Engineering Statistical Analysis and Visualization of Excavated Oil and Gas Pipeline Anomaly Depths to Matched Inline Inspection Estimates using Power BI

William V. Harper[1], Tara P. McMahan[1], Benjamin R. Hanna[1], Adriana Nenciu[2]

[1]DNV GL USA, Inc., 5777 Frantz Road, Dublin, OH 43017
[2]Department of Mathematics and Actuarial Science,
Otterbein University, 1 South Grove St, Westerville, OH 43081

Abstract

In-line inspection (ILI) tools are used to help assess areas of a pipeline considered to have the potential to contain critical defects. As the ILI tools move through the inside of the pipe, signals are sent via transducers to the inside pipe wall. These signals help identify different types of anomalies, such as internal/external corrosion, cracks, and manufacturing artifacts. The depth of the anomaly is a key characteristic estimated by the ILI tool that is used in integrity analyses. How much faith one may place in the ILI results is often assessed through validation efforts that compares ILI to field excavation measurements. The data are used in multiple metrics of an integrity analysis, such as whether the ILI tool passes the industry standard API 1163 or the data should be adjusted. Adjusting ILI data can provide an accurate, but perhaps slightly conservative, safety approach to using all of the ILI calls given that only a small subset of field excavations have been performed. This study used representative ILI results and the Power BI data integration software to create flexible interactive analyses (statistical and visual) that rapidly track pipeline integrity metrics.

Key Words: Oil, Gas, Pipeline, visualization, in-line inspection, pig

1. Introduction

Risk management for hazardous liquid and natural gas pipelines requires that the structural integrity is well understood. In-line inspections (ILIs) provide valuable data that are used in integrity management programs to identify potentially problematic sections of a pipeline. As the ILI tools move through the inside of the pipe, different types of anomalies, such as internal/external corrosion, cracks, and manufacturing artifacts are identified from signals sent via transducers to the inside pipe wall. Assessment of the integrity of pipelines using ILI plays a large role in estimating integrity, with anomaly size being fundamentally important in acceptability analyses. Most pipelines are inspected with an ILI tool every five or seven years, depending on the product being transported and the identified threats to the pipeline.

It is important for pipeline operators to be able to assess the performance of a single ILI survey (i.e. how well the characteristics of each anomaly match the true condition of the buried pipeline). One common method to assess tool performance is to excavate selected

portions of the line and make direct measurements of the anomaly to be compared to the measurements and characteristics of the anomaly as reported by the ILI tool. If there is reasonable agreement between the ILI and direct measurements, then the remainder of the ILI data and results are relied upon to make integrity-based decisions. If there is not reasonable agreement, detailed comparison of the ILI and direct measurements is required to understand the adjustments necessary for the ILI results to be able to rely upon them for integrity-based decision making. Microsoft Power BI is a visualization software that can be used to filter these large databases and allow for easy comparison.

DigIt is one such Microsoft Power BI application providing visualization and analytical capabilities. As will be illustrated, a variety of visualizations and statistical analyses are used to provide an overall assessment of the ILI tools' performance in terms of anomaly depth. Various anomaly types may be analyzed by the DigIt application, but the presentation in the paper focuses on metal loss (i.e. corrosion) in an effort to be succinct. DigIt has many commonalities with another Power BI application that has been developed called CGReal Bias. Where DigIt is used to collectively compare anomaly depths across numerous pipe segment lines to assess the overall performance of a specific ILI tool, CGReal Bias focuses generally on just one line and two ILI surveys. Because pipelines are inspected every five or seven years, comparing successive surveys can show if there is any anomaly growth. The first step in this analysis is to ensure that each survey is directly comparable and each ILI tool has reported accurately. CGReal Bias uses whatever direct data is available but is more focused on matched ILI anomalies from the two surveys. Through a statistical process, comparison of the survey and direct measurement results are used to estimate potential bias in one of the ILI runs.

2. DigIt Power BI Visualizations and Filtering

Figure 1 is the Metal Loss page in Power BI for the DigIt application. A key goal underlying DigIt is to provide pipeline operators with both visual and analytical results that assess how well the ILI results match the direct excavation measurements. Power BI provides a framework that allows users to easily filter data and see results dynamically. The key components shown in Figure 1 are discussed in more detail in this paper.
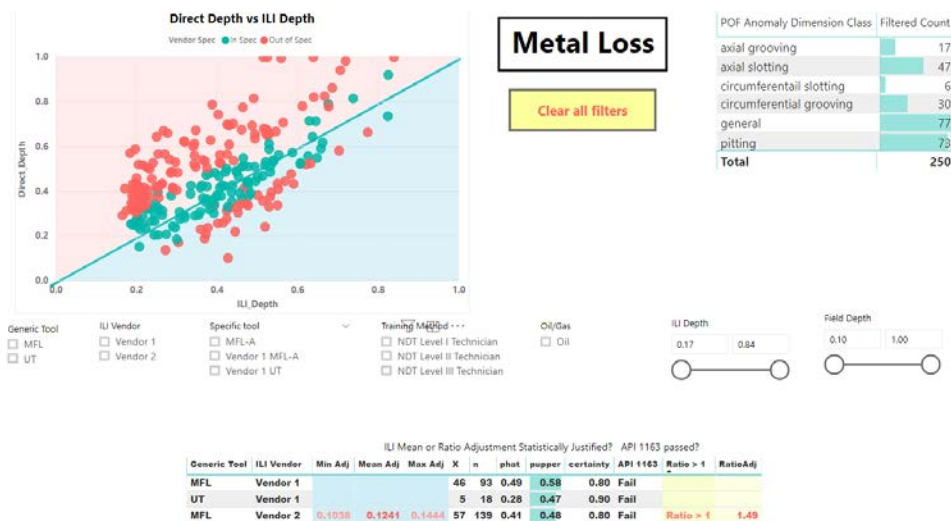

**Figure 1:** DigIt Power BI page focusing on metal loss

2.1 Unity Plot

Typically, to assess ILI tool performance a comparison of the ILI reported and direct measured depths is plotted and often referred to as a "unity plot." The ILI reported depths are plotted on the x-axis, as they are similar to predictor or independent variables used in a standard statistical regression analysis, and the direct measured depths are on the y-axis. The scale on these unity plots reflects the common practice of reporting metal loss depth as a ratio or percentage of the wall thickness (WT) of the pipe.

Figure 2 is an example of the unity plot within the DigIt application. Hovering near a given point results in a box popping up which gives the information recorded for that data point. In addition to displaying the variable in the legend of the plot (Vendor Spec) and the x and y values (ILI_Depth and Direct_Depth), the pop-up can include any variables (called Tooltips in Power BI) that further describe the given point. If the ILI tool has perfect sizing accuracy, all matched Direct-ILI pairs would fall on a 45 degree (1:1) line indicating complete agreement between the two data sets. The data point featured in Figure 2 had an ILI_Depth of 0.70 and a Direct_Depth of 0.58, which indicates that the anomaly depth reported by the ILI (70%WT) was greater than what was directly measured (58%WT). This discrepancy in the depths is indicated in the subsequent "Direct minus ILI Depth" entry, which displays the difference. This value is used to indicate how well the ILI tool adhered to its sizing specification.

Depending on the ILI vendor specification, a given matched pair is binarily assessed as being within or out of specification. Each ILI vendor sets its own specifications for a given ILI tool based upon rigorous testing during tool development. Many of the tools used in North America employ magnetic flux leakage (MFL) and/or ultra-sonic (UT) non-destructive testing (NDT) technologies. The particular ILI tool that reported the example data point was an MFL-A where the "A" stands for axial, which indicates the orientation of the magnetic flux during the inspection. Within their respective performance specifications, ILI vendors specify a tolerance that applies to the ILI reported depth of all anomalies. This states that 80% of the time, when compared to the actual anomaly depth, the ILI reported depth will be within the tolerance value. This number is most commonly 10%WT, which equates to +/- 0.10 for many of the data in Figure 2.

The data point with the pop-up in Figure 2 has a Direct_Depth minus the ILI_Depth of -0.12. Because the absolute value of the difference is greater than the tolerance (0.10), the point is labeled "Out of Spec" for the variable Vendor Spec and set to a different color. Note in Figure 2 that there are some Out of Spec points mixed in with In Spec points. Such data points were reported by UT ILI tools that had a tighter specification than a 0.10 tool tolerance.
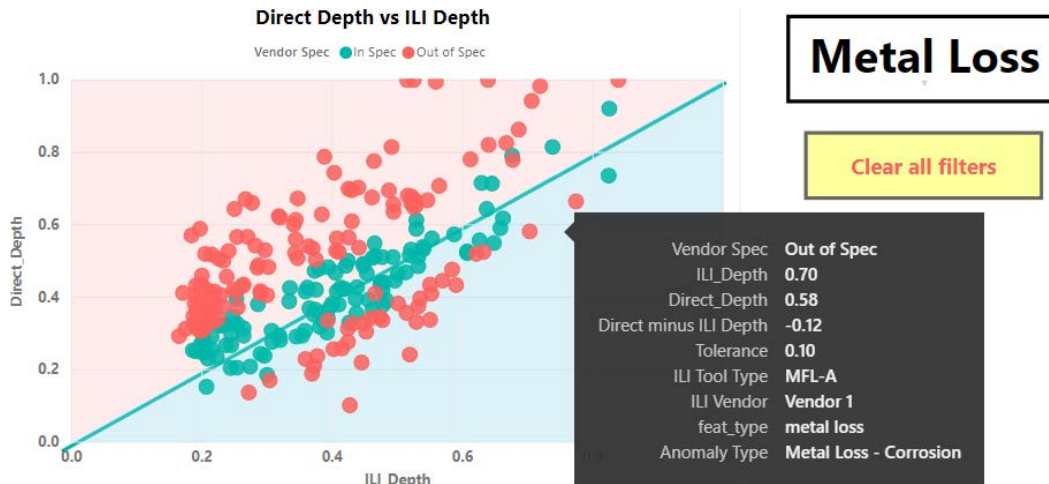
Figure 2: Unity plot of Direct versus ILI depth for the full set of data

2.2 Pipeline Operators Forum (POF) Anomaly Dimension Class

Although Figure 2 is informative as-is, it is examining the full set of data. With Power BI, it is easy to filter the data and have the results change almost immediately. The unity plot can be filtered by "slicer" controls to only include certain anomaly types. The anomaly classification is based on the Pipeline Operator Forum (POF), which categorizes based on length and width of the anomaly. In Figure 3, pitting has been selected in the POF Anomaly Dimension Class[1] table in the upper right of the image. Clicking on any dimension class (or any combination using the control key) fades the non-selected classes. At the same time, the rest of the metal loss page is automatically filtered for only the classes selected. This filtering may be combined with slicers as to further control the data analyzed. The highlighting of a subset of this table filters the data in both the unity plot as well as for the statistics described later.
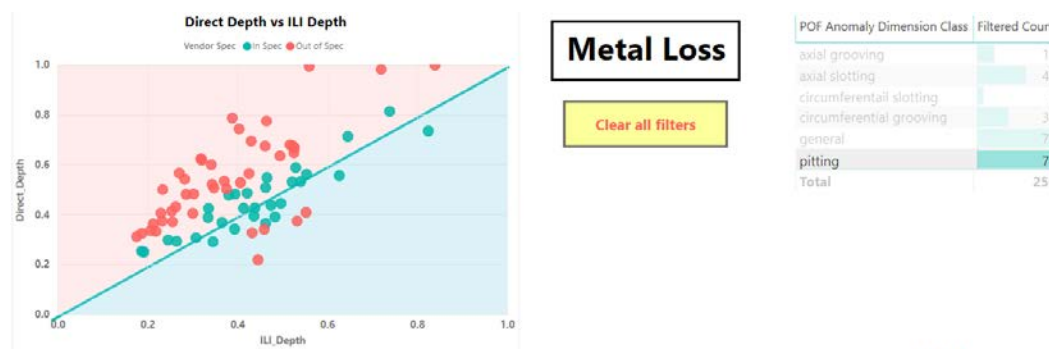
Figure 3: Unity plot of Direct versus ILI depth for just pitting anomalies

The POF Anomaly Dimension Class table in Figures 1, 3, and 4 summarizes the number of metal loss anomalies by the POF dimension class. This provides useful information

---

[1] "Specifications and requirements for in-line inspection of pipelines" Pipeline Operators Forum. Version 2016. Publicly available: https://www.pipelineoperators.org/downloads-links/

because tool performance often varies based upon dimension class. As illustrated in Figure 4, conditional formatting is used for the Filtered Count background color which gives a visual comparison among the anomaly class totals. The dimension class "general" has the most observations (77) and the cell is fully colored in a light green color. Smaller counts are shown with proportionally less of their respective Filtered Count cell colored.

POF Anomaly Dimension Class	Filtered Count
axial grooving	17
axial slotting	47
circumferentail slotting	6
circumferential grooving	30
general	77
pitting	73
Total	**250**

Figure 4: Pipeline Operators Forum (POF) Anomaly Dimension Class

In addition to a general tolerance specified by the ILI vendor for a given tool, such as the +/- 0.10 for many MFL tools, the tolerances may be increased based on the POF class. For example, axial grooving depths reported by MFL-A tools typically have a vendor-specified tolerance of +/- 0.15. Thus, a larger difference between the Direct Depth and the ILI Depth is used to assess if the ILI Depth is within or out of specification. This will impact the binomial-based API 1163 results shown near the bottom on the Power BI page in Figure 1 and later detailed in section 3.2.

**2.3 Categorical Filtering using Power BI Slicers**

The DigIt Power BI app permits numerous data filtering options in addition to the anomaly subset demonstrated in Figure 3. Various slicers allow for dynamic filtering of the data set which can be incorporated into the visual display. As any combination of slicing operations or other filtering options are chosen, the unity plot and the statistical analysis are automatically updated. This quick combination of both visual and analytical displays is valuable to pipeline engineers.

There are options for how slicers show their categories that are not covered here but can easily be seen when entering a new slicer into Power BI. Each of the five categorical slicers shown below will be briefly explained. Figure 5 is a demonstration for three of these five. The Generic Tool slicer shows all the generic ILI tool types in the database. Selecting either the MFL or UT option filters the data and updates the analyses to only include data reported by the representative tool type.

Tool performance can also vary by ILI vendor. The DigIt application is designed to facilitate integrity-based decision making according to ILI technology and, when there is sufficient data, according to ILI vendor. Understanding tool performance at the ILI vendor level allows pipeline operators to select the best performing tool vendor for future inspections. This slicer provides an easy way to compare the overall ability of different vendors to accurately size anomalies with their various ILI tools.

The third slicer in Figure 5 breaks down the generic tool type into more specific categories. For example, some tools can be axial oriented (e.g., the two MFL-A tools in

Figure 5), circumferentially oriented, or use different technology. This is beneficial to assess when a specific type of anomaly is targeted for detection. If there is a need to accurately size cracks, for instance, the three slicers in Figure 5 can be used to determine which ILI vendor, tool technology, and tool type produces the best results.

Generic Tool
☐ MFL
☐ UT

ILI Vendor
☐ Vendor 1
☐ Vendor 2

Specific tool
☐ MFL-A
☐ Vendor 1 MFL-A
☐ Vendor 1 UT
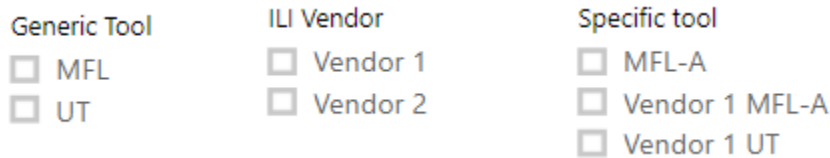
Figure 5: DigIt Categorical Slicers: Generic Tool, ILI Vendor, Specific tool

The last two categorical sliders are depicted in Figure 6. The Training Method slicer allows selection of the level of training of the technician making the direct excavation measurements. Measurement errors occur both with the ILI tool assessment of the anomaly features and the direct measurements. Anecdotal experience within the industry suggests that the higher level of NDT training and certification of the technician, the more accurate the characterization of the anomaly. The second filter in Figure 6 allows the filtering of the data based on whether the line transports oil or gas. The product transported can also play a significant role in tool performance. Most commonly, ILI tools are "pushed" through with the product being transported and velocity excursions (tool speeds that exceed the specified threshold) can be common for gas pipelines due to the compressibility of the gas. Exceeding the specified velocity threshold can result in degraded tool performance.

Training Method
☐ NDT Level I Technician
☐ NDT Level II Technician
☐ NDT Level III Technician

Oil/Gas
☐ Gas
☐ Oil

Figure 6: DigIt Categorical Slicers: Training Method, Oil/Gas

2.4 Continuous Filtering using Power BI Slicers

The filters covered in the previous section are for categorical data. The next two slicers were designed for continuous data, as shown in Figure 7. As with categorical slicers, users have different options on how to display the range of values. The ones shown in Figure 7 by default display the minimum and maximum values in the data set for the variable shown. Users can slide either end of the slicer to reduce the range shown or more directly type in the value desired.

Leveraging Power BI strength in analytics, when either or both of the continuous slicer values are changed, the full statistical analysis and associated unity plot are updated. This is valuable when users want to study a subset of the data, such as only the ILI anomalies with depths 40%WT or greater. This removes the smaller depth calls from the ILI and allows the engineer to study in more detail those that are likely to require remediation.

Figure 7: DigIt Continuous Slicers: ILI Depth, Direct Depth

3. Three Statistical Tests for Direct versus ILI Depth

One goal of an ILI run is to estimate the condition of an entire pipeline segment that might be hundreds of miles in length. Because only a few excavations are made, an evaluation of how well the directly measured depths match the ILI measurements is important. If adjustments should be made to the ILI data, this is accomplished by using statistical analyses of the limited excavation results. For the many unexcavated ILI calls, good estimates of the actual condition of the pipeline is required. In the following sections each of three statistical methods is discussed separately.

Figure 8 is more readable version of the same set of statistical results illustrated in Figure 1. Any filtering changes the unity plot and the statistics to match the subset of values selected. Two (detailed in Section 3.1) of the three statistical approaches suggest possible adjustments to the ILI data to better match the direct depths. Based on the data within the DigIt application, there are three different tool types seen in Figure 8: MFL Vendor 1, UT Vendor 1, and MFL Vendor 2. As will be covered below, none of these pass the API 1163 industry guideline and MFL Vendor 2 has recommended adjustments to the ILI data for better matching with Direct data.

			ILI Mean or Ratio Adjustment Statistically Justified?					API 1163 passed?			
Generic Tool	ILI Vendor	Min Adj	Mean Adj	Max Adj	X	n	phat	pupper	certainty	API 1163	Ratio > 1	RatioAdj
MFL	Vendor 1				46	93	0.49	0.58	0.80	Fail		
UT	Vendor 1				5	18	0.28	0.47	0.90	Fail		
MFL	Vendor 2	0.1038	0.1241	0.1444	57	139	0.41	0.48	0.80	Fail	Ratio > 1	1.49

Figure 8: DigIt Statistical Assessments

3.1 Difference and Ratio Paired t-tests for potential adjustment of ILI Depth

For each Direct-ILI depth pair in the data selected, two different paired t-test comparisons are evaluated. Both are performed at the 95% confidence level and are two-sided tests. Because the paired t-test is well covered in the statistical literature (e.g., Mendenhall, 2009), the step by step analysis sequence is not repeated in this paper.

The first of these t-tests is the traditional paired t-test computed based on the difference of the Direct Depth and ILI Depth, with results shown in the three adjustment (Adj) columns of Figure 9. The statistically suggested adjustment range of the ILI data is shown when the paired mean difference is significantly greater than zero. However, only positive adjustments in ILI data are displayed. The rationale is the required strategic conservatism associated with the potential consequence of the pipelines being analyzed.

It is more conservative to assume the ILI reports smaller anomaly sizes that need to be adjusted in the positive direction.

To aid in the distinction of this traditional paired difference t-test and the subsequent paired ratio t-test, it is worthwhile to give both the null and alternative hypothesis. Below are the null (Ho) and alternative (Ha) hypotheses for the paired difference t-test.

Ho: $\mu_{diff} = 0$ versus Ha: $\mu_{diff} \neq 0$ where

μ_{diff} is the population mean of the paired differences

The Min Adj and Max Adj in Figure 9 are the lower and upper bounds, respectively, on the 95% confidence interval for the mean difference. Statistical significance is based on a two-tailed t-test even though only positive values are shown for conservatism, as discussed earlier.

It is up to the pipeline operator to decide how to act on the results. If the pipeline company wants to adjust the ILI values upwards by a fixed constant, the traditional paired difference t-test is appropriate. The recommended adjustment is the Mean Adj; however, any value between the Min Adj and Max Adj is acceptable from a statistical perspective. Engineering rationale should be used with the viewpoint that the statistical evidence is part of a larger decision support system of the operating company.

A second form of a paired t-test is a paired ratio t-test. Although this form of a paired t-test is straightforward, we have not seen this in the literature. As discussed above, to keep the two paired t-test names distinct, the word difference or ratio is added. For each Direct-ILI depth matched pair, the ratio of the Direct depth to the ILI depth is computed instead of the difference as done in a traditional paired difference t-test. Instead of testing to see if there is evidence that μ_{diff} (population mean of the paired differences) is significantly different than 0, the test is now on $\mu_{ratio} = 1$, as shown below.

Ho: $\mu_{ratio} = 1$ versus Ha: $\mu_{diff} \neq 1$ where

μ_{ratio} is the population mean of the paired ratios

As with the standard paired difference t-test, a two-tailed t-test is performed; however, values above and below 1.0 are shown. If the lower bound of the confidence interval goes negative (not a physical reality), it is set to zero.

Often the two variants of a paired t-test will agree on whether there is a statistically significant difference. Logically, both are assessing if $\mu_{Direct} = \mu_{ILI}$ where μ_{Direct} and μ_{ILI} are the population mean depths of Direct and ILI, respectively. In Figure 9, both paired t-tests show that an adjustment may be warranted for MFL Vendor 2 based on statistics.

Generic Tool	ILI Vendor	Min Adj	Mean Adj	Max Adj		Ratio > 1	RatioAdj
MFL	Rosen						
UT	Rosen						
MFL	Unknown	0.1038	0.1241	0.1444		Ratio > 1	1.49

Figure 9: Paired DigIt t-test Statistical Assessments

Engineers generally favor the paired ratio t-test over the traditional paired difference t-test. One advantage is that adjusting the ILI data by a ratio results in smaller differences for relatively shallow ILI calls and larger differences for deeper ILI calls. Often this mirrors what is seen in the unity plot. Such changes are readily seen and modeled in the related CGReal Bias application in which the user can make multiple adjustments to the original data. In doing so, the CGReal Bias application has side-by-side unity plots of the original and adjusted data and separate statistical output for both.

In the CGReal Bias application, the paired ratio t-test information is expanded from what is shown above in Figure 9 for DigIt. In DigIt, only the sample mean ratio adjustment (RatioAdj for DigIt; Mean Ratio for CGReal Bias) is given and only when the ratio is significantly greater than one. In CGReal Bias, as seen in Figure 10, the two-tailed 95% confidence interval is given in all cases. As mentioned earlier, if the lower bound (Min ratio) goes negative, it is set to zero. Without detailing CGReal Bias, the main intent is to evaluate possible bias in one of the ILI runs. In order to do this, it is important to have the full ratio confidence interval. The data set used in Figure 10 is different from that used in DigIt and involves three different aspects of the two ILI runs that are not discussed here.

Source	Ratio > 1	Min ratio	Mean Ratio	Max ratio
1-1 Pit Match		0.746	0.914	1.082
Earlier FTC	Ratio > 1	1.083	1.364	1.646
UGP		0.875	1.126	1.376

Figure 10: Paired CGReal Bais paired ratio t-test

3.2 API 1163 Statistical Assessment of Direct versus ILI Depth

Section 3.1 examined the matched pairs of direct and ILI depth treating the data as continuous. API 1163 treats each matched pair as binary depending on whether a given ILI depth is within tolerance of the associated direct depth. There are some key points to consider in this approach. Although most MFL ILI runs have a specification of +/- 10% wall thickness, these are not universal. Each ILI call must be assessed to be in tolerance based on its specific case. The value of X as shown below in Figure 11 is the number of successes (ILI calls within tolerance) of n matched pairs. The ratio X/n is the standard phat, $\hat{p}$, or sample proportion within tolerance.

X	n	phat	pupper	certainty	API 1163
46	93	0.49	0.58	0.80	Fail
5	18	0.28	0.47	0.90	Fail
57	139	0.41	0.48	0.80	Fail

Figure 11: API 1163 binary comparison of Direct versus ILI depth

There are multiple issues that should be considered when moving to any binary in/out of specification assessment; however, only one is briefly discussed here. Each Direct to ILI

depth comparison discards much information in any binomial setting as each is reduced from continuous measurements of depth to a simple binary 0-1 value. If a given ILI depth is within its tolerance of the Direct depth, the binary value is set to 1; otherwise it is 0. The value of X in Figure 11 is a sum of all these binary 0-1 values, and hence the number of ILI calls within tolerance. In this binary setting, a match between ILI and Direct depths that is perfect (i.e., both report the same value) and a match that is barely within tolerance each receive a binary value 1. Such information loss should be noted. A full assessment of Direct versus ILI depth should include both continuous comparisons and the unity plot, as covered earlier, in addition to addressing the industry recommended API 1163 specification.

API 1163 (2013) uses the Agresti-Coull (1998) normal approximation that is an improved version of the standard central limit approximation for proportion hypothesis testing taught in typical statistical texts. Harper (2017) covers the API 1163 Agresti-Coull method, including a generalization to finite populations as well as related material on the standard normal approximation to the binomial and the exact Clopper-Pearson method. The standard normal approximation to the binomial confidence intervals is well documented in the literature and not repeated here. Instead, the formulas shown are those specific to the Agresti-Coull approximation.

In place of the typical $\hat{p}$ and n in the standard normal approximation to the binomial, the Agresti-Coull method uses $\tilde{p}$ and $\tilde{n}$ defined below. In the equations below using API 1163 notation, n is the sample size, x is the number of successes (X in Figure 11), and $z_{(\alpha)}$ is the $(100 - \alpha)^{th}$ percentile of a standard normal z along with the other variables defined. This results in the Agresti-Coull upper confidence bound $\hat{p}_{upper}$. Unlike the two paired t-tests discussed earlier, the second edition of API 1163 Agresti-Coull is a one-tailed confidence interval. In this case, it computes only the upper 95% confidence interval, $\hat{p}_{upper}$, of the proportion. If $\hat{p}_{upper}$ is at least as large as what the ILI tool vendors call certainty in Figure 11, then the ILI tool passes the API 1163 assessment. In Figure 11, $\hat{p}_{upper}$ is less than the stated certainty, therefore all three subdivisions fail API 1163.

$$\hat{p}_{upper} = \tilde{p} + z_\alpha \sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}} \text{ where}$$

$$\tilde{n} = n + z_\alpha^2, \tilde{p} = \frac{x + \frac{z_\alpha^2}{2}}{\tilde{n}}$$

4. Summary

The Power BI DigIt application provides users an intuitive responsive environment to select a full data set or a desired subset of Direct-ILI matched depth pairs for fast analysis. The analysis includes visualization and a variety of statistical tests to aid a pipeline operator in comparing different ILI tool types under many different circumstances. Some operators may want to focus only on deeper ILI calls, such as at least 40% wall thickness, which is easily addressed by using the continuous sliders to filter the data to be analyzed. Others may be more focused on how the training level of their direct technicians impacts the comparison of the Direct depths to the ILI depths. In the past, such analyses may have taken considerable sub-setting of data and passage of the data to different software for visualization and the associated statistical analysis.

Additionally, DigIt provides statistical analysis methods not readily available elsewhere. This tool is still in its infancy and recommendations are readily accepted.

References

Agresti, Alan & Brent A. Coull (1998), 'Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions', The American Statistician, 52:2, 119-126.

API Standard 1163, "In-line Inspection Systems Qualification", 2nd edition, April 2013, American Petroleum Institute.

Harper, William V., "Finite Population Agresti-Coull based Binomial Confidence Intervals and Excel VBA Function", Proceedings of the Joint Statistical Meetings (JSM2017 - Section on Physical and Engineering Sciences), Alexandria, VA: American Statistical Association. 725-734.

Mendenhall, William, Robert J. Beaver, and Barbara M. Beaver, Introduction to Probability and Statistics, 13th edition (2009), section 10.5 starting page 412.