Application of Advanced Data Analytics

Dion Phan
Regional Sales Leader, Asia Pacific
Process & Pipeline Services

Copyright 2021 Baker Hughes Company LLC. All rights reserved. The information contained in this document is company confidential and proprietary property of Baker Hughes and its affiliates. It is to be used only for the benefit of Baker Hughes and may not be distributed, transmitted, reproduced, altered, or used for any purpose without the express written consent of Baker Hughes.

Introduction

- Metal loss depth tolerances heavily influence integrity management programs - particularly when using probabilistic techniques and assessing future integrity
- Conservative tolerances lead to unneeded or premature digs
- Industry specifications are not ideal to represent the true influences on sizing tolerances
- Multi-year partnership between Baker Hughes and key customer to develop a means to reduce tolerance conservatism in MFL specifications

Copyright 2025 Baker Hughes Company. All rights reserved.

Baker Hughes

Baker Hughes evolution of MFL sizing algorithms

The old way

Approach

- Manufactured calibration joints of various wall thicknesses pull tested at multiple speeds
- Regression or machined learned models

Statistics

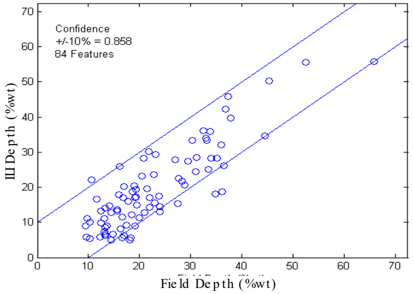
- Typically 100s of defects
- Multiple pulls to acquire larger population and speed influence

Performance

- Established reliable performance within defined confidence levels
- Wider tolerances

Limitations

- Small sample sets
- Highly influenced by "morphology" and shape
- Performance on interacting corrosion is less predictable



The new way

Approach

- Manufactured calibration joints of various wall thicknesses pull tested at multiple speeds
- Training on vast library of correlated laser scanned/AUT field data
- Machined learned models

Statistics

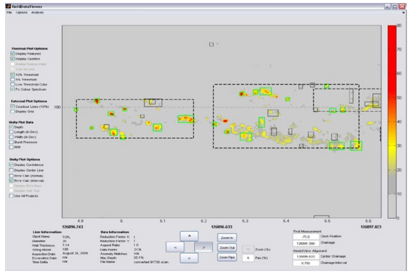
- 10,000s of defects under real operating conditions to develop models
- Pull data ensures bounding of models

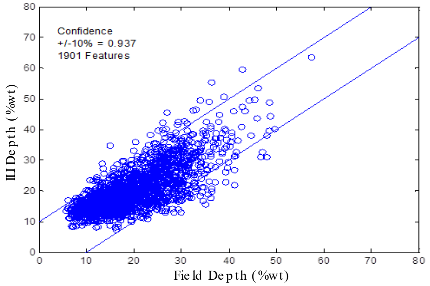
Performance

- Highly robust to "real world" defect shape when corrosion is non- interacting
- Tighter tolerances

Limitations

- Highly skilled process (over training)
- Performance on interacting corrosion improved but remains challenging



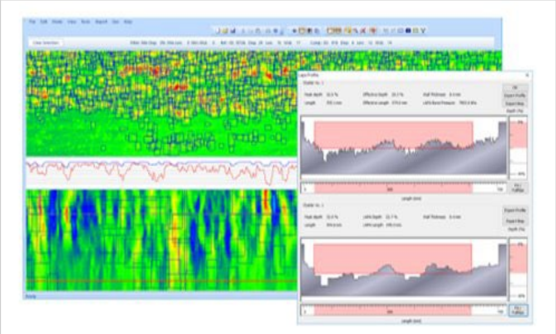


Copyright 2025 Baker Hughes Company. All rights reserved.

Baker Hughes

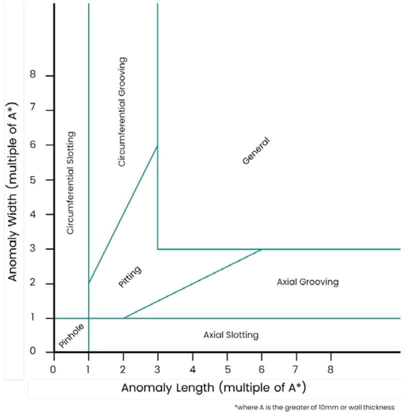
Big Data Library

- Excavated feature data is the best truth data available for training machine learning models
- Extensive repository of truth data aligned with ILI data
  - Enables opportunities for advanced data analytics to step change POD, POI, POS
  - Baker Hughes has 100s thousands of field verified laser scan profiles in the database
- More than just a database of actual results compared to reported results
  - ILI signal data (tri-axial)
  - Full resolution measured defect profiles
  - Tool parameters
  - Operating conditions
  - ...and more
- Flexibility to be utilized for many research initiatives



Copyright 2025 Baker Hughes Company. All rights reserved.

Metal Loss Tolerance Opportunity



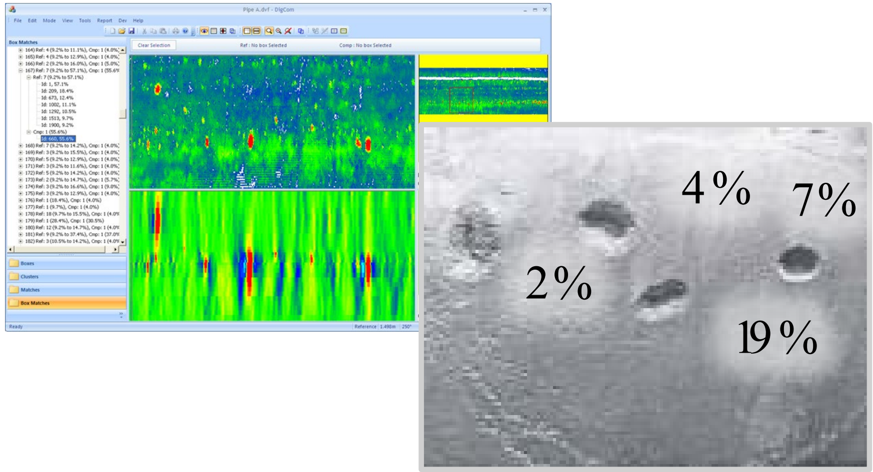
	General	Pitting	Axial Grooving	Circ. Grooving	Pinhole	Axial Slotting	Circ. Slotting
Minimum Dimensions for sizing accuracy[4]	W>3A L>3A	W>A L>A	W>A L>2A	W>2A L>A	W>0.7A L>0.7A	W>0.7A L>A	W>A L>0.7A
Depth at POD=90%	0.05t	0.05t	0.08t	0.05t	0.10t	0.20t	0.10t
Depth sizing accuracy at 80% certainty	± 0.07t	± 0.07t	± 0.10t	± 0.10t	± 0.15t	± 0.20t	± 0.20t
Depth sizing accuracy at 90% certainty	± 0.10t	± 0.10t	± 0.14t	± 0.14t	± 0.20t	± 0.25t	± 0.25t
Width sizing accuracy at 80% certainty	± 0.4-in. ± 9 mm	± 0.4-in. ± 9 mm	± 0.4-in. ± 9 mm	± 0.4-in. ± 9 mm	± 0.3-in. ± 8 mm	± 0.4-in. ± 10 mm	± 0.4-in. ± 10 mm
Length sizing accuracy at 80% certainty	± 0.3-in. ± 7 mm	± 0.2[illegible] ± 5 [illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Standard Reporting Threshold[5]	[illegible]						

[4] For detection, dimensions are at the [illegible]
[5] Lower thresholds available.

	General	Pitting	Axial Grooving	Circ. Grooving	Pinhole[17]	Axial Slotting[18]	Circ. Slotting[18]
Minimum Dimensions for sizing accuracy[19]	W>3A L>3A	W>A L>A	W>A L>2A	W>2A L>A	W>0.5A L>0.5A	W>0.5A L>A	W>A L>0.5A
Depth at POD=90%	0.04t	0.6t	0.6t	0.04t	0.13t	0.13t	0.04t
Depth sizing accuracy at 80% certainty	± 0.08t	± 0.08t	-0.13t +0.08t	-0.08t +0.13t	-0.13t +0.08t	-0.18t +0.08t	-0.08t +0.13t
Depth sizing accuracy at 90% certainty	± 0.1t	± 0.1t	-0.15t +0.1t	-0.1t +0.15t	-0.15t +0.1t	-0.2t +0.1t	-0.1t +0.15t
Width sizing accuracy at 80% certainty	± 0.5-in. ± 12 mm	± 0.5-in. ± 12 mm	± 0.5-in. ± 12 mm	± 0.5-in. ± 12 mm	± 0.3-in. ± 7 mm	± 0.5-in. ± 12 mm	± 0.5-in. ± 12 mm
Length sizing accuracy at 80% certainty	± 0.3-in. ± 7 mm	± 0.2-in. ± 4 mm[20]	± 0.3-in. ±7 mm	± 0.3-in. ± 7 mm	± 0.2-in. ± 4 mm	± 0.3-in. ± 7 mm	± 0.3-in. ± 7 mm
Standard Reporting Threshold[21]				0.10t			

POF Based Tolerances

- Corrosion tolerances have traditionally been built based on arbitrary POF based length and width categories with limited defect populations
- Length and width are not the only influences on accuracy so not ideal as a grouping criteria



Anomaly Specific Tolerances

Specific tolerance for every individual corrosion in a line based on multiple influencing characteristics

Why is this important?

1. Reducing digs - considerable over-conservatism can be removed from integrity assessments
2. Improving safety – better understanding of where true risk lies

Copyright 2025 Baker Hughes Company. All rights reserved.

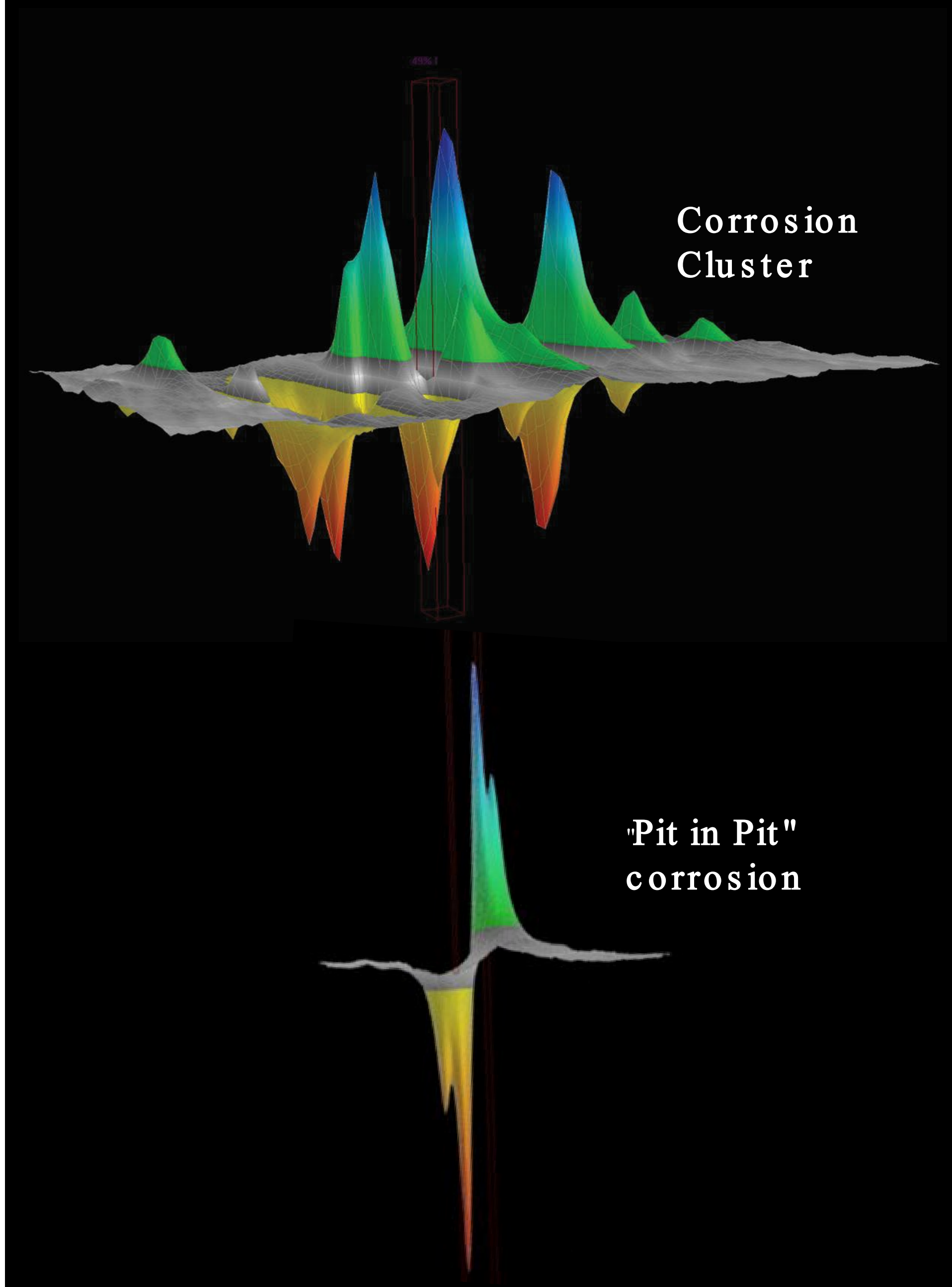
Baker Hughes

What really influences sizing?

- Length and width are not the only important factors related to MFL depth accuracy
- Machine learning technique used to determine greatest influences on sizing
- Reviewed **57** influencing factors against **>200,000** defects
- Highest impacting factors fell into 4 main categories:

Predicted anomaly measurements	Raw signal characteristics	Location and interaction to other pipeline fittings and fixtures	Location and interaction to other defects
The predicted length, width and depth output of anomaly sizing models.	The parameterizations of the anomaly ILI raw signal including both the triaxial MFL data and other supplemental data.	The location and interaction of target anomalies to other pipeline fittings and fixtures..	The location and interaction of target anomalies to other neighboring anomalies..

- The specific factors within and across these four categories interact together in complex ways to define metal loss accuracy
- **Length and width were not the dominant influence**



 Copyright 2025 Baker Hughes Company. All rights reserved.

Baker Hughes

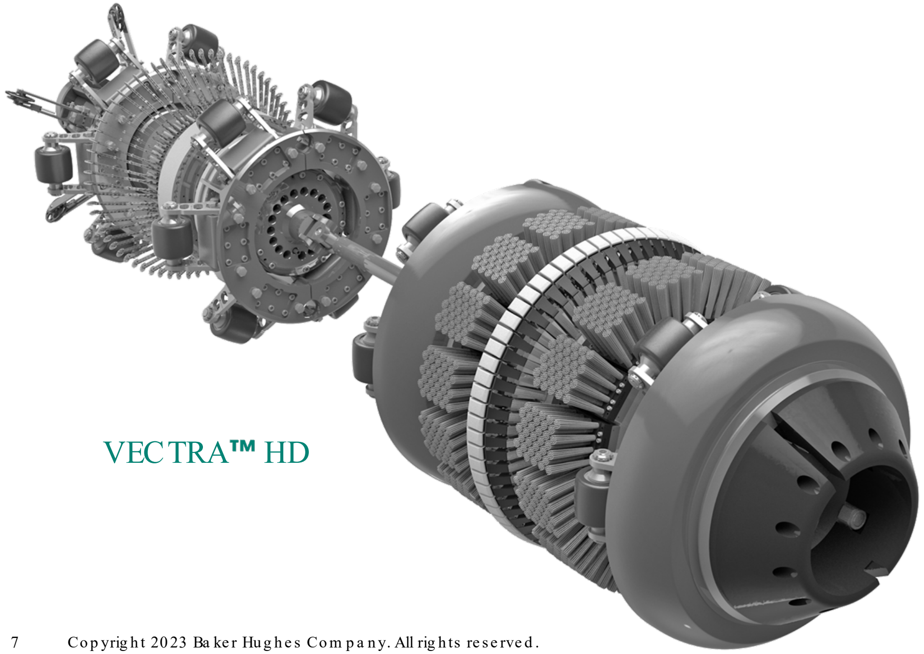
Proof of Concept – initial validation

- Big data library utilized to improve tolerances at an individual line level
  - Machine learning model developed using wide range of influencing factors
  - **Tri-axial** MFL with standard and highest resolutions considered
  - Model applied to 8 lines with high corrosion populations and significant dig programs

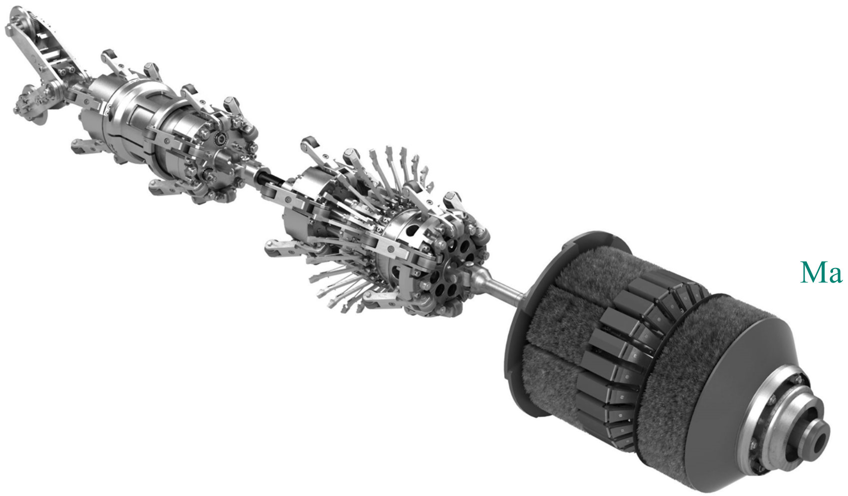
- Improved tolerances relative to the original POF-based tolerances for 89% of defects

- Other defects where wider tolerance were predicted gives better understanding of true risk

Pipe Diameter	Percent of defects with tightening of tolerance
8"	67%
12"	99%
20"	84%
24"	98%
30"	95%
30"	94%
30"	73%
42"	100%
Average	**89%**



VECTRA™ HD

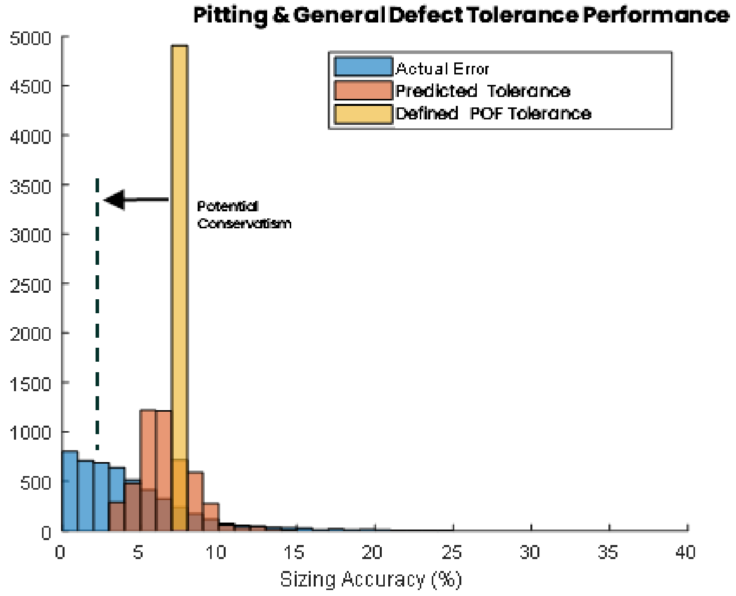


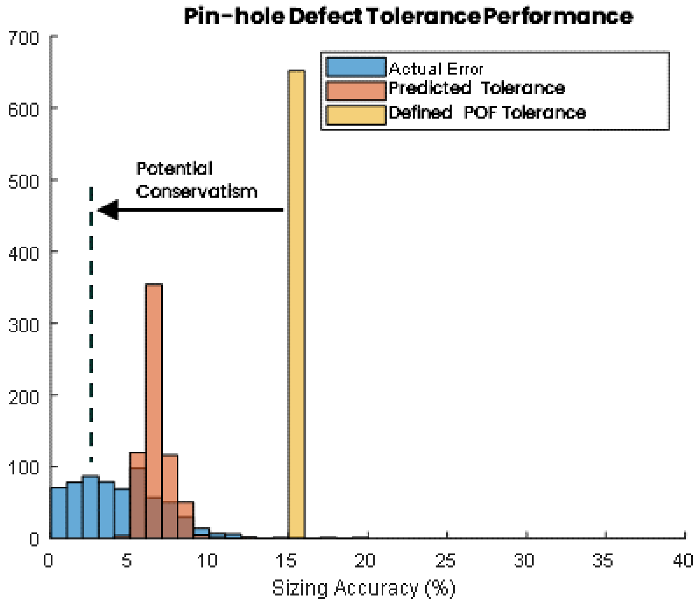
MagneScan™ SHR+

Copyright 2023 Baker Hughes Company. All rights reserved.

Baker Hughes

Understanding the value of predicted tolerances





The tolerance predictor significantly reduces the over-conservatism hidden in current specifications

- Actual error distributions better than POF and predicted tolerances
- Predicted tolerance error distributions remain conservative

Copyright 2023 Baker Hughes Company. All rights reserved.

Baker Hughes

Benefits

- Potential impact to our client's dig program was assessed
- Individual tolerances supplied for all anomalies in pipe tallies for future integrity assessment
  - Comparison made of results using POF and predicted tolerance methods for:
    - Number of digs over a 10-year growth period
    - Number of digs prior to the next re-inspection
    - Re-inspection year

	10-year growth period	Prior to re-inspection year	Re-inspection year
NPS	Reduction of digs	Reduction of digs	Increase in re-inspection year
8	12	4	1
12	-3	-1	0
20	294	14	0
24	43	6	3
30	90	2	0
30	117	-2	1
30	40	0	0
42	0	0	0
Total	593	23	–

Theoretical reduction of 62% of digs to next inspection (from 37 to 14) and possible deferral of 38% of pipeline re-inspections

Copyright 2025 Baker Hughes Company. All rights reserved.

Baker Hughes

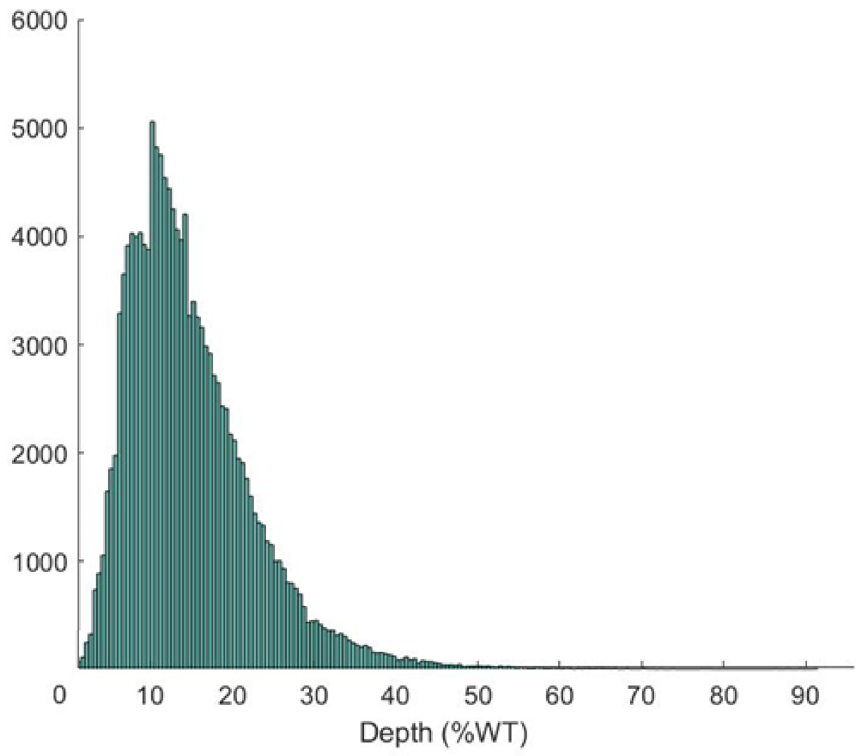
How do we apply this more generally?

- Previous work generated line-specific optimization to establish predictive tolerance models
- The challenge is to be able to predict tolerances on any line with any tool

Generic Predictive Tolerance Model

- Expanded training and test data to entire big data library (100s thousands features)
  - VECTRA HD and MagneScan SHR+ in multiple diameters
- Highly representative database
  - Full joints scanned for many years (100s to 1000s per joint)
  - Realistic skewed population distribution
- Re-optimized machine learned model

Depth Distribution Of Laser Scanned And MFL Signal Model Training Data

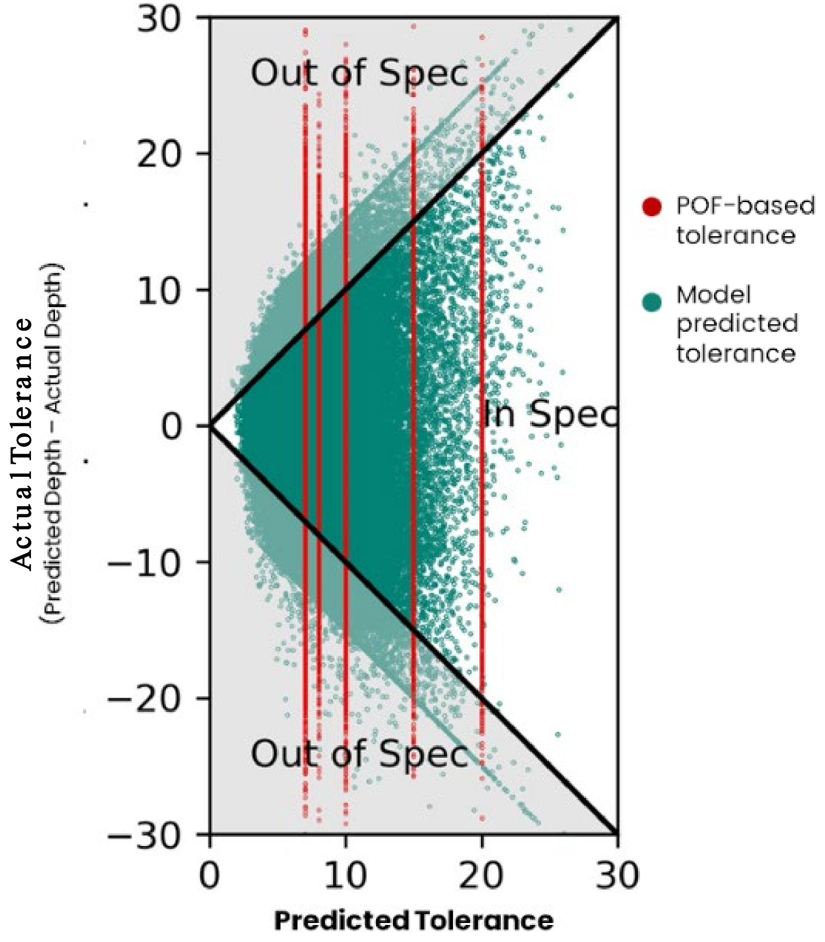


Copyright 2025 Baker Hughes Company. All rights reserved.
Baker Hughes

Generic Model Performance and Assessment

Goals

1. Predict unbiased (symmetric) 80% confidence bounds for performance results
2. Provide tolerances that are more representative of actual depth
3. Remove conservativism from POF specification

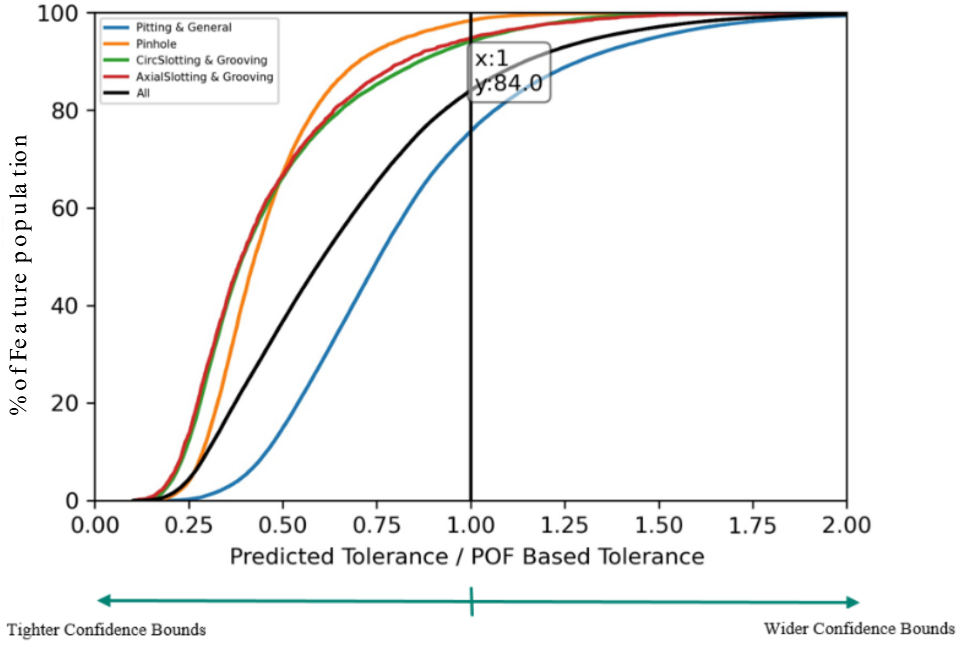


Copyright 2025 Baker Hughes Company. All rights reserved.

Baker Hughes

Did we reduce over-conservatism?

- Cumulative distribution function (CDF) plot is used to demonstrate relative change in predicted tolerance v specified
- POF classifications used to simplify interpretation
- Tolerance reduces in the majority of cases
- Pin-holes show that specification is highly conservative versus actual performance >95% improvement
- Other influencing factors - predicted tolerances increased relative to the published specification on 16% of the population.



84% of defects had tighter tolerances than formal specification

Wider tolerances means we are identifying where greater risk lies on accuracy of reported depths

Copyright 2025 Baker Hughes Company. All rights reserved.

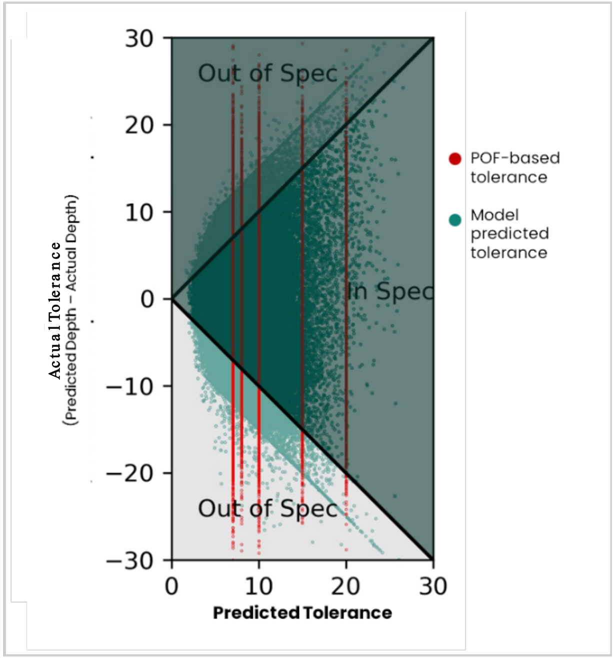
Baker Hughes

Other Key Performance Criteria

Safety Outliers

As we remove conservatism we need to ensure we do not increase risk that critical features will not be significantly undercalled

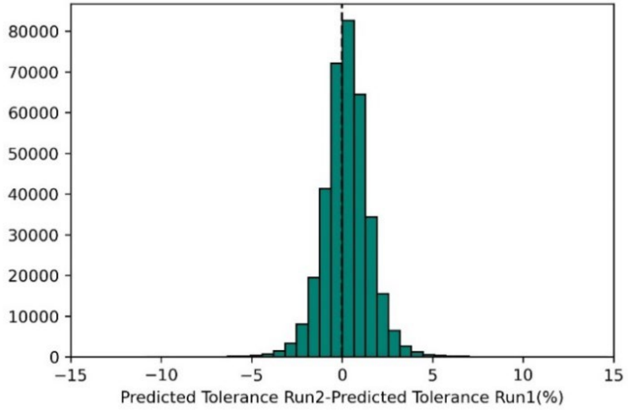
- Predicted tolerances (green) closer to 45° lines (unity) indicate reduced likelihood of underestimating correct tolerance



Repeatability

It is important to determine whether the model consistently predicts tolerances

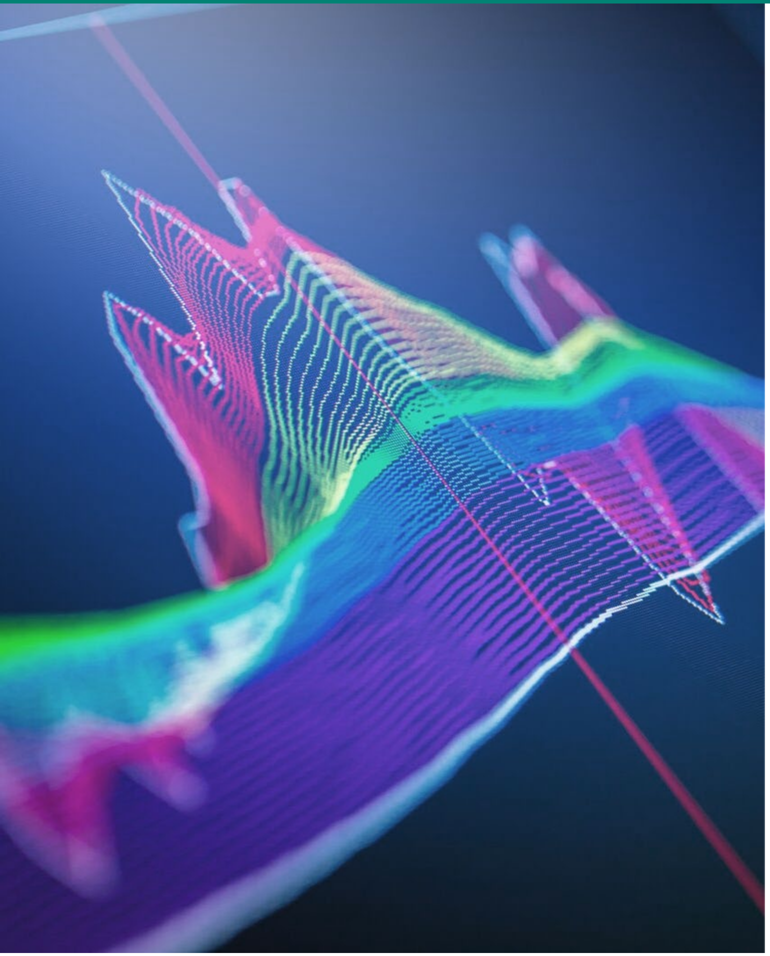
- Runcom software used to compare 350k features to assess repeatability
- Simulated repeat data using inspections on the same lines with defects known to have no growth



Confidence Interval	Predicted Bound Repeatability
80%	±2.23%
90%	±2.86%

Copyright 2025 Baker Hughes Company. All rights reserved.

Conclusion & Acknowledgments



Innovative, high accuracy tolerance predication model generates individual depth tolerances **for every anomaly**

Minority of anomalies with wider tolerance are better indicators of true risk leading to **improved safety**

Less conservative tolerances compared to conventional POF-based tolerances means **reduced dig program costs**

Enables **more efficient and effective** maintenance & dig programs

Copyright 2025 Baker Hughes Company. All rights reserved.

Baker Hughes

Application of Advanced Data Analytics

Q & A

 Copyright 2025 Baker Hughes Company. All rights reserved.

Baker Hughes