The State of Integrity

Sheri Baucom
Irth Solutions

PPIM 2025
PIPELINE PIGGING & INTEGRITY MANAGEMENT
JANUARY 27-31 • HOUSTON

Organized by
CLARION
TECHNICAL CONFERENCES

https://doi.org/10.52202/078572-0052

Proceedings of the 2025 **Pipeline Pigging and Integrity Management Conference.**

https://doi.org/10.52202/078572-0052

Abstract

With over 11,000 assessments on 410k miles of pipeline from 17 pipeline operators and 34 inline inspection (ILI) companies, OneBridge Solutions (now Irth Solutions) is believed to have amassed the largest and most diversified set of ILI data in the industry. This paper presents insights gleaned from the analysis of said ILI data, with the focused aim of determining what variables contribute to high ILI system performance, specifically for technologies that characterize metal loss, i.e., ILI company, technology, metal loss depth, metal loss classification, etc. These insights are presented to achieve the foremost goal of shared learning and provide valuable information that all stakeholders can utilize to improve pipeline integrity.

What is the State of Integrity?

Since its inception in 2017, pipeline operators have uploaded inline inspection (ILI) reports from natural gas and hazardous liquid pipeline assessments into OBS's Cognitive Integrity Management or CIM platform. This application integrates and analyzes said data to meet an operator's requirements for their pipeline systems' integrity and risk management. In addition to inline inspection reports, other data sets are housed in CIM via integration with an operator's GIS (to provide pipeline properties and location information), user input, and actions undertaken within the platform, e.g., analyses.

As of November 1, 2024, CIM contained data from 11,676 inline inspections on 412,719 miles of natural gas and hazardous liquid pipelines in the US, Canada, and South America from 17 different pipeline operators. Comparing the mileage contained in CIM to the mileage reported to PHMSA in 2023, this represents 32% of the assessable US pipeline miles, while CIM customers operate 50% of assessable US pipelines.[a]

Inline inspections were conducted by 34 different inline inspection companies and utilized 230 different ILI technologies or combination of technologies, resulting in roughly 225 million features and 165 million anomalies ingested into CIM. A unity plot showing the ILI predicted depths versus the field-found depths, organized by the vendor, is provided in **Figure 1**.

---

[a] Assessable pipelines were defined as any pipeline identified as liquid gathering and liquid transmission (228,568 miles) and gas transmission (300,473), for a total of 529,041 miles. Miles under CIM subscription are 171,048; miles operated by CIM customers are 262,143.)

https://doi.org/10.52202/078572-0052

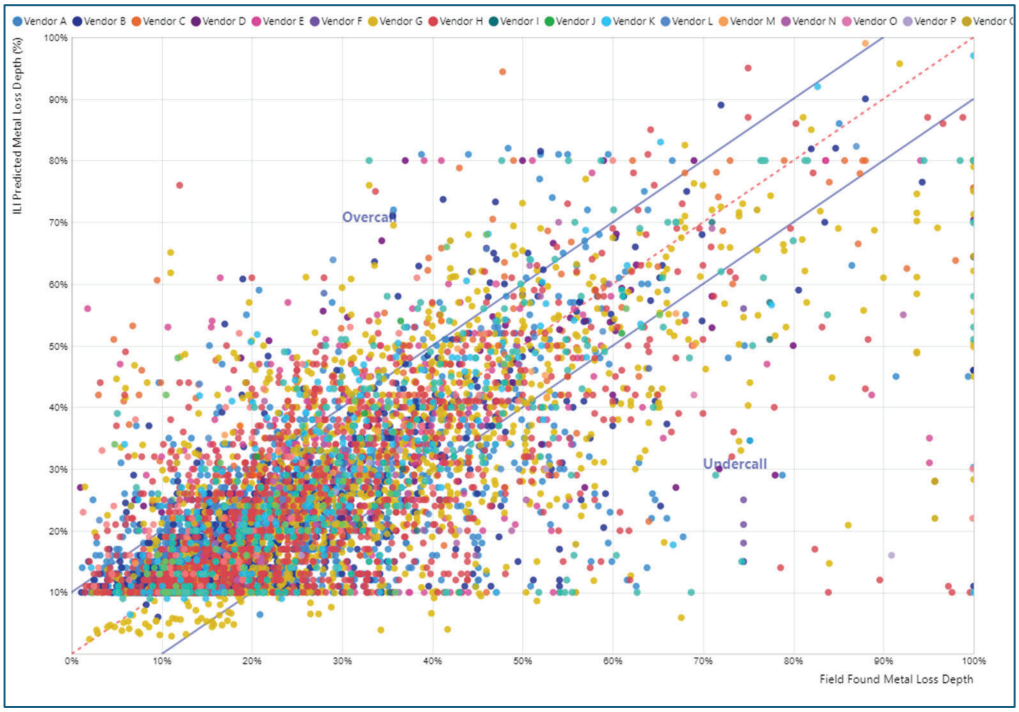


Figure 1. Unity plot for all anomalies utilized in this study, categorized by ILI company (Vendor).

The State of Integrity is a project and subsequent report that presents the findings of analyses conducted on said data contained within CIM. The goal of this project, which started in earnest in 2023, is to provide valuable insights regarding the integrity management of pipelines. This inaugural report explores the key variables contributing to the performance of ILI technologies that characterize metal loss.

How is ILI performance quantified?

For several reasons, ILI technologies designed to identify and quantify metal loss were the primary focus. Firstly, CIM has the highest amount of data on these technologies. This is most likely because metal loss, in the form of internal and external corrosion, is deemed a significant threat to a pipeline's integrity. Additionally, the methods for determining the depth of metal loss "in the field" are associated with less error than assessing the depth of anomalies with more complex morphologies like cracks.

The depth of an anomaly[b], typically deemed the most critical dimension of any metal loss on a pipeline, was utilized to quantify performance. The depth identified by the inline inspection report is compared with the depth reported by the pipeline operator/CIM user, typically from a physical

---

[b] It's recognized that "anomaly" per API 1163[1] means an unexamined deviation which, upon evaluation, is re-classified as an imperfection, defect, or feature. However, in this paper, an anomaly refers to unexamined and examined deviations with metal loss.

examination using some form of Non-destructive Examination (NDE).[c] Therefore, how well the ILI system predicted the depth of the metal loss anomaly can be analyzed. (Depth as a percentage of wall thickness was utilized for this study.) This study quantified the ILI system's performance in predicting depth by calculating certainty within a specified tolerance. Certainty and tolerance, as defined by API 1163,[1] are:

Certainty: the probability that a reported anomaly characteristic is within a stated tolerance.

Tolerance: the accuracy with which an anomaly dimension or characteristic is reported with a specified certainty, typically expressed as a range, i.e., +/- 10%.

Certainty provides the percent of metal loss anomalies whose predicted depths were within a tolerance. The higher this value, the better the performance. This utilizes the exact quantification of sizing accuracy recommended by API 1163[1] where the depth sizing accuracy at a certainty is stated for a certain tolerance, e.g., at 80% certainty, the depth sizing accuracy of a general corrosion anomaly is within +/- 10% (Denoted as 10%t or 20%t, when 20% is the tolerance in the visuals.)

To arrive at the certainty, the absolute (non-negative) difference is first calculated for each metal loss anomaly (ILI depth minus "field" depth). For example, an anomaly found in the field with a depth of 30% of the pipeline wall thickness that matched a predicted ILI measurement of 35% would have a difference of 5 and, therefore, be within a 10% tolerance. The number of times the absolute (non-negative) difference is less than 10 or 20 is then divided by the total number of matched anomalies to provide a percent certainty within a 10% tolerance and a 20% tolerance. The calculated certainties for a 20% tolerance by ILI vendor can be seen in **Figure 2**.

---

[c] This depth is commonly called the truth data or "actual" depth. However, it is recognized that an error is associated with measuring depth in the field, although this is not accounted for in this study.

https://doi.org/10.52202/078572-0052

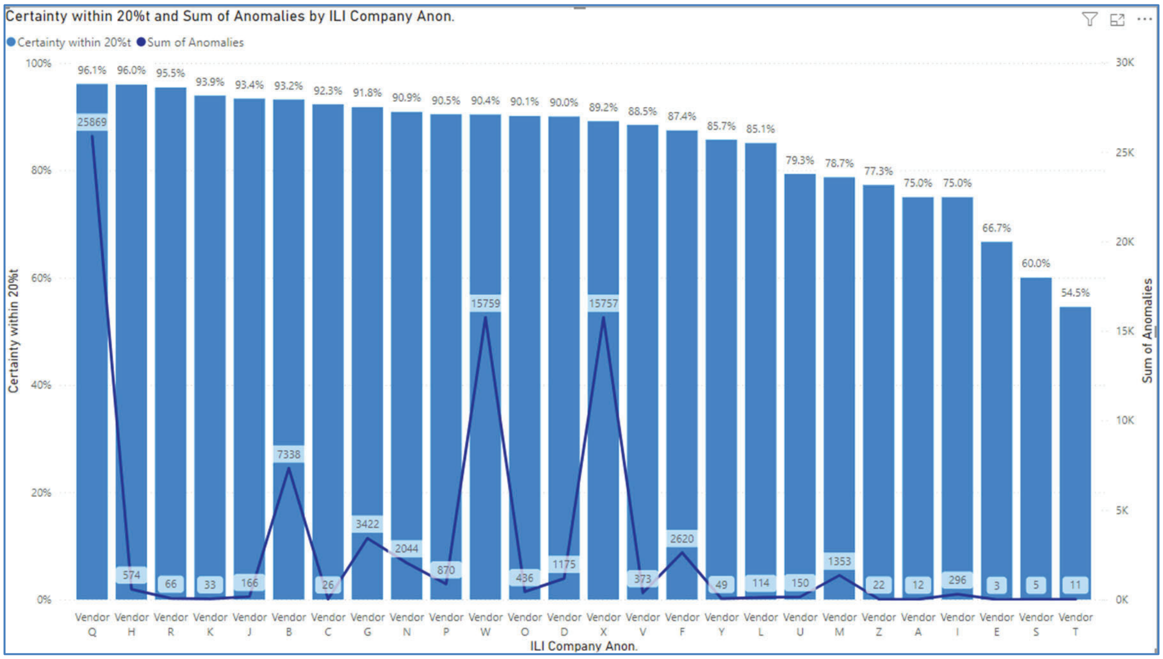


Figure 2. The certainty within 20%, by anonymized ILI company. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis. (As explained later, those certainties with < 23 data points are shown in this figure but excluded from the analysis.)

What variables were analyzed?

Variables were analyzed to determine their impact on performance, termed variability. Additionally, some (but not all) variables were analyzed to determine what had the highest performance and compare performances across variables.

ILI company

This variable encompasses the people, processes/procedures, algorithms/analysis technology, ILI technology proprietary to the company, etc. Although axial magnetic flux leakage (MFL-A) technology, for example, may not be proprietary to the company, the way that the ILI company employs it, the processes and procedures that are utilized to complete the inline inspection, the hardware and software, i.e., the algorithms and other analysis technologies used to analyze the raw data, and the people employed by the company, are specific to that company. Although 34 ILI companies have data in CIM, 21 companies have 23 or more anomalies to quantify metal loss depth sizing accuracy.

ILI technology

This variable is the technology deployed during the inline inspection to detect, identify, and characterize the metal loss anomaly. For over 70% of the inline inspections, a combination of ILI technologies, e.g., Deformation and MFL, was employed on one tool or deployed separately but later combined into one report. Therefore, to explore this variable, a dominant technology was identified. The ILI reports identified 230 different technologies/combinations of technologies. These were then grouped into 10 "dominant technology" categories, 5 of which were metal loss technologies:

- **MFL-A** = axial Magnetic Flux Leakage (MFL) standard resolution. This is MFL-A that was not identified as high resolution and the most common MFL technology. This category includes any combination of MFL and other ILI technology that does not detect metal loss, i.e., Hard Spot, EMAT, UT Crack, etc.
- **High Res. MFL** = axial MFL-A identified as high-resolution or ultra-high resolution axial MFL and any combination containing high-resolution axial MFL. High-resolution MFL tools have a higher density of sensors and are typically employed to detect and size pinholes more accurately. There does not seem to be an industry-wide consensus on when “high resolution” and “ultra-resolution” terms are used to describe MFL.
- **MFL-C** = any MFL technology that applies magnetization in a non-axial direction, i.e., in a circumferential or transverse direction. This category includes circumferential MFL (CMFL, MFL-C), transverse MFL (TFL, TFI), and spiral MFL (SMFL), either high resolution or ultra-high resolution.
- **MFL Combo** = a combination of MFL-A and MFL-C. This includes high-resolution MFL-C but not high-resolution MFL-A.
- **UT Metal Loss** = ultrasonic technology (UT) that measures the pipeline wall thickness and, therefore, any metal loss. This category also includes any combination of technologies that has this technology. Some of the common identifiers or acronyms are UT ML, UT WM for UT Wall Measurement (UT WM), etc.

Pipeline operator

Analyzing this variable analyzes the effect the client or recipient of the ILI report has on ILI performance. There are 14 pipeline operators with 23 or more anomalies with field depth data.

Depth of metal loss

The performance as a variable of the depth of metal loss predicted by the ILI system is explored with this variable. For this analysis, metal loss anomalies were placed into depth bins at 10% intervals, e.g., 20% to 30%.

Number of assessments

The number of assessments that a pipeline operator has conducted was also analyzed. Unlike the number of assessments for an ILI company, most (if not all) inline inspections from a pipeline operator are assumed to be uploaded into CIM. Pipeline operators were grouped into quartiles based on the number of assessments in CIM. Quartile 1 has the least assessments, and Quartile 4 has the most.

Metal loss classification

This variable analyzes the performance as a variable of the size or dimension class of the metal loss anomaly, as described by the Pipeline Operators Forum (POF).[2] (Only 215 anomalies of the 81,517 were missing a classification; therefore, they were excluded when metal loss classification was being explored.) The POF classifications are General (GENE), Pitting (PITT), Axial Grooving (AXGR),

https://doi.org/10.52202/078572-0052

Circumferential Grooving (CIGR), Pinhole (PINH), Axial Slotting (AXSL), and Circumferential Slotting (CISL).

Assessment year

Lastly, the certainties and average errors by assessment year were analyzed to determine performance over time.

What data was utilized?

There are over 112,216 anomalies that have repair data in CIM, with 88,200 anomalies that have both an ILI-predicted depth of metal loss and a field-found depth of metal loss. After data prepping and cleanup, **the total number of anomalies that had both an ILI metal loss depth and a field-found metal loss depth and, therefore, utilized for this analysis was 81,497.**
Figures 3, 4, 5, and 6 provide an overview of the makeup of the data set.

- 54% of the anomalies are from inline inspections conducted by liquid pipeline operators versus 46% by gas pipeline operators.
- 73% of the anomalies are supplied from 3 ILI companies
- 77% of the anomalies are supplied from standard resolution axial MFL
- 81% of the anomalies are labeled external corrosion or external metal loss.
- 74% of the anomalies are either general corrosion or pitting, utilizing the POF metal loss classification system.

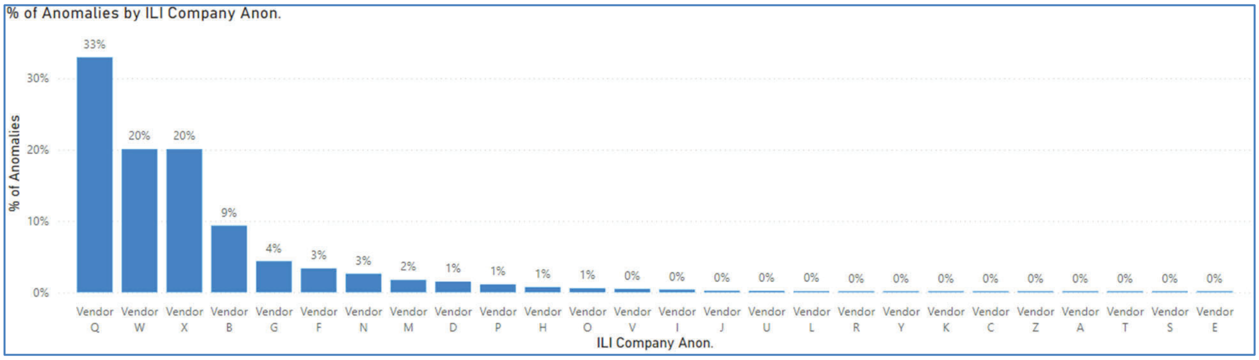


Figure 3. Anomalies categorized by anonymized ILI company. (The percentages compare to the total number of anomalies included in the study, i.e., 81,497).

https://doi.org/10.52202/078572-0052

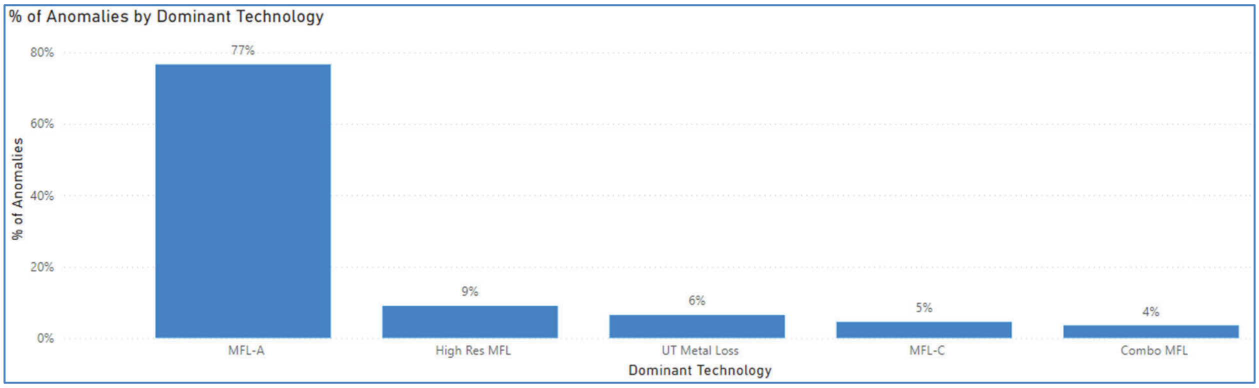


Figure 4. The percentage of anomalies included in this study, by ILI technology.

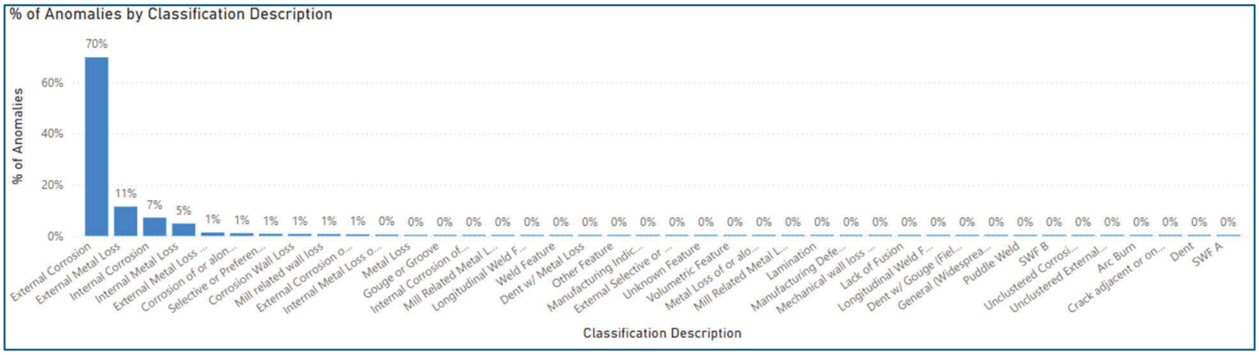


Figure 5. The anomalies that are included in this study by classification description. (CIM classifies all anomalies to a standardized CIM classification so that inline inspection reports can be compared across years, companies, technologies, etc.)

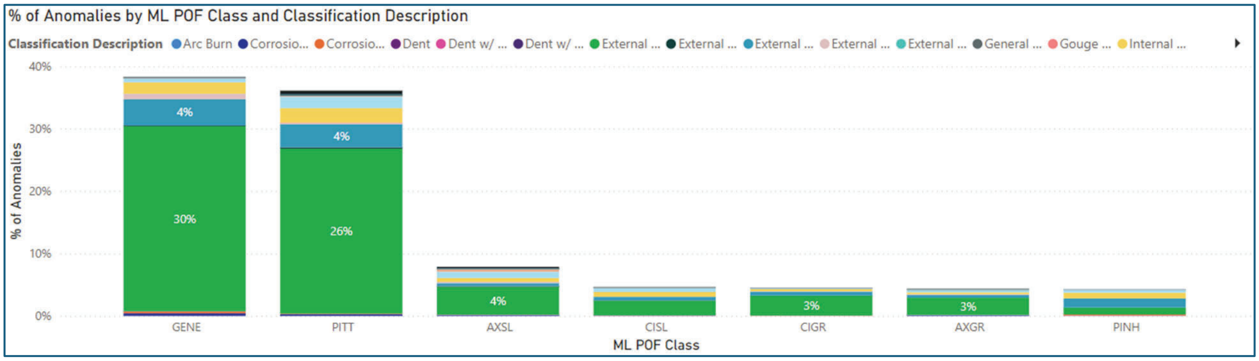


Figure 6: The anomaly dimension class of the anomalies included in this study.

Data prepping

Data cleanup included excluding data points when:

https://doi.org/10.52202/078572-0052

- The field-found metal loss is less than 1%.
    - **Example A**: To remove entries of 0%. Sometimes, a field-found metal loss depth of 0% was entered when the predicted anomaly was not found in the field. This scenario is more indicative of an issue with probability of detection (POD) or probability of identification (POI) and is not the performance characteristic being explored in this analysis.
    - **Example B**: To remove instances where the field depths entered were likely typos, e.g., the ILI-predicted metal loss was 54%, and the field-found depth was entered as 0.54% or a negative, i.e., -54%.
- The field found metal loss is greater than 100%. This is similar to Example B above. In some instances, a field metal loss of 200% was erroneously entered; however, it was unclear if this meant 2% or 20%, so these measurements were also excluded.
- The ILI metal loss = 0. This is similar to Example A above. Sometimes entries of metal loss would be assigned to an anomaly where the ILI did not predict metal loss.
- The ILI metal loss < 1%. These were assumed to be typo errors. MFL-A technologies before 2015 (where these data points came from) cannot size anomalies with a depth of less than 1%.

Minimizing the effect of the number of data points

A positive correlation was found between the certainty calculated and the number of anomalies utilized to produce the certainty value; however, the correlation was not strong, as seen in **Figure 7**. Additionally, a low number of measurements can lead to outliers on the low and high end of the spectrum, as seen in **Figure 8**.

To minimize the effect of the number of measurements on calculated performance, those categories with less than 22 or less anomalies were excluded. This number was determined by analyzing the certainty within 10% for ILI company and pipeline operator, the two variables that experienced the highest correlation with the number of measurements. Those certainties in the lowest plateau (many of which were also statistical outliers) had less than 23 measurements. For reference, this number is higher than the values recommended by PRCI'S API 1163[3] guidance document for a Level 2 analysis and a Level 3 analysis, which are 6 and 10, respectively.

This excluded 5 ILI companies when the ILI company was the primary variable and 3 pipeline operators when the pipeline operator was the primary variable. For example, certainty is not provided in Figure 21 for UT metal loss anomalies with a depth of 80 – 90% because there were less than 23 data points for this particular category.

The effect of the anomaly data coming from 1 pipeline operator versus multiple pipeline operators was also explored, as seen in **Figure 9**. However, after removing those with less than 23 data points, there were no outliers (on either the low or high end.)

Removing Outliers

Additional care was taken to calculate the interquartile range to remove statistical outliers; however, once those categories with less than 23 data points were excluded, no outliers were found in any of the data sets.

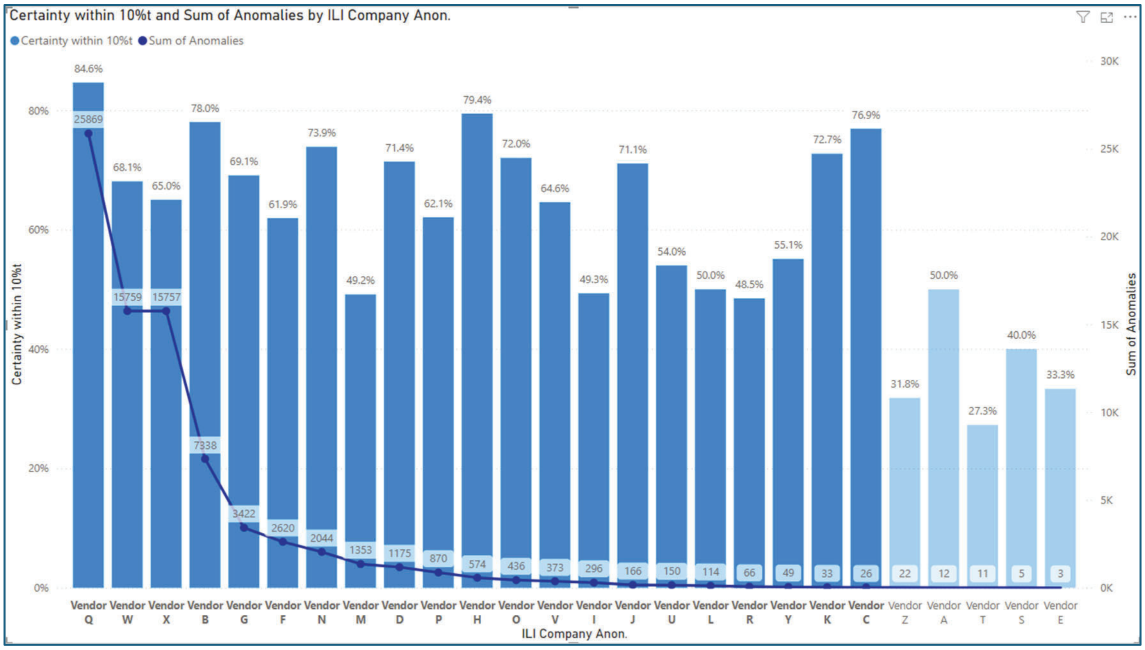
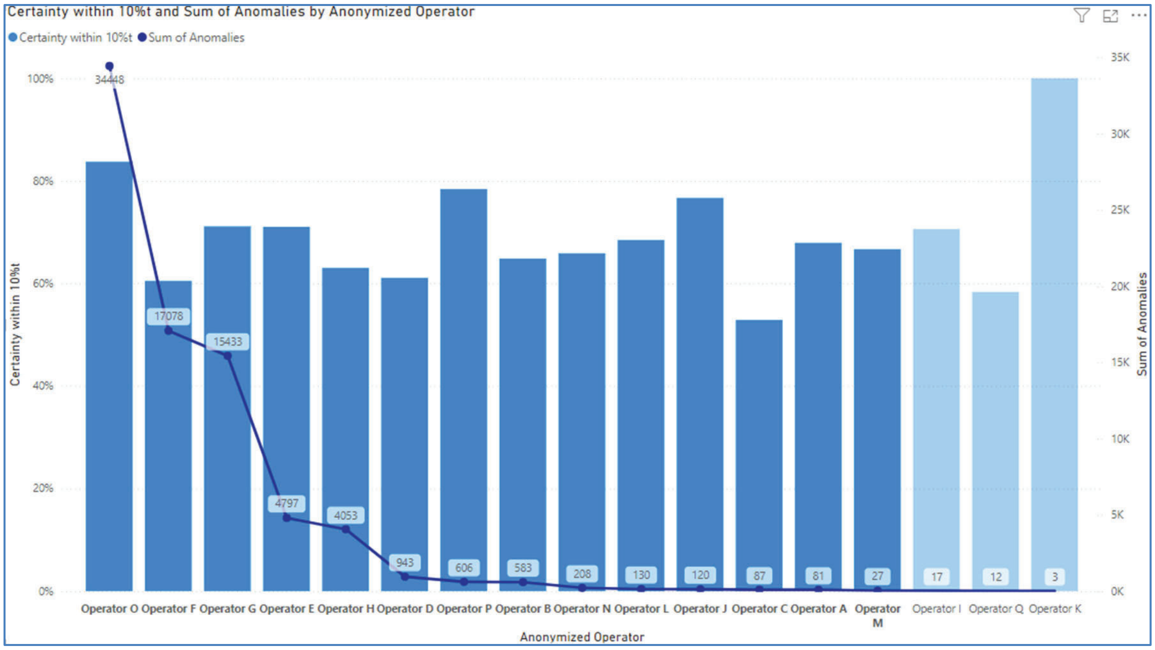


Figure 7. The certainty within 10%, by anonymized ILI company, sorted by the number of measurements, with those ILI companies with less than 23 measurements (far-right) low-lighted. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.



https://doi.org/10.52202/078572-0052

Figure 8. The certainty within 10%, by anonymized pipeline operator, sorted by the number of measurements, with those operators with less than 23 data points low-lighted. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.

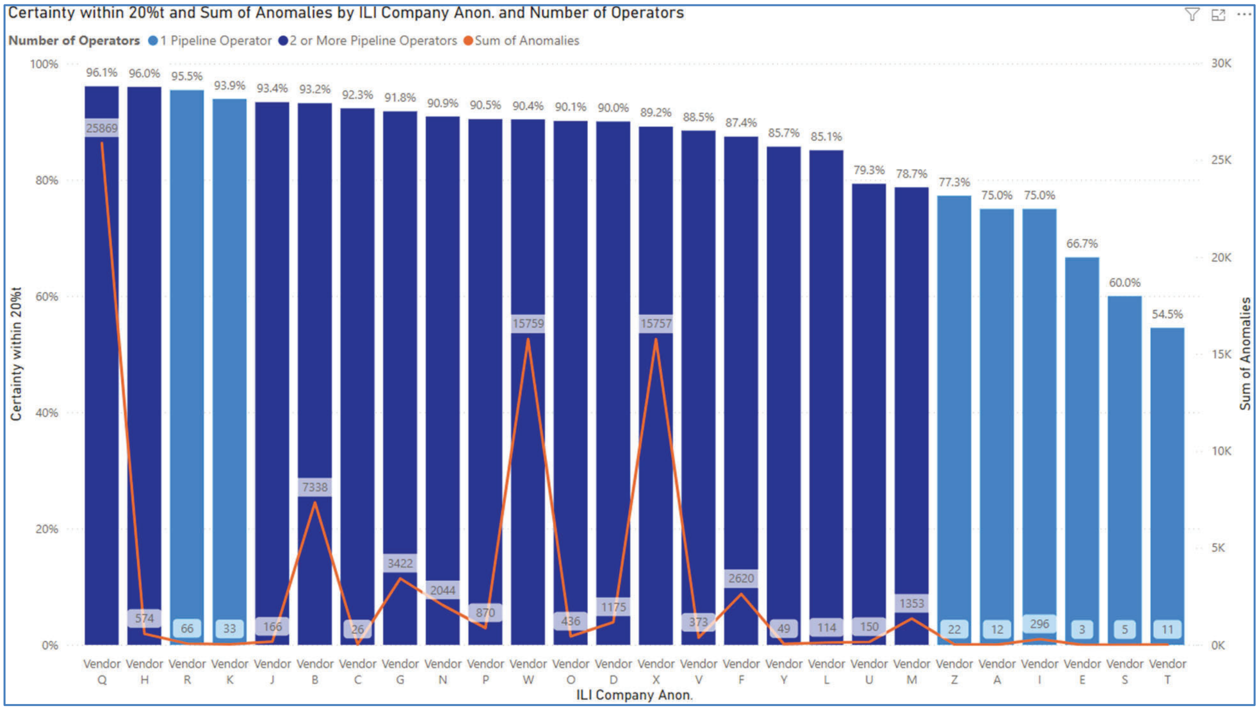


Figure 9. The certainty within 20%, by anonymized ILI company, with those ILI companies with measurements from just one pipeline operator highlighted in light blue. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.

What provides the highest variability in performance?

The effect that the above-mentioned variables had on ILI performance was analyzed by:

1. Calculating the variability of performance for that variable by standard deviation, i.e., for technology, how does the performance vary from MFL-A to MFL-C to UT, etc?
2. Calculating the performance for the variable (presented in the next section)

The certainty within a 10% and 20% tolerance was calculated for each category. Figure 10 shows each ILI company's certainty for a 10% tolerance, which illustrates the variability that can exist in performance when the ILI company is the primary variable, i.e., the certainty can range from 49% to 85%.

The standard deviation (variability) was calculated for the following variables: ILI company, pipeline operator, technology, metal loss depth, and metal loss classification. **Figure 11** illustrates the

https://doi.org/10.52202/078572-0052

variability in the calculated certainties for a tolerance of 10% for each variable, while **Figure 12** provides the variability in the certainties for a tolerance of 10% and 20%.

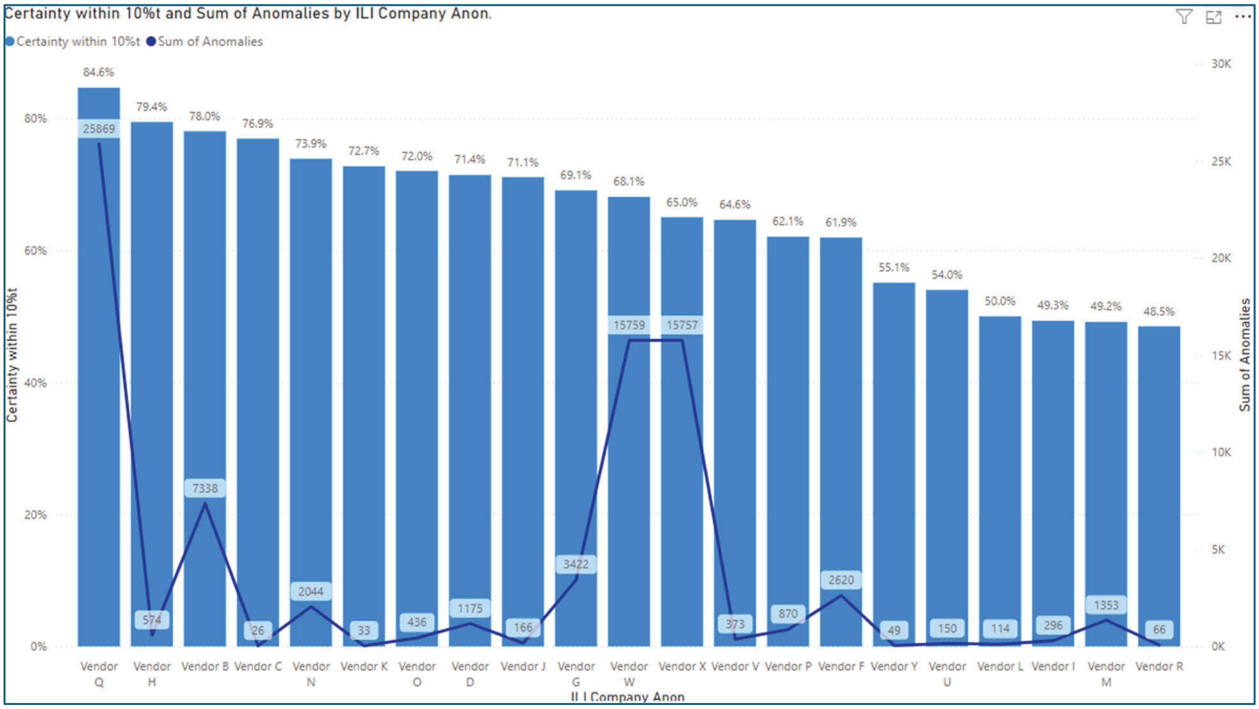


Figure 10. Certainty within 10%, by anonymized ILI company. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.

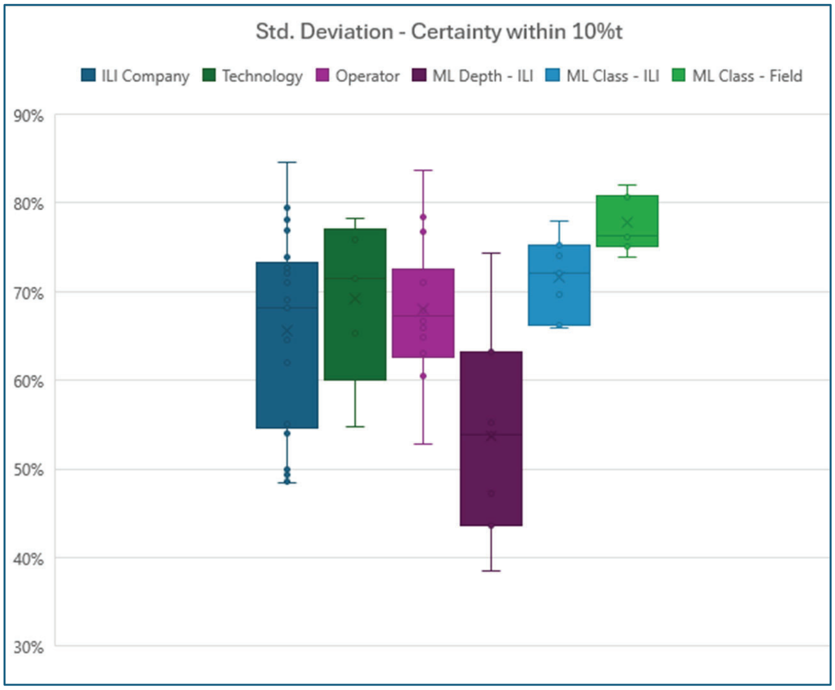


Figure 11. Visual representation of the variability of the calculated certainty within 10%.

https://doi.org/10.52202/078572-0052

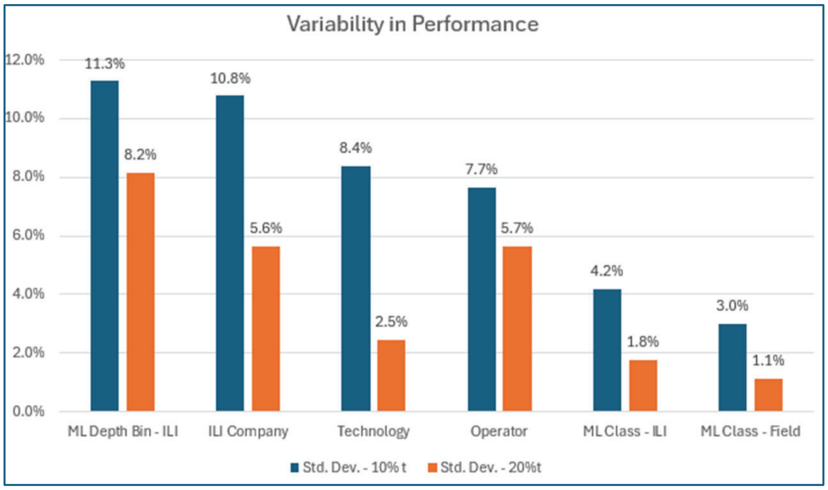


Figure 12. The standard deviation, i.e., variability of the calculated certainties with a tolerance of 10% vs. 20%.

Firstly, the variability is higher across all variables for a 10% tolerance (higher accuracy) than a 20% tolerance. Looking at the tolerances separately, if a tolerance of 10% is needed, the depth of the metal loss anomaly has the highest contribution to the performance, either low or high, followed by the ILI company and technology. For a wider tolerance of +/- 20%, the metal loss depth of the anomaly is again the most impactful factor, followed by the ILI company and pipeline operator, which are essentially tied.

The calculated variability diverges the most for technology and ILI companies when comparing a 10% tolerance to a 20% tolerance. **If a tolerance of 10% or better (high depth sizing accuracy) is needed, selecting the ILI company and technology is most important. If a 20% tolerance is sufficient, the ILI company and operator are more critical than the type of technology utilized.**

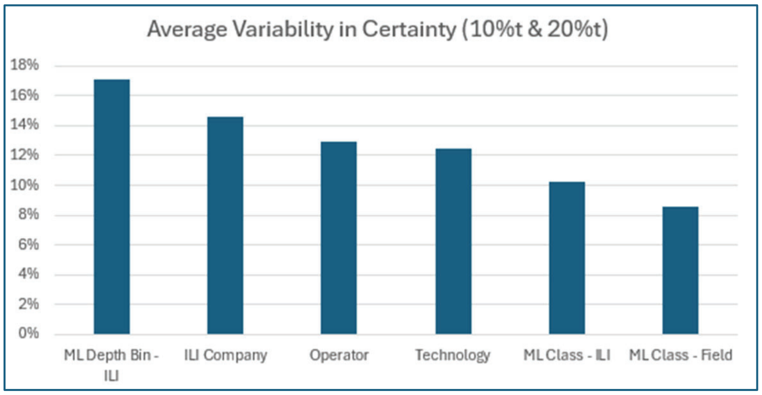


Figure 13. The average variability, i.e., an average of the values, is presented in **Figure 12.**

If averaging the certainties, the depth of the metal loss is the most critical factor, as seen in **Figure 13**, followed by ILI company, pipeline operator, and technology. Overall, **depth has the highest effect on performance, whereas metal loss classification had the lowest.** Interestingly, ILI systems typically have different performance specifications for metal loss classifications, not for anomalies at different depths.

Variability within ILI company & pipeline operator versus technology

When there's a change in technology type, a change in performance is not surprising; one would expect varying performance for depth sizing when utilizing an axial MFL versus UT ML technology, for example. However, ILI companies and pipeline operators having a higher variability than technology type may defy expectations.

Knowing that MFL-A makes up 77% of the data points utilized in this study, certainties were then calculated for just MFL-A technology by ILI company. As seen in **Figure 14**, **high variability in the performance of MFL-A exists, with certainties ranging from 49% to 85%, which is the same range when all technologies are included.** A similar pattern was seen with pipeline operators as well. The variability in performance within the ILI companies and pipeline operators is supported by the variability in the MFL technology.

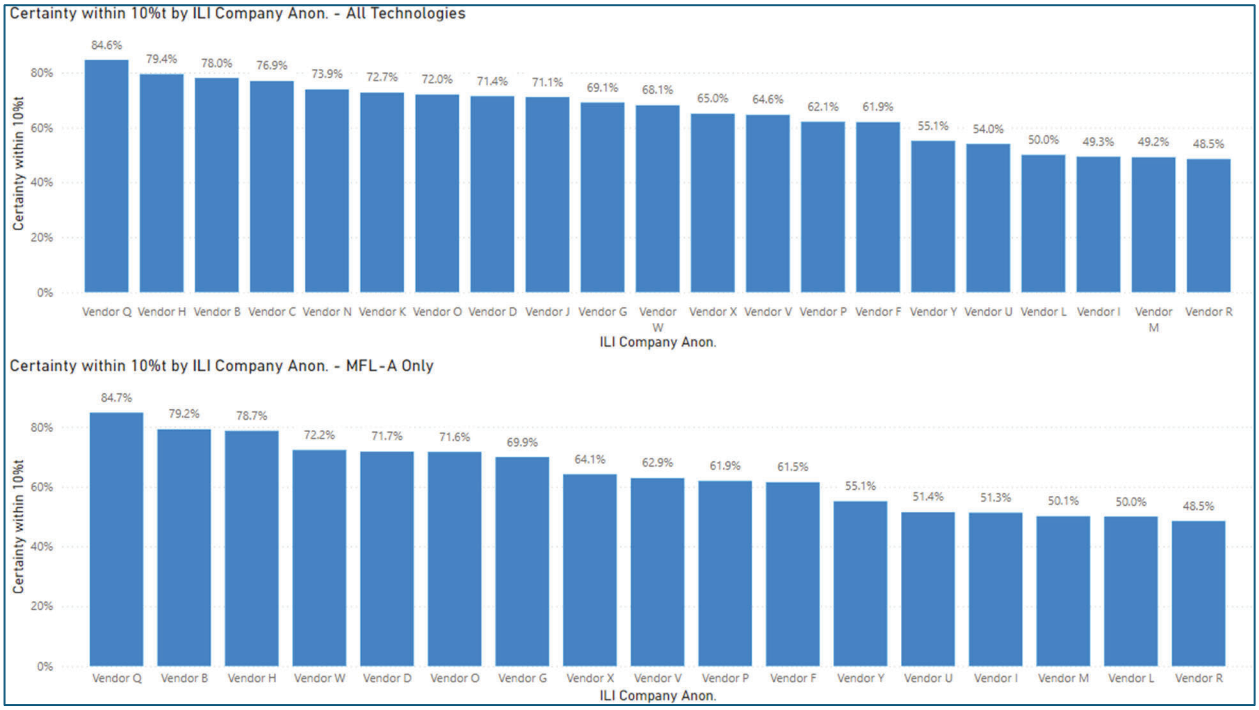
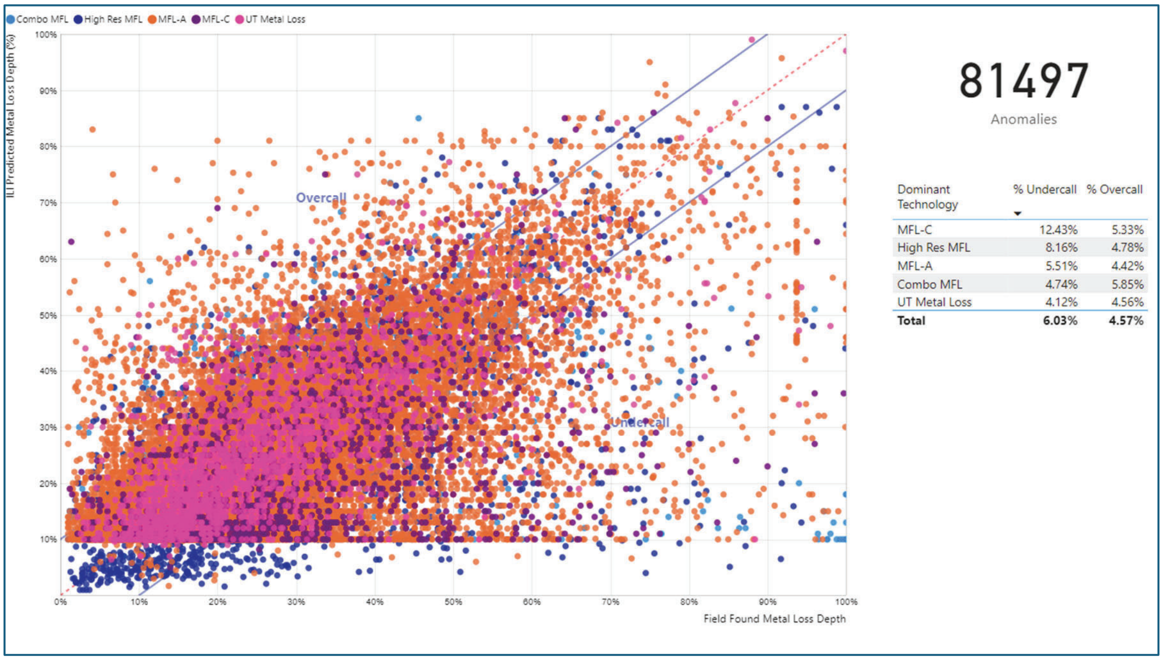


Figure 14. The certainty within 10%, by anonymized ILI company, for all technologies (top) vs. MFL-A (bottom).

Performance per technology

https://doi.org/10.52202/078572-0052

Each data point was categorized by technology, as seen in the unity plot in **Figure 15**. The % Undercall and Overcall shown in Figure 15 represents the tendency of the ILI technology to under-report or over-report the depth, utilizing a 10% tolerance. For example, MFL-C tended to under-report the depth the most compared to all other technologies.



81497

Anomalies

Dominant Technology	% Undercall	% Overcall
MFL-C	12.43%	5.33%
High Res MFL	8.16%	4.78%
MFL-A	5.51%	4.42%
Combo MFL	4.74%	5.85%
UT Metal Loss	4.12%	4.56%
Total	**6.03%**	**4.57%**

Figure 15. Unity plot of all anomalies included in this study, categorized by technology.

Because ultrasonic technology is a direct measurement of wall thickness, the performance specifications for UT metal loss are usually specified at 90% certainty, with a depth sizing tolerance provided in absolute units of depth, i.e., mm. Performance specifications for MFL, an indirect measurement technology, usually have a specification at 80% certainty with the tolerance as a percentage of depth, i.e., +/- 10% or +/-20%.

Ultrasonic ILI tools use piezoelectric sensors to generate a sound wave through the coupling medium into the pipe wall, and the returned signals are recorded and measured. The wall thickness is a direct measurement since the distance is calculated from known or measured physical properties.[4] MFL in contrast, utilizes magnets to induce a magnetic field in the steel pipeline. A volumetric metal loss caused by corrosion, for example, causes magnetic flux leakage. That leakage is then measured and converted to dimensions of wall loss. UT directly measures the pipeline wall thickness, whereas MFL indirectly measures the pipeline wall loss.

The certainties for a 10% and 20% tolerance were calculated for each technology, as seen in Figure 16. As expected, UT metal loss has the highest performance for depth sizing of metal loss anomalies compared to the other technologies. What may not be expected is that a) the performance of UT is not significantly higher than that of MFL-A but also b) MFL-A has a higher performance than a combination MFL that contains both MFL-A and MFL-C and also a higher performance than high-resolution MFL.

It's posited that the performance of high-resolution MFL-A and combination MFL may be negatively impacted by what we refer to as "selection effects," i.e., these two technologies are selected to inspect pipelines with complex or difficult-to-assess anomalies, and therefore, their calculated performance is lower than that of MFL-A. After all, each technology may not be run on the same pipelines but on different pipelines with anomalies of varying complexities and operational challenges. (The performance of UT ML may also suffer from "selection effects.")

Additionally, MFL-A may benefit from these "selection effects" as new and "clean" pipelines, i.e., pipelines with low defect populations, are often inspected by standard resolution MFL-A, not a combination of MFL or high-resolution MFL, as pipeline operators may think the benefit does not outweigh the cost.

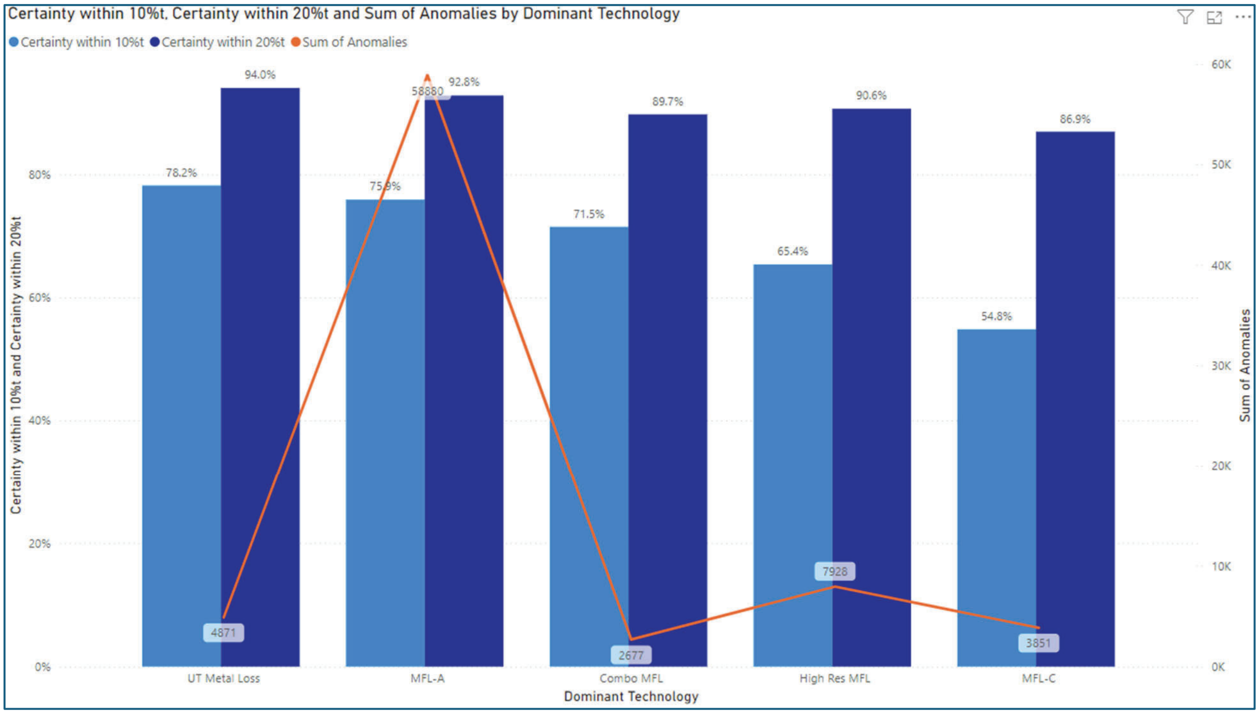


Figure 16. Certainty within 10% and 20%, by technology. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.

Knowing that there's wide variability in MFL-A systems, the same certainties were calculated utilizing only the middle quartile of ILI companies or "mid-pack," i.e., those ILI companies that fell between the first and third quartile for certainties calculated using a 10% tolerance, seen in **Figure 17**. When doing so, certainty percentages calculated for MFL-A decreased, and the calculated performance for UT ML increased, showing combination MFL to be the second-highest performer behind UT metal loss. This also indicates an increased disparity between the depth sizing performance of UT metal loss versus MFL-A, with UT metal loss at 81% for a 10% tolerance and MFL-A at 67%.

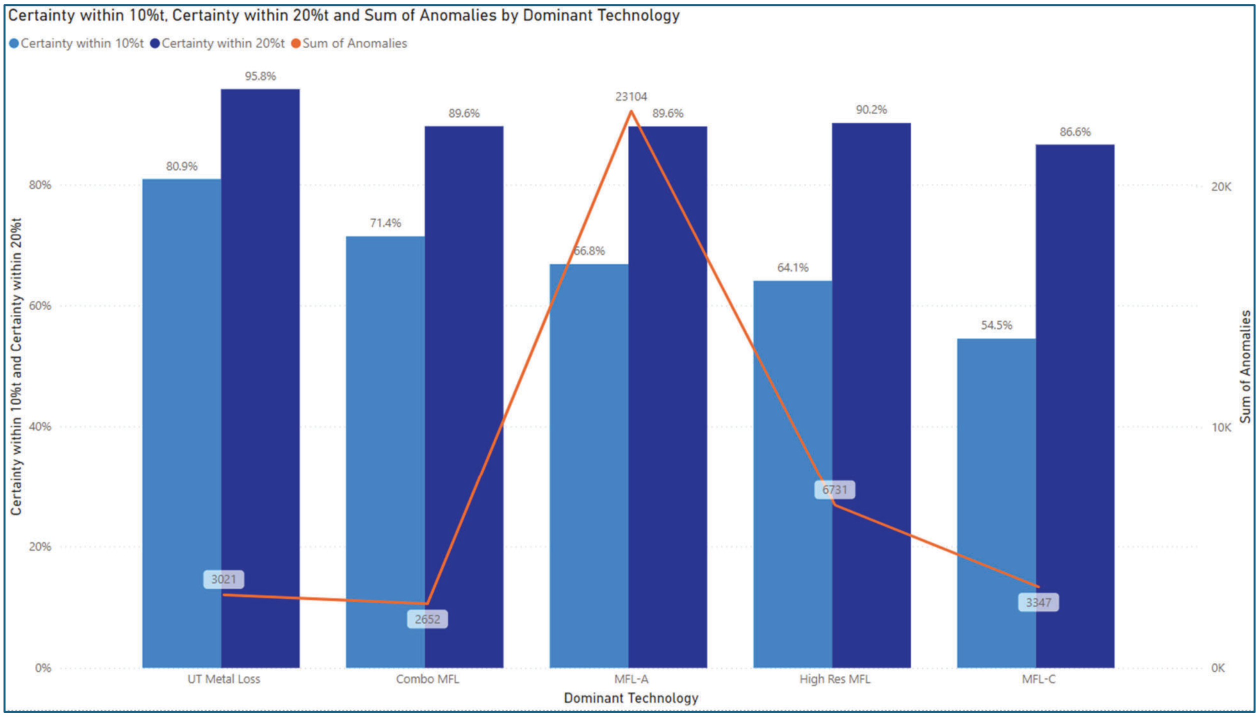


Figure 17. Certainty within 10% and 20%, by technology with only the "mid-pack" ILI companies. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.

If only the certainties for the top quartile of ILI companies were analyzed, MFL-A outperforms UT metal loss, as seen in **Figure 18**. This indicates that **MFL-A from a high-performing ILI company can have depth sizing performance that is as good or better than an ultrasonic ILI system**.

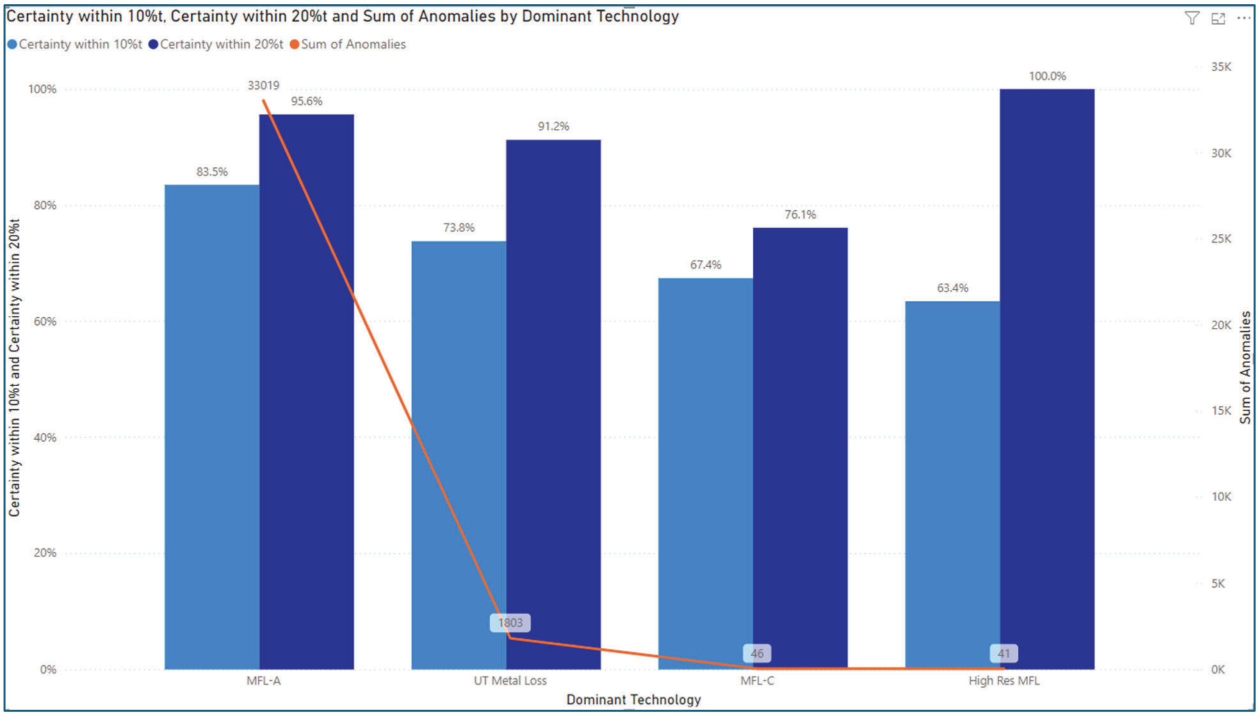


Figure 18. Certainty within 10% and 20% by technology, with only the top quartile ILI companies. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.

Performance per metal loss depth

The second most crucial factor determining whether a metal loss anomaly's depth would be sized correctly (per a 10% tolerance) is the depth itself. Metal loss anomalies were grouped into bins of 10% to calculate the certainty percentage, as seen in the unity plot in **Figure 19**. The categories correspond to the bin's lower limit, e.g., 10% refers to anomalies with depths between 10% and 20%.

As seen in Figure 20, the certainty decreases as the depth of the metal loss anomaly increases.

https://doi.org/10.52202/078572-0052

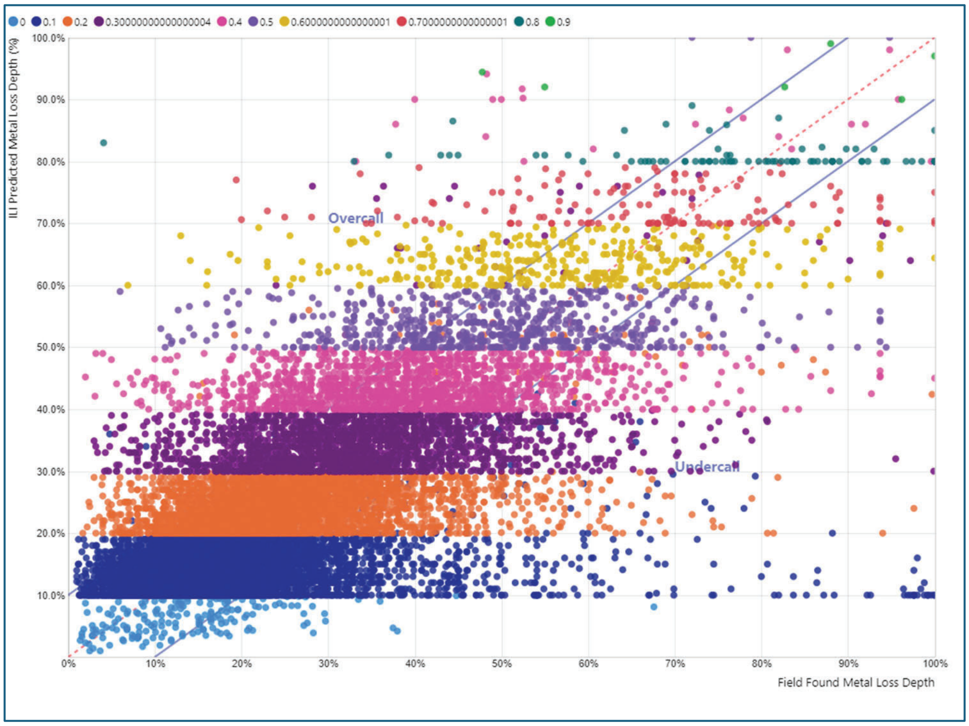


Figure 19. Unity plot of all anomalies included in this study, categorized by depth “bin.”

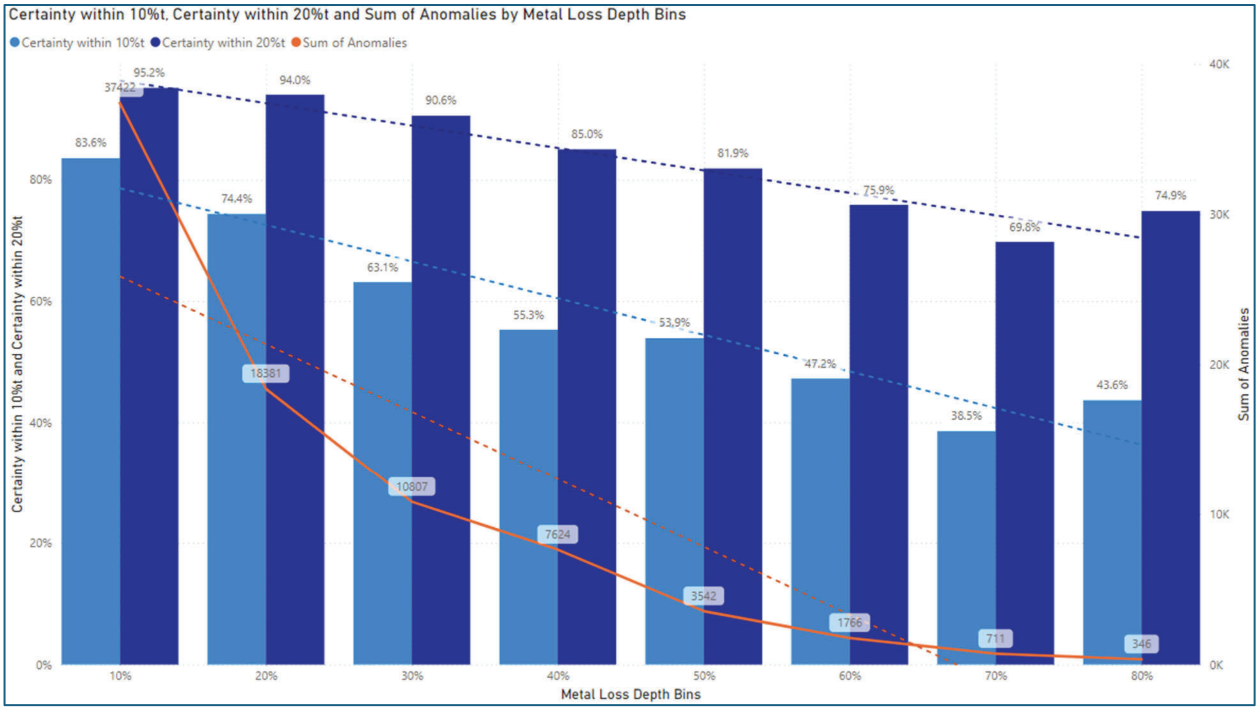


Figure 20. Certainty within 10% and 20% for all technologies by metal loss depth “bins.” The certainties decrease by 5% and 4% as the metal loss depth bin increases for a tolerance of 10% and

20%, respectively. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.

Performance per metal loss depth and technology

MFL-A versus UT ML

The performance as a function of depth and technology was also analyzed. When comparing the certainties calculated for UT vs MFL-A (shown in **Figure 21**), the following were observed:

- For a tolerance of 10%, UT ML outperforms MFL-A (or is within <1%) for all depth bins, except for 70% or deeper, where there's not enough data for UT ML. (The lack of UT data for deeper metal loss anomalies may be due to the inability of standard resolution UT to detect anomalies greater than 4 mm (0.16 in) in depth.)
- For a tolerance of 20%, MFL-A and UT have very similar performances, with MFL-A slightly outperforming UT ML for anomalies 40 to 70% in depth.
- For both MFL-A and UT, the certainty decreases as the depth increases, except for UT metal loss at a 10% tolerance, where the certainty for 60 - 70% is higher than that of 40 to 50% and 50 to 60%.

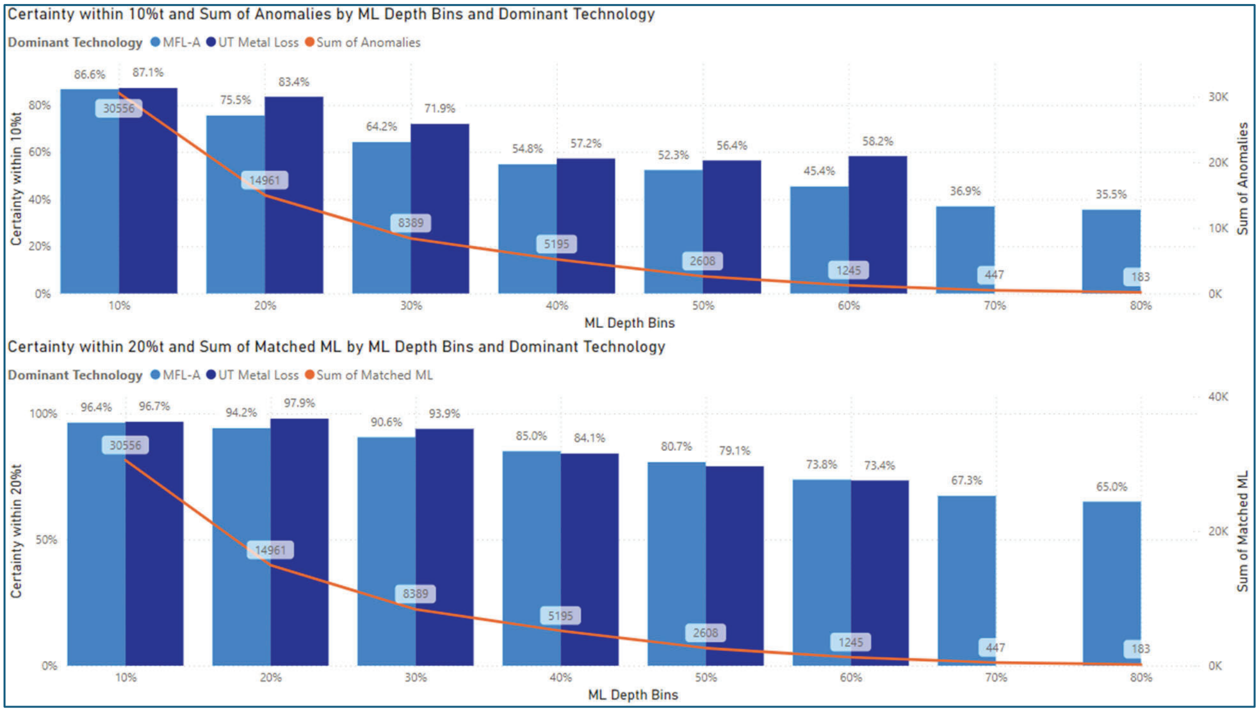


Figure 21. Certainty within a 10% (top) and 20% (bottom) for MFL-A vs. UT ML. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.

MFL-A versus other MFL technologies

https://doi.org/10.52202/078572-0052

When comparing MFL-A to high-resolution MFL-A, MFL-C, and combination MFL, certainty decreases significantly for MFL-A as the depth increases. In contrast, the certainties for the other MFL technologies stay relatively constant, as seen in **Figure 22.** To that end, **when the metal loss depth is 50% or greater, the other MFL technologies outperform MFL-A,** as seen in **Figure 23**.

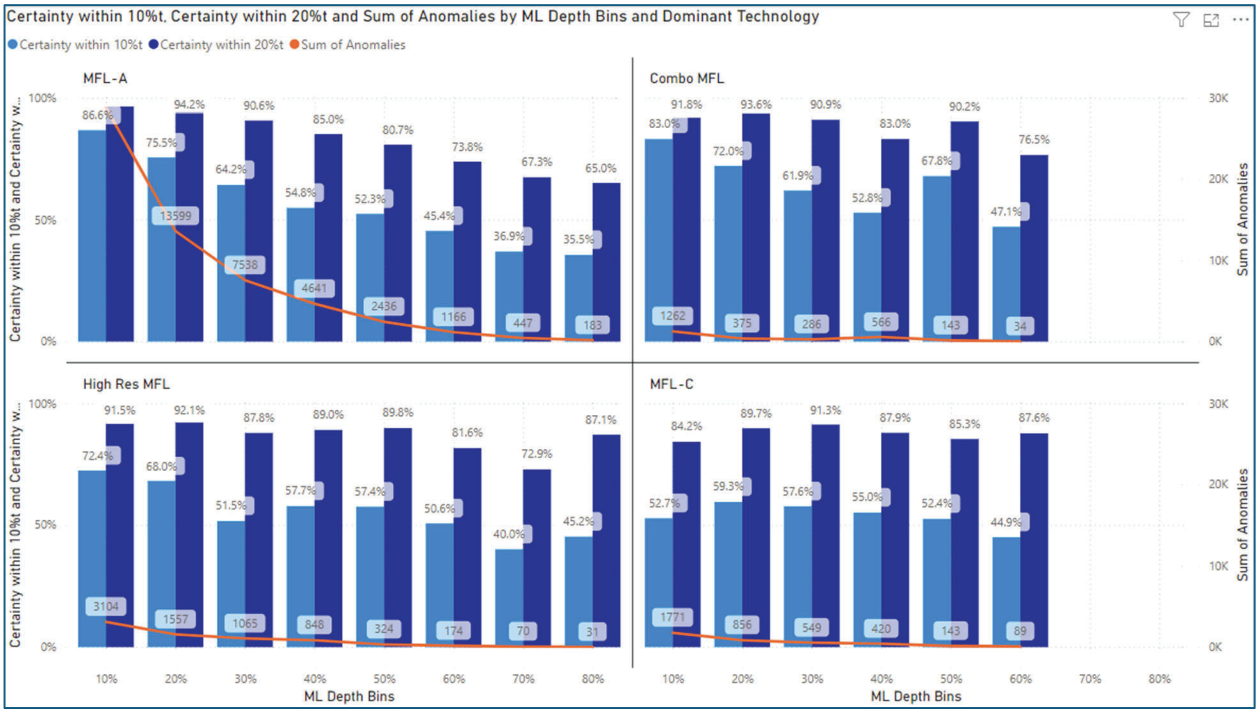


Figure 22. MFL-A versus high-resolution MFL, MFL-C, and combination MFL for certainties within 10% and 20%, by depth bins 10% - 80%. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.

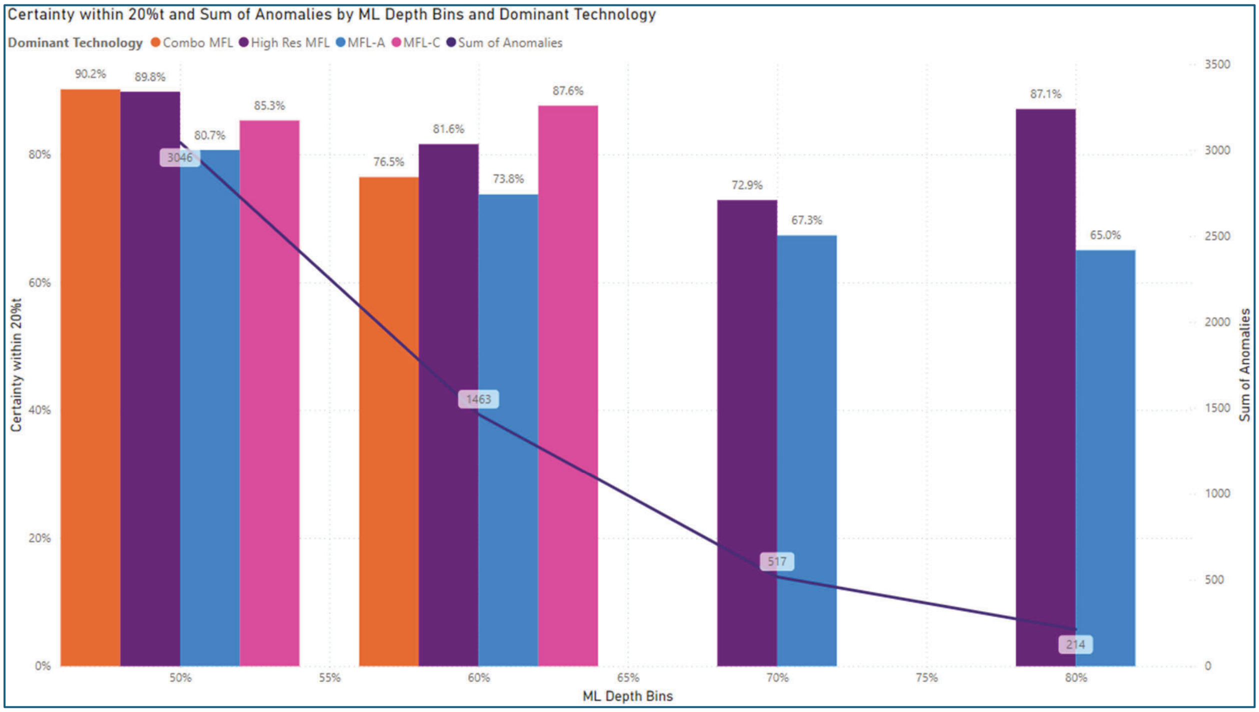


Figure 23. MFL-A versus high-resolution MFL, MFL-C, and combination MFL for certainties within 10% and 20%, by depth bins 50% - 80%. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.

Performance per pipeline operator experience

Pipeline operators were grouped into four quartiles based on the number of assessments completed to understand the variability and performance amongst the categories. Pipeline operators with the least number of assessments had the combined highest performance. This could indicate that the number of assessments / the experience of the pipeline operator is unimportant. However, fewer assessments may lead to fewer measurements, and a low number of measurements could indicate a much higher performance (or a much lower performance) than the actual performance. A similar analysis was conducted for ILI company to compare to the pipeline operator quartiles. Both sets of calculated certainties can be seen in **Figure 24**.

https://doi.org/10.52202/078572-0052

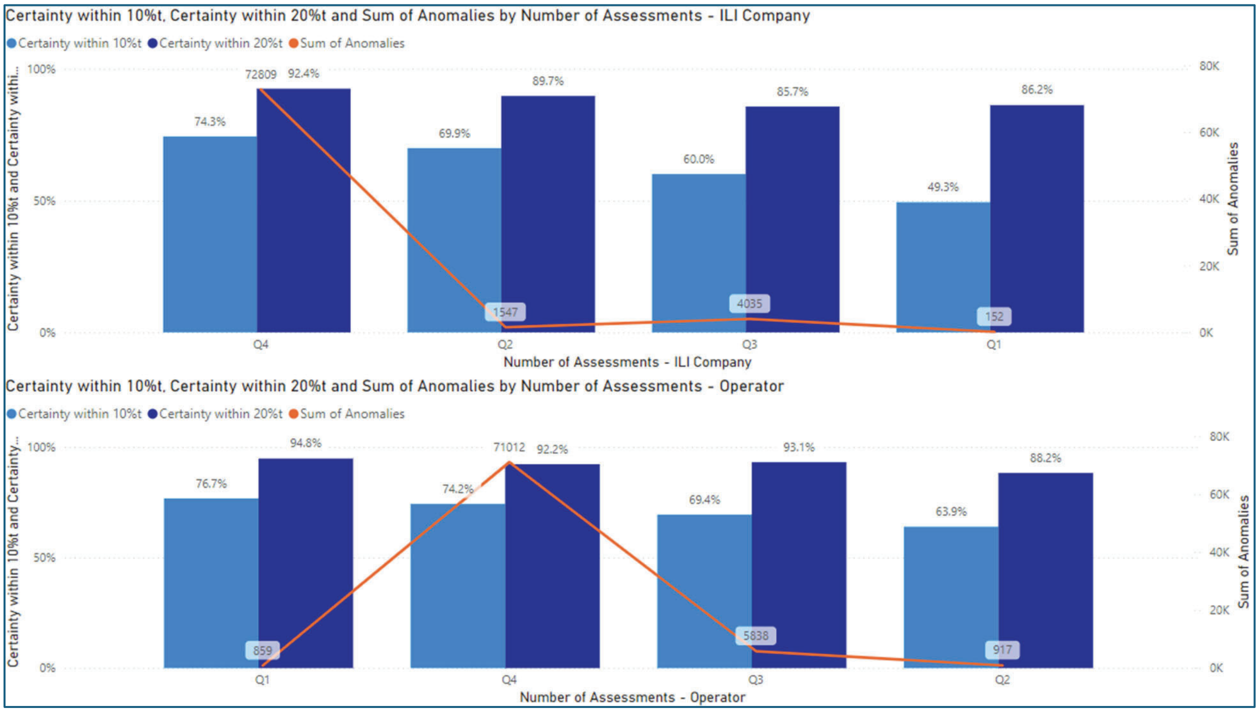
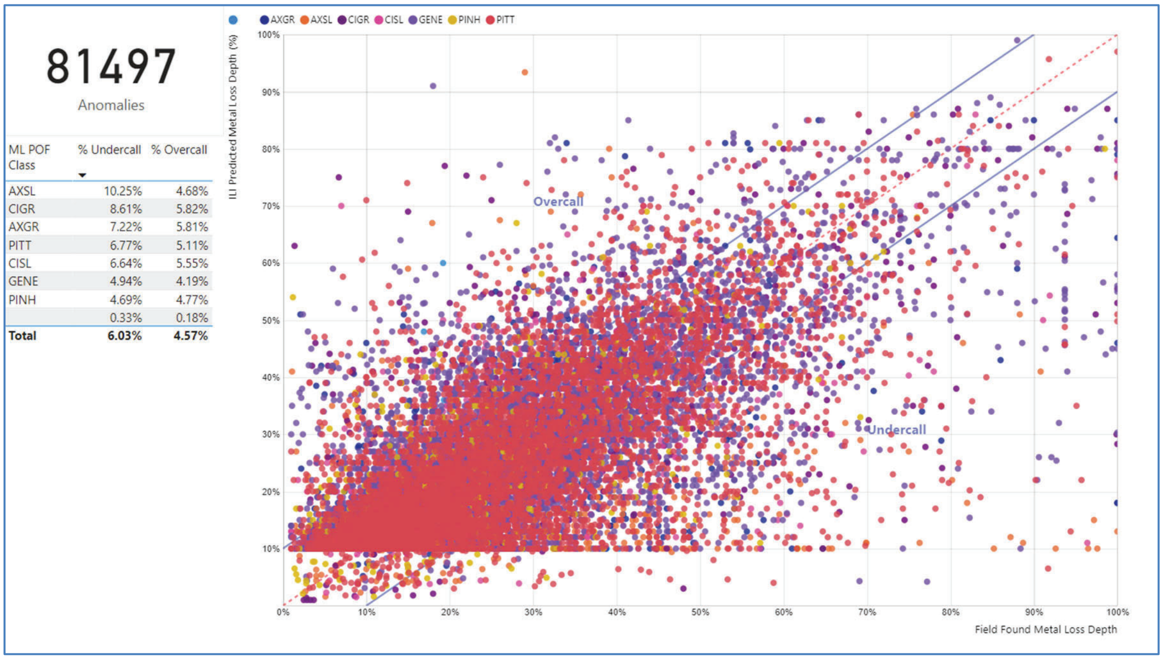


Figure 24. Certainties within 10% and 20%, by quartiles derived from the number of assessments, for ILI company (top) and pipeline operator (bottom). Q1 has the lowest number of assessments; Q4 has the highest. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.

Performance per metal loss classification and technology

Depth sizing performance regarding metal loss classification, as defined by the Pipeline Operators Forum (POF), was also analyzed for each technology. According to the *Specifications and requirements of in-line inspection of pipelines,*[2] published by POF: "The measurement capabilities of non-destructive examination techniques, in particular the MFL technique, depend on the geometry of the metal loss anomalies." This is why there are often different performance specifications for different anomaly classifications.

Axial MFL, for example, typically has a decreased specification for axial grooving than for circumferential grooving and no specification for axial slotting. The difficulty in sizing axial slotting can be seen in the unity plot presented in **Figure 25,** where the tendency to under-call the depth is highest for axial slotting.



81497

Anomalies

ML POF Class	% Undercall	% Overcall
AXSL	10.25%	4.68%
CIGR	8.61%	5.82%
AXGR	7.22%	5.81%
PITT	6.77%	5.11%
CISL	6.64%	5.55%
GENE	4.94%	4.19%
PINH	4.69%	4.77%
	0.33%	0.18%
Total	**6.03%**	**4.57%**

Figure 25. Unity plot of all anomalies included in this study, categorized by depth anomaly dimension class.

MFL-A versus UT

Because of the physical limitations of ultrasonic technology, detecting and sizing pinholes is a challenge because "the size of the metal loss is small relative to the diameter of the sound field generated by the sensor," i.e., UT tools cannot detect anomalies that are smaller than the ultrasonic beam. According to NDT Global[4], a standard-resolution UT tool has 3 mm axial and 8 mm circumferential resolutions. In comparison, a standard-resolution MFL tool has a 2.5 mm axial resolution and a 5.9 mm circumferential resolution. (Axial resolution is the axial distance between two consecutive measurements of the ultrasonic sensors. Circumferential resolution is the circumferential distance.)

The strength of ultrasonic technology lies in its ability to measure wall thickness directly and, therefore, measure the wall thickness of long axial corrosion, especially when there is a gradual reduction of wall thickness. In contrast, MFL technology relies on the change of magnetic flux leakage to estimate wall loss. Therefore, MFL tools may only "see" the beginning and end of long shallow anomalies.

MFL-A vs MFL-C

Circumferential MFL should perform better for axially aligned anomalies than axial MFL due to the alternate magnet orientation. To that point, axial MFL typically has a decreased specification for axial grooving than for circumferential grooving and no specification for axial slotting. In contrast, MFL-C typically doesn't have a specification for circumferential grooving or circumferential slotting. **Table 1** below shows performance specifications for standard resolution MFL-A and MFL-C technologies.

https://doi.org/10.52202/078572-0052

Table 1: Depth sizing specifications for MFL-A and MFL-C

MFL-A at 80% certainty					
GENE	PITT	AXGR	AXSL	CIGR	CISL
+/- 10%	+/- 10%	+/- 15%	n/a	+/- 10%	+/- 10%
MFL-C at 80% certainty					
+/- 15%	+/- 19%	+/-15%	+/- 15%	n/a	n/a

Performance per metal loss classification for MFL-A

Figure 26 shows the certainties for the tolerance of 10% and 20% for each metal loss class for axial MFL. For a 10% and 20% tolerance, MFL-A has the highest performance for pinholes and general corrosion.

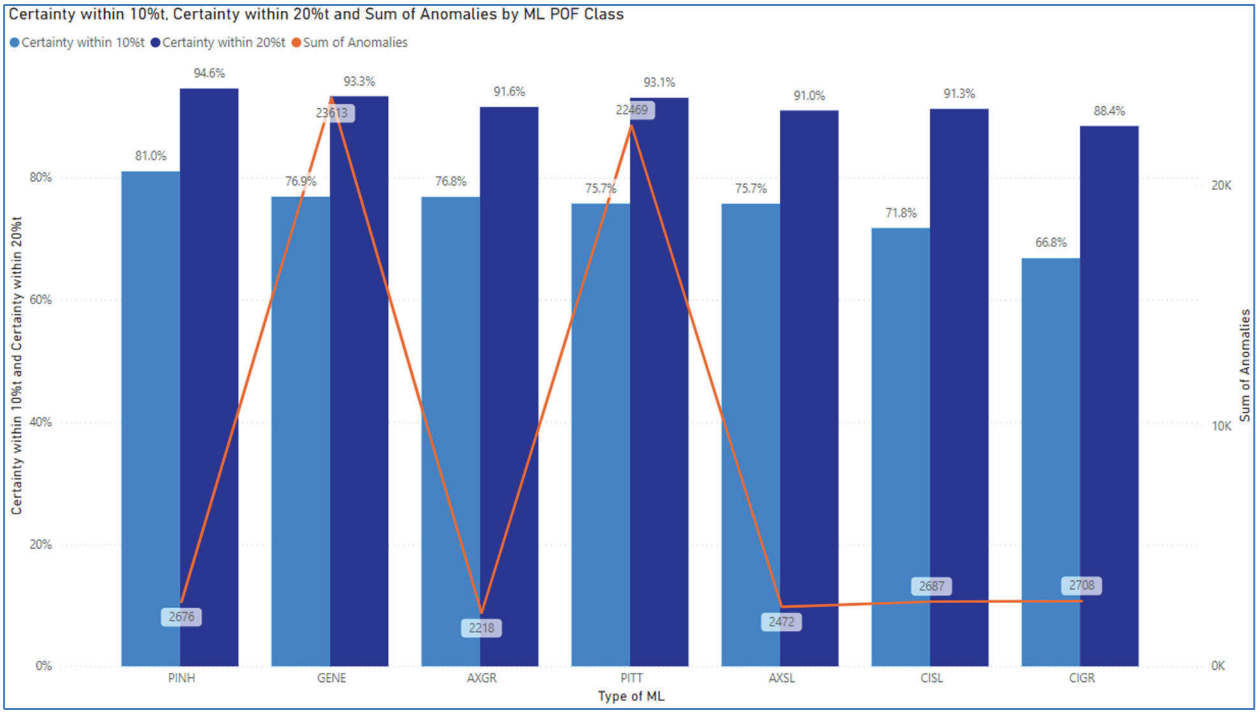


Figure 26. Certainty within 10% and 20% for MFL-A by anomaly dimension class. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.

Performance per metal loss classification for UT metal loss

A similar analysis was conducted for UT metal loss, as seen in **Figure 27**. It can be noted that there are significantly fewer anomalies to calculate certainty for UT metal loss than for MFL-A when categorizing anomalies by POF classification.

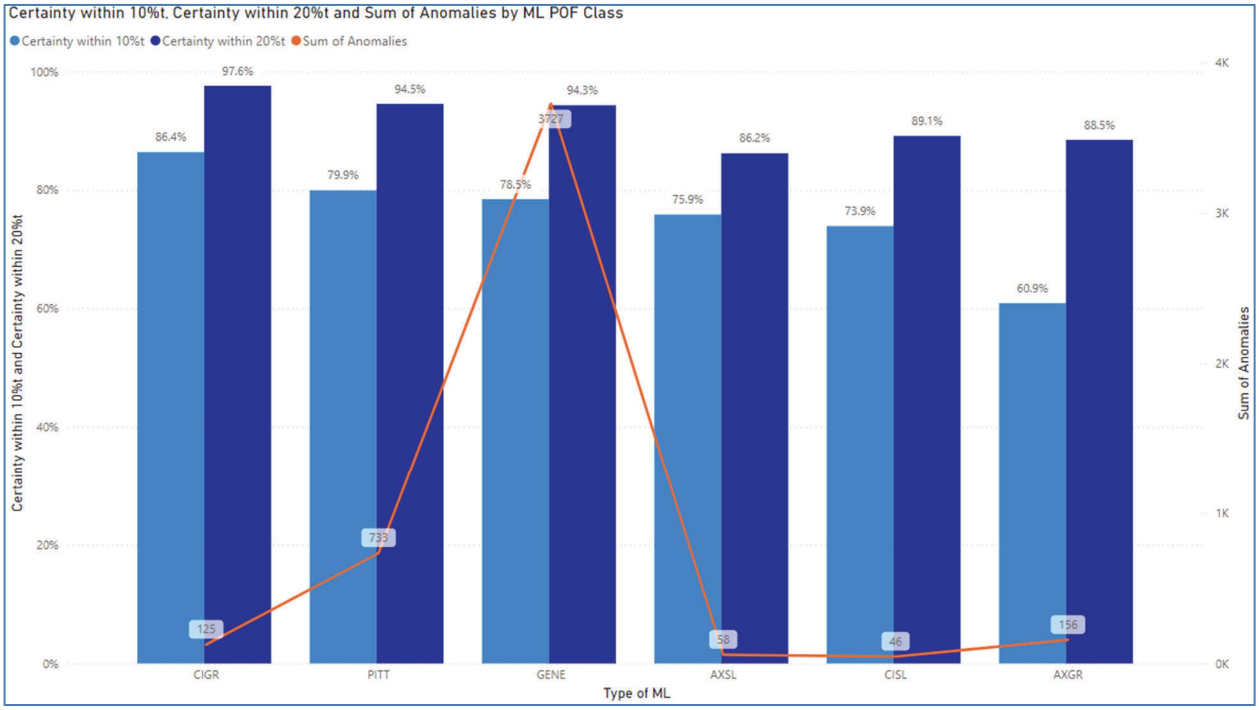


Figure 27. Certainty within 10% and 20% for MFL-A by anomaly dimension class. Pinholes are missing due to lack of data, i.e., < 23 data points. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.

Performance per metal loss classification for MFL-A versus UT

Figure 28 shows the certainties as a function of metal loss classification for MFL-A and UT. The most significant differences occur for grooving with a 10% tolerance:

- For circumferential grooving, UT metal loss significantly outperforms MFL-A (20% difference).
- For axial grooving, MFL-A significantly outperforms UT metal loss (16% difference).

The ultrasonic technologies that sized the anomalies in this study showed higher performance in the circumferential direction than in the axial direction.

https://doi.org/10.52202/078572-0052

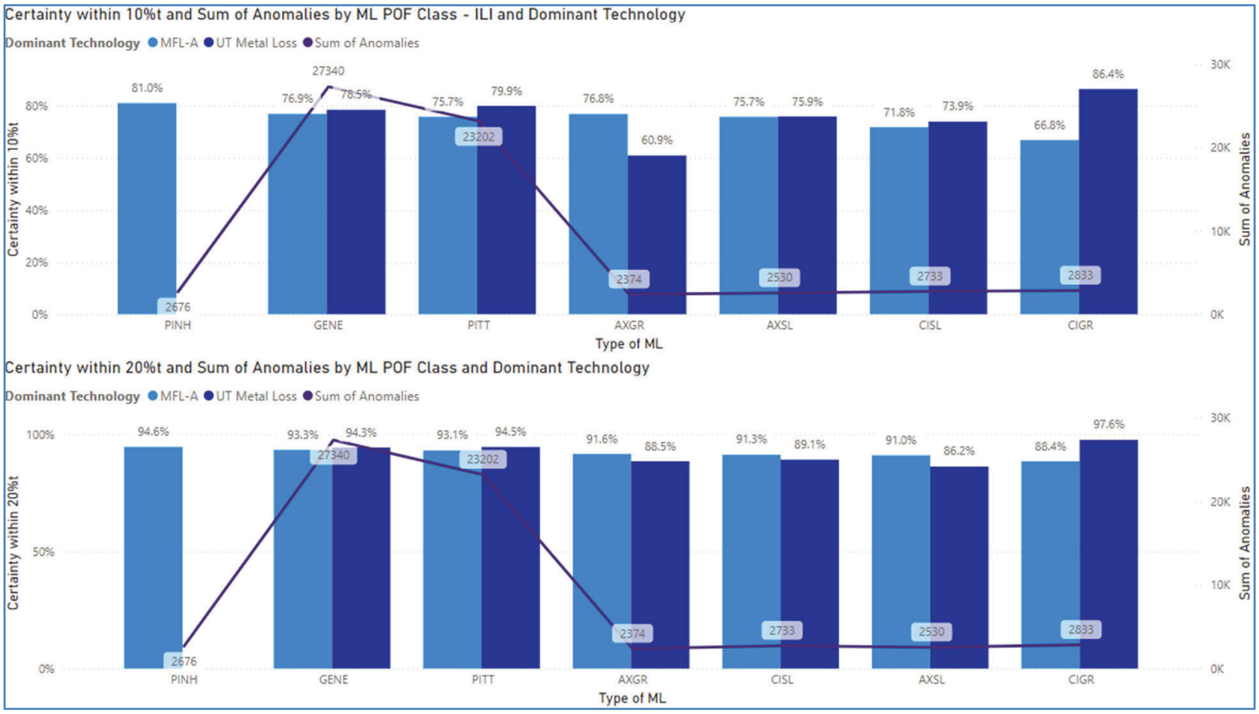


Figure 28. Certainty within 10% and 20% for MFL-A vs. UT metal loss by anomaly dimension class. The number of anomalies utilized to calculate certainty is plotted as a line on the secondary y-axis.

Performance per metal loss classification for MFL-A versus high-resolution MFL

MFL-A outperformed high-resolution MFL-A for all classifications except for circumferential slotting, which defies published performance specifications, as seen in **Figure 29**. This may be due to reporting/user errors, i.e., a high-resolution MFL was utilized but not reported in the CIM platform as high resolution. It also may be due to "selection effects," as discussed previously. High-resolution MFL-A did outperform non-high-resolution MFL-A for deeper anomalies, as was shown earlier in **Figure 23**.

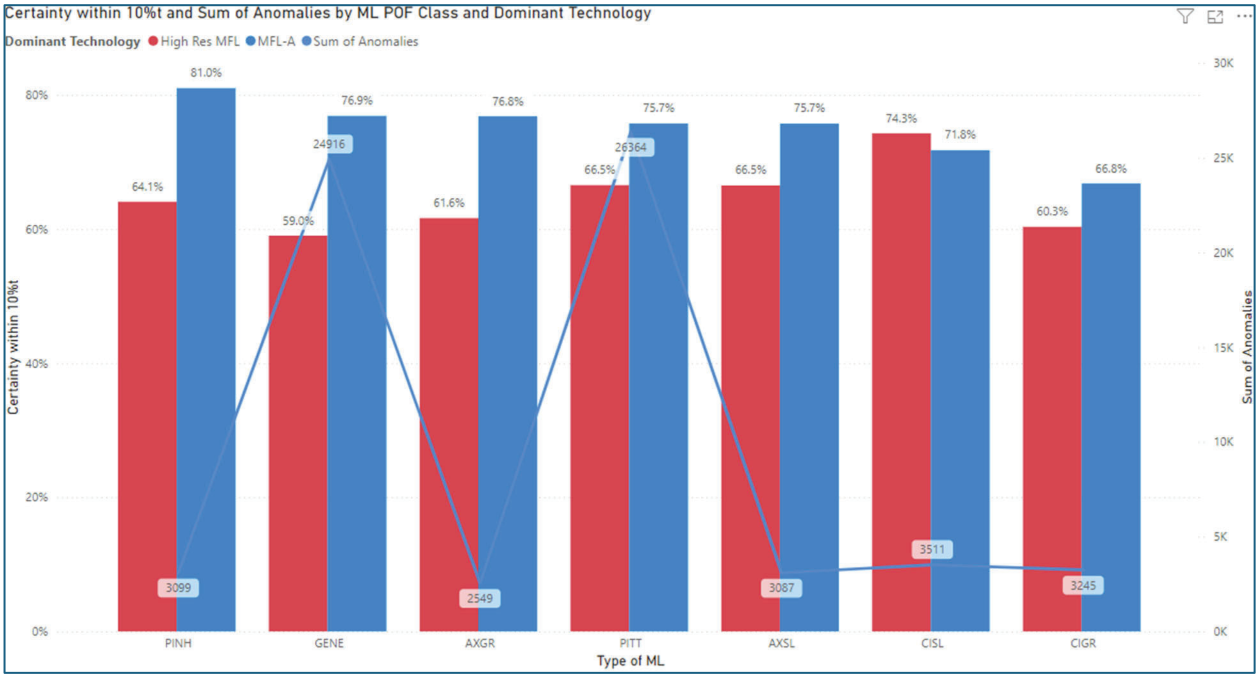


Figure 29. Certainty within 10% for MFL-A vs. high-resolution MFL-A. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.

Performance per metal loss classification for MFL-A versus MFL-C

Although MFL-C is typically deployed to detect axially oriented anomalies, the data shows that MFL-A performed significantly better at sizing the depth of axial grooving and axial slotting, as seen in **Figure 30**, which also defies published performance specifications.

https://doi.org/10.52202/078572-0052

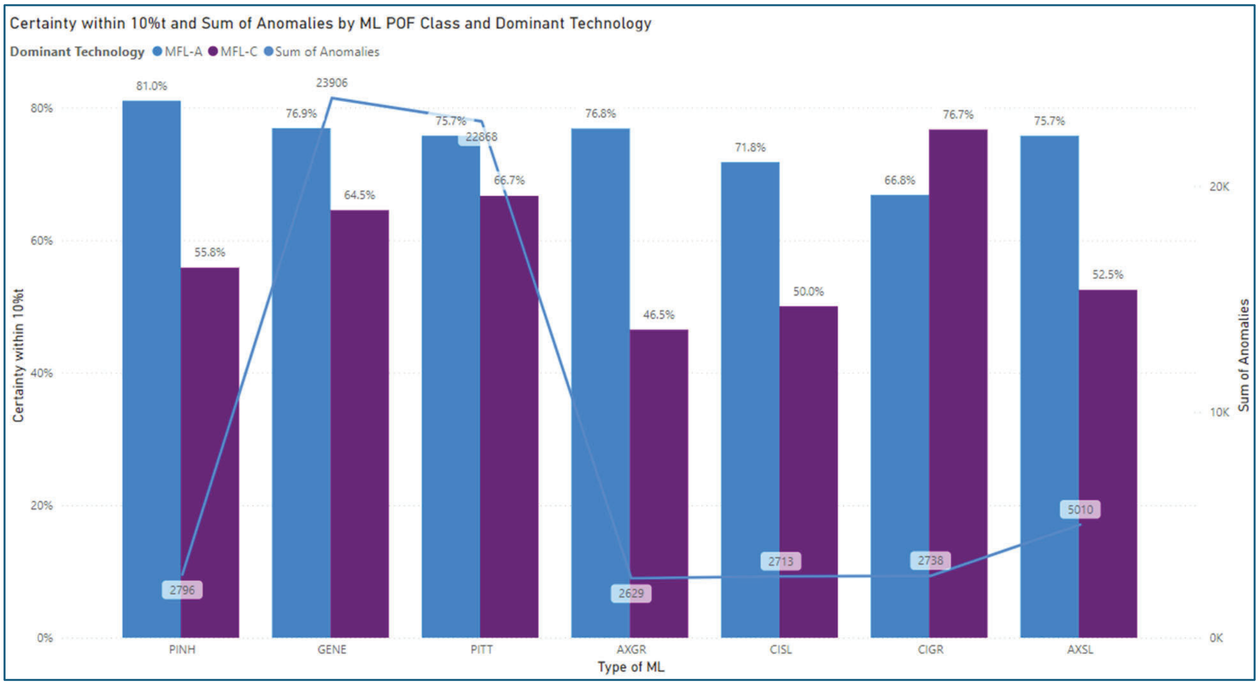


Figure 30. Certainty within 10% for MFL-A vs. MFL-C. The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.

Performance per metal loss classification, ILI-derived versus field-derived

The certainties for all technologies when the ILI system identified the metal loss classification were compared to the classification calculated from field measurements, and there were no noticeable differences. However, when MFL-A and UT ML were being compared, the large discrepancy in performance for grooving (axial and circumferential) was minimized when utilizing the classification calculated from field-provided dimensions, as seen in **Figure 31** below.

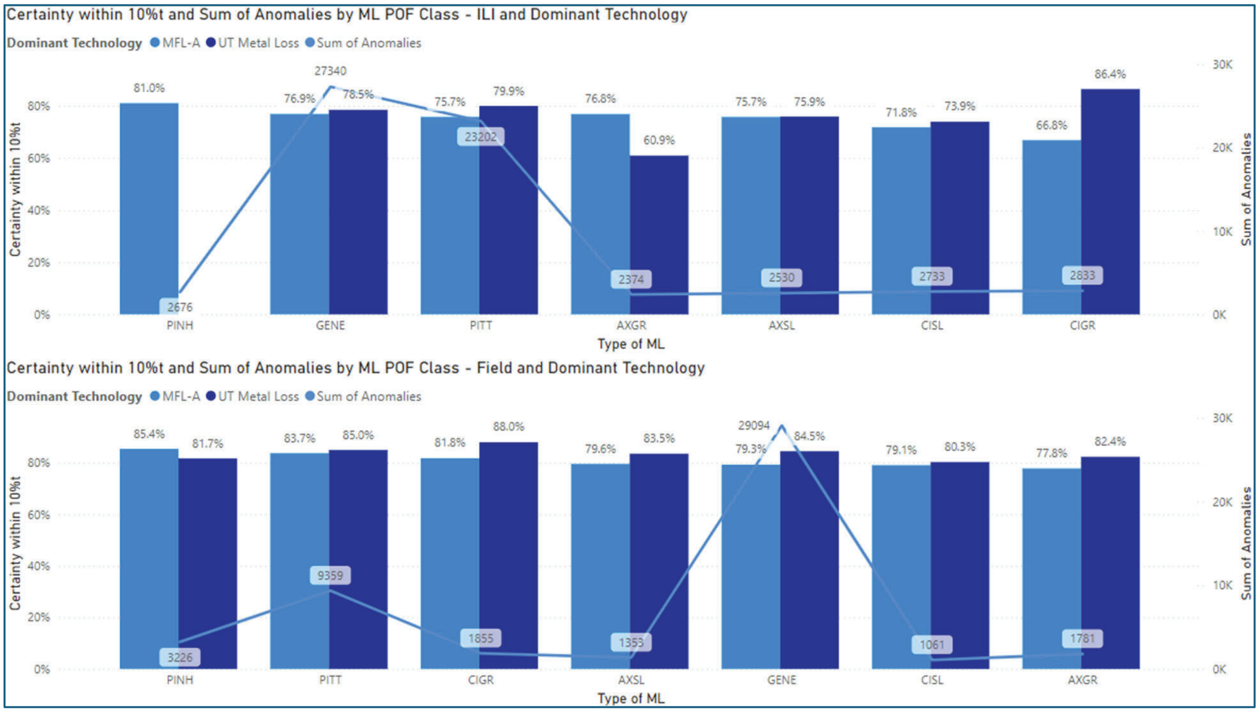


Figure 31. Certainty within 10%, MFL-A vs. UT ML for metal loss classification reported by the ILI report (top) vs. metal loss classification calculated by field-provided lengths and widths (bottom). The number of anomalies to calculate certainty is plotted as a line on the secondary y-axis.

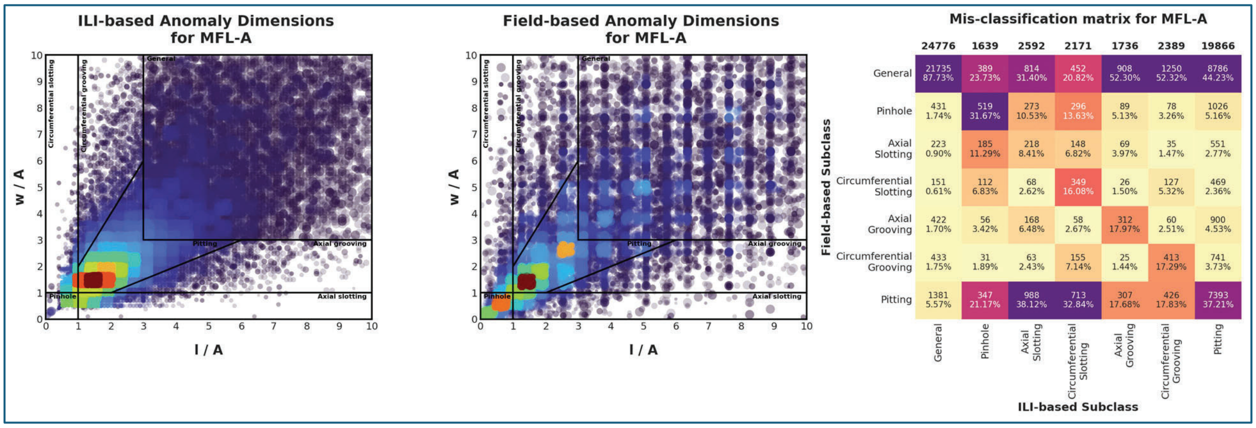
Metal loss misclassification: ILI versus field

The tendency for an ILI system to identify an anomaly as a different anomaly dimension class than what was found in the field for all metal loss technologies was explored. The field anomaly dimension class was calculated from the field-provided length and widths of the anomaly. Therefore, this analysis relies on the accuracy of the reported length and width dimensions from the field. Thus, the error associated with measuring the size of a metal loss anomaly in the field is doubled compared to only analyzing depth. **All graphs indicate the limitations of the tool technologies in identifying pinholes.**

Figure 32 shows that MFL-A has the most trouble with axial slotting; only 8% of the time is axial slotting identified by the ILI system found to be axial slotting. Anomalies identified as axial slotting and circumferential slotting by MFL-A have an 11% and 14% probability of being sized as a pinhole in the field.

Knowing how MFL-A technology works, it is unsurprising that 44% of pitting identified by MFL-A was reclassified as general corrosion in the field. Not only is the ILI system limited by the lower probability of detection (POD) when determining if pits are interacting to form general corrosion (as metal loss <10% might be in between deeper pits), but different interaction rules utilized by the ILI system versus the NDE system may also play a factor.

https://doi.org/10.52202/078572-0052



ILI-based Anomaly Dimensions for MFL-A

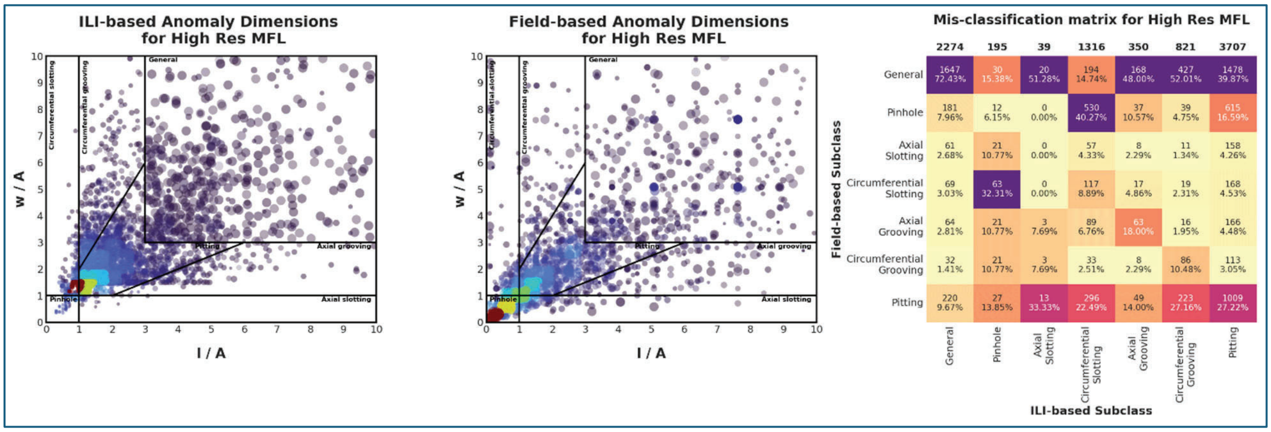
Field-based Anomaly Dimensions for MFL-A

Mis-classification matrix for MFL-A

	General	Pinhole	Axial Slotting	Circumferential Slotting	Axial Grooving	Circumferential Grooving	Pitting
	24776	1639	2592	2171	1736	2389	19866
General	21735 87.73%	389 23.73%	814 31.40%	452 20.82%	908 52.30%	1250 52.32%	8786 44.23%
Pinhole	431 1.74%	519 31.67%	273 10.53%	296 13.63%	89 5.13%	78 3.26%	1026 5.16%
Axial Slotting	223 0.90%	185 11.29%	218 8.41%	148 6.82%	69 3.97%	35 1.47%	551 2.77%
Circumferential Slotting	151 0.61%	112 6.83%	68 2.62%	349 16.08%	26 1.50%	127 5.32%	469 2.36%
Axial Grooving	422 1.70%	56 3.42%	168 6.48%	58 2.67%	312 17.97%	60 2.51%	900 4.53%
Circumferential Grooving	433 1.75%	31 1.89%	63 2.43%	155 7.14%	25 1.44%	413 17.29%	741 3.73%
Pitting	1381 5.57%	347 21.17%	988 38.12%	713 32.84%	307 17.68%	426 17.83%	7393 37.21%

Figure 32. ILI vs. field-determined POF classification of anomalies for MFL-A.

The same figures are presented for high-resolution MFL-A (**Figure 33**), MFL-C (**Figure 34**), combination MFL (**Figure 35**), and UT metal loss (**Figure 36**). For high-resolution MFL-A, 40% of anomalies classified as circumferential slotting were found to be pinholes.



ILI-based Anomaly Dimensions for High Res MFL

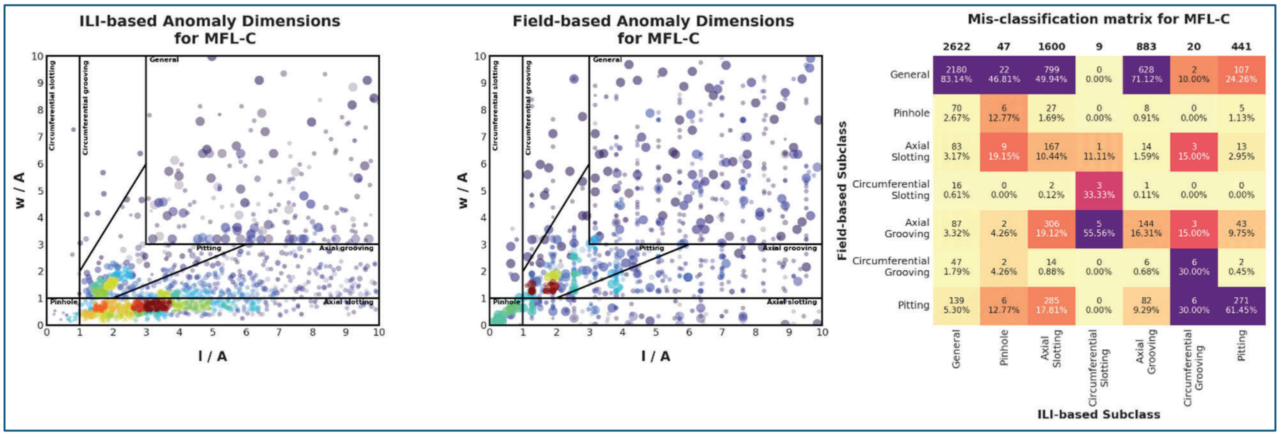
Field-based Anomaly Dimensions for High Res MFL

Mis-classification matrix for High Res MFL

	General	Pinhole	Axial Slotting	Circumferential Slotting	Axial Grooving	Circumferential Grooving	Pitting
	2274	195	39	1316	350	821	3707
General	1647 72.43%	30 15.38%	20 51.28%	194 14.74%	168 48.00%	427 52.01%	1478 39.87%
Pinhole	181 7.96%	12 6.15%	0 0.00%	530 40.27%	37 10.57%	39 4.75%	615 16.59%
Axial Slotting	61 2.68%	21 10.77%	0 0.00%	57 4.33%	8 2.29%	11 1.34%	158 4.26%
Circumferential Slotting	69 3.03%	63 32.31%	0 0.00%	117 8.89%	17 4.86%	19 2.31%	168 4.53%
Axial Grooving	64 2.81%	21 10.77%	3 7.69%	89 6.76%	63 18.00%	16 1.95%	166 4.48%
Circumferential Grooving	32 1.41%	21 10.77%	3 7.69%	33 2.51%	8 2.29%	86 10.48%	113 3.05%
Pitting	220 9.67%	27 13.85%	13 33.33%	296 22.49%	49 14.00%	223 27.16%	1009 27.22%

Figure 33. ILI vs. field-determined POF classification of anomalies for high-resolution MFL-A.

For MFL-C, 50% of the anomalies identified as axial slotting were found to be general corrosion.


ILI-based Anomaly Dimensions for MFL-C

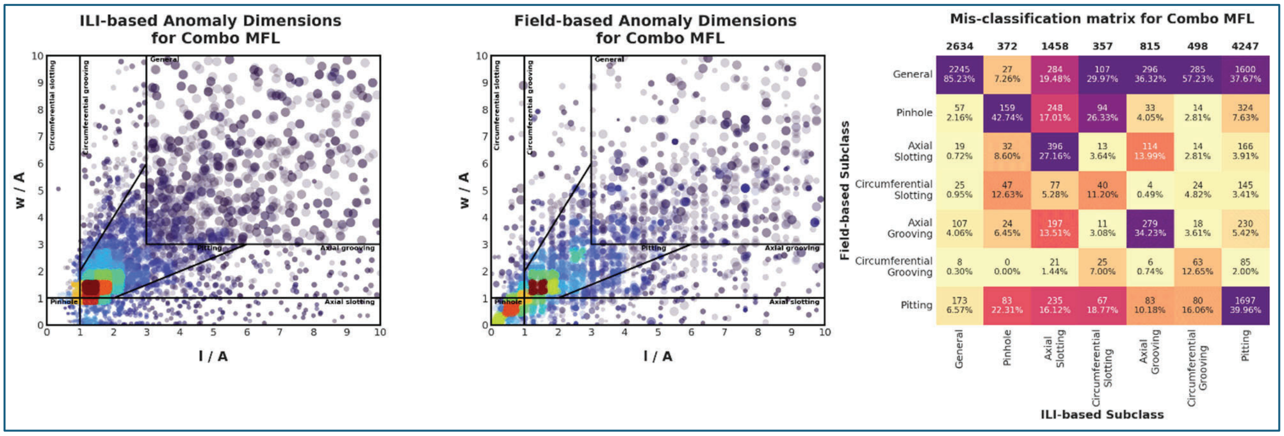
Field-based Anomaly Dimensions for MFL-C

Mis-classification matrix for MFL-C

Field-based Subclass	General	Pinhole	Axial Slotting	Circumferential Slotting	Axial Grooving	Circumferential Grooving	Pitting
	2622	47	1600	9	883	20	441
General	2180 83.14%	22 46.81%	799 49.94%	0 0.00%	628 71.12%	2 10.00%	107 24.26%
Pinhole	70 2.67%	6 12.77%	27 1.69%	0 0.00%	8 0.91%	0 0.00%	5 1.13%
Axial Slotting	83 3.17%	9 19.15%	167 10.44%	1 11.11%	14 1.59%	3 15.00%	13 2.95%
Circumferential Slotting	16 0.61%	0 0.00%	2 0.12%	3 33.33%	1 0.11%	0 0.00%	0 0.00%
Axial Grooving	87 3.32%	2 4.26%	306 19.12%	5 55.56%	144 16.31%	3 15.00%	43 9.75%
Circumferential Grooving	47 1.79%	2 4.26%	14 0.88%	0 0.00%	6 0.68%	6 30.00%	2 0.45%
Pitting	139 5.30%	6 12.77%	285 17.81%	0 0.00%	82 9.29%	6 30.00%	271 61.45%

ILI-based Subclass

Figure 34. ILI vs. field-determined POF classification of anomalies for high-resolution MFL-C.


ILI-based Anomaly Dimensions for Combo MFL

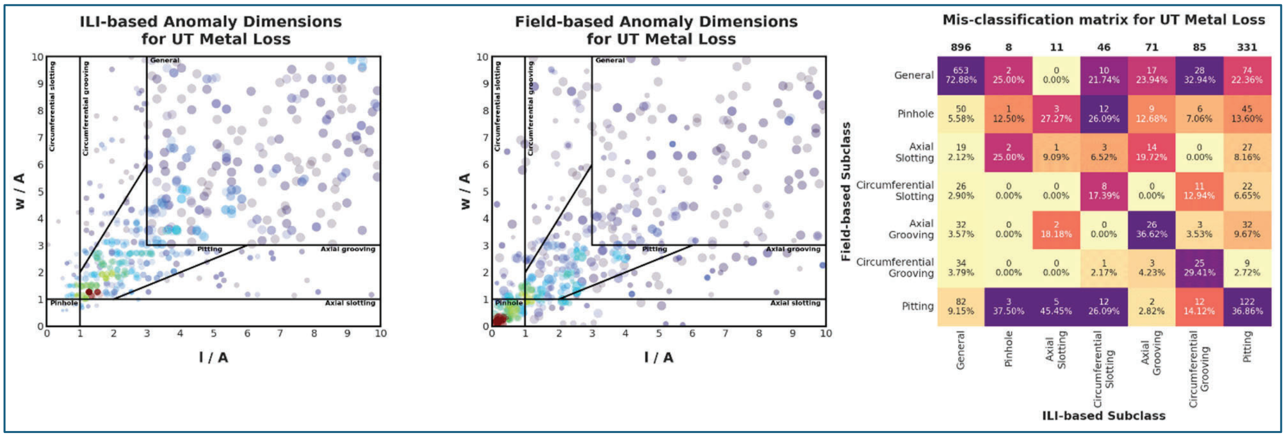
Field-based Anomaly Dimensions for Combo MFL

Mis-classification matrix for Combo MFL

Field-based Subclass	General	Pinhole	Axial Slotting	Circumferential Slotting	Axial Grooving	Circumferential Grooving	Pitting
	2634	372	1458	357	815	498	4247
General	2245 85.23%	27 7.26%	284 19.48%	107 29.97%	296 36.32%	285 57.23%	1600 37.67%
Pinhole	57 2.16%	159 42.74%	248 17.01%	94 26.33%	33 4.05%	14 2.81%	324 7.63%
Axial Slotting	19 0.72%	32 8.60%	396 27.16%	13 3.64%	114 13.99%	14 2.81%	166 3.91%
Circumferential Slotting	25 0.95%	47 12.63%	77 5.28%	40 11.20%	4 0.49%	24 4.82%	145 3.41%
Axial Grooving	107 4.06%	24 6.45%	197 13.51%	11 3.08%	279 34.23%	18 3.61%	230 5.42%
Circumferential Grooving	8 0.30%	0 0.00%	21 1.44%	25 7.00%	6 0.74%	63 12.65%	85 2.00%
Pitting	173 6.57%	83 22.31%	235 16.12%	67 18.77%	83 10.18%	80 16.06%	1697 39.96%

ILI-based Subclass

Figure 35. ILI vs. field-determined POF classification of anomalies for combination MFL.


ILI-based Anomaly Dimensions for UT Metal Loss

Field-based Anomaly Dimensions for UT Metal Loss

Mis-classification matrix for UT Metal Loss

Field-based Subclass	General	Pinhole	Axial Slotting	Circumferential Slotting	Axial Grooving	Circumferential Grooving	Pitting
	896	8	11	46	71	85	331
General	653 72.88%	2 25.00%	0 0.00%	10 21.74%	17 23.94%	28 32.94%	74 22.36%
Pinhole	50 5.58%	1 12.50%	3 27.27%	12 26.09%	9 12.68%	6 7.06%	45 13.60%
Axial Slotting	19 2.12%	2 25.00%	1 9.09%	3 6.52%	14 19.72%	0 0.00%	27 8.16%
Circumferential Slotting	26 2.90%	0 0.00%	0 0.00%	8 17.39%	0 0.00%	11 12.94%	22 6.65%
Axial Grooving	32 3.57%	0 0.00%	2 18.18%	0 0.00%	26 36.62%	3 3.53%	32 9.67%
Circumferential Grooving	34 3.79%	0 0.00%	0 0.00%	1 2.17%	3 4.23%	25 29.41%	9 2.72%
Pitting	82 9.15%	3 37.50%	5 45.45%	12 26.09%	2 2.82%	12 14.12%	122 36.86%

ILI-based Subclass

Figure 36. ILI vs. Field-determined POF classification of anomalies for UT ML

https://doi.org/10.52202/078572-0052

Has performance improved over time?

The effect that assessment year has on certainty was also explored in various ways. The certainty within a 10% and 20% tolerance from 2003 (the earliest year we have contiguous data) to 2024 is provided in **Figure 37**.

- Overall, the certainty has trended up for a 10% and 20% tolerance.
- The certainty within a 10% tolerance trended up between 2003 and 2019 with a rise of 32%. **There's a recent trend change that started in 2020, where the certainty decreased by 10% from 2020 to 2024.**
- The certainty for a 20% tolerance has remained relatively steady, with the most sustained period of growth being between 2003 and 2019.

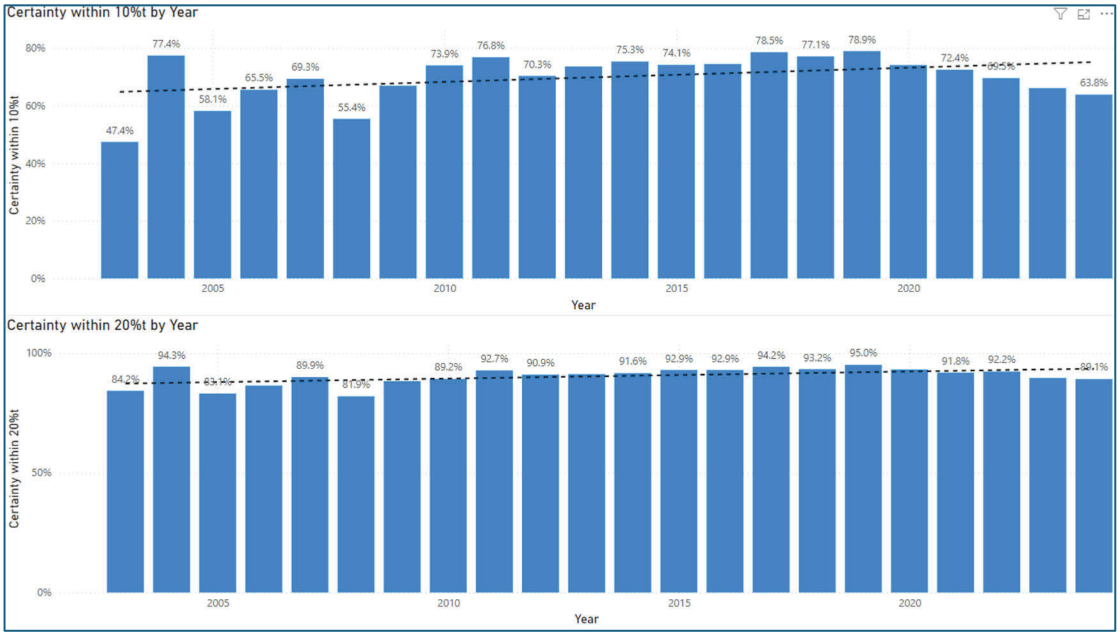


Figure 37. Certainty within a 10% and 20% tolerance by assessment year, 2003 to 2024.

A similar trend can be seen when looking at the average error - the average of the absolute differences of field-found depth and ILI depth. The error also started trending up in 2020, seen in **Figure 38**, indicating that ILI system performance regarding depth sizing has decreased recently.

https://doi.org/10.52202/078572-0052

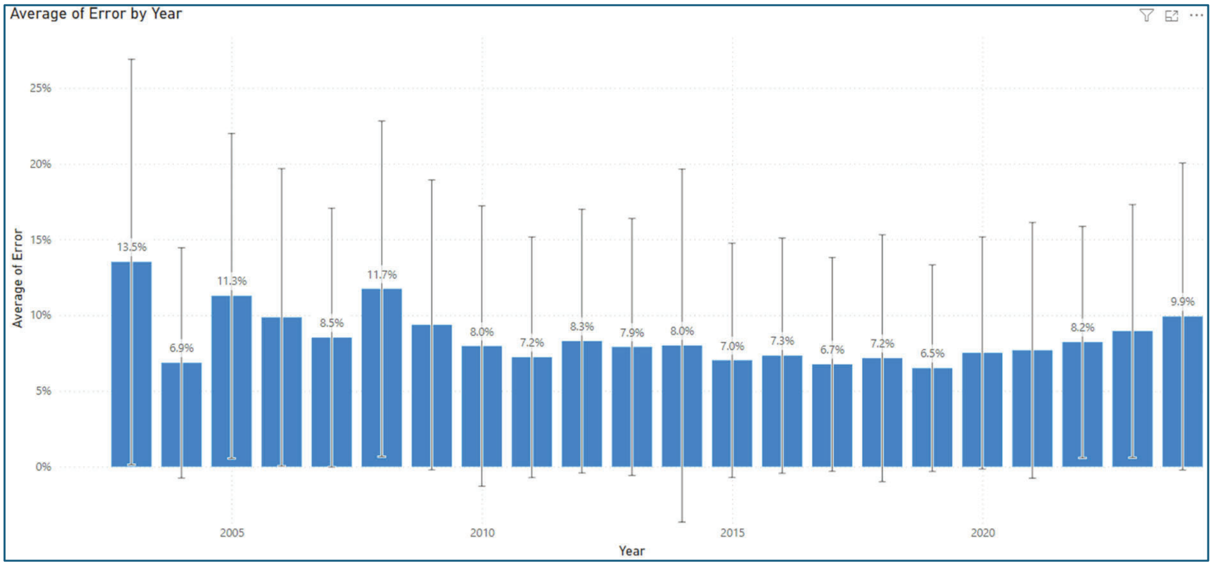


Figure 38. Average error by assessment year, 2003 to 2024.

Why is recent performance trending down?

One hypothesis is that ILI systems have plateaued in their progress, as significant technological gains have yet to happen as they did between 2003 and 2019. However, a plateau of technological advancement would not necessarily cause a decline in performance.

Another cause might be the change in NDE technology utilized to determine the depth of anomalies "in the field." As manually operated pit gauges are being replaced by automated laser scans of corrosion fields, the field-produced depths are becoming more accurate. Is the advancement of NDE technology increasing the discrepancies between the ILI-predicted depths and the "truth data"?

Another possible explanation is that the anomalies are changing or in other words, the anomalies that pipeline operators choose to evaluate are changing. Could the anomalies in recent years be of a morphology that's more difficult to size? To investigate whether anomalies identified for evaluation have changed recently, the metal loss classification was plotted over time for those anomalies included in this study, as seen in **Figure 39**. One can see a significant change in the makeup of metal loss classes that pipeline operators are remediating. Between 2005 and 2024, circumferential slotting had the most significant increase (2500%), while general metal loss had the largest decrease (62%). General metal loss started trending down in 2020, falling by 93% in 4 years.

https://doi.org/10.52202/078572-0052

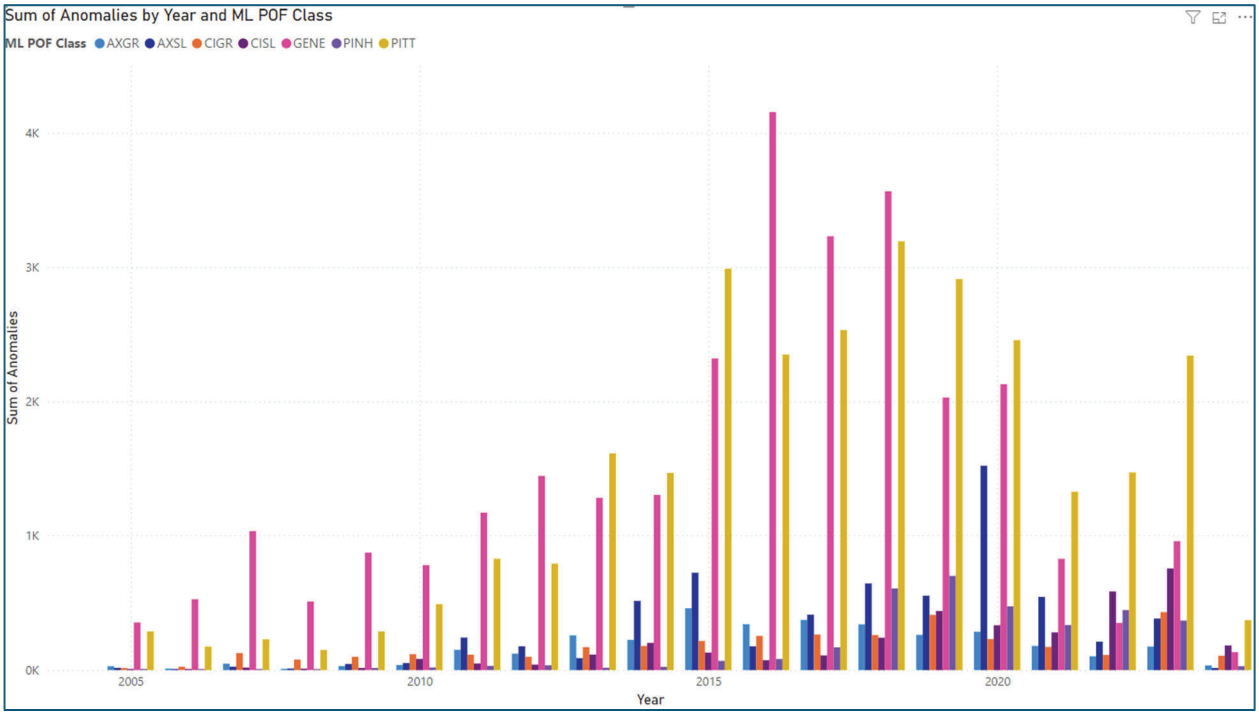


Figure 39. Anomalies included in this study by metal loss classification by assessment year.

The makeup of the anomalies in this study was also analyzed per the classification description, seen in **Figure 40**. Although it's known that external corrosion and external metal loss make up the bulk of the anomalies included in this study, internal corrosion started trending up around 2020, including internal corrosion along the long seam. External corrosion along the long seam also started trending up around 2020.

https://doi.org/10.52202/078572-0052

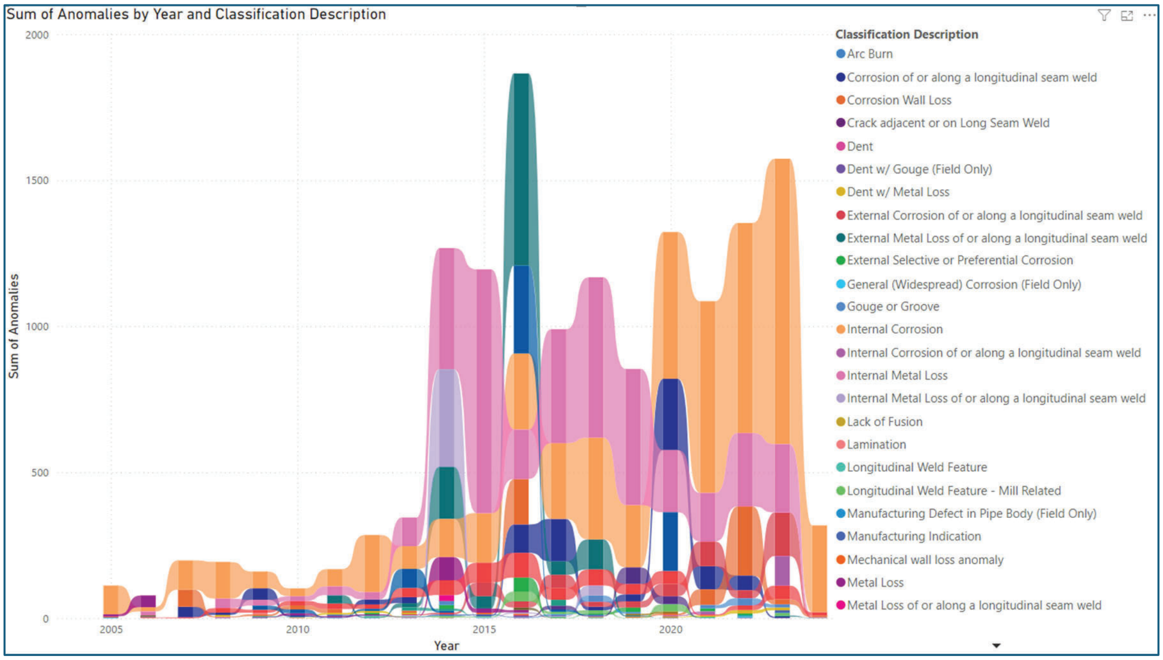


Figure 40. A description of the classification of the anomalies included in this study, excluding external corrosion and external metal loss.

This increase in the number of metal loss anomalies associated with the long seam may be linked to the fact that more anomalies from the pipe vintage "bin" 1950 – 1960 are being inspected and remediated, as seen in the top visual in **Figure 41**. Longitudinal weld anomalies make up the third-highest anomaly type identified on pipelines installed between 1950 and 1960, behind external and internal metal loss, as seen in the bottom visual of **Figure 41**.

https://doi.org/10.52202/078572-0052

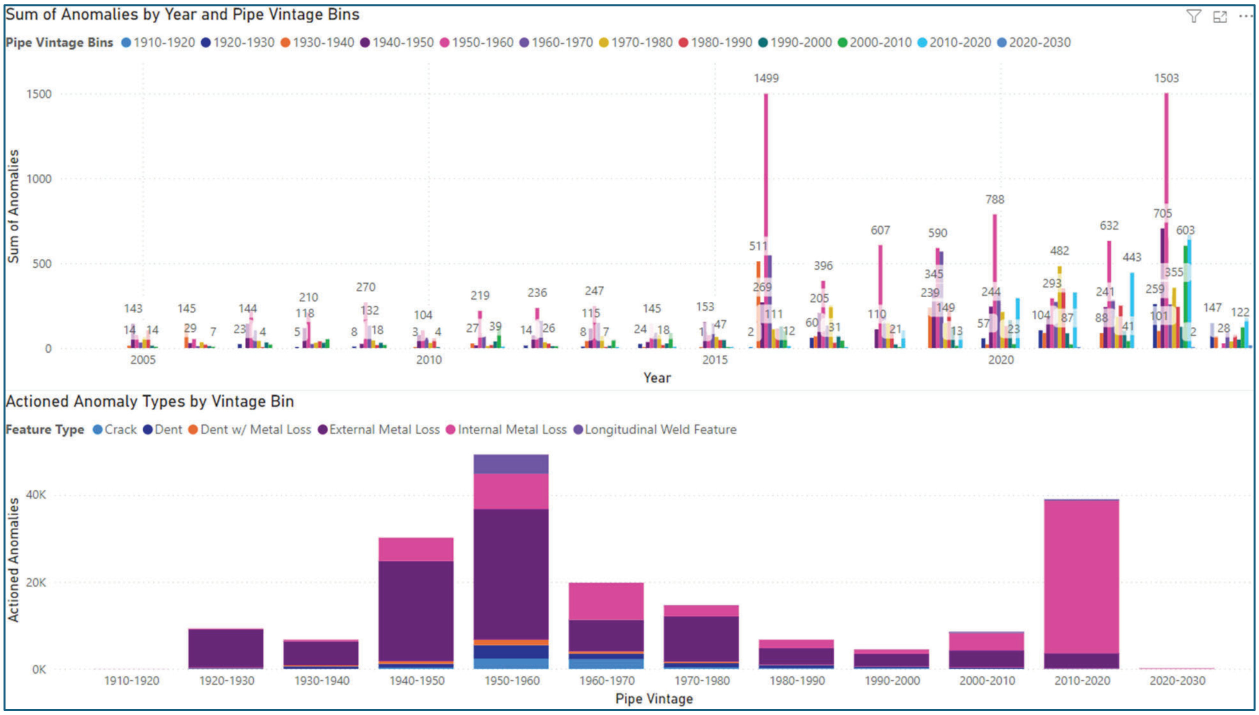


Figure 41. The pipe vintage "bin" where anomalies included in this study were found (top) and the top 5 anomaly types per vintage bin (bottom).

Conclusions

Of the 112,216 anomalies with both an ILI depth and field depth, 81,497 were utilized in this study. The following takeaways were gleaned from analyzing the depths of these anomalies: ILI versus field.

What drives performance?

- If a tolerance of 10% or better (high depth sizing accuracy) is needed, selecting the ILI company and technology is most important. If a 20% tolerance is sufficient, the ILI company and operator are more critical than the type of technology utilized.
- The depth of the metal loss anomaly has the highest effect on whether the ILI system will predict the correct depth, whereas metal loss classification had the lowest. Interestingly, ILI systems typically have different performance specifications for metal loss classifications, not for anomalies at different depths.

On technologies

- There's a high variability in the depth sizing performance of MFL-A technology, with certainties ranging from 49% to 85% for a 10% tolerance. (Including outliers, the lowest certainty was calculated to be 27%.)

- When comparing the top quartile of ILI companies, MFL-A outperformed UT metal loss. This indicates that MFL-A from a high-performing ILI company can have depth sizing performance as good or better than an ultrasonic ILI system.
- When analyzing anomalies of all depths and dimension classes, UT metal loss outperformed all other technologies (but not by much.)

On the depth of the metal loss anomaly

- In general, depth sizing performance decreases as the depth of the metal loss anomaly increases.
- This trend is more pronounced for MFL-A than high-resolution MFL, MFL-C, and combination MFL, which all outperform MFL-A when the metal loss depth is 50% or greater.
- For a tolerance of 10%, UT ML outperforms MFL-A (or is within <1%) for all depth bins, except for 70% or deeper, where there's not enough data for UT ML.
- For a tolerance of 20%, MFL-A and UT have very similar performances, with MFL-A slightly outperforming UT ML for anomalies 40% to 70% in depth.

On metal loss classification

- When comparing MFL-A versus UT:
    - Using a 10% tolerance, UT ML significantly outperformed MFL-A by 20% in circumferential grooving, and MFL-A significantly outperformed MFL-A for axial grooving (16% difference).
    - For all other metal loss classifications, the difference in calculated certainties at a 10% and 20% tolerance was less than 5%.
- Pinholes are challenging to detect for all technologies, most likely due to the physical limitations of the tools.

On time/year of assessment

- Overall (from 2003 to 2024), ILI performance has improved. However, recently, performance has decreased between 2020 and 2024.
- One hypothesis supported by the data in this study, is that anomaly types being evaluated in recent years are more challenging to size by ILI systems due to their morphology, i.e., a decrease in general metal loss and an increase in circumferential slotting as well as internal corrosion and corrosion associated with the long seam.

https://doi.org/10.52202/078572-0052

What’s next?

With a vast amount of data comes an infinite number of questions that can be asked of the data. In many ways, the most challenging part of this project was determining the right questions to ask. As with any research project, the more questions answered, the more questions were created. Some ideas for additional questions or research are as follows.

- Analyzing the performance of different technologies on the same pipelines could support or refute the findings in this study and be particularly helpful in further investigating the performance of MFL-A versus UT metal loss.
- Conducting a Level 3 API 1163 analysis to more rigorously quantify performance and determine a confidence interval would be helpful, specifically for those insights that seemed at odds with expectations. (The confidence level defined by API 1163 is a statistical term that describes the mathematical certainty with which a statement is made.)
- Extend this analysis to other defect types, i.e., cracks, dents, hard spots, etc.
- Interrogate the data behind the recent trend in recent decreased performance.

(It should be noted that other integrity-related questions were analyzed as part of this project but left out for brevity, i.e., what anomaly types are being identified for action or remediation the most, whether anomaly type and pipeline properties correlate, and how, etc.)

Acknowledgments

Much gratitude goes to our pipeline operator clients. Without them, this paper wouldn’t exist. Not only have they been generous with their data, but generous with their time answering questions and reviewing findings. Our sincerest thanks go to them and their drive to move the industry forward.

References

1. “API 1163 In-line Inspection Systems Qualification." *API 1163*, 2013.
2. "Specifications and Requirements for In-line Inspection of Pipelines." *Pipeline Operators Forum*, November 2021
3. API 1163 Performance Validation Guidelines Project Number IM-1-06." *Pipeline Research Council International, Inc.*, no. March 17, 2023, 2023.
4. Jager, Christoph. "Enhancing Pipeline Integrity: A Comparative Study of Ultrasonic and Magnetic Flux Leakage Inspection Technologies." 19th Pipeline Technology Conference, Euro Institute for Information and Technology Transfer, 8 Apr. 2024, Berlin. Presentation.

https://doi.org/10.52202/078572-0052