How Automated Data Normalization & Alignment Helps Operators Make Better Decisions

Presented at DPSF on May 18-19, 2022

by Jordan Dubuc, CTO
Colin Scott PhD, Pipeline Integrity Engineer

Abstract

Inline Inspection (ILI) reports are the basis for much integrity work, providir
pipeline at one point in time. While some tasks can be accomplished with jus
information can be gained by a detailed alignment of current and past ILIs. Today's most advanced alignment software will align a full history of ILIs file to each other automatically, compensating for any repair sections, routing changes, or changes in flow direction. Additionally, complete pit-to-pit matching of every single anomaly call provides a detailed history of each defect on a pipe. With this complete alignment and matching, whole layers of information and error correction are possible instantaneously that would have been impossible in the past. Growth or apparent nucleation trends can be examined in granular detail across multiple ILIs instead of just the coarse difference between two ILIs in an area. Complete, automated alignment across multiple ILIs extracts much more information out of users' existing ILIs than has previously been possible.

Introduction

A modern pipeline integrity data management and analysis platform can support the entire integrity management business process and lifecycle. The process begins with ingesting and normalizing all the available data through a seri LEAVE THIS PAGE BLANK rning classification models. This means the ability to elevate unst clean, structured dataset to support cognitive learning and analysis. These machine learning models can be trained to understand integrity datasets like ILI and can interpret a wide variety of different reporting formats, naming schemes, and vendor conventions and help bring volumes of historical data into a standardized data structure. From there, APIs or interfaces that can synchronize data from existing systems such as repair / NDE data, and GIS, PODS, or other asset management systems, support data integration and enable improved decision-making by leveraging more of the available data. This can also help with data verification, highlighting where different data sources are inconsistent with each other.

Core datasets like the inline inspection vendor report or "ILI tally" information can be statistically validated and automatically aligned against other linear and spatial datasets, including the complete history of every reported feature or anomaly across many different inline inspection tool technologies.

This allows development of a comprehensive model of the condition of the asset over time and ultimately to determine a corrosion growth rate for every anomaly in order to project a remaining life based on wall loss or burst pressure. Interacting threats can also be identified by aligning anomalies from different inline inspection tool technologies and incorporating data from other systems such as PODS or other GIS.

All of this integrated data can then drive predictive analytics and business intelligence and supports the execution of the integrity management lifecycle from assessment planning and tracking, to integrity compliance and dig selection, dig program planning, producing dig sheets and packages for the field, then finally to threat monitoring, planning the next re-assessment & measuring of the effectiveness of the program.

In this work, we illustrate how pipeline operators can create value and confidence in their integrity programs through data ingestion, alignment, and analytics. Better understanding of data leads to better decision making [1].

Data Ingestion and Normalization

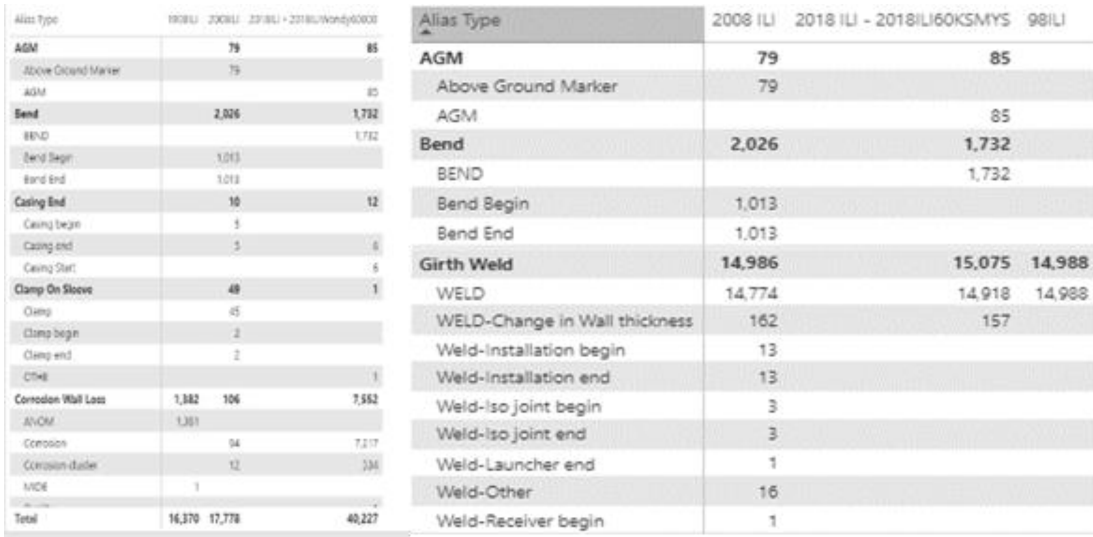
The first area where machine learning can be applied is in the ingestion and normalization of pipeline integrity data from many different types of vendor reports. The method used in this work is based on a series of Bayesian classification models trained on over 5,000 ILI pipe tallies with hundreds of different formats and over 40 million reported anomalies completed in proprietary software, Cognitive Integrity Management. This allows the ingestion process to accept a wide variety of different vendor report structures without having to manually format them into a template and to accurately identify and classify anomalies based on how they've been described by vendors. Pipeline operators and users of the platform are not limited to a single vendor or tool type.

One specific example of how machine learning can be used during this process is in interpreting all the different ways that ILI vendors can describe different types of pipeline features and anomalies. A feature classifier looks at all the available data in the pipe tally, the anomaly type or description and any additional comments provided by the vendor as well as the associated attributes of each anomaly to determine what is being described by the vendor and how it should be classified. This enables operators to take a wide variety of historical ILI reports, each with different standards and conventions into an apples-to-apples structure which then allows for more advanced data analysis and business intelligence.

Operators are often challenged with historical data which may not be organized or consistent enough to take advantage of an approach such as this, and so another critical part of this ingestion process is data validation, with hundreds of data quality checks performed as part of ingestion to help highlight gaps or inconsistencies in the data before moving on to the analysis process. With a machine learning based data ingestion process, operators can much more quickly and easily take unstructured or semi-structured data from a variety of different sources or spread across disconnected systems or thousands of Excel spreadsheets into a standardized integrity data management system.

Data such as inline inspection is loaded through the machine learning ingestion and classification process which understands the formats of different ILI tally files and the way in which ILI vendors describe the assortment of anomalies and pipeline features through their types, descriptions, additional vendor comments and attributes – position, length, width, depth, orientation, and so on.

One can also look at the number of features reported in each assessment and how the machine learning based classification identified each of them and categorized them into a standard taxonomy. For example, as illustrated in *Figure 1*, there are multiple ways that vendors can describe a given feature and the system can group all of these into one classification system for analysis. This can provide an apples-to-apples understanding of what has been reported over time on this line, what has changed, and whether a given inline inspection report is fit for use in further analysis.



Alias Type	[illegible]	2008ILI	[illegible]
AGM		**79**	**85**
Above Ground Marker		79	
AGM			85
Bend		**2,026**	**1,732**
BEND			1,732
Bend Begin		1,013	
Bend End		1,013	
Casing End		**10**	**12**
Casing begin		5	
Casing end		5	6
Casing Start			6
Clamp On Sleeve		**49**	**1**
Clamp		45	
Clamp begin		2	
Clamp end		2	
CTMB			1
Corrosion Wall Loss	**1,382**	**106**	**7,552**
ANOM	1,381		
Corrosion		94	7,217
Corrosion cluster		12	334
MDE	1		
Total	16,370	17,778	40,227

Alias Type	2008 ILI	2018 ILI - 2018ILI60KSMYS	98ILI
AGM	**79**	**85**	
Above Ground Marker	79		
AGM		85	
Bend	**2,026**	**1,732**	
BEND		1,732	
Bend Begin	1,013		
Bend End	1,013		
Girth Weld	**14,986**	**15,075**	**14,988**
WELD	14,774	14,918	14,988
WELD-Change in Wall thickness	162	157	
Weld-Installation begin	13		
Weld-Installation end	13		
Weld-Iso joint begin	3		
Weld-Iso joint end	3		
Weld-Launcher end	1		
Weld-Other	16		
Weld-Receiver begin	1		

Figure 1 Bayesian classification of ILI features into a common taxonomy.

Data Alignment

Once data has been ingested into a data model and interpreted into a standardized structure, the next stage is alignment. This can be accomplished via an automated algorithmic alignment of linear data like inline inspection odometer, PODS linear referencing, or pipeline stationing information and spatial data such as close interval survey. Alignment enables data integration across these different datasets and comparison of individual anomalies over their complete measurement history.

A modern digital integrity platform can provide the ability to align data from inline inspection, including all historical logs, as well as GIS, close interval survey, and historical repair records. Unlike manual or semi-automated alignment methods which often filter down and align just a candidate set of the deepest metal loss, such an approach can align each and every anomaly and provide a discrete corrosion growth rate for each based on that alignment. Matching every anomaly is critical for identifying shallow but fast-growing corrosion, and as part of optimizing efficiency with long-range planning and forecasting.

Alignment algorithms can use distinctive patterns in joint lengths and anomaly geometry to isolate and resolve the drift in tool odometer measurements and the roll of the tool that results in orientation offset. This ultimately allows the alignment algorithm to determine the appropriate alignment and pit-to-pit matching of individual anomalies across their complete reported history.

Through this alignment process, each joint of pipe can be identified and maintained as part of a master joint listing, and this provides complete traceability for every joint in the system. This also includes the ability to automatically detect flow direction and align datasets where for example the tool was run in the opposite direction (as seen in *Figure 2*), as well as to handle re-routes and changes to line configuration. Such algorithms are able to identify the common portions of the pipeline even in cases when ILI tools are launched from different locations or when traps have been added or removed, as long as there is enough distinctive pattern in the remaining joint lengths to identify the common segments.

Assessment Name	ILI2010			ILI2020		
Master Joint ID	Joint No.	Log Distance	Joint Length	Joint No.	Log Distance	Joint Length
500,005,700.00	68070	1,972.39	58.56	68070	389,849.05	58.61
500,005,800.00	68060	2,030.95	60.16	68060	389,790.43	60.17
500,005,900.00	68050	2,091.11	57.96	68050	389,730.26	57.96
500,006,000.00	68040	2,149.07	59.73	68040	389,672.31	59.39
500,006,100.00	68030	2,208.80	46.11	68030	389,612.91	46.50
500,006,200.00	68020	2,254.92	59.42	68020	389,566.41	59.55
500,006,300.00	68010	2,314.34	57.44	68010	389,506.86	57.58
500,006,400.00	68000	2,371.78	59.53	68000	389,449.28	59.57
500,006,500.00	67990	2,431.31	59.48	67990	389,389.71	59.56
500,006,600.00	67980	2,490.79	58.55	67980	389,330.15	58.55
500,006,700.00	67970	2,549.34	57.66	67970	389,271.60	57.74
500,006,800.00	67960	2,607.00	57.69	67960	389,213.86	57.74
500,006,900.00	67950	2,664.69	57.50	67950	389,156.12	57.74

Figure 2 Weld alignment for a flow reversal scenario.

Data Integration

Each of the previous steps, data ingestion and normalization, followed by alignment of each of those datasets, is ultimately pressing toward data integration. Data integration means putting all of the relevant data with respect to integrity decision-making right at the fingertips of the integrity analyst. By combining data from multiple ILI tool runs including different tool technologies, GIS, CP data like close interval survey, NDE and historical repair data and others this allows integrity specialists to assemble the broader context from each of these individual pieces of data.

The first opportunity is in a deeper understanding of corrosion growth, enabled by aligning each metal loss anomaly from every ILI tally to its complete reported history and accounting for ILI measurement error and bias. That can then be correlated with close interval potential survey, for example where ILI measurements indicate external corrosion growth in areas with potentially insufficient cathodic protection. We can then bring in coating information from the GIS, soil characteristics, historical digs and field measurements in that region, and the picture as to what's happening at those locations becomes even more clear. These are just some examples of how data integration can support integrity data analysis and more informed decisions.

Data integration can bring practical benefits for pipeline operators in the form of improved integrity decision-making, predictive maintenance and forecasting, and increased efficiency across their entire integrity business process. Operators also benefit from business optimization and risk management as incorporating more data can help provide context and support around integrity decisions.

Operator involvement has proven that this approach can provide cost savings, improved performance of the dig program, and reduced overall risk in their integrity management program by leveraging more of the available data in making those integrity decisions. Modern analytics tools based on data science and machine learning help integrity teams shift from a prescriptive integrity program to a performance-based program, to ultimately select fewer ineffective digs and to identify threats and defects that can otherwise go undetected by legacy analysis methods and tools.

Through shared learning across the industry, vast libraries of integrity criteria, operator best-practices, algorithms and machine learning models become even more accurate and effective as more operators embrace this digital transformation. This can be done without sharing any individual operator's sensitive data. The learning, knowledge, and advancement can be shared while the underlying integrity data remains private and secure in the cloud.

Ultimately, the goal is to accelerate the industry's push toward zero pipeline failures by leveraging more of the available data, by applying the unprecedented scale and resources of cloud computing, and by bringing modern data science analysis and machine learning technologies to the world of integrity management.

Data Analytics

Inline Inspection Data

The first analysis that many pipeline operators would perform upon completion of an inline inspection tool run would be to align reported features with data from previous tool runs. *Figure 3* provides examples of multiple aligned data on a "demonstration" pipeline. Analysts can review areas of concern along the pipeline and see how pits or pit clustering have changed over the years between inspections. This visual check confirms the platforms algorithms are performing correctly and helps the analyst understand the nature of the corrosion threat acting on the pipeline.

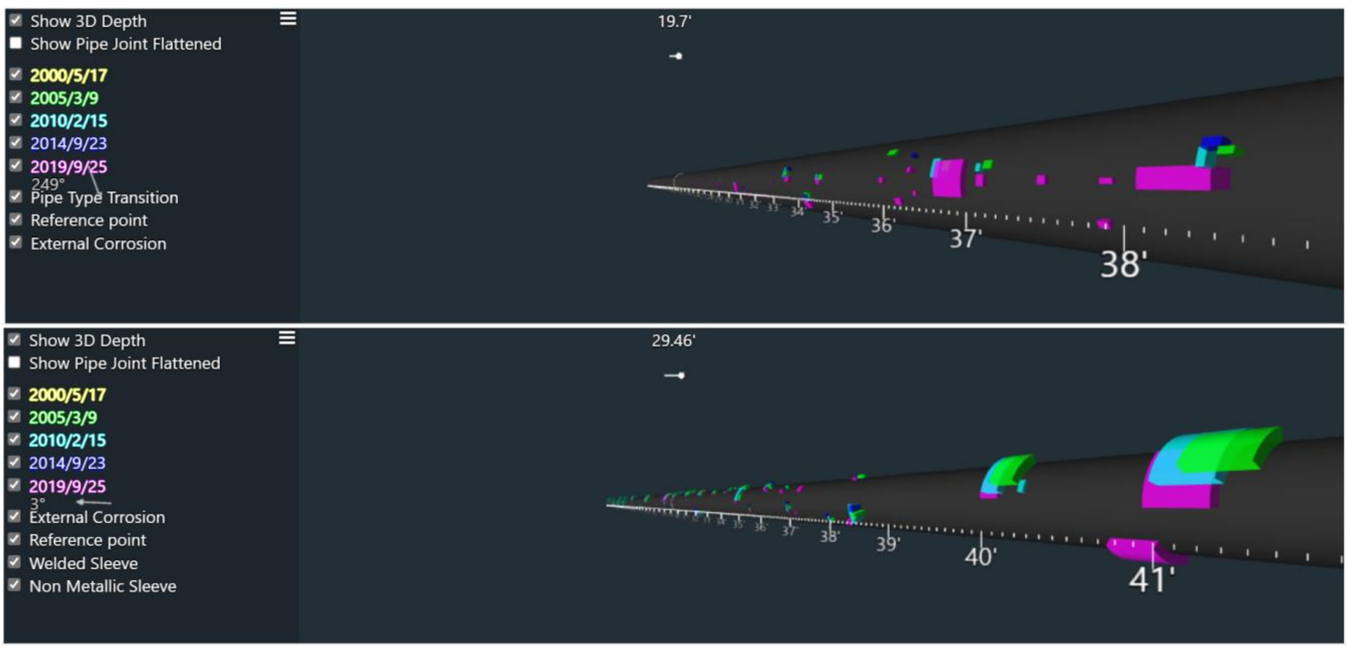


Figure 3 Examples of pit-to-pit matched ILI reported corrosion depth data used to confirm pit clustering.

Well-aligned data can also be analyzed on a pit-by-pit basis. Reviewing reported ILI feature depths over multiple tool runs can provide increased confidence in the ongoing analysis and highlight problem areas for further study. Corrosion growth rate estimates can be calculated using the techniques described by recent PRCI studies [2]. More advanced analyses allow growth rates to be calculated from multiple depth measurements using linear regression analysis [3].

Figure 4 provides examples of pit depth measurements aligned over multiple tool runs, again from our demonstration pipeline. The variation in depth measurements implying either metal loss or gain remind analysts of the depth sizing error implicit in ILI tools and the need for caution in interpreting inspection results. Increasing the number of inspection data sets provides more accurate growth rate estimates. The plus or minus errors from each depth measurement offset each other and narrow the precision of the estimates. Given the known depth error, analysts should consider their estimates probabilistically. Various methodologies, based on either probability integration or Monte Carlo simulations, demonstrate the value of increasing the number of datasets included in these analyses [3].

Figure 5 provides a comparison of two complete ILI data sets. It allows analysts to identify differences in reported depths at a high level. Data sets may be consistent or inconsistent from run to run. Analysts must consider that depth sizing errors come into play and any conclusions drawn from these unity plots are supported by review of all available data. In the example below, it could be concluded that the tool runs were consistent in their accuracy, and there was no overall corrosion growth. However, it is also possible that significant corrosion growth was masked by an equal and offsetting depth sizing error.

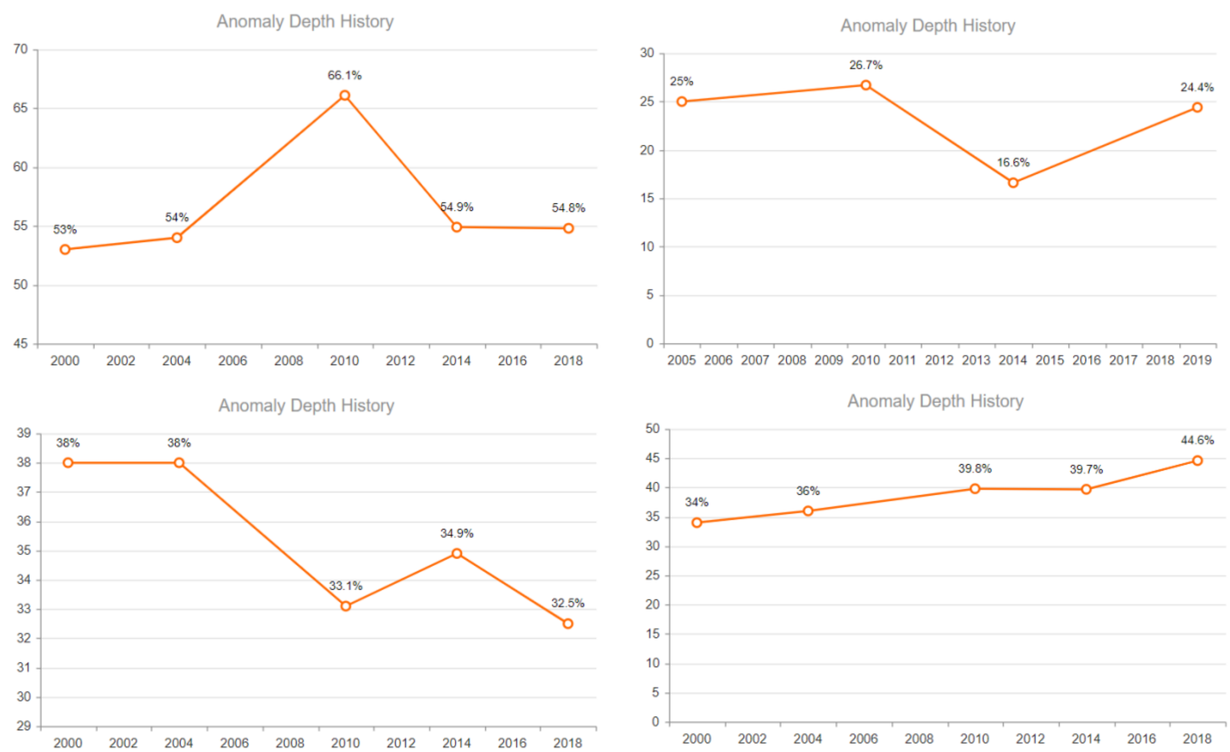


Figure 4 Examples of pit-to-pit matched ILI reported corrosion depth data used to estimate corrosion growth rates.

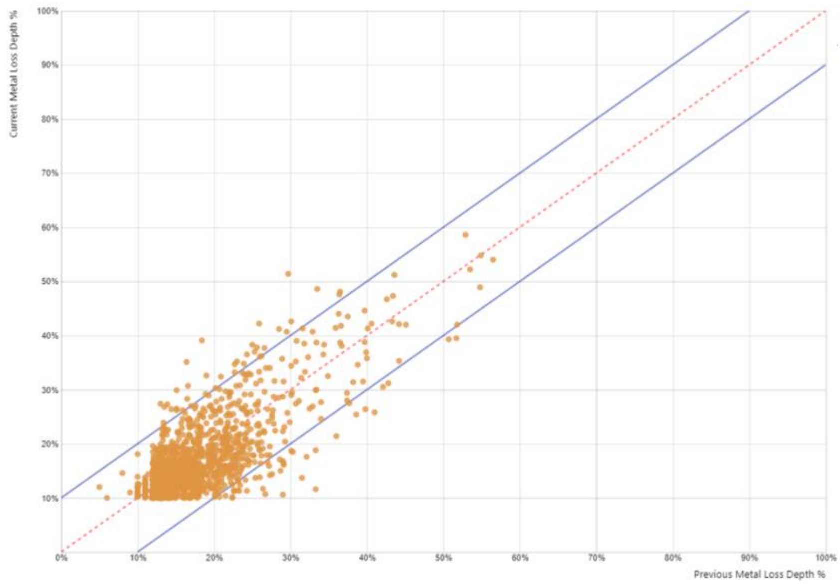
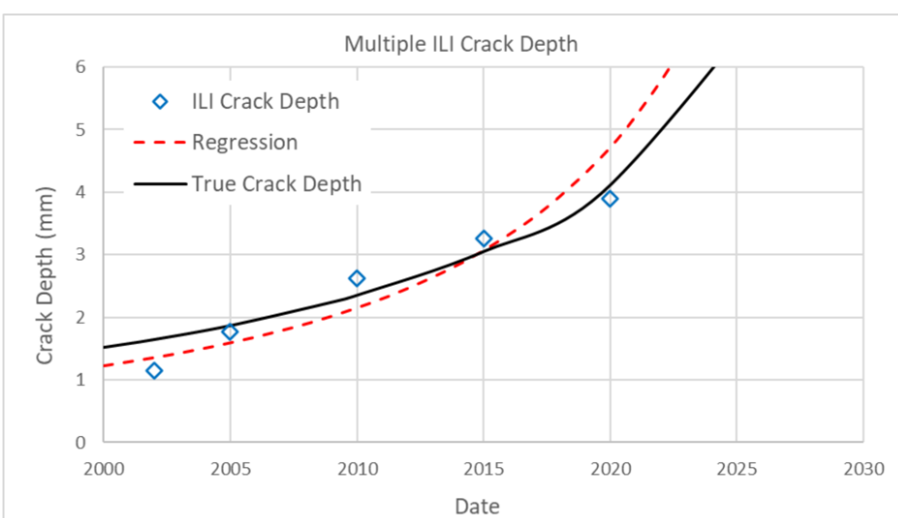


Figure 5 Comparison of two ILI data sets to demonstrate run consistency or implied corrosion growth.

Many pipeline operators will be familiar with analysis of pit-to-pit corrosion growth. These analyses are relatively straight forward, as the growth is assumed to be approximately linear. However, not all flaw growth is linear. Fatigue crack growth is generally assumed to follow Paris Law, and the growth is a function of the square root of the crack depth to an exponent, typically three. If two ILI data sets are aligned, and the value of Paris Law exponent is assumed, it becomes possible to estimate remaining lives for cracks, in the same way that is currently done for corrosion. If three or more ILI datasets are available and aligned, it also becomes possible to estimate the Paris Law exponent from the ILI depth data.

Figure 6 illustrates how these analyses can be performed using simulated data. In one case, a Paris Law exponent is assumed. In the other, it is determined by regression of the data. Reasonably accurate remaining life predictions are possible even if there are sizing errors in the depth measurements, provided there are multiple data points to offset and mitigate the problem. In principle, these techniques can also be used on strain data to determine the behavior of geohazards.



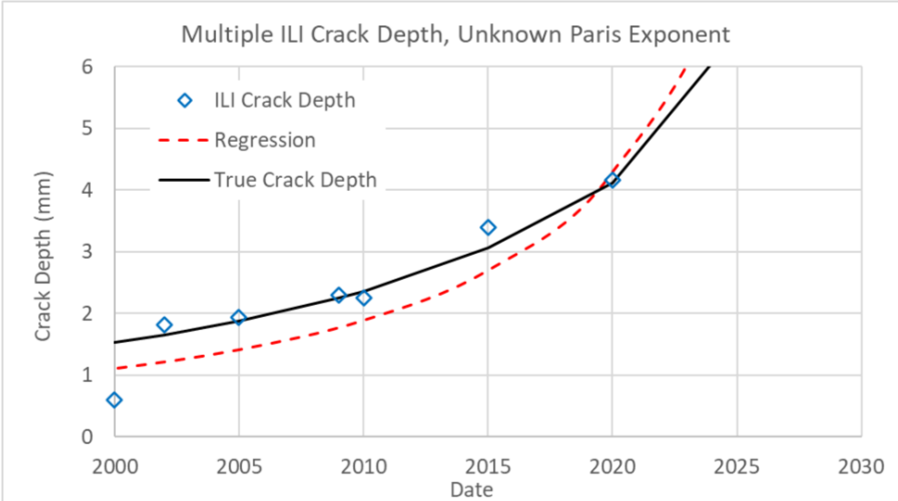


Figure 6 Analysis of non-linear flaw growth is possible if multiple data sets are available.

Field Evaluation Data

The discussion above has focused on ILI data, the core of many pipeline operators' integrity programs. A significant amount of information can be inferred from these data. However, field evaluations are typically required to confirm ILI detection and sizing. If ILI data is biased, then corrections can be made to improve fitness for service analyses. This is done using unity plots of the ILI and field data.

Most operators will perform a dig program following receipt and analysis of their ILI data. The flaws are analyzed and their current and future risk to the pipeline evaluated. Flaws of concern are excavated and measured in detail by technicians before being repaired. The field technicians' data is aligned with the ILI data and any statistical adjustments necessary are applied to the ILI data. This process typically takes several months to produce sufficient field data to validate the ILI data. However, if advanced algorithms are available, current ILI data can be aligned with field data from previous dig programs, provided some flaws remain in the pipeline due to recoat or sleeve. This allows analysts to be proactive. Preliminary analysis of past dig data allows operators to decrease their ongoing dig budgets by leveraging data already available in the database.

Figure 7 illustrates the use of historical unity plots. In these cases, two different ILI data sets are compared to the same field data. The ILI data sets are shown to be non-conservative relative to the field data, allowing analysts to compensate for the tool sizing error prior to full analysis of the pipeline and launching of a dig program.

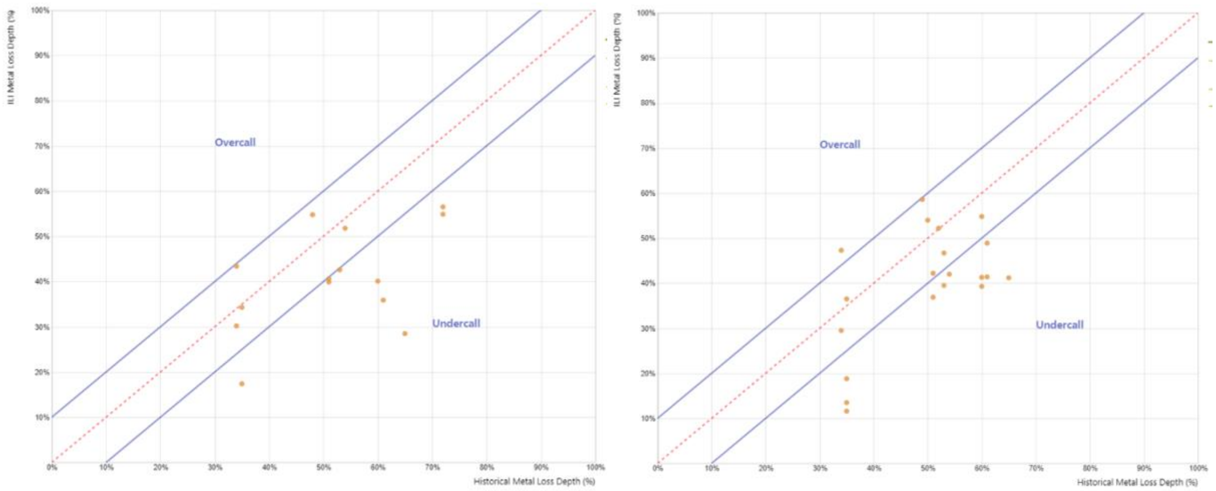


Figure 7 The use of historical field data sets allows analysts to determine ILI run quality prior to completing a dig program.

Additional Data Sets

Additional data sets become of increasing value if ingestion and alignment are efficient. Data importance is typically threat specific, so selections of appropriate data are made on a case-by-case basis. A more efficient alignment allows analysts to review more and more data sets, even if correlations are not expected.

Figure 8 illustrates additional data sets aligned for the analysis of a cathodic protection system. The ON- and instant-OFF potentials are aligned with the calculated corrosion growth rates. Areas with problematic potentials can easily be identified, leading to a more efficient and focussed integrity program.

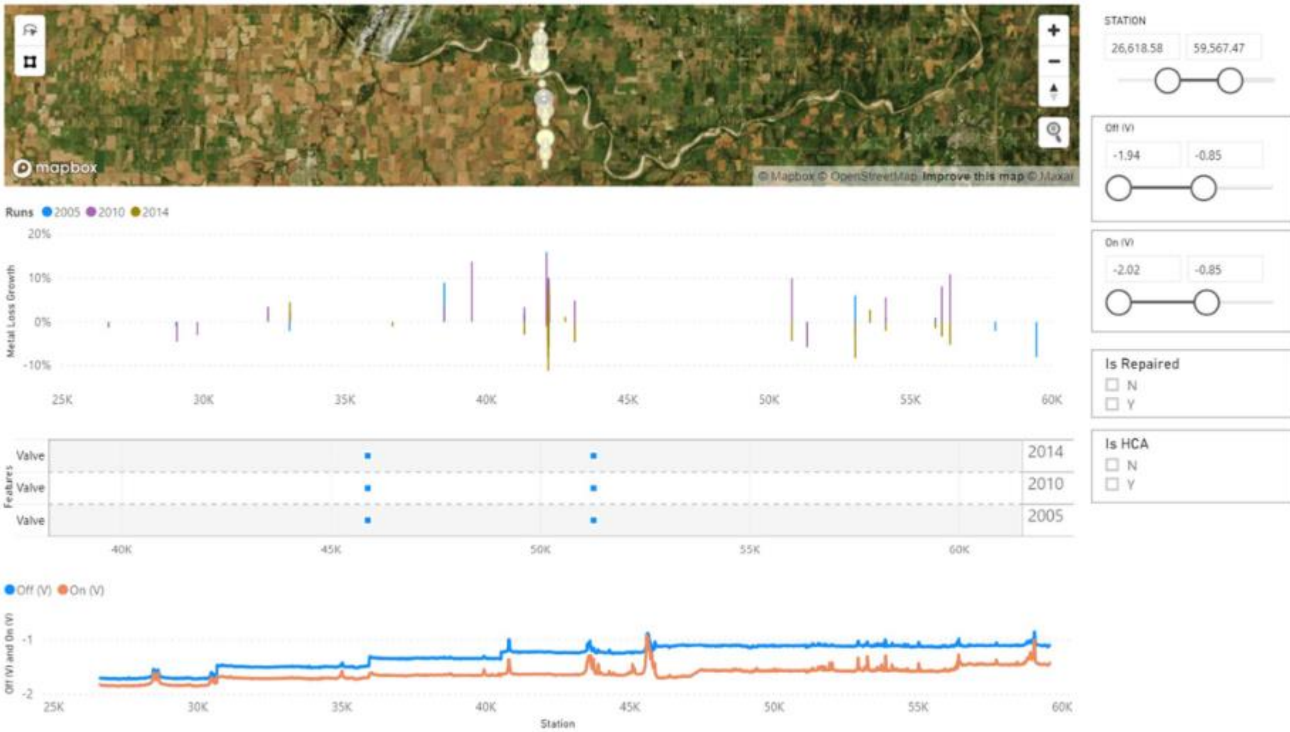


Figure 8 Integration of different inspection results

Figure 9 illustrates the alignment of elevation profiles, in this case the alignment is geographical. Elevation profiles are typically reviewed during internal corrosion analysis. Local low elevations in the pipeline tend to accumulate water and sediment, and these exacerbate internal corrosion issues. Inclination angles can influence inhibitor coverage or corrosion mechanisms associated with multi-phased flow regimes.

Figure 10 illustrates the alignment of a spatial data set, in this case, soils data. The soils database included data on soil type (clay, sand, silt, loam, etc), drainage, slope, electrical conductivity, corrosion susceptibility and mechanical properties. Color-coding helps analysts understand how these properties are distributed along their pipelines. Soil types, drainage, and electrical conductivity are all important to external corrosion studies. These data type can also be aligned with ILI metal loss data and the potentials provided by cathodic protection surveys. Soil topography, slopes and mechanical properties are all important to geohazard studies. These data can be aligned with ILI calliper and strain data sets to help analysts understand local deformations and geohazards.

Conclusion

Advanced ingestion, alignment and integration algorithms provide pipeline operators with increasingly efficient and valuable analyses. Even dissimilar data types can be managed with relative ease. The examples provided illustrate how starting with ILI data sets, operators can include field evaluation and additional data sets to provide more robust analyses that lead to higher confidence in decision making.

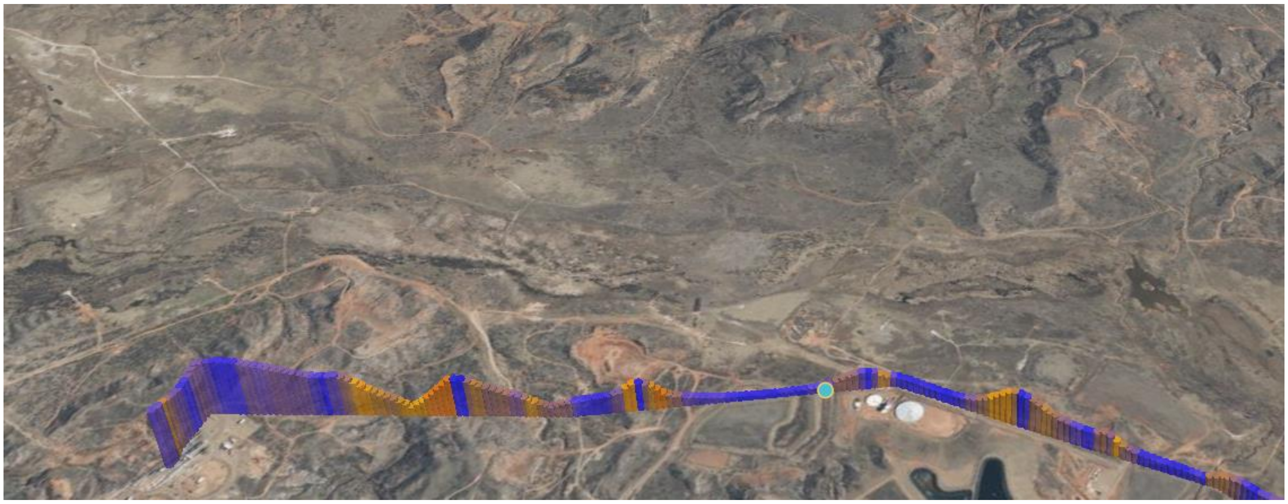


Figure 9 Elevation profiles help analysts understand patterns of internal corrosion in their pipelines.

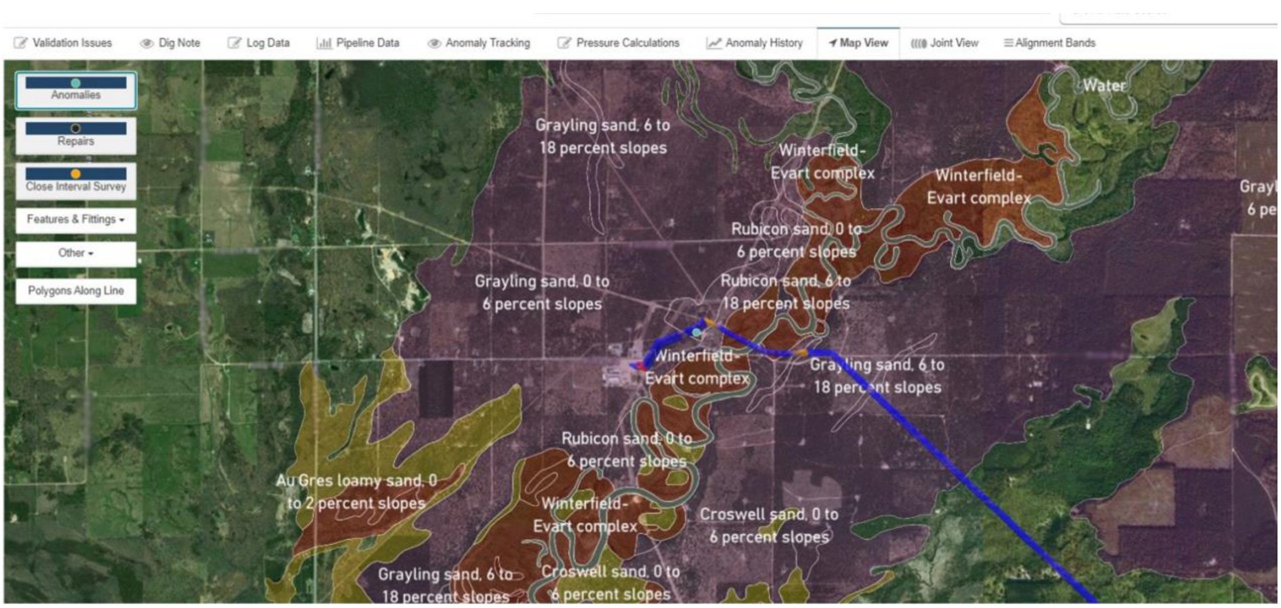


Figure 10 Soil properties mapping help analysts understand patterns of external corrosion or geohazard threats to their pipelines.

References

1. Colin Scott, "The snowball effect of integrity data compilation", World Pipelines Magazine March 2022, Page 13.
2. PR-331-063525-S01, "Development of Detailed Procedures for Comparing Successive ILI Runs to Establish Corrosion Growth Rates", Pipeline Research Council International, July 2009.
3. Scott, C. and Philpott, K., "Probabilistic Flaw Growth Rate Estimates Using Multiple Inline Inspection Tool Run Data Analyses", submitted to ASME for IPC 2022.

May 18-19, 2022

Organized by

CLARION
TECHNICAL CONFERENCES