In-field crack validation data: can we trust it?

Tom Oldfield, Service Manager – Field Verification, ROSEN, UK, finds the most efficient way to understand the quality of the field verification while validating ILI system performance.

Validating inline inspection (ILI) system performance is a critical element of any ILI campaign. It ensures that the integrity decision-making based on the data collected is realistic, considering both the classification and the sizing tolerances stated in the inspection system's performance specification. Nowhere is this task more important than in validating crack sizing, due to the difficulty of predicting the behaviour of crack like-defects and the challenges in measuring these features.

The ILI crack sizing technologies of ultrasonic (UT) and electromagnetic acoustic transducer (EMAT) are validated and qualified technologies, but the industry is continuously learning and gaining experience. The quality of results across different vendors can therefore vary due to the level of experience with particular feature types, the robustness of the processes, technological limitations and numerous other factors. Therefore, from a safety perspective, it is essential that ILI system validation is performed considering pipeline specifics in order to understand performance on a run-by-run basis and thus confirm that the ILI system performance is within its specification. This validation is now a regulatory requirement in the US, under legislation that explicitly references API 1163. Validation can be done in a number of ways, including benchmarking against previous data, cut-outs and validation spools. The most common, however, is in-field investigations of features reported by ILI. The question we will be looking to answer: can we trust the data we get from the field to validate ILI system performance?

Contrary to popular belief, the feature sizes measured in the field are not absolute and do come with a level of tolerance – either known or, more worryingly, unknown. The accuracy and repeatability of measurement techniques used for sizing cracks are heavily dependent on user knowledge and skill, meaning that two technicians using the same equipment, working to the same procedure, can get two very different results depending on their level of experience and competence. Understanding uncertainty in field measurement accuracy is a fundamental challenge to ILI system validation and the quantification of ILI system performance.

In-field crack sizing

The term 'crack sizing' encompasses numerous cracking mechanisms with different surface and subsurface morphologies. As a result, and due to the limitations of the various techniques, there is no single field-deployable technology that is appropriate for all cracking types. Further complicating matters, results collected in-field are heavily reliant on the skill and experience of the technician taking the measurements.

Combining these factors means that there can be considerable variation in in-field sizing accuracies. If these measurements are then used without understanding their tolerance or accuracy to prove or disprove ILI performance, this can clearly cause challenges in understanding the real problem.

ILI system performance

The published ILI system performance specification of a given ILI service provides a statistical basis for its detection, classification and sizing capability. ILI service performance specifications are created and refined through extensive testing of representative anomaly samples in small-scale (laboratory trials), full-scale (pull tests) and real-pipeline environments. There are, however, only a finite number of scenarios that can be considered as part of a development programme. Real-life variations in run condition, line cleanliness and defect morphology, to name a few, can all affect the ability to meet the detection and sizing performance stated in the ILI service specification.

API 1163

API 1163 gives guidance on how to use field results to validate ILI system performance. It states, "ignoring field measurement inaccuracies is generally conservative but may be overly conservative when evaluating ILI sizing performance". This phrase alludes to the fact that tolerance of field measurement is understood to be an issue, but that it can be difficult to quantify.

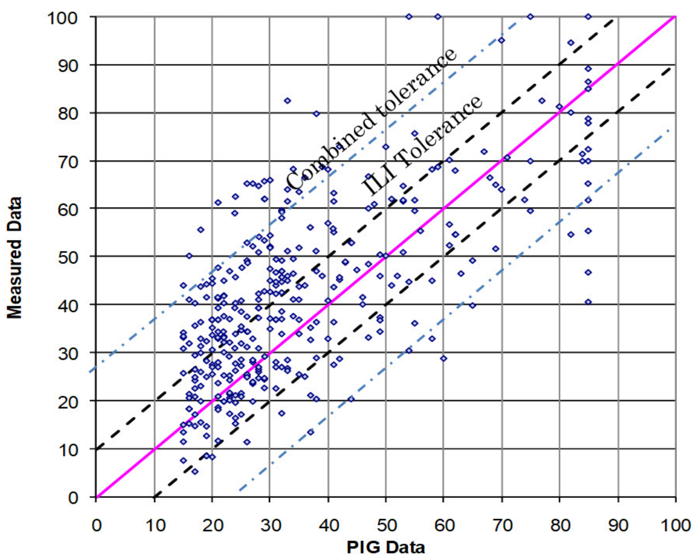

Figure 1. Example of a unity plot from a validation campaign.

Ignoring tolerance from the field-measurement may well indicate that the ILI results are not in the ranges stated in the performance specification – potentially incorrectly so, because only one of the tolerances is considered, which will narrow the acceptable window and make any ILI measurement more likely to be out of specification. This relationship is illustrated in Figure 1 in the form of a unity plot contrasting ILI-tolerance-only data with the combined tolerance of field and ILI results.

Importantly, however, API 1163 does not give or suggest tolerance values that can be used for in-field validations. Equation 8 in API 1163 demonstrates how using a combined tolerance for comparing two datasets with a defined level of uncertainty is a less-conservative way to assess ILI-sizing performance if the field tolerance is understood.

The following equations are taken from API 1163:

$$\boldsymbol{\delta e_{comb}} = \sqrt{[\boldsymbol{\delta}\left(\tfrac{d}{t}\right)_{ILI}]^2 + [\boldsymbol{\delta}\left(\tfrac{d}{t}\right)_{field}]^2}$$

where:

$$\boldsymbol{\delta}\left(\tfrac{d}{t}\right)_{ILI} \text{ and } \boldsymbol{\delta}\left(\tfrac{d}{t}\right)_{field}$$

represent the tolerance on the relative depth measurements associated with the ILI and field measurements, respectively, both at the certainty level associated with the stated ILI tolerance. An individual measurement can only be considered out of tolerance when:

$$|\boldsymbol{e}| > \boldsymbol{\delta e_{comb}}$$

otherwise, the measurement is within tolerance.

What this means is that the measurement should consider both the ILI tolerance at the defined confidence as reported in the specification (typically 80%), with the field measurement tolerance at the same confidence interval. Only if the measurement is outside this combined tolerance should it be considered out of specification. This is shown graphically in Figure 2.

In the example in Figure 2, if the tolerance of the field measurement is ignored, the ILI measurement falls outside of the stated tolerance at 80% confidence of the field measurement. If the tolerance of the field measurement is known and a combined tolerance is calculated in accordance with equation 8 from API 1163, then the field measurement falls within the now wider acceptable range from the ILI system measurement. This example demonstrates the possibility for both values to be correct within the limits of the stated tolerances. But the value can clearly be acceptable anywhere in the combined tolerance region.

For direct measurements like pit gauges or laser scanning, field tolerances are generally so small – an order of magnitude smaller than ILI – that their influence on the combined tolerance is minor. However, crack-sizing tolerances can have a more significant influence, as they sometimes exceed ILI tolerances.

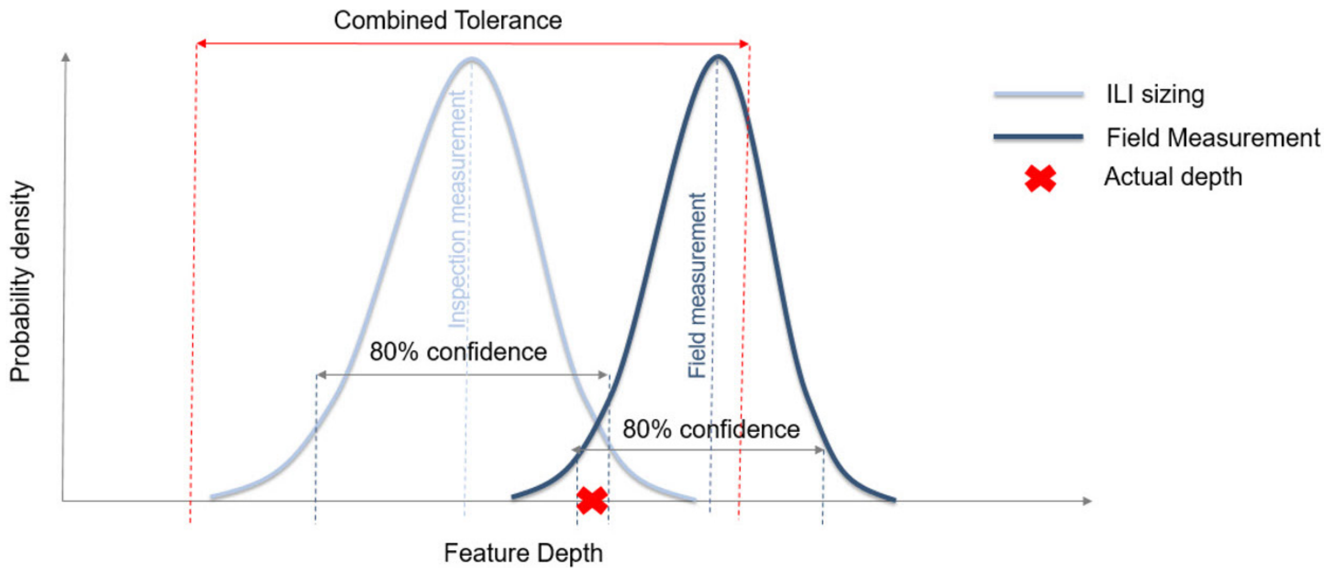

Figure 2. Combined tolerance of inspection.

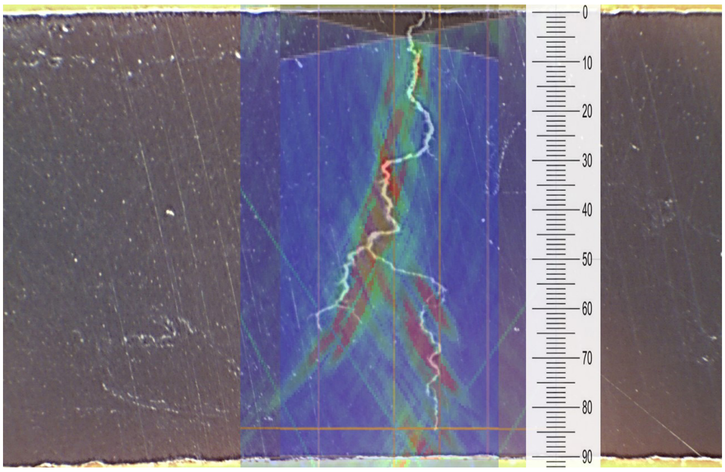

Figure 3. SCC cracking cross-section with PAUT data superimposed.

Is there a solution?

If there is significant uncertainty in field-crack sizing measurement, are the measurements worth taking? In short, the answer is yes, but there are several factors or options that need to be in place before these measurements can be used to understand ILI tool performance.

One way to manage this risk is to assume that the technicians have a general, acceptable level of competence from a review of procedural documentation and applying conservative tolerances to the field measurement. ROSEN's field verification requirements document outlines the checks that should be in place to assume this basic level of competence. Once the document review is successful, conservative tolerances for various technologies can be used that have previously been suggested.[1] However, this approach still assumes that the technicians are practically as well as theoretically competent. This may not be a justified assumption and, hence, non-conservative results may be acquired.

The preferred option is to understand the individual technician's performance on the sizing of features with a predetermined procedure and process in order to apply a less conservative tolerance with higher confidence. This tolerance can be calculated through a blind-trial assessment of technicians on representative samples, assessment of in-field sizing prior to destructive testing in the laboratory, or through in-field buffing of surface-breaking features.

Do operator qualifications not guarantee quality?

Typically, in-field technicians should have qualifications that are managed and controlled by the requirements of ISO 9712 or ASNT, which ensures that technicians have a minimum level of competence to be able to use the equipment and be familiar with the technique. However, as NDT is a broad subject, these qualifications typically cover heavier-wall materials, while cross-country pipelines are usually thinner wall, which poses some unique challenges; plus, any error as a percentage of wall thickness is more significant. The exams also typically have a lower importance on the critical sizing element of the measurement and do not give a tolerance of inspection for the exam, which means that the technician sizing accuracy is still unknown following the qualification.

Until the criticality of the accuracy of the measurement is acknowledged within the pipeline industry, and there is a move towards understanding individual inspector performance assessed by a centralised body, the responsibility for understanding technician competence remains with the pipeline operator, who must request evidence of performance or test for it.

How to test competence

Artificial defects can be created with high levels of sizing accuracy in representative samples in order to understand technician competence in blind trials. These can also be used by the technicians in-field as reference blocks for improving sizing accuracies.

An efficient way of proving technician competence but also validating run-by-run performance of ILI tools is by manufacturing and implementing validation spools. These are typically spools of the same diameter and wall thickness as the pipelines being inspected by ILI; they have a number of representative features manufactured into them and are sized with a high degree of confidence. This allows technician performance to be benchmarked against a depth value with high confidence under blind trial conditions. These spools can then tie into the pipeline loop so they can be measured during the ILI inspection under live conditions to understand inline performance – as opposed to relying on the pull-through data or only field data. It is unlikely that synthetic defects will ever be able to truly replicate the complexities of natural features, but advances in manufacturing techniques can produce highly realistic crack-like defects, as well as notch-like defects produced by EDM.

Once a statistically significant sample of defects of a known size has been validated, probability of detection and sizing tolerances can be calculated. This will allow for a comprehensive understanding of operator competence and tolerance, ideally before getting into the ditch. Classification of features using NDT methods remains a very challenging topic. Realistically, for high-confidence feature classification, metallography is still a necessary evil.

Validating tool performance

Once there is an understanding of in-field technician competence and tolerance, the data used for the validation of ILI system performance has more clarity and will increase the confidence in ILI system performance.

There are still factors that the blind-trial process cannot control, such as the human element of the inspection (mood, weather, aliments, etc.), which can affect the calculated tolerances from blind trials usually held in laboratory conditions. As the human factor cannot be accurately quantified, the tolerance from the blind trial is probably the most appropriate starting point, because assuming a more conservative (wider) tolerance for the field measurements may result in some ILI features being assessed as acceptable. And that means integrity decisions may be non-conservative overall.

Conclusion

Understanding in-field competence and tolerance is a critical part of validating ILI system performance, particularly for crack sizing. But it is not always done in practice, possibly misrepresenting the actual situation of the pipeline and leading to integrity choices that could be overly conservative or, conversely, unsafe. The benefit of considering the measurement accuracy of all involved systems has been explained. It is important to establish a process to ensure that all sources for tolerances are discovered and considered. For now, the most efficient way to understand the quality of the field verification while validating ILI system performance is the manufacturing and installation of validation spools with artificial defects. The goal: to build an industry where high-quality field data is widely tested and valued.

References

1. OLDFIELD T., FOWLER, S. TORRES, D., "Tolerance of ILI Validation Inspections, Why Is It Important, and How to Reduce It," Pipeline Pigging and Integrity Management Conference, No.2023.127:43, Houston, Texas, February 2023.