OPTIMISING MAGNETIC FLUX LEAKAGE INSPECTION ACCURACY THROUGH THE APPLICATION OF AI TECHNIQUES AND ONGOING PERFORMANCE TRACKING

Philip English, Stuart Clouston, Anthony Tindall, Baker Hughes

Abstract

Magnetic Flux Leakage (MFL) continues to be the most used in-line inspection technology since its widescale introduction in the 1980's. Thousands of MFL inspections are completed each year by pipeline operators as a reliable and effective assessment method to determine current pipeline condition on which their monitoring, future integrity assessments and repair programs will be based.

Over the years the technology has advanced with combination inspections such as caliper and IMU data gathered efficiently in the same inspection. Advances have also been made in both resolution and the measurement of the three leakage fields that occur in response to corrosion defects. Measurement improvement has benefited detection, characterization and sizing of smaller and more challenging corrosions that otherwise might lead to outliers from specification leading to potential safety risks.

This paper will briefly explain some of the advantages tri-axial measurement has brought to standard MFL analysis, but importantly it will look deeper into how the multiple measurements taken at defects have enabled a step change in the application of AI to improve corrosion measurement sizing accuracy. In addition, this AI capability has introduced the possibility of a defect-by-defect tolerance prediction, and improved corrosion growth estimation. Together, these advances will both reduce conservatism, leading to unnecessary repairs, while at the same time predict where tolerances may be greater and remove unknown risk.

Specifically, the paper will provide insight into the technical development work Baker Hughes has conducted by applying deep learning techniques to large areas of raw MFL data to predict complex defect morphologies and improve burst pressure estimations in situations typically challenging for standard analysis. Additionally, we will discuss how AI is being used to predict new corrosion during run-to-run assessment to consider corrosion cluster growth more accurately as a basis for future integrity assessment.

Lastly, the authors will describe the techniques being adopted to effectively monitor ILI performance to specification, how it is used to identify when inspections may have not met operator expectation and how that data can be monitored to target continuous improvements to system performance.

Introduction

The in-line inspection industry has measured the success of an in-line inspection service in the same way since the introduction of the Pipeline Operators Forum defect classifications and the API 1163 definitions on how to measure statistical population performance. However, these traditional methods pose challenges to pipeline operators when real life situations occur, which do not fit into standard dimension class categories. Let us consider the following two familiar questions:

- *“The corrosion on my pipeline does not fall into the simplified 7 POF categories; how do I measure performance on complex of defects like pit in pit and other interacting defects?”*

- *“It is great that the ILI tool was within specification on the 98 features from infield verification on my pipeline, but I have 2 outliers, not in specification which nearly caused a leak. How do I manage this effectively?”*

If technology is only challenged to conform to this overall statistical approach, then the opportunity to truly solve real world problems and take advantage of technological improvements is lost.

To fully drive change and improve quality and ultimately safety, then we must look at the problem/challenge in a multi-phased approach:

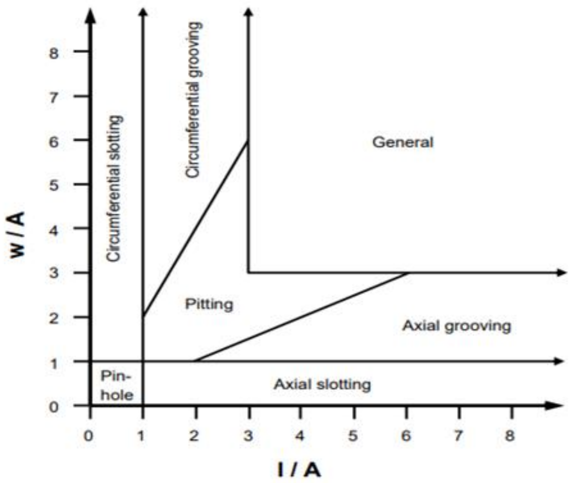
1. Understand the limitations of a simplified specification and integrity management.
2. Collect the ILI data you need: Triaxial MFL – Axial, Radial and Transverse
3. Build a comprehensive Dig Database
4. Utilise Artificial Intelligence (AI) for :
   a. Improved defect profiles
   b. Specific individual defect tolerances
   c. Statistical dynamic growth
5. Consider ILI performance as not just meeting specification but minimised risk of significant outliers.

Within this paper we will revisit some of the principal benefits of triaxial data and how deep learning (Artificial intelligence) utilising Baker Hughes’ vast dig database can be utilised effectively to improve pipeline integrity and safety, whilst also looking at how we can utilise a dig data management system and Data Accuracy Score to ensure that technology advancements are focused on what will have the most impact to dig efficiency and effectiveness.

Understanding the Limitations of a Simplified Specification and Integrity Management

Pipeline corrosion can take on various shapes and sizes, both internally and externally. For over 30 years, the Pipeline Operators Forum (POF) specification has served as the industry standard for classifying corrosion defects based on their length and width (Figure 1). While this simplified approach is widely used, it becomes inadequate when dealing with complex corrosion features, such as defects within defects, intersecting axial slots, or corrosion surrounded by background material. These irregular defect shapes do not fit neatly into the length-and-width-based framework, potentially misrepresenting their severity and morphology.

Figure 1: POF Metal Loss Categories



The geometrical parameter A is linked to the NDE methods in the following manner:

- If t < 10 mm then A = 10 mm
- If t ≥ 10 mm then A = t

Simplified specifications are often based on idealised parabolic shapes, failing to account for the full range of real-world corrosion scenarios. Pipeline operators dealing with real life complex corrosion features can often face the two main challenges of:

1. **Safety Outliers**: These occur when the true severity of a defect is underestimated, potentially leading to catastrophic failures like leaks or ruptures.
2. **Resource Outliers**: These arise when the severity of a defect is overestimated, leading to unnecessary excavations and repairs.

To mitigate these risks, advanced technologies, such as Baker Hughes triaxial Magnetic Flux Leakage (MFL) sensors, offer enhanced data capture. The Triaxial sensors better characterise complex corrosion profiles and minimise the risk of both safety and resource outliers.

Challenging these traditional specification concepts further, it is prudent that other considerations are considered for defect tolerances. Simply applying a static specification to a defect POF category limits the usefulness of the In-line Inspection report and does not truly reflect reality. If you consider a simple specification for +/-10% for all pitting, even those untrained in Magnetic data analysis can see that applying this to a 10% defect and a 70% does not logically fit. This is a simple example and if you consider interacting defects compared to isolated features, this concept of a static specification applied for all pits rapidly becomes illogical.

Only by taking the concept of pipeline integrity into the 4th dimension of time, then a pipeline operator can truly manage the threats to their pipelines effectively. Historical time-based pipeline assessments have been based on predicting corrosion growth between inspections and applying that linearly in the future years. However, how do we ensure that we can correctly consider new initiation sites for corrosion.

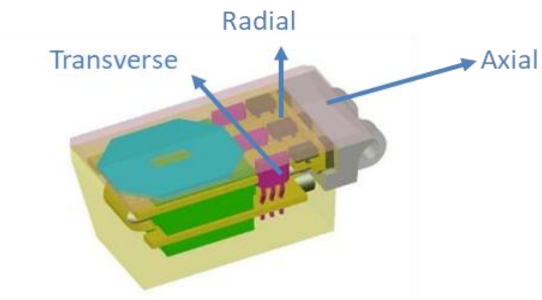
Collect the Data You Need: Benefits of Triaxial MFL – Axial, Radial and Transverse

Magnetic flux leakage is a vector quantity, meaning it has both magnitude and direction. Unlike traditional MFL tools that measure only one component (axial), **triaxial MFL sensors** capture three orthogonal components of the leakage: **Axial**, **Radial**, and **Transverse**. This additional data provides deeper insights into the morphology of corrosion defects, enabling more accurate and detailed assessments.

Defects that are difficult to assess with single-axis measurements—such as axial slots, pinholes within larger corrosion areas, or complex interacting defects—are better understood through the combination of all three vector components.

In addition to the three primary leakage components (Axial, Radial, and Transverse), the same sensor head can also record **Eddy Current** data for internal and external (IDOD) measurements, providing a comprehensive understanding of wall loss.

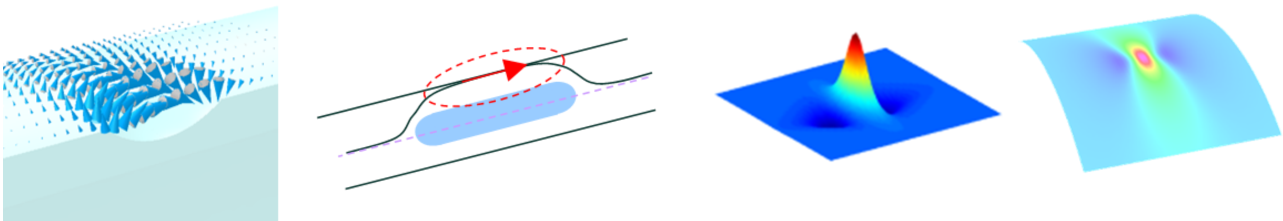
Figure 2: High density multiple "Triaxial" sensor head



As illustrated in Figure 2, the high-density triaxial sensor head enables the recording of detailed leakage data from every angle. The data captured by these sensors are not just theoretical; they provide practical applications for differentiating between complex defect morphologies. To gain a deeper understanding of how the three independent leakage components offer unique identifiers, we will explore the applications.

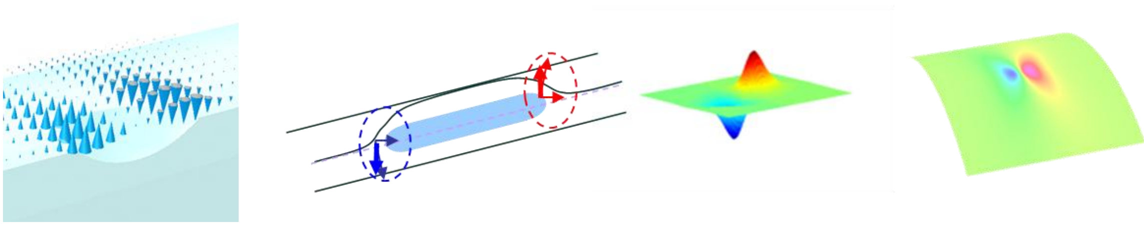
The axial sensor measures the flux leakage i.e., volume of corrosion and its key components. For a simple pit, the signal consists of a single peak (Figure 3):

Figure 3: A typical defect signal in the axial direction



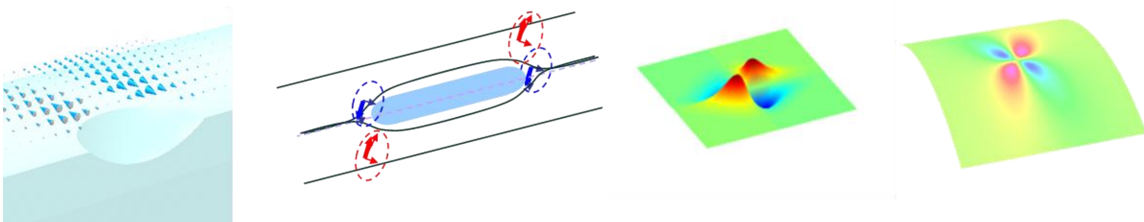
The radial sensor measures the points and the rate of change where the field leaves the pipe and returns. For a simple defect, the signal consists of two peaks of opposite polarities (Figure 4):

Figure 4: A typical defect signal in the radial direction



The transverse sensors measure the movement of the field around the defect. This highlights the corners or extent of a defect, and the rate of change measured circumferentially around the pipe. For a simple defect, a typical signal will have four poles surrounding the peak of the defect (Figure 5). Added sensitivity to in-plane shape thus improves width measures and the identification of background corrosion (larger underlying areas of corrosion) and corrosion in corrosion situations.

Figure 5: A typical defect signal in the transverse direction

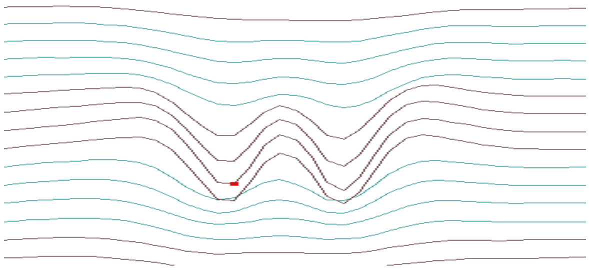


Examples in Figure 6 consider the interaction between two isolated axially aligned pits:

Figure 6: Axially aligned pits

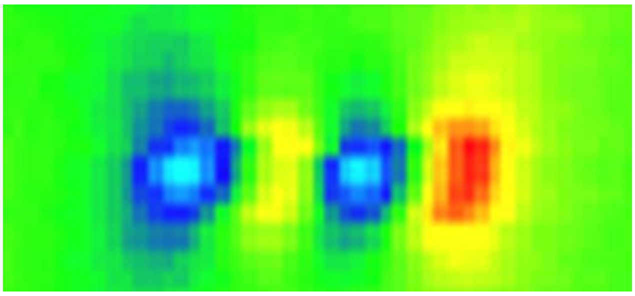
Figure 7 shows the magnetic response of Axial only MFL from the two pits, which can be clearly seen in trace view.

Figure 7: Axially aligned pits – axial ILI signal



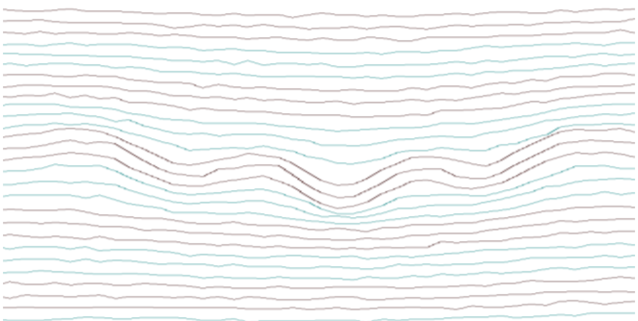
When analysing the radial measurement of the triaxial sensor data, it is beneficial to utilise a "false colour or rainbow/heat map" view. Figure 8 shows the radial signal for the two pits. In the false colour, radial measurement from the radial sensor, the start is represented by the blue on the heat map, while the yellow/red represents the end of the defect. Two pairs of "red-blue" poles can be seen, indicating that whilst the pits are close, they are not interacting.

Figure 8: Axially aligned pits – radial ILI signal



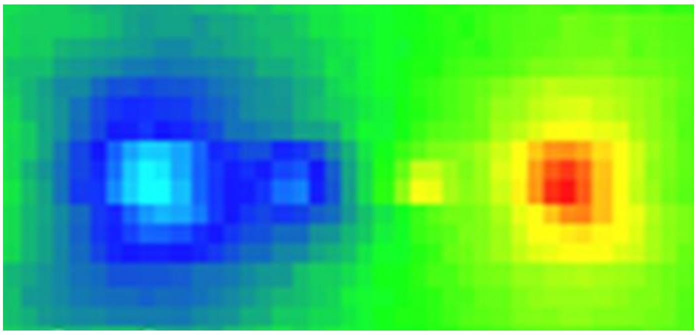
However now consider the axial data represented in Figure 9. Utilising the axial data only, these signals could easily be mis-classified as three pits, with some interaction.

Figure 9: Potentially 3 pits – axial ILI signal



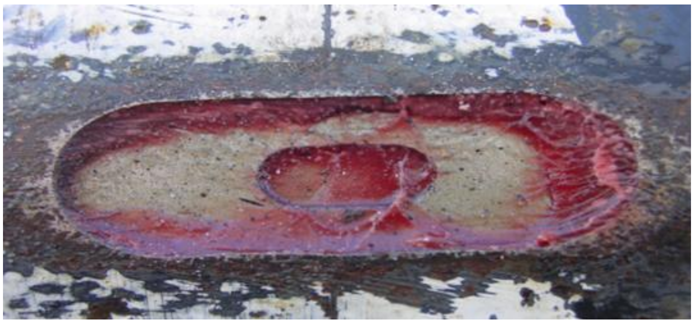
The radial signal of the triaxial sensor measures the start and end of the defect and the rate of change as the flux leaves the pipeline. The rainbow signals clearly indicate that there is one large start, followed by a smaller start (the two blue peaks). This is followed by a smaller end and a larger end (the two yellow/red peaks). From this it is clear from the signal that the defect is not three pits but a defect within defect.

Figure 10: Defect within defect – Radial ILI signal



Looking at the radial signal further, as the starts are separated, then it is not only a defect within defect, but a slot within a slot.

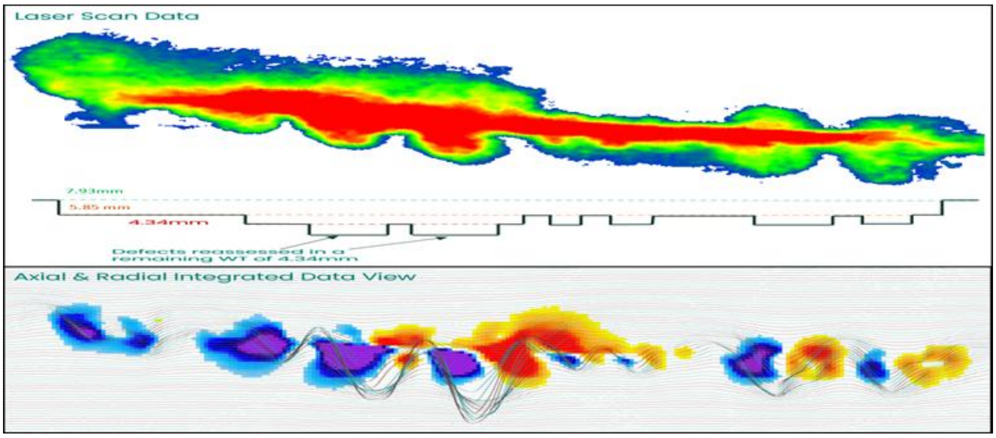
Figure 11: Defect within a defect



This demonstrates how important it is to deconstruct complex morphologies and use this in the analysis process to improve POD, POI, and POS. These identifiers not only ensure correct corrosion interpretation, but also that these defect types are identified correctly within the often hundreds of thousands of features within many miles of pipeline.

Real-world applications of triaxial MFL Data are shown in Figure 12. An overlap view of axial and radial sensor readings clearly illustrates how complex corrosion morphologies, such as pitting corrosion within an axial corrosion area, can be deconstructed for more accurate analysis. By utilising the triaxial sensor data, applying background corrosion, and remaining WT techniques, the technology was able to assess the deepest points as 69 %wt with a total length of 663 mm by 221 mm width within 10% of the actual measurements in field.

Figure 12: MFL triaxial Sensor Overlap View (MagneScan™ tool)



However, what is clear is that even with comprehensive rules, multiple signal responses and advanced techniques to manage complex signals resulting from the overlapping features and subsequent super positioning magnetic flux signals is very challenging. So, an alternate means to reconstruct the triaxial MFL signals directly into their true profiles on the pipeline would be highly advantageous and indeed necessary to reach accuracy entitlement of the technology. It also follows that any AI technique will be more successful if provided three different "views" of the signal on which to train again truth data in the form of LaserScan or other high resolution and accuracy profile information.

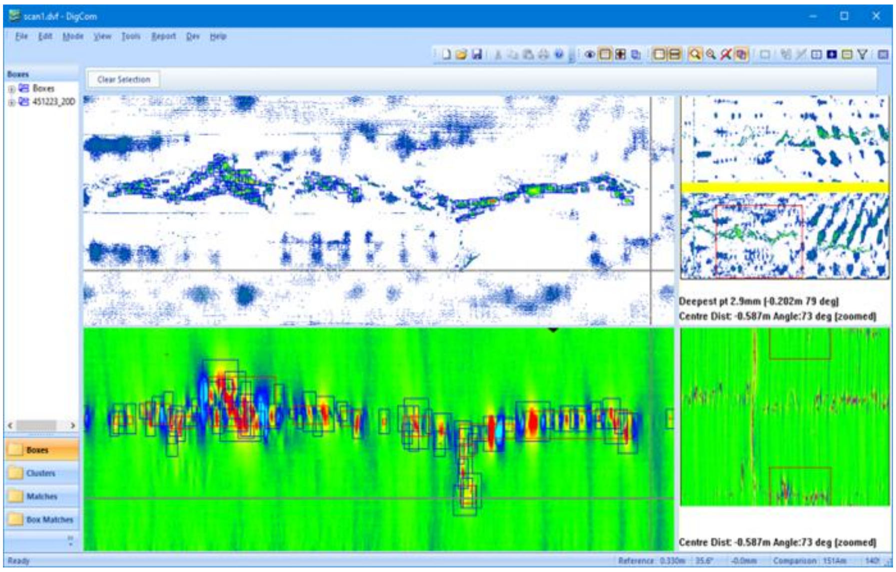
Building a Comprehensive Dig Database

Baker Hughes has implemented a cloud-based Dig Database system from thousands of ILI runs. This allows for the streamlined integration, alignment, and storage of signal data, with analysis and dig results. The data can be mined for performance metrics (at the defect level), identify trends, and target improvements.

Baker Hughes' dig database allows for the ingestion of not only numerical data, but also the ILI inspection data and LaserScan Data.

The LaserScan Data is additionally stored matched with the axial, radial and transverse data, along with all the pipeline and ILI tool data.

Figure 13: LaserScan data aligned to Baker Hughes ILI data ready to import into dig database

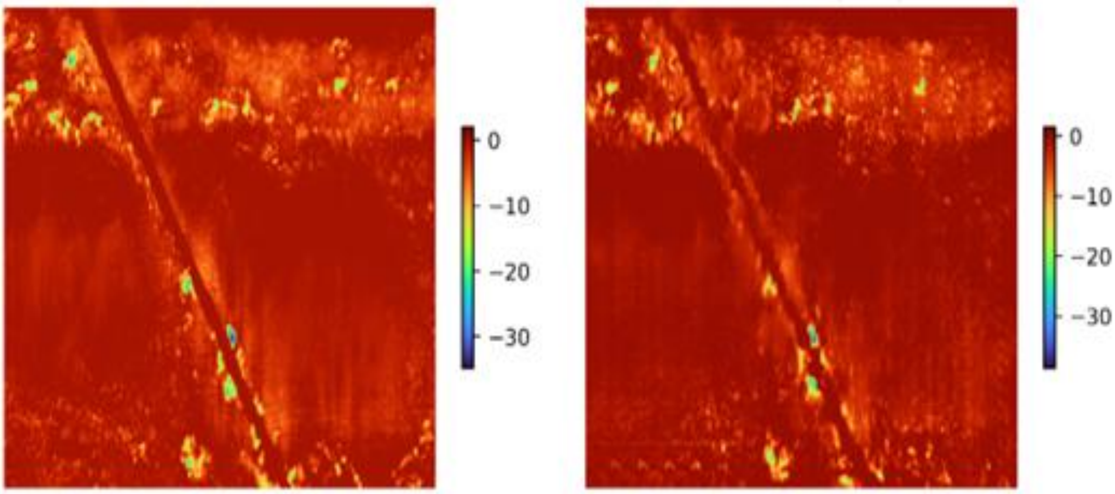


Establishing the alignment of the LaserScan (truth data) to the triaxial MFL data allows for the three distinct leakage responses of the triaxial data to be compared to the profile from the LaserScan data. LaserScan data normally contains areas of pipeline data between 2-6m; this data is vital as it holds information on how each defect interacts with each other and any other pipeline component. Due to the completeness, it is ideal as the truth data needed to train machine learning models and of course traditional engineering methods of algorithm development and refinement.

Generating "True Defect Profiles" from MFL Triaxial Data

To push the boundaries of MFL sizing accuracy, Baker Hughes has incorporated machine learning and deep learning techniques into its analysis workflows. By leveraging the vast datasets acquired in the dig database and advanced computational models, AI techniques are improving defect characterisation, especially for complex morphologies. Figure 14 demonstrates a comparison of laser-mapped defects and AI-generated reconstructions of corrosion from the raw MFL data, highlighting the potential of deep learning to improve the accuracy of defect sizing and burst pressure assessments.

Figure 14: Laser mapped defects, and the output of a generative AI deep learned model for the same area. An example of laser (left) and predicted laser profile (right)



When trained on high volumes of representative truth data, it can produce highly accurate predictions of defect depth and burst pressure from raw ILI data. The data shown is generated predicted without analyst input. This technology is still developing, but as can be clearly seen represents a significant leap forward, particularly when, as previously highlighted, assessing complex areas that are difficult, and often subjective, when applying traditional analysis techniques.

Figure 15 shows the infield measured depth versus the reported depth generated by the deep-learning model. A reference tolerance was chosen to be +/- 10% and 92.8% of the points sit within this boundary.

Figure 15: Unity plot comparing the depths of defects found in measured laser data and in the deep-learning model output.

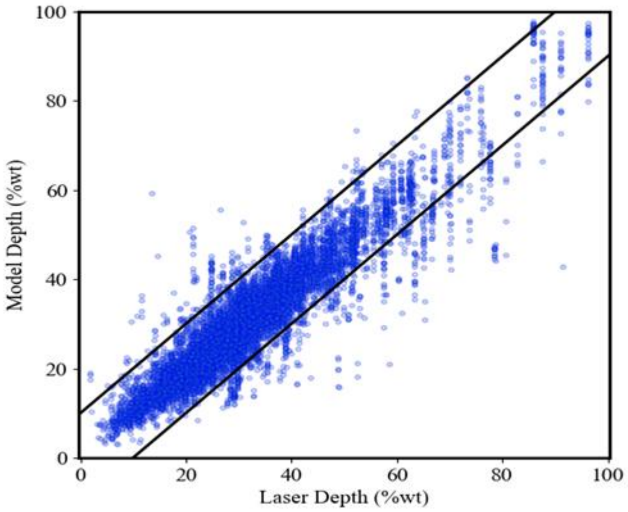
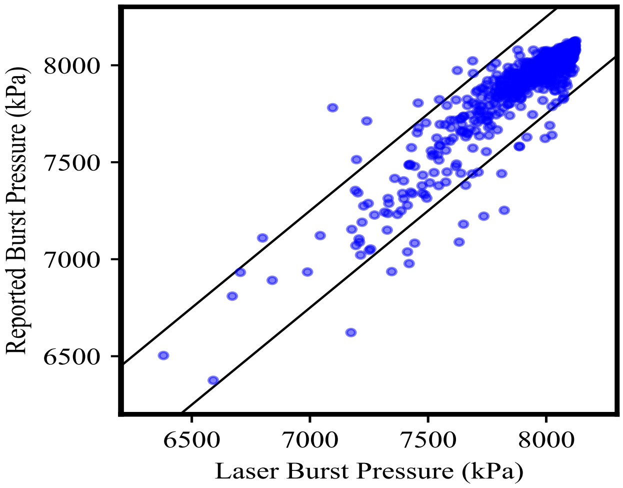


Figure 16 shows the calculated burst pressures for the infield measured versus the reported burst pressure generated by the deep-learning model. The reference tolerance chosen was +/- 250 kPa and 99.8% of the points sit within this boundary.

Figure 16: Unity plot of burst pressures determined from the measured laser profile and the reported defects from the current sizing method.



Defect Specific Individual Tolerances

Due to varying exposure to moisture and oxygen, corrosion is naturally irregular. Therefore, there are significantly contributing factors which feed into how accurate ILI tools can predict depth tolerance. In simple terms, all pits will not have the same tolerance, However the pipeline inspection industry has used length and width dimensions to classify all defects into the 7 POF categories and associated tolerances to each category.

Consider the simple question: “Will a pinhole be more accurately sized if it is isolated within a spool or within a large area of corrosion?” The answer is obvious; however, even this simple consideration is not reflected in tolerances.

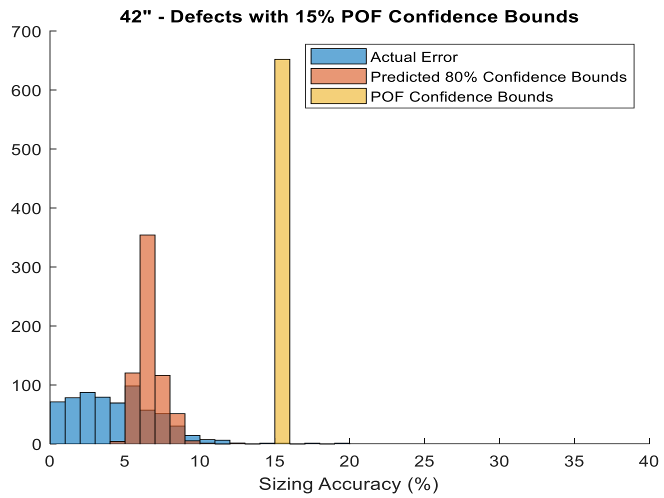
Therefore, other factors need to be considered:

- Depth
- Is the defect within another defect (e.g., pit in axial slot)
- Neighbouring defects
- Number of neighbouring defects
- Shape of neighbouring defects
- Depth of neighbouring defects
- Proximity to girth weld
- Proximity to wall thickness changes
- Proximity to field changes (casings, sleeves, etc.)
- Etc.

Baker Hughes has developed Artificial Intelligence models that read in the MFL triaxial signal data and analyse multiple factors to produce **Defect Specific Tolerances** for each defect within the pipeline. This is applied to each sub-defect within a cluster (grouped corrosion defects) and allows a focussed driven integrity and dig programme. This model allows for each defect (cluster and individual parts) to have its own tolerance based multiple contributing factors, thus making the management of corrosion threats more effective.

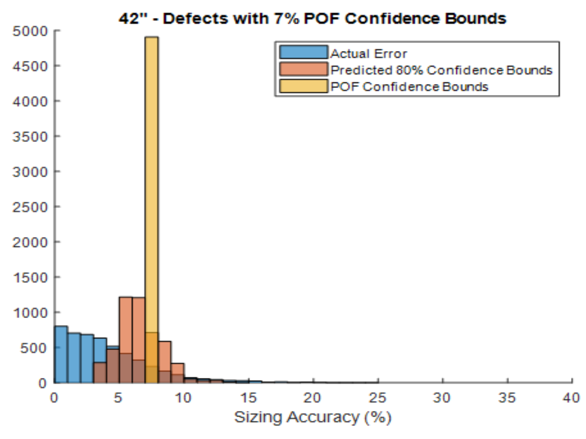
In a real-life example, the histogram in Figure 17, shows a specification +/-15% for pinholes versus the actual sizing specification achieved from LaserScan data (the blue bars). This highlights that the actual sizing tolerance on this pipeline for pinholes was significantly greater than the published +/-15%. This demonstrates that utilising the standard specification on these pinholes would lead to significant conservatism. The "Defect Specific tolerances" (the orange bars) show the models output for tolerances. Utilising this would be more representative and remove the conservatism built into the traditional static POF specifications.

Figure 17: Defect Specific Tolerance versus LaserScan for Pinholes



Similarly in Figure 18, the chart shows the different sizing tolerances for pits. It should be noted that there are some defects where the defect specific tolerance prediction is greater than the published specification of +/-8%. This might initially be interpreted as a negative outcome of the model However, this highlights to the ILI operator that this defect has a wider tolerance due to it potentially being in a very complex area of corrosion interacting with different components. Therefore, a more conservative tolerance should be applied into the integrity management of the pipeline resulting in greater safety.

Figure 18: Defect Specific Tolerance versus LaserScan for pits

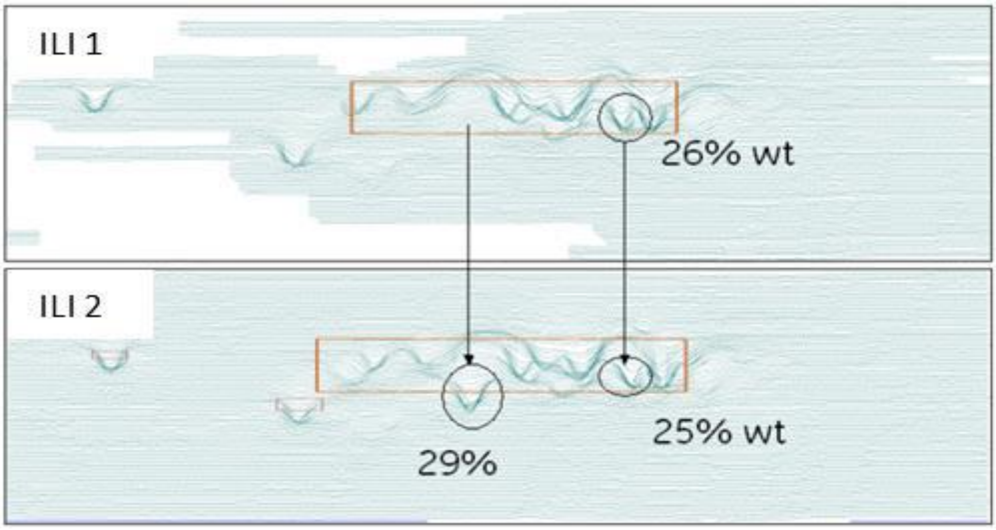


In a test case with a pipeline operator, it was determined that by applying defect specific tolerances, across a series of pipelines, there was the potential to reduce scheduled digs by up to 62% (Ref 7) and extend re-inspection intervals while maintaining safety.

Statistical Dynamic Growth – RunCom™ Cluster Growth 3D

The application of corrosion growth rates to pipelines to assess their future integrity in terms of the number of repairs required versus time allows for the reinspection interval to be calculated.
During a signal-to-signal comparison, it is clear that corrosion growth rates within an area of corrosion (cluster) will have significantly different growth rates. In the example below, accelerated depth growth of one defect compared to the other defects within the feature is shown.

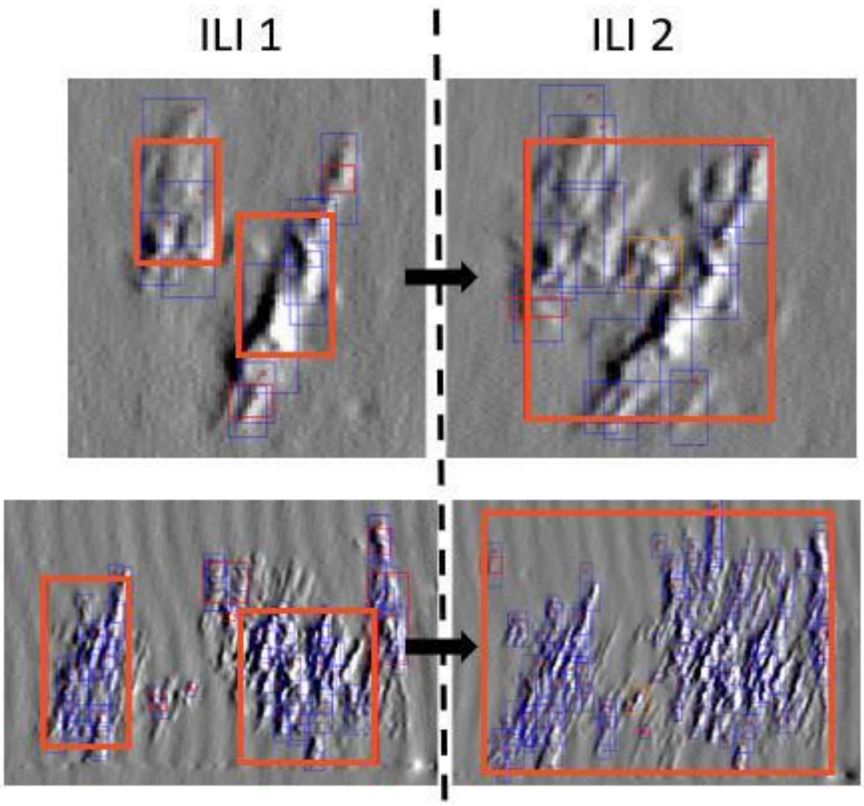
Figure 19: Magnetic ILI repeat inspection data showing new feature with fastest growth



Growth rate ILI based assessments only account for growth in the depth dimension, ignoring length and width growth, along with probably the most important aspect of all: “the potential interactions between adjacent corrosion areas over time”.

Figure 20 shows that often cluster count and dimensions can change dramatically between inspections. When new (often low-level) features form between existing defects or the peripheral edge corrosion of one area (cluster) extends, then very rapidly two or more low priority defects can combine into one critical defect.

Figure 20: Magnetic ILI repeat inspection data demonstrating the increased cluster defect areas due to interconnecting new corrosion features

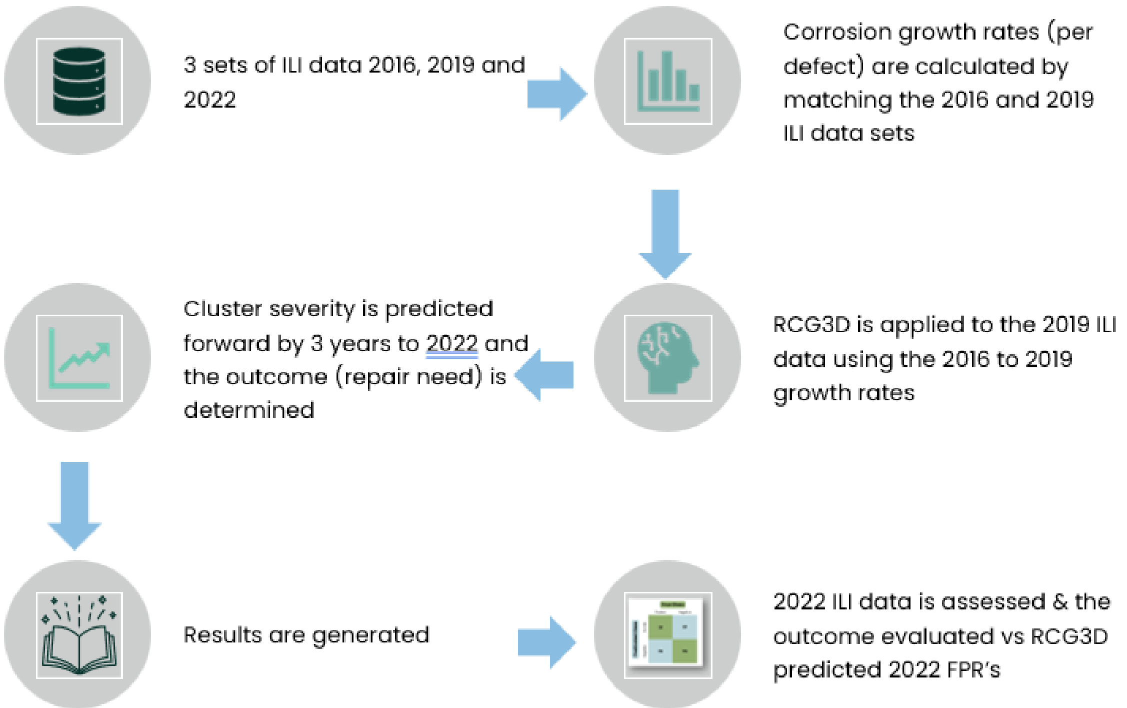


With the historical data collected by Baker Hughes over the vast number of repeat MFL inspections, this data can be used to train machine learning models to build statistical models of growth. These models were trained across thousands of corrosion areas in multiple pipelines with varied initiation patterns of corrosion, combining the spatial relationship between new and existing corrosion sites, including subtle indications of low-level corrosion between features which were often well below the reporting specification.

By looking at the pattern changes of corrosion between runs and the growth of the existing and new features, then statistical probability models could be generated.

With the advantage of many pipelines having repeat inspections over multiple runs and years (more than two), Baker Hughes was then able to validate the statistical projection of new features and compare it to the actual 3rd or 4th inspection run. This allowed for the validation of not only quantitative growth, but also the prediction of qualitive projections of new features and growth in length and width. An example of this approach is presented in Figure 21.

Figure 21: Validation cycle from RunCom Cluster Growth 3D



The ability to predict the future state of clusters and how they not only grow in the future but start to interact and anticipate where new interlinking corrosion will likely initiate is a significant advancement to a well-established signal comparison methodology that has been demonstrated to result in better corrosion rate prediction and will lead to more effective future integrity assessment.

What Does Good Performance Look Like?

To assess the accuracy of MFL inspections, operators and vendors typically use a Depth (or burst pressure) Unity Plot, which compares the infield data with the reported data from in-line inspection (ILI) tools. Unity plots do not fully capture the complexities of real-world corrosion, where both safety outliers and resource outliers need to be carefully managed. As discussed in Section 1, the performance on a unity plot is linked to the standard ILI specifications; however, it does not look at the complex real-world situations found with corrosion on a pipeline.

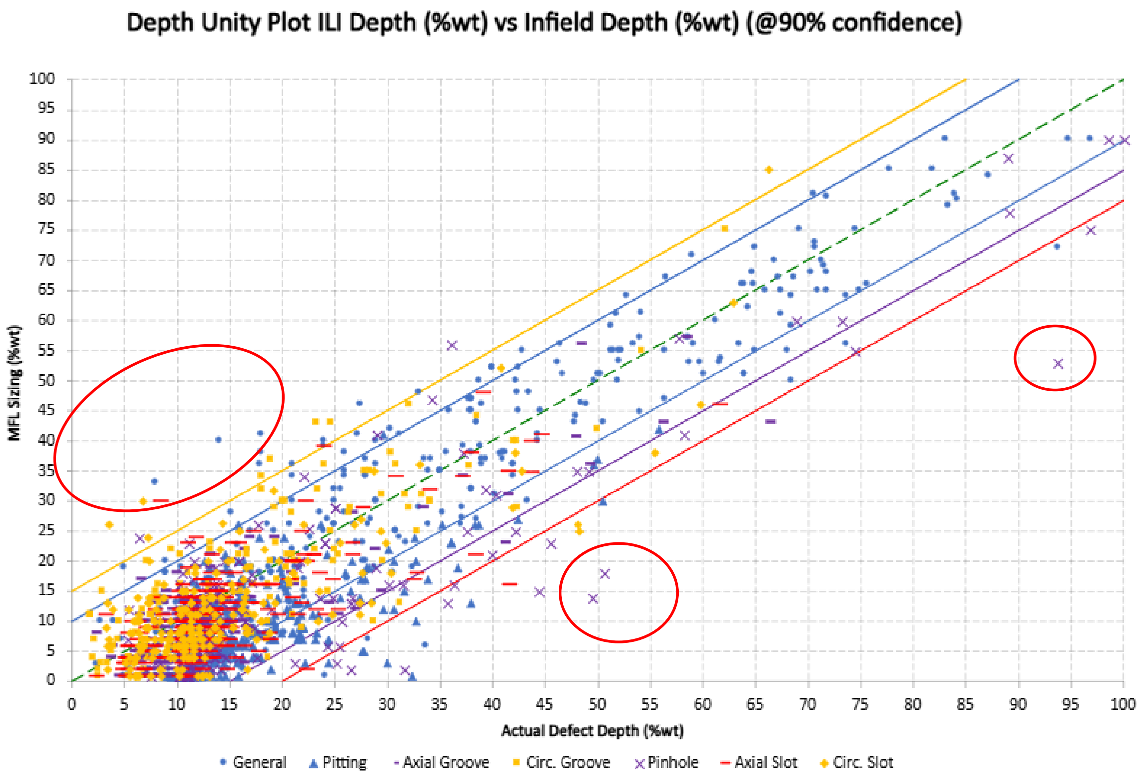
Both operators and ILI vendors need to ask the question: *"If the population is 98% within specification but there is a leak on the pipeline, is that really an acceptable outcome?"*

Pipeline operators are focused on ensuring zero failures with cost effective repair programmes. To do this they need to minimise both:

- Safety outliers – under-calling defect significance (potential of leak/rupture)
- Resource outliers – over-calling defect significance (unnecessary digs)

In the example below the overall statistical performance to the contract specification is achieved (Figure 22).

Figure 22: MFL Depth Unity Plot



All operators would agree that even though the contract specification was achieved, they would have concerns about the range of outliers (in red). Therefore, a better way to describe acceptable performance would be:

- The general population must be statistically significantly correct.
- The number of outliers and their extent is limited.

Even if statistical performance on a unity plot meets specification, the presence of outliers—whether underestimating the significance of defects or overestimating their severity—can still pose risks to safety or unnecessary cost. As Figure 22 shows, while the overall statistical performance may be within specification, the outliers (highlighted in red) suggest areas for further improvement.

Establishing Performance Indicators to Drive Up Measurement Accuracy

To ensure that data driven performance improvements are proactively made, Baker Hughes has implemented an internal key performance indicator "Data Accuracy Score" (DAS) that focuses directly on what matters to the users of the inspection technology.

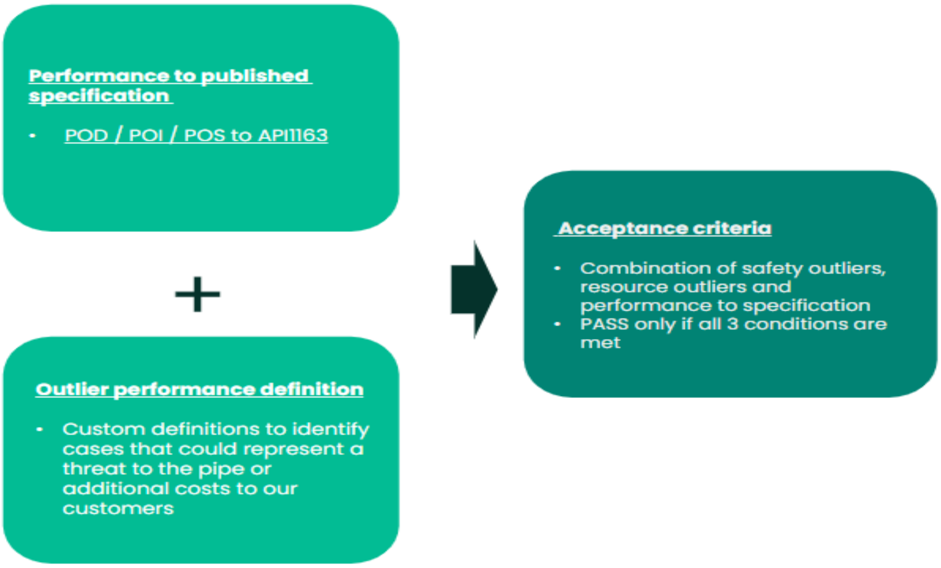
The Data Accuracy Score is a measure believed to be more indicative of when inspection dig results are truly meeting user expectation. It considers two performance criteria, "Performance to Specification" and "Outlier Performance". It is not considered good enough to simply meet published specifications.

This is key, because it removes the effects of favourable statistics and highlights unwanted outlier performance that may not present a significant customer concern but to the ILI vendor is an opportunity for improvement.

The principle is the same for all technologies, but this paper specifically focusses on MFL data (Figure 23).

Figure 23: Baker Hughes Data Accuracy Score Criteria.

The Data Accuracy Score combines the Specification Performance with the Outlier Performance to give an overall score.



1) "Performance to specification" is calculated by the API 1163 Level 2 assessment methodology.
    - Calculation of standard published metrics (POD, POI, POS).
    - Specification verification.
2) "Outlier performance" is calculated based on prescribed definitions of Safety and Resource outliers that were determined based on client feedback and to identify outliers (previously not considered) that could represent either a threat to the pipe or additional cost to our clients.
    - Independent Pass/Fail for Safety and Resource Outlier is calculated.

The advantages of the adoption of the DAS metric are two-fold:

- Firstly, from a customer's perspective there is an assurance that performance is being monitored from a perspective of **"what matters"** and not just contractual obligation.

- Secondly, it provides a means to promote dig feedback discussion, assess and trend performance and identify common areas for targeted process, algorithm, or technology refinement.

Conclusions

This paper emphasises the limitations of simplified specifications in pipeline inspection and highlights the potential of advanced technologies to overcome these challenges. By leveraging triaxial Magnetic Flux Leakage (MFL) measurement and AI-driven analysis, operators can adopt a more data-driven approach and achieve improved accuracy and less conservative integrity assessments.

By combining the traditional statistical population performance and Safety/Resource outlier statistics into a more meaningful performance indicator operators can be confident that ILI vendors are committed to continuous improvement and all operators to drive more reliable and efficient pipeline management.

Improving technology and techniques through a continuous feedback process that integrates data from inspections, analysis, and field digs is essential, but more importantly that monitoring and proactively targeting data led improvement areas is a way of life for the ILI vendors and an expectation of the users of the technology.

References

1. POF 100 Specifications and requirements for in-line inspection of pipelines, Standard Practice, November 2021.
2. American Petroleum Institute, In-Line Inspection Systems Qualification, API STD 1163, Second Edition, April 2013, reaffirmed August 2018.
3. Anthony Tindall, David Classen, Identifying and Characterizing Corrosion Morphologies More Effectively with Triaxial MFL Signal Analysis, Pipeline Pigging & Integrity Management Conference, 2025.
4. Jane Dawson, Steve Farnie, Accounting for Corrosion Growth and Interaction in Future Severity, International Pipeline Conference, 2024.
5. Steve Farnie, Anthony Tindall, Addressing complex and axially aligned Corrosion with MFL technologies and benefits of triaxial methods, International Petroleum Conference, 2024.
6. Scott Miller, Dustin Mattio, James Rodway, The application of deep learning and computer vision techniques to magnetic flux leakage data to accurately detect and predict corrosion profiles, International Petroleum Conference, 2024.
7. Melissa Gurney, Scott Miller, Samaneh Sadeghi, Geoff Hurd Baker Hughes, Keila Caridad, Terry Huang,Vincent Tse, Billy Zhang, TC Energy, Calgary, AB, Canada Application of Advanced Data Analytics to Generate MFL Anomaly Specific Sizing Tolerance Specification PRCI-REX2022-093.
8. Keila Caridad, Terry Huang, Vincent Tse, Billy Zhang, TC Energy; Melissa Gurney, Geoff Hurd, Scott Miller, Baker Hughes, Exceeding Performance Standards of Pipeline Inspection and Integrity with Advanced Data Analytics, Pipeline Pigging & Integrity Management Conference, 2022.
9. Martin Bluck, Justin Pearce, Jeff Sutherland, Eric Quick, Validation of Latest Generation MFL In-Line-Inspection Technology Leads to Improved Detection and Sizing Specification for Pinholes, Pitting, Axial Grooving and Axial Slotting, International Petroleum Conference, 2010.
10. Cassidy Ryan, Anthony Tindall, Optimising Magnetic Flux Leakage Inspection Accuracy through the application of AI techniques and Ongoing Performance Tracking, Pipeline Technology Conference, 2025.