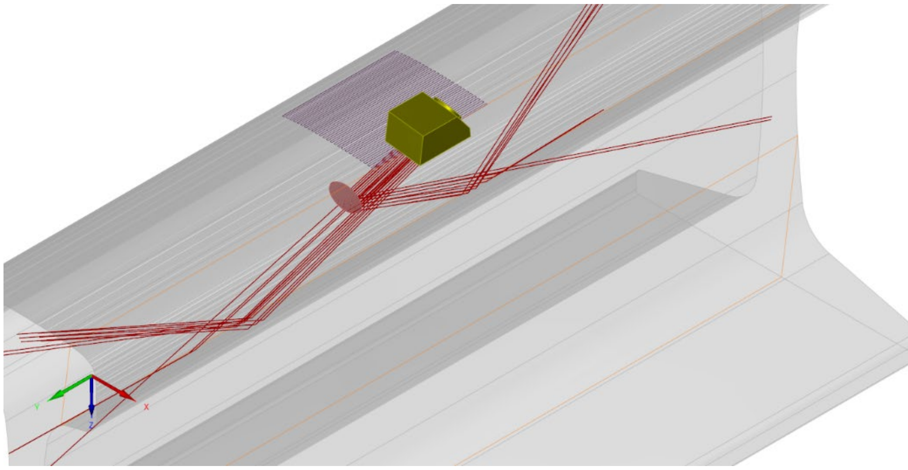


Figure 36. CIVA simulation for Flaw A

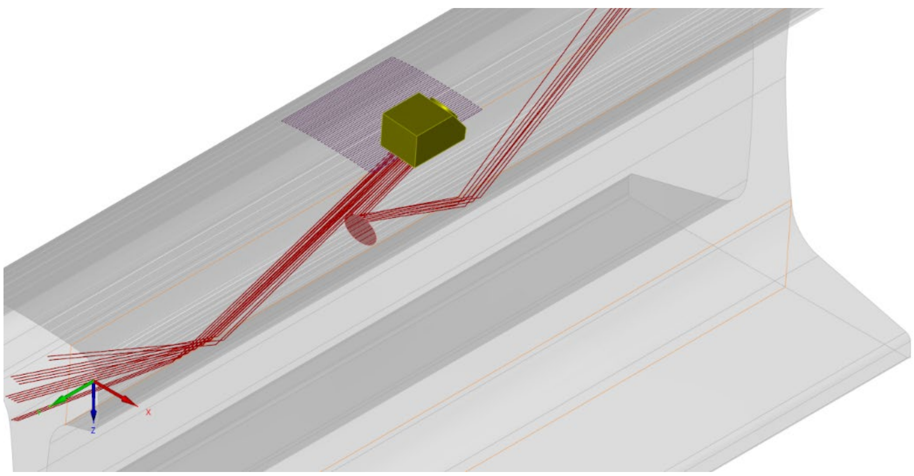


Figure 37. CIVA simulation for Flaw B

Figure 36 shows numerous rays from the probe reflect from the flaw and the underside of the rail head. Figure 37 shows fewer rays follow this path. The density of rays received by the probe is related to the ultrasonic energy received. Thus, the simulation for Flaw A can be expected to show a higher a_{max} than that for Flaw C.

8.2 Wedge Height Sensitivity

Flaw A in Figure 35, being a small flaw, was used to test the sensitivity of results to the height of the wedge. The calibration tests reported in Section 6 showed that wedge height can vary from 12.8 to 19.7 mm. Figure 36 shows the variation in a_{max} over this range of wedge heights while flaw size, position, and refraction angle were held constant.

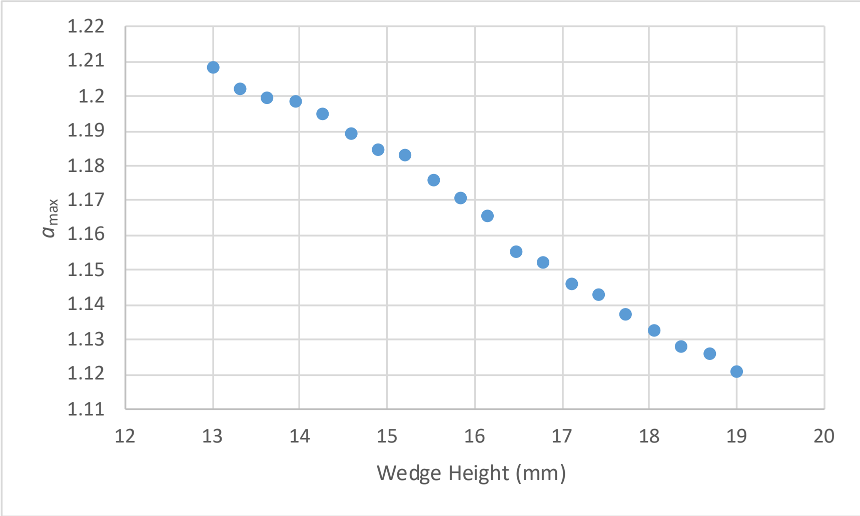


Figure 38. Wedge height sensitivity – Flaw A, Rail Profile S7a

Figure 38 shows a_{max} reduced approximately linearly as wedge height increased. Over the range of wedge heights studied, a_{max} changed from 1.121 to 1.208 (i.e., by approximately ±3.8 percent).

8.3 Refraction Angle Sensitivity

Flaw A in Figure 35 was also used to test the sensitivity of results to the refraction angle. The calibration tests reported in Section 6 showed that refraction angle can vary from 67.5° to 72.5°. Figure 39 shows the variation in a_{max} over this range of refraction angles while flaw size, position, and wedge height were held constant.

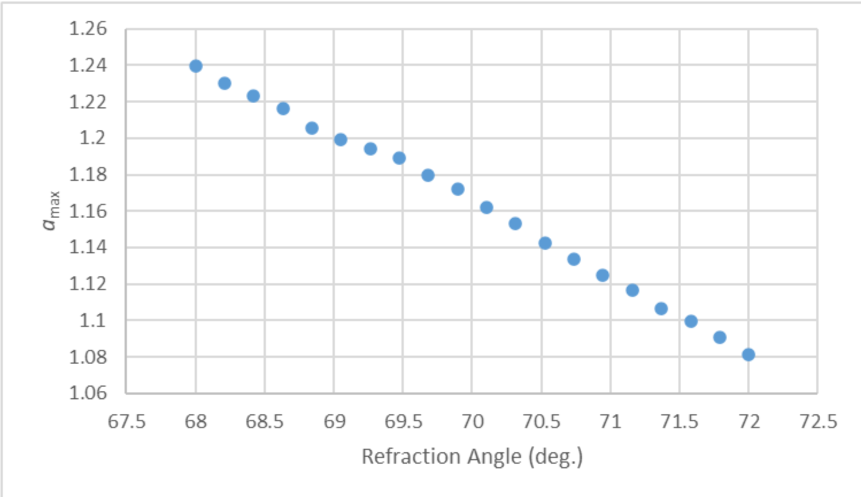


Figure 39. Refraction angle sensitivity – Flaw A, Rail Profile S7a

Figure 39 shows a_{max} reduced approximately linearly as refraction angle increased. Over the range of refraction angles studied, a_{max} changed from 1.081 to 1.240 (i.e., by approximately ±6.8 percent).

8.4 Effect of Fixed Wedge Properties

The effect of excluding variance in calibration was analyzed by repeating an example batch of simulations with a fixed probe height and refraction angle. The team used the S7a rail profile with a 17 mm wide rail flaw for this analysis. Table 21 shows the effect of fixed probe properties on the results. The values of a_{max} using the variable wedge heights and refraction angles shown in Table 20 are compared to those with the refraction angle fixed at 70° and the wedge height fixed at 16 mm.

Table 21. Example S7a results with variable and fixed wedge properties

Flaw Z (mm)	Flaw X (mm)	a_{max} – variable wedge properties (points)	a_{max} – fixed wedge properties (points)
17.6	19.0	0.986	1.015
16.5	28.8	0.191	0.212
23.0	8.0	0.555	0.635
20.2	10.7	0.831	0.775
23.6	30.0	0.086	0.111
17.1	25.8	0.614	0.629
19.9	18.2	0.867	0.812
18.3	20.8	0.892	0.910
15.0	20.5	1.230	1.167
20.8	17.2	0.699	0.739
20.0	19.4	0.749	0.809
19.3	-16.2	0.815	0.762
21.2	3.8	0.674	0.708
22.3	35.4	0.017	0.018

Figure 40 shows box and whisker plots for the example results found in Table 21.

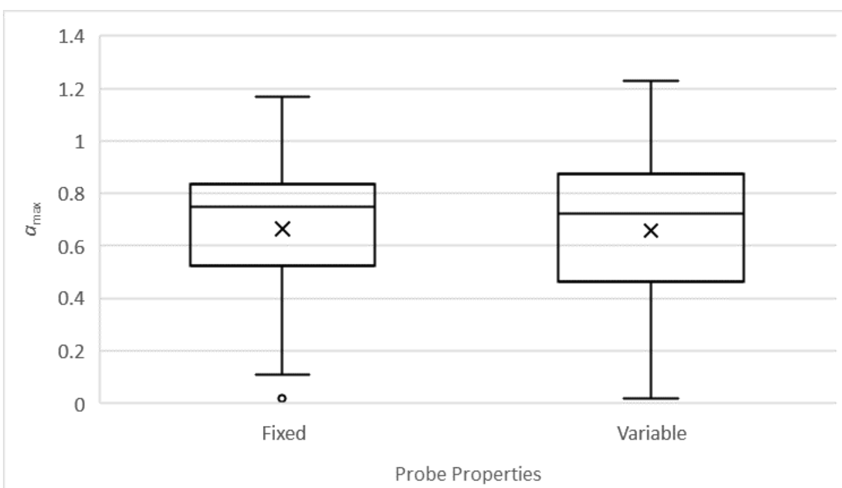


Figure 40. Example S7a results with variable and fixed wedge properties

Figure 40 shows slightly less variation in the results when the wedge properties were fixed. The variance for fixed wedge properties was 0.111 points, and that for the variable wedge properties was 0.120 points. Thus, the variance was approximately 10 percent higher when the effects of variation in calibration were included. Fixing the wedge properties also reduced the highest a_{max} result from 1.230 to 1.167 points.

8.5 S7a POD

Figure 41 shows the a_{max} results for the S7a Rail Profile's 224 shots (14 shots for each of 16 flaw widths).

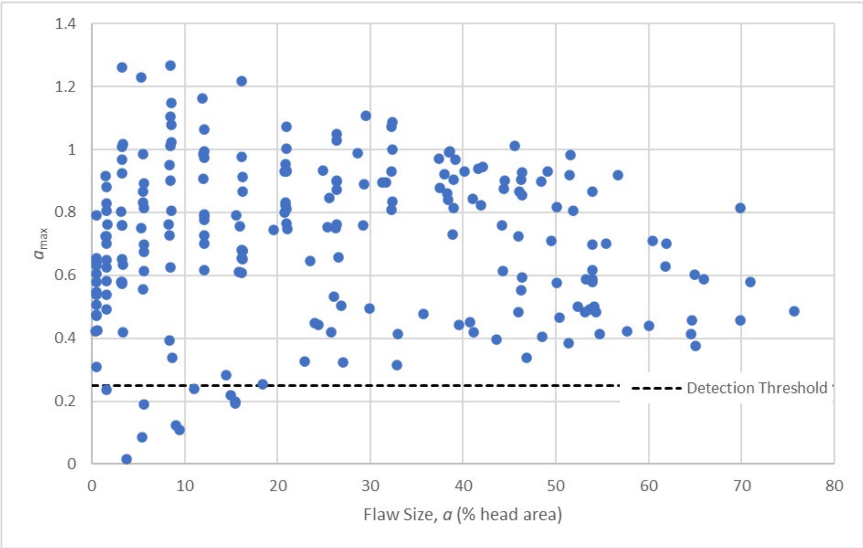


Figure 41. S7a a_{max} results

Figure 41 shows a detection threshold at a_{max} = 0.25. This value was assumed to be reasonable since it would result in a small number of false negatives. The sensitivity to detection threshold is evaluated in Section 8.9 and discussed further in Section 9.1.

Figure 42 shows the hit-miss POD calculated from the results in Figure 41 with a detection threshold of 0.25.

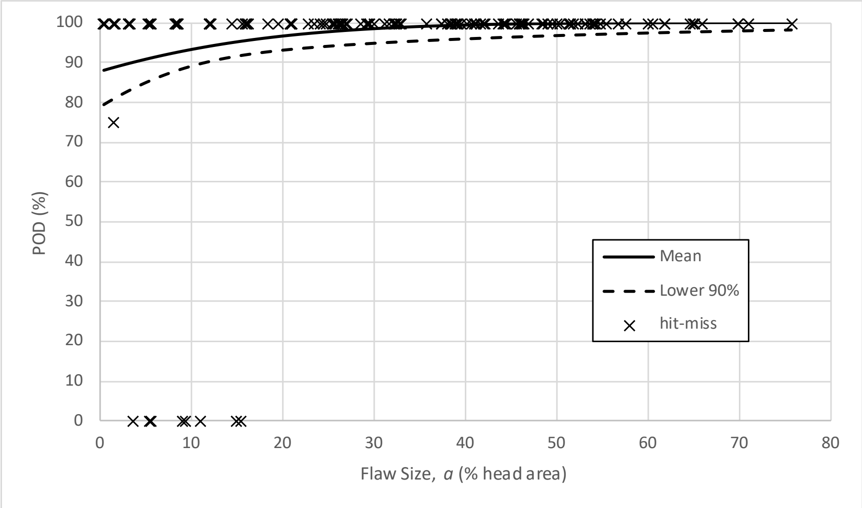


Figure 42. S7a POD results – 0.25 detection threshold

The solid curve in Figure 42 shows the mean POD, and the dashed curve shows the POD for the lower 90 percent confidence interval. The crosses on the x-axis indicate flaws that have been missed and crosses on the top of the chart show hits. The cross at 75 percent POD indicates three out of four results for this flaw size were hits.

Figure 42 shows small flaws can be detected with a mean probability higher than 88 percent. Also, no flaws occupying more than 16 percent of the rail head area were missed.

For the S7a Rail Profile and a detection threshold of 0.25, the mean POD for a flaw occupying 25 percent of the rail head area is 97.7 percent. This means that on average approximately 5 out of 200 flaws of this size will be missed. The POD for the lower 90 percent confidence interval drops to 94.2 percent. This means there is only a 90 percent probability that a flaw of this size will not be missed 188 times or more out of 200.

8.6 Rail Profile Sensitivity

Table 22 shows how POD varies with rail profile using three values of POD for a rail flaw occupying 25 percent of the rail head area.

Table 22. POD results by rail profile – 25 percent rail head area

Profile	Condition	Shots	Mean POD (percent)	Lower 90 percent POD (percent)	Lower 95 percent POD (percent)
S3a	Lightly worn tangent	48	95.6	85.8	83.0
S7a	Moderately worn tangent	224	97.7	94.2	93.2
E2a	Heavily worn tangent	112	99.2	90.8	87.1
N11a	Lightly worn curve	80	93.0	84.5	82.4
W4a	Moderately worn curve	48	86.4	74.9	72.2

Table 23 shows the mean POD varies with rail profile. The heavily worn tangent rail, E2a, produced the highest mean POD of 99.2 percent. This is likely due to the flaws being closer to the surface of the worn rail and easier to detect. The moderately worn curve rail, W4a, gave the lowest mean POD of 86.4 percent. This is likely due to the loss of reflection from some flaws when the probe moved to the gage side of the worn rail.

8.7 Full Dataset POD

The full dataset of results combined the simulations from five different rail profiles and 16 flaw sizes. This resulted in 512 separate shots (32 shots for each of 16 flaw widths). Figure 43 shows these results with a_{max} plotted against the known size of the flaw in terms of percent rail head area. A detection threshold at $a_{max} = 0.25$ is shown.

Figure 43 shows significant variation in results. This reflects the effect of variations in physical and human factors used in the simulations. Each of the 512 shots used a weighted choice of rail profile and random choices of flaw location and probe properties.

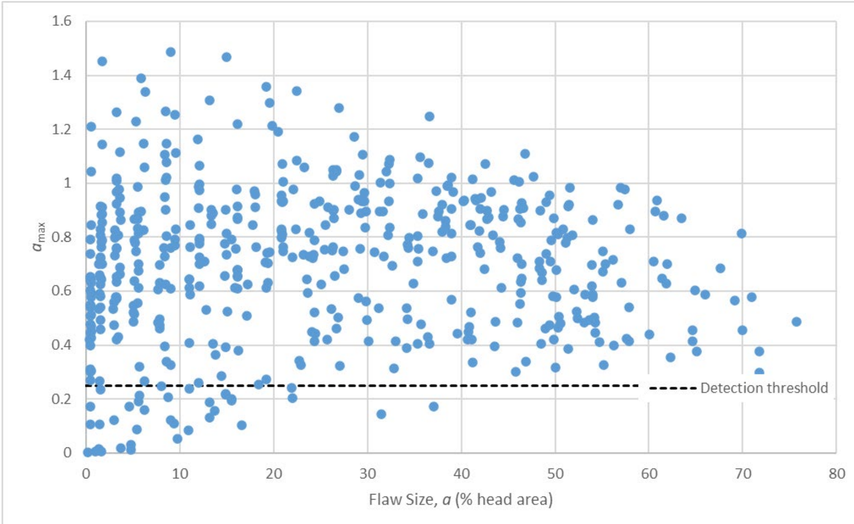


Figure 43. Full dataset a_{max} results

Figure 44 shows the hit-miss POD calculated from the results in Figure 43 with a detection threshold of 0.25. Crosses that appear between 0 and 100 percent indicate a combination of hits and misses.

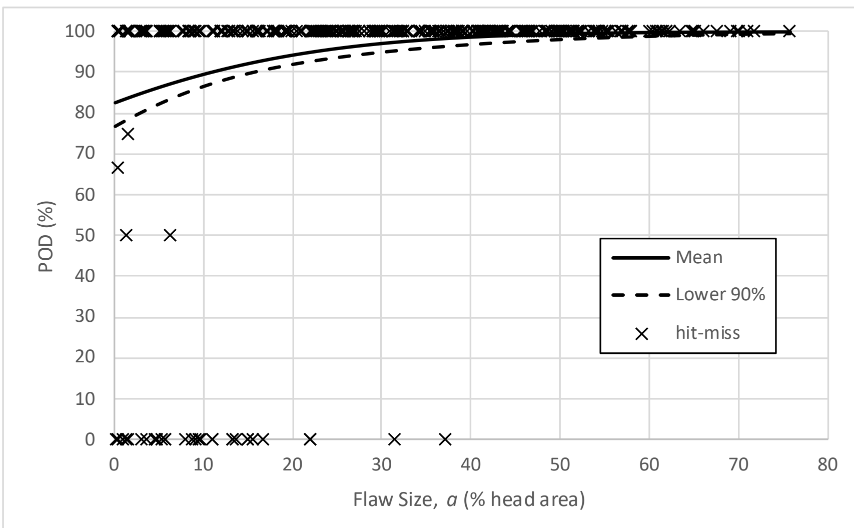


Figure 44. Full dataset POD results – 0.25 detection threshold

Figure 44 shows small flaws can be detected with a mean probability higher than 82 percent. Also, no flaws occupying more than 38 percent of the rail head area were missed. The POD for the lower 90 percent confidence interval is closer to the mean when compared to the results for the S7a profile in Figure 42 that only used results from 224 shots.

For the full dataset of rail profiles and a detection threshold of 0.25, the mean POD for a flaw occupying 25 percent of the rail head area is 95.8 percent. This means that on average approximately 4 out of 100 flaws of this size will be missed. The POD for the lower 90 percent confidence interval drops to 93.6 percent. This means there is only a 90 percent probability that a flaw of this size will not be missed 187 times or more out of 200.

The POD approaches 100 percent for larger flaws. For example, the mean and lower 90 percent confidence interval PODs for a flaw occupying 60 percent of the rail head area are 99.8 percent and 98.8 percent, respectively.

8.8 Sample Size Sensitivity

A subset of results was produced that only used one shot from each flaw size for each rail profile. This resulted in 80 a_{max} versus flaw size data points. Figure 45 shows the PODs calculated from this subset of results with a detection threshold a_{max} = 0.25.

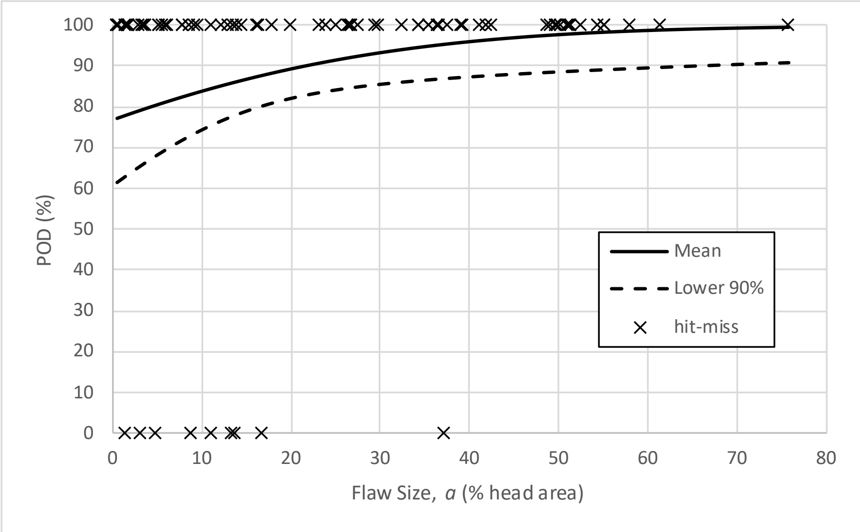


Figure 45. Sub dataset POD results – 0.25 detection threshold

Comparing Figure 45 to Figure 44 shows the confidence interval moves further from the mean when there are fewer results. It also shows differences in the mean POD. For example, the estimated mean POD for a flaw occupying 25 percent of the rail head area is 95.8 percent for the full dataset and 91.6 percent for the subset of results.

Table 23 compares PODs from the full dataset of 512 results with those from the subset with 80 results.

Table 23. Sample size sensitivity – 0.25 detection threshold

Shots	Flaw Size (percent head area)	Mean POD (percent)	Lower 90 percent POD (percent)	Lower 95 percent POD (percent)
512	25	95.8	93.6	93.1
80	25	91.6	83.9	82.0
512	60	99.8	98.9	98.5
80	60	98.8	89.1	84.9

Table 23 shows the benefits of large sample sizes that result in more accurate estimates of the mean POD with narrower confidence intervals. This effect is more significant for small flaws. For the example of a flaw occupying 25 percent of the rail head area and the lower 95 percent POD, at least 7 in 100 flaws were estimated to be missed when the full dataset was used. If the smaller dataset is used to calculate the POD, the estimate is at least 18 in 100 missed flaws.

8.9 Detection Threshold Sensitivity

The results in Figure 43 show that flaw detection depends on the detection threshold. A low detection threshold results in only small flaws being missed, and a high one causes some large flaws to be missed. Table 24 shows the effect on POD when changing the detection threshold for a flaw occupying 25 percent of the rail head area.

Table 24. Detection threshold sensitivity – 25 percent rail head area

Detection Threshold	Mean POD (percent)	Lower 90 percent POD (percent)	Lower 95 percent POD (percent)
0.1	99.9	98.2	97.3
0.15	98.7	96.8	96.3
0.2	97.2	95.1	94.6
0.25	95.8	93.6	93.1
0.3	93.8	91.4	91.0
0.35	90.0	87.4	86.9
0.4	86.7	84.1	83.5

Table 24 shows the POD reduces as detection threshold increases. Thus, a low detection threshold is desirable. However, in practice, due to noise and external factors, too low a detection threshold can result in an unacceptable number of false positives. Conversely, setting the detection threshold too high can result in unacceptable false negatives.

8.10 Lateral Scan Region Sensitivity

Figure 46 shows the size and location of a small flaw on the gage face of the W4a Rail Profile. The solid, horizontal red line shows the part of the rail head scanned to obtain the full dataset of results. This scan region was moved horizontally in steps of 1 mm to study the effect on a_{max}. The dashed, horizontal red line shows the last position of the scan region. In this position the probe reached the lower corner of the worn gage face.

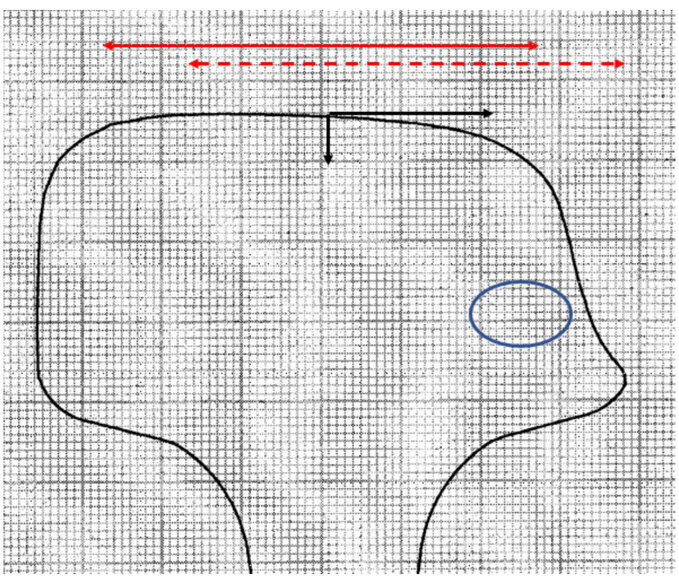


Figure 46. Gage face flaw – W4a Rail Profile

Figure 47 shows the effect of scan region on a_{max}. Results are shown for two probe configurations: a nonconformal probe for which any gaps between the probe and the rail are filled with lubricant, and a conformal probe which is adjusted to fit the shape of the rail at each scan location.

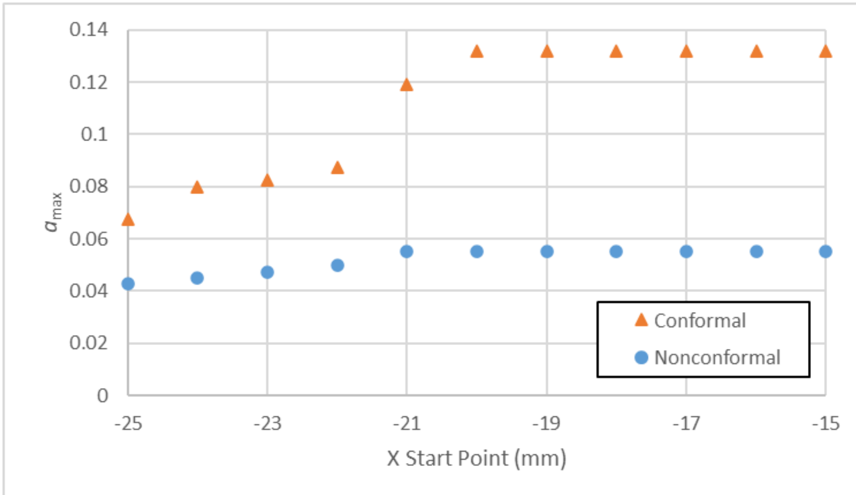


Figure 47. Effect of scan position on a_{max} – W4a Rail Profile

Figure 47 shows the standard scan region, with the zone centered on the rail head, results in a_{max} of 0.043 for the nonconformal probe. a_{max} increases as the scan region moves toward the gage face to reach a maximum value of 0.055. These values are likely to be significantly lower than any practical detection threshold.

When the probe is made to conform to the shape of the rail, a_{max} increases. For the standard scan position a_{max} = 0.068, increasing to 0.132 for the farthest scan position. This value could potentially be above a practical detection threshold.

9. Discussion

9.1 POD results

This research successfully demonstrated the application of MAPOD to a particular track inspection technology, handheld ultrasonic inspection. Computer simulation generated 32 estimates of detection for each of 16 flaw sizes using various rail profiles. The resulting hit-miss POD had a narrow confidence interval. This makes it suitable to compare with other technologies and for evaluation against requirements in federal regulations.

FRA's track safety standards (Subpart D 213.113 Defective Rails) require railroads to take significant actions if transverse flaws in rails exceed 25 percent of the rail head area. Using an appropriate detection threshold $a_{max} = 0.25$, this research showed the lower 90 percent confidence interval POD of detecting this size of flaw is 93.4 percent. This means there is only a 90 percent probability that a flaw of this size will be detected more than 187 times out of 200. Knowing this POD allows railroads to decide on inspection frequency to ensure the timely testing of any missed flaws.

After the team set up the computer model, it spent four months continuously processing the simulations. To achieve the same results through physical testing would have been impractical, requiring more than 512 samples of rails with various profiles and flaws of different sizes. After repeated testing with different wedges by three operators, the rails would have to be broken open to confirm the flaw type and size. Samples found not to have transverse flaws would be discarded, and additional tests would be needed if each flaw size was found not to have 32 samples.

PODs using the full dataset of 512 results were compared with those using a subset of 80 results. Different estimates of the mean POD were found and the confidence intervals for the subset were twice as wide as for the full dataset. Even as few as 80 results from physical samples would be very costly and time-consuming to collect. Attempts to calculate PODs from 16 samples (one for each flaw size) produced statistically insignificant results. This demonstrates the significant benefits of MAPOD when it is impractical to conduct physical testing of sufficient samples to get meaningful results. The level of confidence achieved with the full dataset is narrow enough to compare handheld ultrasonic inspection with other rail flaw inspection technologies.

The results showed considerable scatter. Small flaws can be detected unless they are deep in the head or near the field and gage faces of the rail. Some large flaws were found to give low a_{max} values. This scatter demonstrates the danger of drawing conclusions from a small number of physical tests.

The scope of this research was limited to the physical factors of rail profile and flaw position, and the human factors involved in equipment calibration. Table 25 lists some factors not considered that could also influence POD. The columns show if the factor is expected to increase or reduce POD. The effects of these factors could be researched in a future stage of this work.

A practical extension of the hit-miss analysis reported here is to calculate flaw size from the CIVA results. Figure 48 illustrates how post-processing of the C-scan results could be used to size flaws in a similar way to field testing. The C-scan shows the maximum signal response from each scan in the simulation as the probe is moved across and along the rail. Once a_{max} has been calculated, the locations of the scans producing half the a_{max} value can be identified. These are the locations of the edge of the flaw since half the maximum ultrasonic energy is reflected to the receiver. The horizontal distance between these locations in the X direction would indicate the width of the flaw. Similar post-processing in the Z direction and adjusting for refraction angle would indicate the height of the flaw. This method would emulate practice used by operators in the field.

Table 25. POD factors for future consideration

Could Increase POD	Could Reduce POD
Operator skill in manipulating probe to find flaws	Variation in lubrication
Use of probes with alternative incidence angles	Rail surface damage
	Operator fatigue
	Presence of rail welds and joints
	Flaw orientation

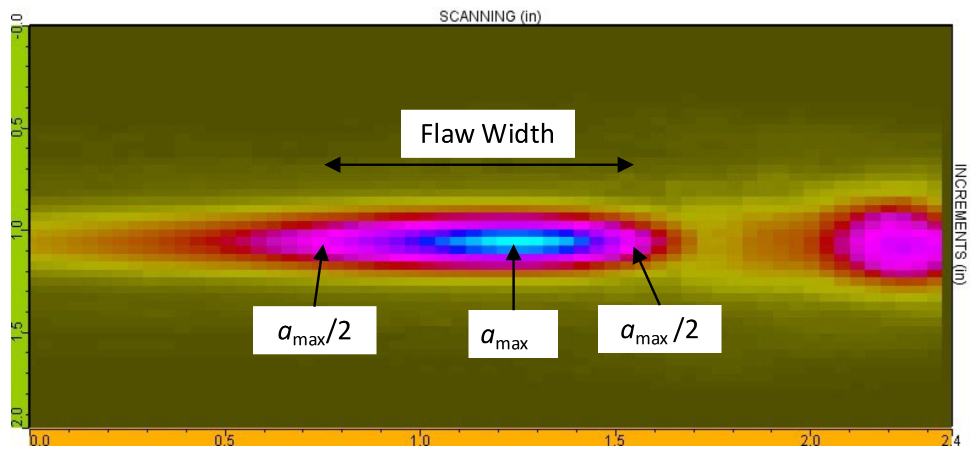


Figure 48. Estimating flaw width from C-scan results

9.2 CIVA Modeling

The team found the CIVA modeling software to be very sophisticated and more than capable of simulating handheld ultrasonic rail flaw inspection. CIVA's post-processing functionality was useful in calculating PODs and allowing the effects of thresholds and confidence levels to be investigated. The ability to export results for external post-processing to show percent rail head area, then to import the results in the modified format was important. It also allowed results from simulations with different rail profiles to be combined.

CIVA simulations are computationally intensive and require significant time. A sensitivity zone was specified to avoid processing when the flaw was outside the area covered by the ultrasonic rays. Since small flaws were often outside the sensitivity zone, this meant they could be simulated relatively quickly, often in less than five minutes. However, larger flaws took up to 12 hours to process. Batch processing was used to avoid constant supervision of the simulations.

CIVA is capable of modeling some of the physical factors that were not included in this study. These include rail surface condition, flaw orientation, and the presence of welds and rail joints. Future research could investigate the sensitivity of the calculated POD to these factors. CIVA also allows the probe to be skewed away from the longitudinal axis during scanning. This would more closely represent the technique operators use in the field to find flaws close to the gage face of the rail.

The calibration option in CIVA was not used in this study. This meant that the results in terms of a_{max} could not be converted to a physical value, and it was assumed that CIVA's results were valid enough to be used to calculate hit-miss PODs. This is a reasonable assumption given the extensive validation of CIVA by other researchers (Foucher, Lonne, Toullelon, Mahaut, & Chatillon, 2018) and previous FRA research (Poudel). However, it meant that the choice of detection threshold could not be made based on physical analysis. Future research could explore the benefits of using CIVA's calibration functionality.

9.3 Human Factors

This research considered the role of human factors in equipment calibration and the influence on POD. The human factors calibration tests produced useful data on the variability of three calibration factors: zero, refraction angle, and gain. Variations in zero and refraction angle were represented in CIVA by adjusting the geometrical properties of the wedge, since the calibration function in CIVA was not used in this study. It assumes the calibration is accurate, but the wedge properties vary. Future research using the calibration option in CIVA could explore the effects of incorrect calibration.

Variation in calibration was found to have a measurable effect on results. The variance in a_{max} results for a typical small flaw was found to increase by approximately 10 percent when calibration variation was included. The full set of simulations performed for this study could be repeated without the variation in calibration to determine the overall effect on POD.

Future research could investigate other human factors that might have significant effects on POD. This would likely involve human factors tests on a small number of physical samples. The tests would need to be carefully designed to produce results that would complement the computer simulations. Further human factors to consider include:

- Skill in maintaining correct lubrication between the probe and the rail. Results from physical tests could be simulated in CIVA.
- Skill in detecting flaws near the gage face by changing probe direction and using probes with alternative incidence angles. CIVA is also capable of simulating these factors.
- Accuracy in sizing flaws, which would require samples with known flaws or samples being broken open after testing. Results could then be compared to those from CIVA flaw size calculations.
- Level of training and degree of experience. Human factors tests with operators from different parts of the industry may use different procedures and have a different skill level than those that participated in this study.

9.4 Other Track Inspection Technologies

The handheld ultrasonic rail flaw inspection studied here is typically used to follow up from continuous testing by vehicle-mounted systems. These onboard systems often use probes embedded in wheels. FRA has sponsored the development of CIVA models for this type of system (Poudel). Those models could be used to calculate POD in a similar way to that reported here.

The feasibility study (Bruzek, Feasibility of Applying Model-Assisted Probability of Detection to Track Inspection, 2021) showed track geometry was another good candidate for MAPOD analysis. Commercially available software for modeling vehicle-track interaction could be used to simulate the dynamic motion of a geometry measurement system moving along a track with known geometry flaws. The onboard analysis of measured signals could be duplicated by post-processing the outputs from the dynamic simulation. This would require development of models that simulate the responses of optical and laser systems typically used to measure displacements on track geometry inspection vehicles.

The modeling of track geometry inspection could be validated by comparing results with those from testing at the high-speed adjustable perturbation slab track at TTC. The validated model could be used to generate *a*-hat versus *a* data and produce POD curves for various track geometry parameters.

10. Conclusion

Previous research (Bruzek, Feasibility of Applying Model-Assisted Probability of Detection to Track Inspection, 2021) identified MAPOD as a method for increasing confidence in POD estimates by using computer models to simulate track inspection technologies. The research in this study successfully demonstrated MAPOD applied to handheld ultrasonic rail flaw inspection. This technology was chosen from a feasibility study of six different track inspection technologies due to the availability of sophisticated software (CIVA) for simulating ultrasonic inspection.

A CIVA model was developed from previous FRA-sponsored research and used to generate 512 simulated inspection results from a variety of rail profiles, with flaws ranging from 1 to 76 percent of rail head area. Each result was generated using a random selection of input variables that described the location of the flaw and the properties of the ultrasonic probe. Probe properties were established from calibration tests with ultrasonic operators.

Human factors calibration tests showed significant variation in results. With one small exception, no two out of the three operators produced statistically similar calibration factors. A single operator calibrating the same equipment on six different occasions produced inconsistent results. Excluding these human factors in an example subset of simulations reduced the variance. The effect on POD could be determined by repeating simulations without the human factors on the full dataset. Human factors tests with other cohorts of operators could quantify the effects of different calibration procedures and levels of operator experience.

The POD estimates from the full set of 512 results had sufficient levels of confidence to allow meaningful comparison with limits in federal regulations. They could also be used to compare rail flaw detection technologies. This demonstrates the benefit of MAPOD when physical testing of more than a few samples is impractical.

This research focused on several physical and human factors known to affect POD. Other physical factors, such as rail surface condition and the presence of welds and joints, could also be studied. Human factors beyond those considered are likely to have significant effects on POD. Operators' skills in detecting flaws near the gage face of the rail and sizing flaws could be evaluated and used in the MAPOD simulations.

A logical extension of this research would be to post-process the simulation results to estimate the size of the flaws. This would enable the calculation of signal-response (*a* versus *a*-hat) PODs. Conclusions could then be drawn about the technology's ability to accurately size flaws. This would simulate the results achieved by operators when they use handheld equipment to verify indications and size flaws reported from onboard systems.

Since the team has successful developed an approach to applying MAPOD to one track inspection technology, it could be applied to others. Track geometry measurement would be a practical technology to investigate next.

11. References

AREMA (2022). *Manual for Railway Engineering.* American Railway Engineering and Maintenance-of-Way Association.

ASNT (2022, August 24). *ASNT NDT Certification Programs*.

Banerjee, A., Morrison, K., Witte, M., & LoPresti, J. (2021). *FAST High Strength Rail Test: Results through 325 MGT.* Pueblo: Transportation Technology Center, Inc.

Bruzek, R. (2021). *Feasibility of Applying Model-Assisted Probability of Detection to Track Inspection* (Report No. RR 21-18). Federal Railroad Administration.

Bruzek, R. (2021). *Model-Assisted Probability of Detection for Ultrasonic Rail Flaw Inspection* (Report No. RR 21-19). Federal Railroad Administration.

Bruzek, R., Drape, G., Einbinder, D., & Tunna, J. (n.d.). *A Procedure for Evaluating the Performance of Track Inspection Technology.*

Bruzek, R., Maymand, S., Drape, G., Smart, K., & Tunna, J. (2020). *Methods for Evaluation of Track Inspection Technology Effectiveness* (Report No. DOT/FRA/ORD-20/03). Federal Railroad Administration.

EXTENDE (2022). *Ultrasonic Testing with CIVA*.

Foucher, F., Lonne, S., Toullelon, G., Mahaut, S., & Chatillon, S. (2018). *An Overview of Validation Campaigns of the CIVA Simulation Software*.

FRA (2015). *Track Inspector Rail Defect Reference Manual*. U.S. Department of Transportation.

Nordco Rail Services (2021, January 14). *Rail Flaw Defects Identification Handbook*.

Office of the Federal Register (2021). *Code of Federal Regulations, Title 49 Part 213 - Track Safety Standards*.

Poudel, A. (n.d.). *Quantification and Evaluation of Rail Flaw Inspection Practices and Technologies – Phase I.* Federal Railroad Administration.

Schmerr, L., & Song, S. (2007). *Ultrasonic nondestructive evaluation systems: models and measurements.* Springer.

Sheehan, R. (2018). *Recharacterization of Rail Defect Test Facility at the Transportation Technology Center* (Report No. DOT/FRA/ORD-18/12). Federal Railroad Administration.

Stabler, L., Irani, F., Li, D., & Tajaddini, A. (2016). Advancements in testing facilities for railways. *World Congress on Railway Research.* Milan: Rail Safety & Standards Board, UK.

U.S. Department of Defense (2009). *Nondestructuve Evaluation System Reliability Assessment.* ASC/ERNS.

Appendix A.
Feasibility Considerations

Table A.1 Ultrasonic rail flaw detection

Controlling Factor	Experimental/Theoretical Considerations	Experiment or Theory
Electronic signal conditioning noise	A few experiments on clean rail could quantify background noise level.	Experiment
Surface condition	Difficult to keep the condition of stored samples stable.	Theory
Flaw geometric properties	Large range of possible sizes, locations, orientations, and shapes. Can be modeled in CIVA.	Theory
Rail type	Large range of possible cross-section and chemical properties. Can be modeled in CIVA.	Theory
Non-flaw features	Large range of possible features (e.g., rail welds and joints).	Theory
Human factors	Influenced by fatigue and skill of operator. Affects calibration error, probe positioning and contact, interpretation of scans, etc. TTC rail defect test facility available, but sample size is limited.	Experiment
Accuracy target	More false positives likely if low false negatives required. TTC rail defect test facility available, but sample size is limited.	Experiment

Table A.2 Inertial track geometry measurement

Controlling Factor	Experimental/Theoretical Considerations	Experiment or Theory
Electronic signal conditioning noise	May be affected by measurement direction and car orientation. May only require a few tests to establish noise level.	Experiment
Flaw type	Large range of possible flaw types. A vehicle dynamics model of the measurement car and instrumentation would need to be developed.	Experiment
Longitudinal position error	e.g., encoder and GPS errors. Could use published GPS accuracy data.	Theory
Measurement speed	TTC test track available. Safety is a concern for large flaws. This and other vehicle-related factors could be modeled with currently available vehicle dynamics software.	Theory
Measurement direction	TTC test track available (see above).	Theory
Leading or trailing orientation	TTC test track available (see above).	Theory
Vehicle-track interaction	Likely to vary with speed (see above).	Theory
Non-flaw features	e.g., switches (see above).	Theory
Equipment mounting	Can come loose over time.	Theory
Track movement	Including the effects of curvature.	Theory
Laser image quality	Including effects of sunlight and obstructions. Would require a database of images.	Theory
Human factors	Editing error, configuration error, calibration error (e.g., gage, encoder). Could be tested without track runs.	Experiment
Automated editor algorithm error	Would require a database of raw measurement datasets.	Theory

Table A.3 Gage restraint measurement

Controlling Factor	Experimental/Theoretical Considerations	Experiment or Theory
Electronic signal conditioning noise	May only require a few tests to establish noise level.	Experiment
Pressure control system	i.e., applied force consistency. May not require on-track testing.	Experiment
Longitudinal position error	e.g., encoder and GPS errors. Could use published GPS accuracy data.	Theory
Measurement speed	Would require on-track testing.	Experiment
Platform weight and configuration, and car geometry	This and other vehicle-related factors could be modeled with currently available vehicle dynamics software.	Theory
Truck dynamics	For truck-mounted axles (see above).	Theory
Force level normalization	Involves PLG and GWP calculation equations.	Theory
Track structure	e.g., timber or concrete.	Theory
Mechanical noise	e.g., bearing condition.	Theory
Wheel profile	Profile wears over time (see above).	Theory
Curvature	Effect on unloaded gage (see above).	Theory
Image quality	Including effects of sunlight and obstructions. Would require a database of images.	Theory
Human factors	e.g., editing, axle position, configuration file, and calibration.	Experiment

Table A.4 Ballast condition (ballast fouling index) by ground penetrating radar

Controlling Factor	Experimental/Theoretical Considerations	Experiment or Theory
Electronic signal conditioning noise	From cabling and circuit boards. May only require a few experiments.	Experiment
Electro-magnetic interference	This is a principal source of noise. Modeling may be possible in CIVA.	Theory
Surface condition	Difficult to reproduce consistent test conditions. Difficult to build and validate a model (GPR Max research project is the only known software).	Theory
Sub-surface obstruction	Large range of possible sizes, orientations, and shapes. Modeling may be possible in CIVA. Need to determine the range of obstructions.	Theory
Human factors	e.g., calibration error, system initiation checklist. Could be tested without track runs.	Experiment
Human factors	Post-processing set up	Experiment

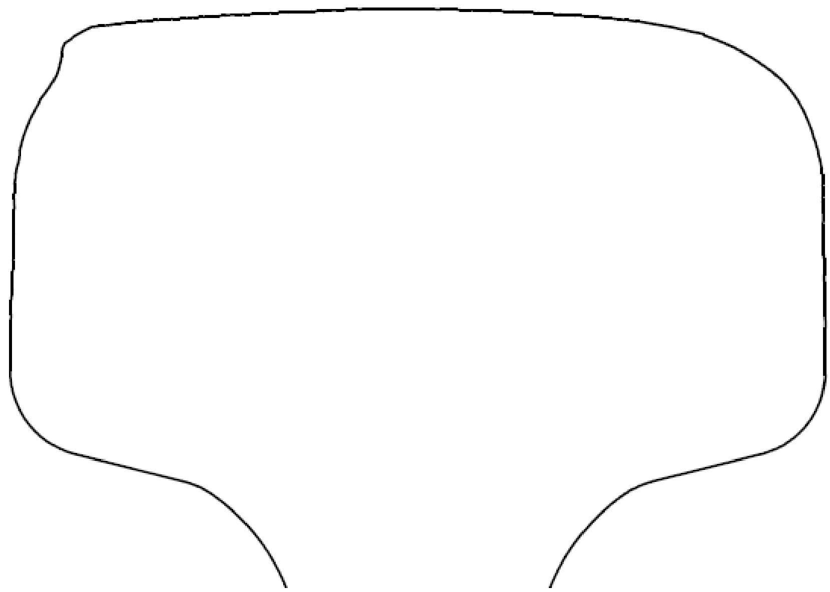
Table A.5 Grade crossing profile measurement by LiDAR

Controlling Factor	Experimental/Theoretical Considerations	Experiment or Theory
Electronic signal conditioning noise	Including CMOS noise (black and white intensity), and inertial package noise. Could require many tests.	Experiment
Mechanical noise	Including vibrations and mechanical damage to mechanical rotating mirrors. Could require many tests.	Experiment
Signal obstruction	e.g., from fog, rain, snow, smog, dust. Model would have to exist to describe laser wave propagation.	Theory
Profile geometric properties	Large range of possible sizes, orientations, and shapes. Database of geometries required.	Theory
Presence of road vehicles, etc.	Crossing should be clear, but vehicles may be on approaches, depends on number of lidar cameras. Adjacent track also causes loss of optical resolution on the sides.	Theory
Target surface obstructions	e.g., by trees. Database of obstructions required.	Theory
Measurement speed and orientation	Slower speed better resolution, obstruction when reversing in hi-rail. Tests needed at several speeds and a wide range of other factors such as obstructions.	Experiment
Human factors	Position sensors must be calibrated and systems may have multiple LiDARs. Could be tested without track runs.	Experiment
Mechanical noise	Sensors can be knocked out of position and misaligned. Database of sub-factors required.	Experiment

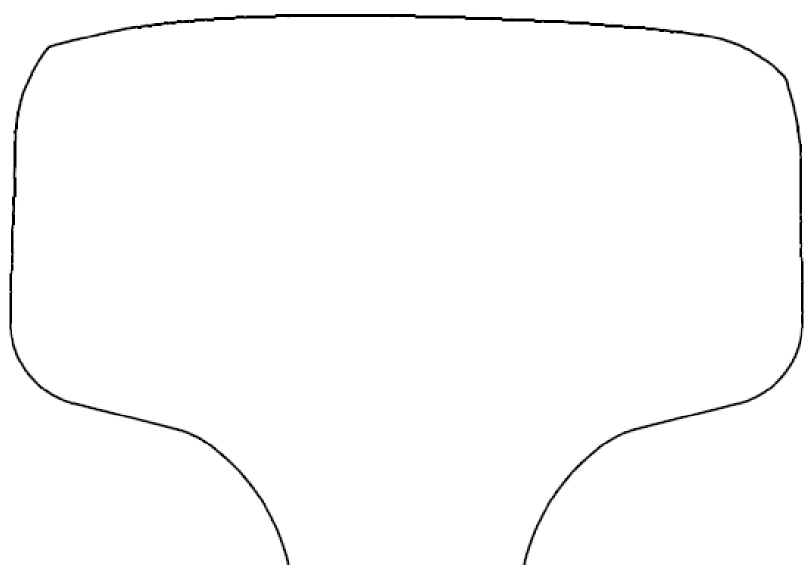
Table A.6 Joint bar condition by camera image

Controlling Factor	Experimental/Theoretical Considerations	Experiment or Theory
Camera image quality	Lighting, shading, environmental conditions, focus. Could create edited images for automatic processing. Some experiment needed to determine range of conditions. Library of training images are available.	Theory
Image obstructions	Joint bar not detected because obstructed by ballast, etc. Like image quality assessment. Database of obstacles required.	Theory
Flaw geometric properties	Large range of possible sizes, locations, orientations, and shapes. Library of flaw types available. Other examples may be needed.	Theory
Joint bar type	Affects joint bar detection. Small number of types in practice.	Theory
Non-joint bars detected	Some crossing surfaces and features can look like joint bars.	Experiment
Artifacts that look like cracks	e.g., grass, shadows, cracks in insulation. Database of artifacts required.	Theory
Operating speed	Could affect image quality. Tests needed over a range of environmental conditions.	Experiment
Human factors	Exceptions dismissed incorrectly. May not require field testing.	Experiment

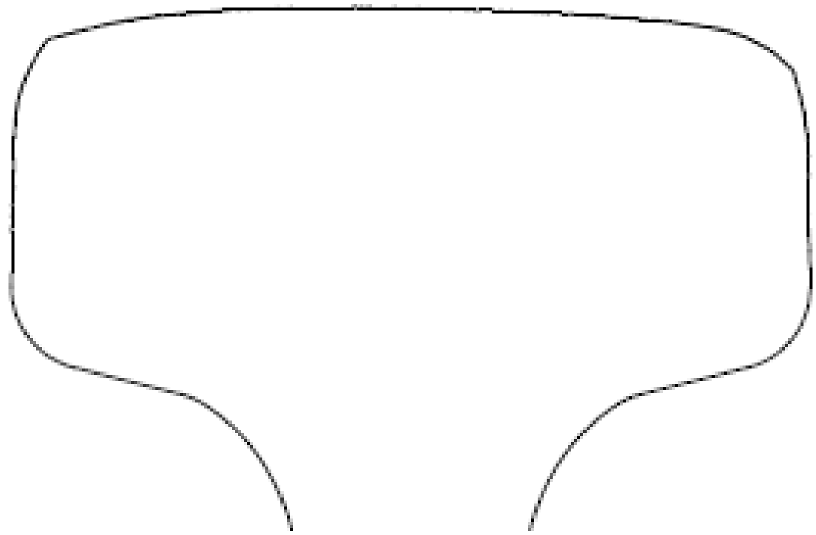
Appendix B.
Selected Rail Profiles



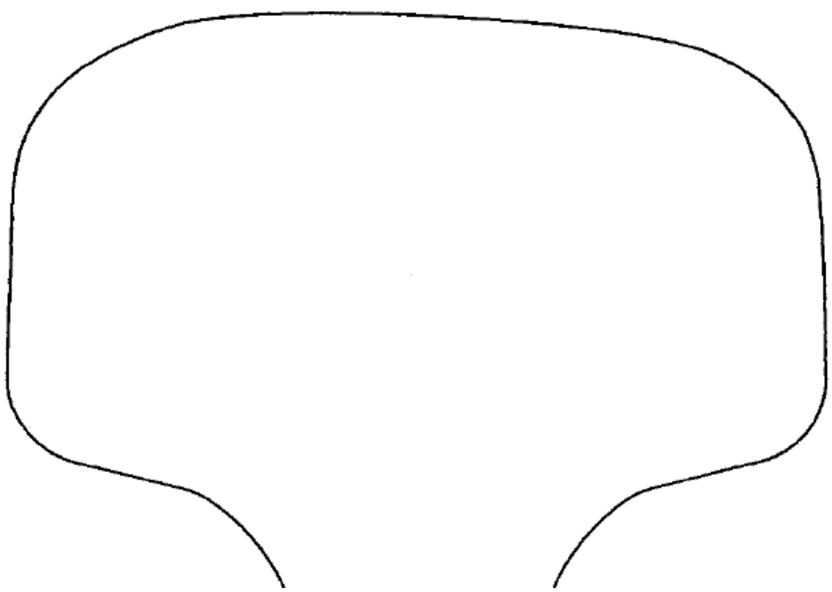
S3a – Lightly worn tangent rail



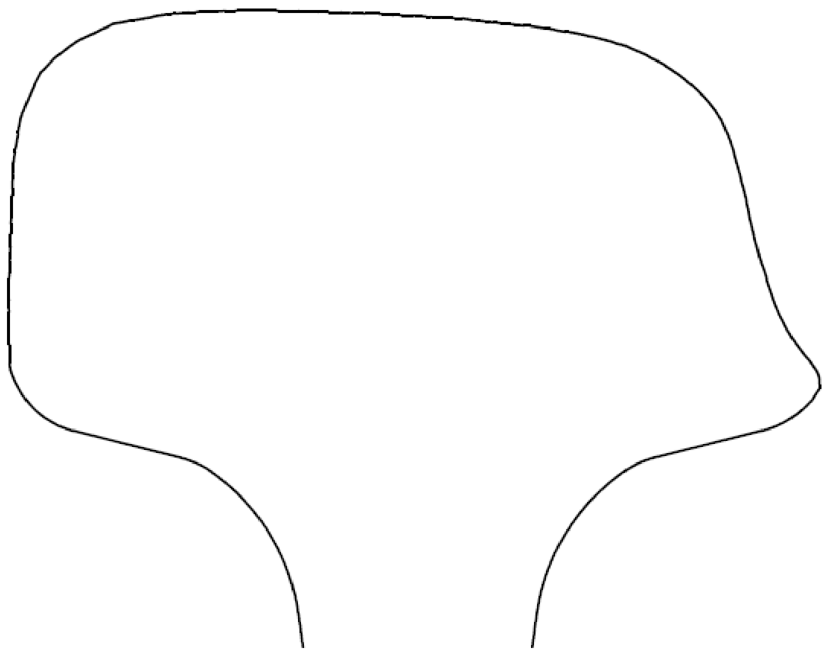
S7a – Moderately worn tangent rail



E2a – Heavily worn tangent rail



N11a – Lightly worn curve rail



W4a – Moderately worn curve rail

Appendix C.
Human Factors Test Results

Table C.1 list the results from human factors tests in the order the calibrations were performed.

Table C.1 – Calibration Results from Human Factors Test

#	Operator	Wedge	Condition	Device	Zero (µs)	Angle (Deg.)	Gain (dB)
1	B	6	N	600	11.622	68.0	60.2
2	B	7	N	600	11.940	68.0	60.4
3	C	3	W	650	10.236	69.0	64.5
4	A	5	W	650	11.998	70.0	65.5
5	B	2	W	600	10.905	70.0	64.8
6	B	4	N	600	12.088	68.0	60.9
7	A	9	W	650	11.644	70.5	69.7
8	B	5	W	600	10.943	68.5	63.8
9	C	10	W	650	10.842	68.0	64.2
10	A	1	W	650	10.872	69.8	66.8
11	B	2	W	600	12.037	70.0	64.0
12	A	7	N	650	12.511	69.5	62.7
13	C	8	W	650	12.504	67.5	65.0
14	A	4	N	650	12.064	70.0	62.6
15	B	9	W	600	11.165	68.5	67.4
16	B	6	N	600	12.542	69.0	61.1
17	A	2	W	650	11.517	71.8	65.5
18	B	5	W	600	11.315	69.0	62.4
19	C	1	W	650	9.411	69.0	68.0
20	A	3	W	650	12.200	69.3	65.4
21	B	8	N	600	12.507	68.5	61.5
22	A	6	N	650	12.298	70.0	63.0
23	B	3	W	600	12.051	69.0	62.3
24	B	4	N	600	11.610	68.0	60.8
25	B	9	W	600	11.366	70.0	67.4
26	C	7	N	650	11.109	69.0	63.9
27	B	2	W	600	11.260	70.0	64.1
28	B	1	W	600	11.214	70.0	64.9
29	C	9	W	650	9.922	69.8	70.4
30	B	3	W	600	12.546	69.0	62.1
31	B	6	N	600	11.629	69.0	61.0
32	C	4	N	650	10.708	69.0	64.1
33	B	7	N	600	12.381	69.0	61.0
34	A	8	N	650	12.316	71.5	62.1

#	Operator	Wedge	Condition	Device	Zero (µs)	Angle (Deg.)	Gain (dB)
35	A	8	N	650	11.654	69.0	62.6
36	A	10	W	650	12.837	67.9	63.1
37	B	9	W	600	11.588	68.5	68.3
38	C	5	W	650	10.335	70.0	66.6
39	A	6	N	650	12.044	69.2	62.9
40	B	8	N	600	12.396	69.0	60.6
41	B	10	W	600	12.408	68.0	61.5
42	A	1	W	650	11.479	70.3	66.6
43	C	2	W	650	9.964	72.0	68.3
44	B	5	W	600	11.914	68.5	62.9
45	A	10	W	650	13.198	67.5	63.8
46	C	6	N	650	10.657	69.0	64.2
47	B	8	N	600	11.914	68.5	60.9
48	A	5	W	650	12.653	69.2	66.3
49	B	7	N	600	12.199	69.0	60.9
50	A	3	W	650	12.089	69.2	64.0
51	C	4	N	650	10.815	69.0	64.2
52	B	1	W	600	10.776	69.0	65.2
53	A	2	W	650	11.905	70.3	66.3
54	B	6	N	600	12.991	68.0	64.3
55	A	4	N	650	12.807	69.0	63.1
56	C	9	W	650	9.823	69.3	70.9
57	B	10	W	600	11.805	68.0	61.9
58	A	7	N	650	12.325	70.0	63.2
59	B	9	W	600	11.353	70.0	67.5
60	B	4	N	600	11.290	70.0	61.3
61	C	3	W	650	10.296	69.0	65.8
62	A	9	W	650	11.936	69.1	69.1
63	A	10	W	650	12.585	69.2	63.1
64	B	2	W	600	11.798	70.5	64.3
65	C	8	N	650	10.669	69.0	64.7
66	A	4	N	650	12.336	69.3	62.8
67	A	3	W	650	11.377	70.0	63.8
68	C	10	W	650	10.976	68.0	64.7
69	A	5	W	650	11.576	70.4	64.6
70	C	6	N	650	10.750	69.0	64.1
71	B	3	W	600	11.811	68.0	62.2
72	B	5	W	650	11.207	68.0	64.0
73	A	1	W	650	11.750	69.0	66.2
74	B	8	N	600	11.933	68.0	61.0

#	Operator	Wedge	Condition	Device	Zero (µs)	Angle (Deg.)	Gain (dB)
75	C	4	N	650	10.925	68.7	64.3
76	A	9	W	650	11.668	69.2	69.3
77	B	7	N	600	12.977	68.0	61.2
78	B	8	N	600	12.214	68.0	61.2
79	A	2	W	650	11.474	71.3	66.0
80	C	1	W	650	9.726	69.0	68.3
81	B	3	W	600	10.748	68.0	62.2
82	B	2	W	600	11.336	71.0	64.4
83	A	6	N	650	12.428	69.5	63.6
84	B	9	W	600	11.363	68.0	67.4
85	C	8	N	650	10.714	68.6	64.9
86	B	6	N	600	11.799	68.0	61.2
87	A	7	N	650	12.462	70.2	62.7
88	B	5	W	600	11.578	69.0	63.2
89	A	8	N	650	12.440	70.0	64.2
90	C	2	W	650	9.815	71.5	68.4
91	B	4	N	600	11.936	69.0	61.3
92	B	1	W	600	10.859	68.0	65.5
93	A	6	N	650	12.601	70.0	63.0
94	B	10	W	600	11.930	67.5	61.9
95	C	7	N	650	10.562	69.0	64.7
96	A	5	W	650	11.912	70.0	65.0
97	B	7	N	600	11.993	69.5	62.4
98	A	4	N	650	12.952	69.3	64.1
99	B	2	W	600	10.578	71.5	64.8
100	C	9	W	650	9.851	69.5	71.3
101	A	1	W	650	11.892	69.1	66.7
102	B	3	W	600	11.794	68.0	62.8
103	C	5	W	650	10.233	69.5	67.1
104	A	2	W	650	11.012	70.8	66.2
105	B	8	N	600	12.214	68.0	60.6
106	B	1	W	600	11.416	68.0	65.4
107	A	3	W	650	11.587	70.4	63.8
108	C	7	N	650	10.724	69.1	64.8
109	B	5	W	600	11.458	69.0	63.4
110	A	9	W	650	12.635	70.0	69.1
111	B	6	N	600	12.638	70.0	61.3
112	A	10	W	650	12.564	70.2	63.6
113	C	1	W	650	9.690	69.0	68.3
114	B	4	N	600	12.311	68.5	61.2

#	Operator	Wedge	Condition	Device	Zero (µs)	Angle (Deg.)	Gain (dB)
115	B	1	W	600	10.844	70.0	65.6
116	A	8	N	650	12.084	69.5	63.8
117	B	10	W	600	13.603	69.0	61.7
118	C	6	N	650	10.694	69.0	64.6
119	B	3	W	600	11.685	69.0	62.4
120	A	7	N	650	12.400	70.0	63.6
121	C	4	N	650	10.871	69.0	64.6

Appendix D.
CIVA Parameters

This Appendix lists the fixed CIVA parameters used to create the base simulation model. References are given to CIVA User Guide (2020, Service Pack 3 version). This information is provided so future researchers can reproduce the model used in this study.

Some of the CIVA inputs were taken from other FRA-sponsored research (Poudel).

The computer used for running CIVA had the following specification:

- CPU: 10th Generation Intel(R) Core(TM) i7-10875H (16MB Cache, up to 5.1 GHz, 8 cores)
- OS: Windows 10 Pro
- RAM: 32GB DDR4-2933MHz, 2x16G
- HD: 1TB M.2 PCIe NVMe Solid State Drive
- GPU: NVIDIA(R) GeForce RTX(TM) 2060 6GB GDDR6 with Max-Q

D.1 SPECIMEN (CIVA 2.2.1)

D.1.1 Geometry (CIVA 2.2.1.1, page 65)

Geometry is defined in a 2D CAD (.dxf) file. This defines the rail's cross-sectional shape. A plane extension is used to give the rail a length of 600 mm.

A separate CIVA model was created for each of five rail profiles. These models were run separately, and the results combined to give an overall POD. The number of simulations for each profile was adjusted so that when combined the results correctly represented the estimated distribution of these profiles in the national network.

D.1.2 Roughness (CIVA 2.2.1.2, page 91)

Roughness of the top surface of the rail was set to 0 µm on the assumption rail in service is clean and smooth.

D.1.3 Material (CIVA 2.2.1.4)

The following material properties of the rail model were used:

1. Type: "Isotropic"
2. Density: 7.8 g/m^3
3. Longitudinal wave velocity: nominally 5,900 m/s
4. Transverse wave velocity: 3,250 m/s
5. Noise type: "None"
6. Attenuation type: "None"

D.1.4 Additional Component (CIVA 2.2.1.5)

There were no additional components.

D.2 PROBE (CIVA 2.2.2, page 129)

The properties of the probe were imported. This defined the following:

D.2.1 Type of Transducer (CIVA 2.2.2.1, page 131)

The probe manufacturer was Olympus, and the transducer model was A540S. The probe was single element (not phased array), piezoelectric (not EMAT).

1. Probe Type: "Contact"
2. Pattern: "Single element"
3. Crystal shape "Circular"
4. Crystal diameter: 12.7 mm
5. Focusing, Surface type: "Flat"

D.2.2 Wedge Details (CIVA 2.2.2.1, page 133)

The transducer was fitted to a 70°, flat-bottomed, acrylic wedge. The wedge manufacturer was Olympus, and the wedge model was ABSA-5T-70. The wedge properties were:

1. Geometry: "Flat"
2. Front Length: 10.0 mm
3. Back Length: 17.0 mm (as measured from a new wedge)
4. Height (defined as the length of the sound path in the wedge): nominally 15.0 mm (the wedge height will be varied in the POD analysis to account for variation in the calibration zero)
5. Width: 18.5 mm
6. Refraction Angle: nominally 70°. The refraction angle in degrees is a variable in the POD analysis.
7. Incidence Angle: calculated by CIVA from the refraction angle and the velocities of ultrasonic shear waves in the wedge and specimen. A new wedge has an incidence angle of approximately 53°, which can be used to check the CIVA calculated value.
8. Wave Type: Transverse
9. Longitudinal wave velocity (in the rail): 5,900 m/s.
10. Transverse wave velocity (in the rail): 3,250 m/s
11. Rotation: 0°
12. Squint Angle: 0°
13. Material: Acrylic with standard properties
14. Wedge density: 1.19 kg/m^3
15. Wedge longitudinal velocity: 2,730 m/s
16. Wedge transverse velocity: 1,430 m/s

D.2.3 Focusing (CIVA 2.2.2.3)

Not relevant to the contact probe (only relevant for immersion probes).

D.2.4 Signal (CIVA 2.2.2.4, page 168)

Signal properties were loaded from an external file that defined the following:

1. Signal choice: Hanning

2. Center Frequency: 2.25 MHz
3. Bandwidth: 46.07 percent at -6 dB
4. Phase: 0°
5. Sampling Frequency: 200 MHz
6. Number of Points: 8,192
7. Temporal position: 20.48 μs

D.2.5 Probes Library (CIVA 2.2.2.5)

The probe and wedge could not be found in the CIVA library, so their properties were input manually.

D.3 INSPECTION (CIVA 2.2.3)

The inspection system was a single transducer operating in the pulse/echo mode (both transmitting and receiving).

D.3.1 Inspection Settings (CIVA 2.2.3.1)

Inspection system: "Single transducer"

Type of trajectory: "Parametric"

D.3.2 Specimen Frames (CIVA 2.2.3.2, page 175)

The rail profile defined in the 2D CAD file is the X, Z plane in the CIVA simulation (it is the X, Y plane in the 2D CAD file). The specimen was created by extending this profile in the Y direction. The origin was also defined in the 2D CAD file.

Table D.2 shows details of the specimen frame.

Table D.2. Specimen Frame Details

Axis	+ve Direction	Origin
X	To the right looking along the rail	Center of the rail web
Y	Along the rail	0.0
Z	Downwards	186 mm above bottom of rail

D.3.4 Configuration (CIVA 2.2.3.4, page 178)

The inspection configuration was set as follows:

1. Inspection plane: "Perpendicular to profile plane"
2. Scanning direction: "Positive"
3. Matched contact: "Adapted probe" was checked except for the sensitivity test with the W4a rail profile
4. New specimen origin: X and Y = zero (the specimen was not moved from its initial origin)
5. The configuration was "Coupling medium".
6. Coupling medium: "EcoGel".
7. Density: 1.02 g/m2
8. Primary wave velocity: 1,550 m/s

9. Attenuation type: “None”
10. Bottom medium: “Air”

D.3.5 Positioning (CIVA 2.2.3.5, page 194)

This is the start position of the probe. Wedge center was defined by: X = -25 mm, Y = 175 mm.

D.3.6 Scanning (CIVA 2.2.3.7, page 197)

The positioning mode was “Impact point” by default for a contacting probe.

The displacement mode was “Along profile mode.” This means the scan was first made across the head of the rail and then moved in increments along the rail.

Translation along X:

1. Direction of scanning: “Tracking along axis”
2. Step: 1 mm
3. Number of steps: 50

Translation along Y:

1. Step: 1 mm
2. Number of steps: 60

Choice of scanning modes:

1. Increment/Scanning reversed: “No”
2. Increment skip: “Raster”

D.3.7 Inspection Mode (CIVA 2.2.3.8)

The inspection mode was “Single transducer.” There was no need to specify the properties of a second probe.

D.4 FLAWS (CIVA 2.2.5, page 274)

A single flaw was modeled. Its size and position were variables in the POD calculation. It could break the surface. The flaw was in the X, Z plane.

D.4.1 Flaw Properties (CIVA 2.2.5.1)

The flaw geometry was “Elliptical planar.”

D.4.2 Flaw List Management (CIVA 2.2.5.2)

There was only one flaw in the list.

D.4.3 Geometry (CIVA 2.2.5.3, page 276)

1. Length: 5 mm to 65 mm in 4 mm steps
2. Height: 70 percent of the flaw length
3. Type: “Full”

D.4.4 Flaw Positioning (2.2.5.4, page 291)

The flaw was positioned in the “Specimen Frame.” The flaw positioning parameters were as follows:

1. The flaw positioning: “Along profile”
2. Positioning mode: “Defect center”
3. Positioning X: Distributed randomly between -22.2 mm and 37.8 mm (this is a variable in the POD analysis)
4. Positioning Y: 250 mm (the flaw is in a fixed longitudinal position)
5. Positioning Z: Distributed randomly between 14 mm and 32.5 mm (this is a variable in the POD analysis)
6. Orientation: Tilt: 0°, Skew 0°, Rotation 0°

D.5 UT INSPECTION SIMULATION (CIVA 2.4, page 354)

D.5.1 Simulation Settings (CIVA 2.4.2, page 354)

Initialization: “Advanced definition”

Involved modes: default checked with “Transversal wave”

Interactions: the active surfaces of the specimen are the back wall with a single reflection simulated.

Skips: Number of half skips was 1 Max (checked)

Flaw: the single flaw was activated (the model was “Kirchhoff & GTD”)

Sensitivity zone: Enable sensitivity zone was checked. The following were specified:

1. Xzone : 100 mm
2. Yzone: 50 mm
3. Zzone: 50 mm
4. Mechanical link: “Probe”
5. Positioning: “In the inspection plane”
6. Orientation: “Arbitrary”
7. Coordinates: “Cartesian”
8. Depth direction: “Along Local Normal”
9. Local cartesian coordinates X: 62 mm
10. Local cartesian coordinates Z: 28 mm
11. Rotation around y: 0°

Modes: List modes was checked with number of modes equal to 3. Automatically compute modes list was checked.

Gates: Gate 1 was set to auto. Bounds was set to auto. Synchronization was set to Detection synchro: Echo max(Abs).

Options: computation mode was set to “3D”. Field/reflector interaction was set to “Plane wave approximation for incident beam.” Accuracy was set to “Field=1” and “Defect=1.” Modes identification was set to “Activate” and Number of modes to return was 5.

D.5.2 Calibration (CIVA 2.4.2.5, page 368)

Calibration: calibration types was set to “None.”

Abbreviations and Acronyms

ACRONYM	DEFINITION
2D	Two Dimensional
ANOVA	Analysis of Variance
AREMA	American Railway Engineering and Maintenance-of-Way Association
ASNT	American Society for Nondestructive Testing
CAD	Computer Aided Design
CIVA	Trade name of ultrasonic modeling software
Df	Degrees of Freedom
EXTENDE	Company supplying CIVA
FAST	Facility for Accelerated Service Testing
FRA	Federal Railroad Administration
IIW	International Institute for Welding
LiDAR	Light (or Laser Imaging) Detection and Ranging
MAPOD	Model Assisted Probability of Detection
MS	Mean Square
POD	Probability of Detection
SS	Sum of Squares
TTC	Transportation Technology Center