Bayesian Calibration of In-line Inspection Tool Tolerance

by

Jeffrey Liang Lee

B.S. Mechanical Engineering, Rice University (2013)

Submitted to the System Design and Management Program
in partial fulfillment of the requirements for the degree of

Master of Science in Engineering and Management

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2020

Author .....
System Design and Management Program
August 19, 2020

Certified by .....
Herbert H. Einstein
Professor, Department of Civil and Environmental Engineering
Thesis Supervisor

Accepted by .....
Joan S. Rubin
Executive Director, System Design and Management

Bayesian Calibration of In-line Inspection Tool Tolerance

by

Jeffrey Liang Lee

Submitted to the System Design and Management Program
on August 19, 2020, in partial fulfillment of the
requirements for the degree of
Master of Science in Engineering and Management

Abstract

Calibration of Magnetic Flux Leakage (MFL) In-line Inspection (ILI) tools is an important part of the overall pipeline integrity management process. Over-called or under-called corrosion features can have significant impacts on safety and resource management.

This thesis examines methods for improving the Validation and Calibration processes using Bayesian Inference. The focus is on improving the tolerance that is applied to undug features to optimize the execution of risk-based repairs.

A simulated data set was generated, with two separate categories, one which represents tool performance on basic features and another for challenging features. The calculated parameters of α, β, and σ, were calculated using a Bayesian model leveraging a Markov Chain Monte Carlo simulator. The σ parameter is used to determine the appropriate tolerance to apply and was compared with a σ calculated via the method recommended by API 1163. Results from the example data set show that in challenged situations, the Confidence Level of the tool performance can be increased from 89% to 95% and the mean average error can be decreased using the Bayesian Inference model.

Opportunities to use the methods outlined to improve other processes in ILI validation are discussed. By appropriately updating the likelihood used in the Bayesian model with dig data, the tolerance can more accurately represent the undug features and risk management decisions can be conducted accordingly.

Thesis Supervisor: Herbert H. Einstein
Title: Professor, Department of Civil and Environmental Engineering

Acknowledgments

I thank Professor Herbert Einstein for providing support and guidance through a unique time to be a graduate student and for helping me formulate and tackle this question. I also thank the MIT SDM program, my professors, and my cohort for helping me to become a better engineer through the sharing of ideas, struggles, and accomplishments. Additionally, I acknowledge the support of my supervisors for the opportunity to take a leave of absence to pursue my education.

Most of all, I thank my family. First, my parents Jana and Jemmy for always believing in me and helping me realize my childhood dreams. And finally, my fiancée Erin for being my rock throughout the inevitable late-night team meetings in year one or the long distance visits to Boston in year two. We are now one more thesis closer to palm trees and ocean breezes!

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

List of Figures

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

1.1 Problem Statement

Pipelines are a critical part of the world's energy and petrochemical infrastructure. Pipelines carrying hazardous liquid and gas in the public domain are regulated in the United States by the Department of Transportation's Pipeline and Hazardous Materials Safety Administration (PHMSA) according to Title 49 of the Code of Federal Regulations Part 192 (Natural and Other Gas) and Part 195 (Hazardous Liquids).

To manage the potential threats for High Consequence Areas (where the route may intersect with populated or ecologically sensitive interests), pipeline operators are required by law to implement integrity inspection programs and assess the lines periodically. In-line inspection (ILI) tools, i.e. devices equipped with sensor packages designed to be propelled through a pipeline while in service, have been run since the 1970s and are one of the assessment methods Congress listed in the 2002 Pipeline Safety Improvement Act. [Ulrich, 1996] Of the three primary candidate inspection techniques specifically named in the Act, ILI tools can supply the most comprehensive picture of the condition of a buried pipeline.

The current generation of ILI tools is dominated by two primary sensing technologies, Magnetic Flux Inspection (MFL) and Ultrasonic (UT). [Caleyo et al., 2007] The MFL technique, which is the focus of this thesis, is based upon the Hall Effect principle, whereby magnets on the tool creates a magnetic field that saturates the

Figure 1-1: ILI Tool being loaded into a pipeline [Offshore Mag., 2012]

ferromagnetic pipe wall (Figure 1-2a). Any surface-breaking defects on the inner or outer pipe wall present magnetic permeability obstacles to the magnetic field, allowing the flux line to leak around the defect (Figure 1-2b). The dimensions and location of the defect can then be inferred from the logged changes to the magnetic leakage field. [Shi et al., 2015]

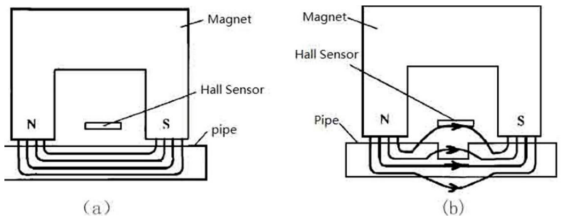


Figure 1-2: Hall effect principle [Shi et al., 2015]
In (a) the induced magnetic flux is undisturbed. In (b), a defect causes the the flux to leak from the pipe body, which is detected by the Hall sensor.

As with other sensors that infer physical properties based on the processing of inverse signal data, MFL tools are subject to uncertainties in their reporting capabilities. These are usually provided by the tool vendor in terms of accuracy associated with confidence levels, i.e. 80% of the reported features will be within ±10% of wall thickness of the true feature depth. (See Figure 1-3 for an example.) Pipeline operators must then integrate this specification with the reported data to make risk justified decisions about where to deploy resources to investigate and repair features.

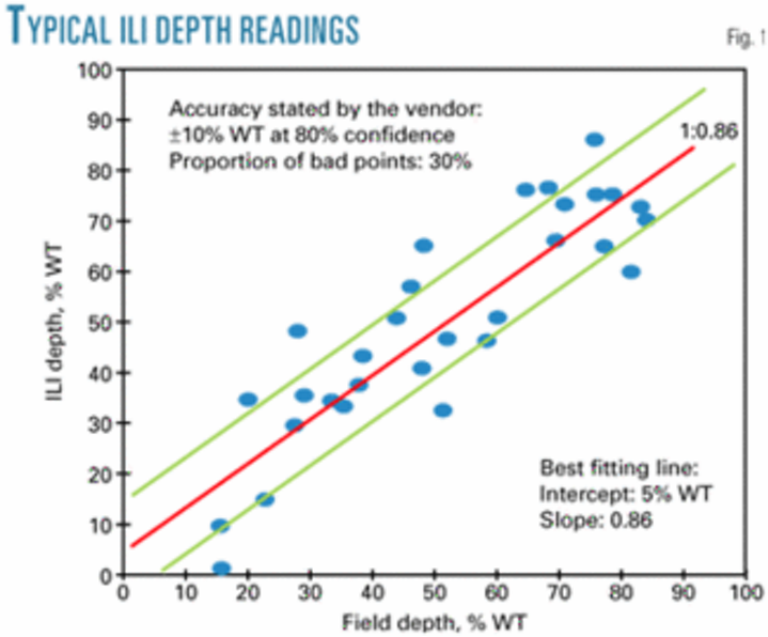


Figure 1-3: Unity Plot Example [Caleyo et al., 2004]
Here, field readings for corrosion depths are plotted against the ILI calls, both as a percentage of wall thickness (WT). The percentage of points that fall in between the green error bands, which are offset from the intercept by 10% on either side, is compared against the stated confidence interval of 80%.

The consequence of having inaccurate tolerances can lead to wasted effort digging anomalies that pose no integrity threat but were over-called by the tool, or missing features that were under-reported and thus allowed to reach failure.

Therefore, a critical portion of the in-line inspection process is the validation process to confirm the accuracy of the inspection tool on a per-run basis. The outcome of this process, which can range from outright rejecting an ILI inspection dataset in favor of a re-run to revising the expected depths for a corrosion population, has immediate effects on the risk management and decision-making processes for pipeline operators.

1.1.1 Corrosion Mechanism

According to data from PHMSA [PHMSA, 2018], corrosion accounts for 18% of pipeline incidents between 1998-2017. Corrosion can manifest as several different morphologies depending on the mechanism that begins it, these include: pitting, uni-

Figure 1-4: Pitting Corrosion [Baker, 2008]

form, selective seam weld, and electrical interference. (Some of these are illustrated in 1-5 Factors that can accelerate the process of corrosion include anything that exposes the bare metal in a pipe body to a surrounding moist environment, including air or soil. Also, corrosion that occurs on the pipe's internal surface may require different responses than externally based corrosion. However, MFL and UT tools are both able to detect both internal and external corrosion anomalies given the proper conditions such as a clean inner diameter free of obstructions and properly sized sensors capable of covering the full wall thickness. [Baker, 2008]

For external corrosion, pipeline operators can use a combination of coatings and cathodic protection to prevent or delay the process of pipe degradation. However, both coatings and cathodic protection each have specifications and design limitations and require proper maintenance. Even on a new pipeline system with modern Fusion Bonded Epoxy coating, the smallest discontinuity in the coating application can serve as a seed for future corrosion pits, which can quickly grow in a damp soil environment. Because of this, inspection methods are a key part of a pipeline integrity management program.

1.1.2 Considerations for MFL Inspections

In recent years, the development of improved processing techniques, higher sampling rates, and novel sensor placement have helped to improve the accuracy of

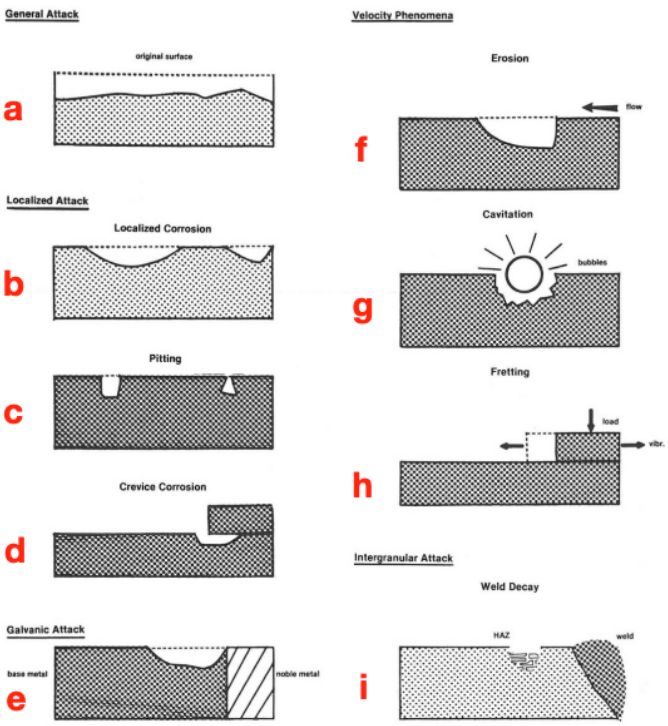


Figure 1-5: Examples of different forms of corrosion [Dillon, 1982] Pitting corrosion (c) can be difficult for MFL tools to detect given the small surface area out of which flux can leak out of. Proximity to dissimilar metals in the Galvanic corrosion (e) and Weld Decay (i) can create noise that obscures the true MFL signal.

MFL-based technologies. [Peng, 2020] However, from a first-principles standpoint, all MFL tools are governed by the same physical limitations of needing to sample the Hall sensors at a high enough frequency while the magnets saturate the pipe wall. [Dutta and Ghorbel, 2008]

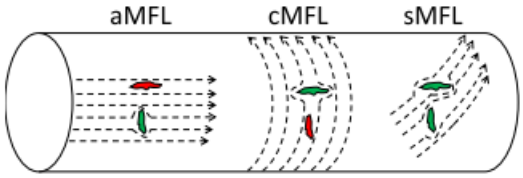


Figure 1-6: The effect of directionality of the MFL signal [Peng, 2020]
The features in shown in green would are detectable by the respective tool. From left to right; Axial, Circumferential, and Spiral MFL.

The following are some of the considerations that must be considered when analyzing MFL results:

- Sensor lift-off: Physical obstructions on the interior of the pipe can prevent the sensor from being within the required proximity
- Nearby magnetic interference: Casings, repairs, or other residual magnetism

can create noise which obscures true defects

- Tool speed: A faster tool travel speed reduces the sampling rate over a given length of pipe and therefore lowers the resolution in that section
- Orientation of the defect's primary axis: Since the direction of the MFL magnets are fixed before the run, and since the magnetic field is a directional property, tools can be "blind" to defects that are oriented parallel to the direction of the field, as shown in figure 1-6

Corrosion anomalies that have one or more of these considerations in play will be referred to as "challenged" features since they can present difficulties for conventional MFL tools.

Finite element modeling and calibration spools (where the performance of the tool can be trained upon engineered features of known dimensions) have also helped to improve the state of the art. However, the primary mechanism for informing tool verification has historically come from in-field excavation and in-the-ditch Non-Destructive Examination (NDE) using pit gauges, UT measurements, and laser profilometry. In the case of pit gauges and UT measurements, these are point values which require the operator to manually select and size the deepest point of a corrosion field and judge what the center of the abscissa is, as in Figure 1-7. Conducting these verification digs can be a substantial undertaking, requiring $20,000 or more per inspection site, on top of the safety considerations associated with having a crew excavating near live lines. [Baker, 2008] The Pipeline Operator Forum Guidance on Field Verification Procedures for In-Line-Inspection [Pipe Operators Forum, 2012] describes the procedures for in-field measurements.

1.1.3 Existing Validation Processes

Several existing industry standards that guide the validation process, including the API (American Petroleum Institute) Standard 1163 In-line Inspection Systems Qualification, CEPA's (Canadian Energy Pipeline Association) Metal Loss Inline Inspection Tool Validation Guidance Document, and the Pipeline Operators Forum Procedures for In-Line-Inspection.

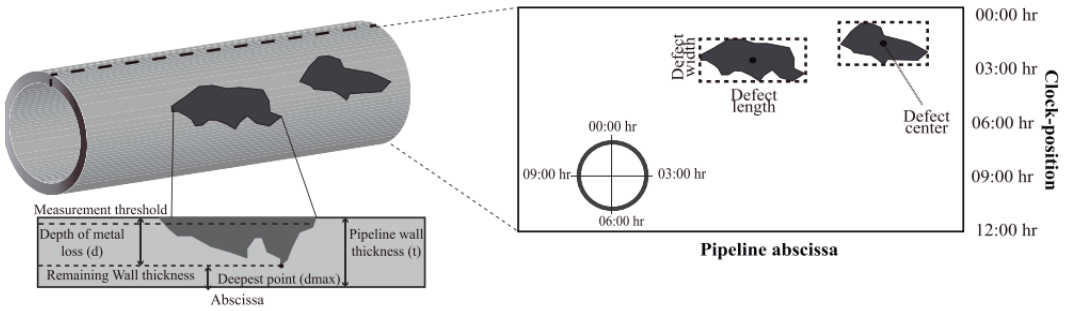


Figure 1-7: Corrosion Parameters [Amaya-Gómez et al., 2019]
Parameters used to describe corrosion anomalies on a ILI tool report and how they correspond to location on the pipe body. Of note are the depth of metal loss, pipe wall thickness, length, and width.

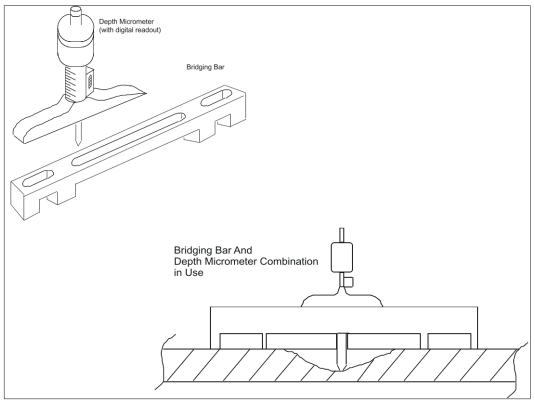


Figure 1-8: Schematic of a micrometer [Pipe Operators Forum, 2012]
A common instrument for manually gauging the depth of corrosion features found in the field. As the gauge proves a point measurement, the burden is on the field operator to correctly identify the deepest pit to size.

API 1163 [American Petroleum Institute, 2013] provides for three methodologies for validation: based on historic data (prior excavations), full-scale testing (running the ILI tool through test pipe with known defects), and small-scale testing, modeling, and analysis (Finite Element Analysis). The performance of the ILI tool established by any of the methods is tied to the essential variables of the tool, which include operational constraints (e.g. tool velocity), inspection parameters (e.g. magnet strength), and sizing system components (e.g. sensor type and spacing).

Section 8 of the standard covers three successively higher levels of System Results Verification:

- Level 1: Use historical data or vendor-stated tool performance. Unable to reject run validity. Only to be used for anomaly populations with low levels of risk.
- Level 2: Use field measurements to check stated tool specifications. Allows for stating whether the tool performance is worse than the specification and therefore can be rejected, but unable to state with confidence that tool performance is within specification.
- Level 3: Extensive measurements available to state as-run tool performance. Statistical processing available.

These levels, and the actions required to achieve them, are outlined in Figure 1-9 below.

API 1163 lists Performance Specifications that must be provided by the service provider, which covers anomaly types, detection thresholds, sizing characteristics, location accuracy, and other limitations. The standard acknowledges that for corrosion features, geometry and orientation can be factors that interrelate with probability of detection and sizing capabilities. The standard requires that the performance specification indicate, as a function of anomaly type, the *tolerance* (e.g. $\pm 10\%$ of wall thickness and a certainty 80% of the time, as shown in Figure 1-3).

1.1.4 Unity Chart

Annex C of API 1163 introduces the concept of a "unity" graph to plot the relationship between ILI-reported characteristics (including depth) as compared with the measurements found in the field. A unity line as well as stated confidence bounds are added to show where ideal measurements should fall if both tools are performing as expected and in agreement. Evaluating the percentage of paired data points that fall between the dashed lines provides pipeline operators with data about whether to accept the ILI run. Specifically, features that fall significantly under the line are over-called (ILI reported the depths deeper than they were found in the field) while features that fall above are under-called (field data show the feature to be deeper than the ILI reported). In either case, the error can be an additive constant (common to

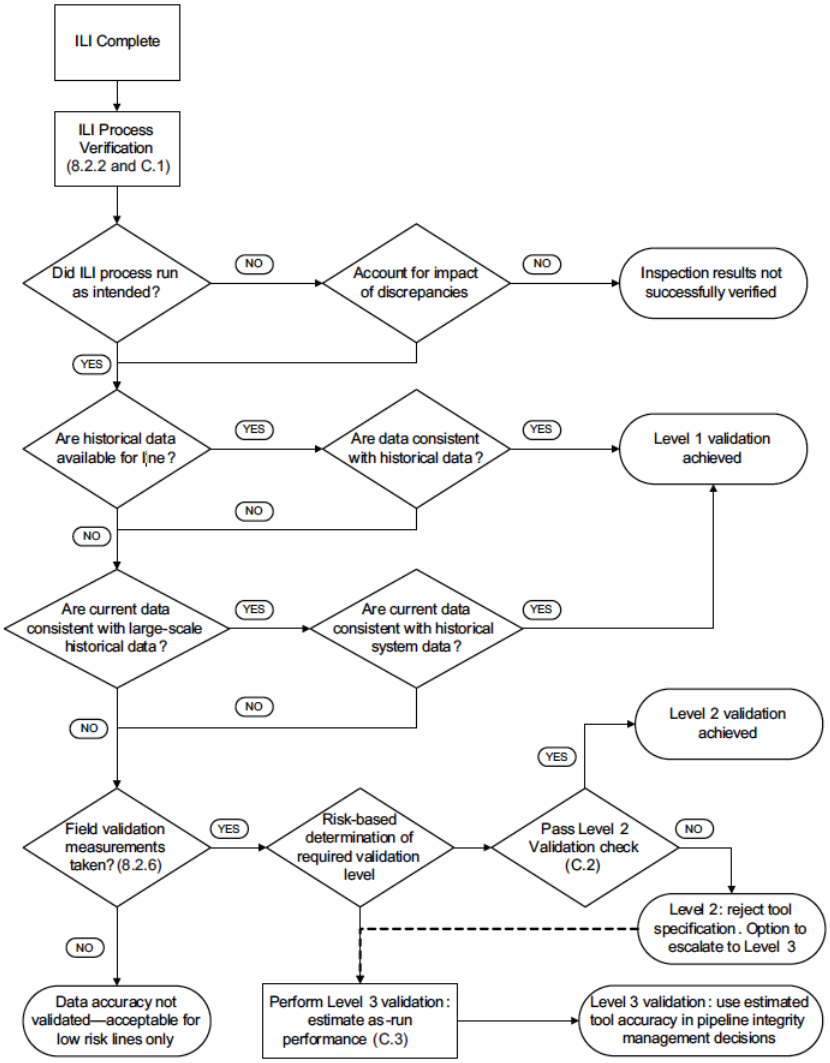


Figure 1-9: API 1163 Validation process with the corresponding requirements for data consistency to achieve Levels 1-3

all data points) or multiplicative bias proportional to the depth. These are equivalent to the parameters of the best-fit line that matches the data points, and for ideal tools the constant bias is close to zero and the multiplicative bias is one, reflecting a 1:1 match between field observed data and the ILI-reported depths.

CEPA's Guidance Document [CEPA, 2016] follows the same processes, but Appendix A includes additional discussion on criteria for handling systemic bias, determining target tolerance, and guidance for sample size. While providing rigorous definitions of these terms, the ultimate focus of the document is primarily on how to

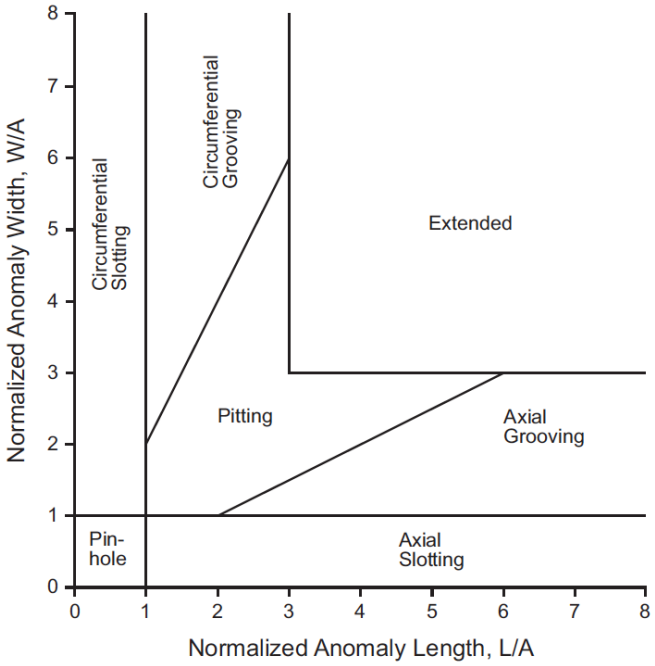


Figure 1-10: API 1163 Dimensional Classes for Metal Loss Indications
Features that are narrower in one dimension (grooving, slotting), or are very small (pinhole), can be difficult for ILI tools to detect and size.

	Extended Metal Loss	Pitting	Axial Grooving	Circumferential Grooving
In Pipe Body	±10%	±10%	+10%/-15%	+15%/-10%
In Vicinity a Weld	±15%	±15%	+15%/-20%	+20%/-15%

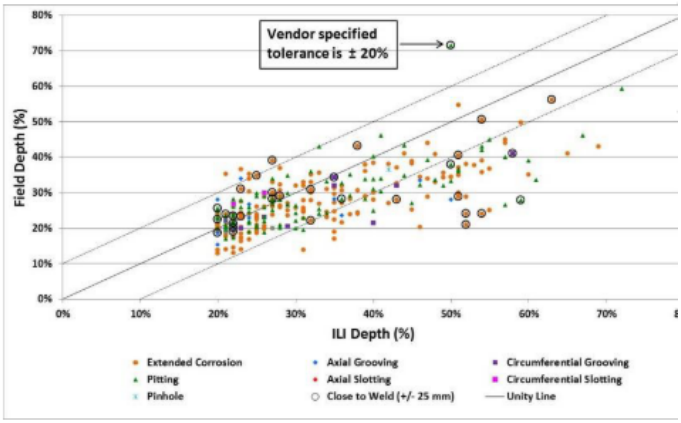


Figure 1-11: Example of a Unity Chart with different performance specifications [Li et al., 2016]
Another example of a Unity Chart, showing different classification of features per the API 1163 recommended classifications shown by Figure 1-10. As can be seen in the table above, sensor lift-off near welds and narrow features result in larger tolerance ranges.

accept the parameters provided by the tool vendor instead of a statistical treatment of determining the data from field verification. Applications of both the API 1163 and CEPA processes have been described in [Salama et al., 2012], [Mora et al., 2004], and [Li et al., 2016].

1.2 Problem and Goal

While there are established methods for using distributions to determine whether or not correlated field data can be used to reject or accept a population of corrosion anomalies generated by an MFL ILI, individually calibrating the population of ILI defects based on the performance of a small sample of field verified anomalies remains an open area of study. However, a more tractable related sub-problem is determining the worst-case tolerance that should be applied to the ILI-provided population. Due to the high-risk nature of pipelines, tolerances must be factored in when making risk-based decisions. Confidently being able to reduce the tolerance applied to select features can help direct operator resources to address the right features.

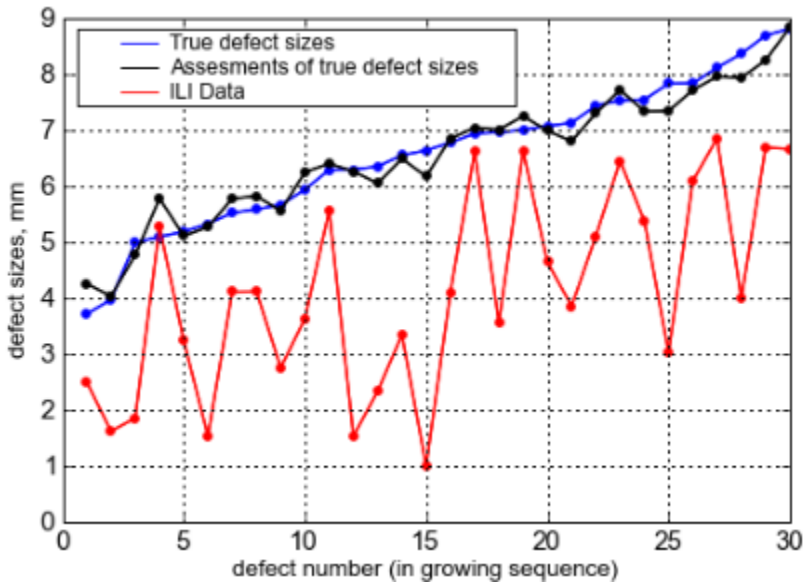


Figure 1-12: Calibration Example [Timashev and Bushinskaya, 2009]
In this case study, the true defect depth (blue) is compared with ILI measurements (red). The ultimate goal for calibration approach is to allow operators to "scale" the ILI Data to recover the black line approximation of the true depths.

1.3 Thesis Overview and Contribution

The main contribution of the present work is integrating a Bayesian Inference model into the validation process and examining how the model works on both base data as well as a data set representing challenging conditions. Chapter 2 presents a review of existing literature. Chapter 3 discusses the proposed Bayesian Inference method and the method of generating simulation data. Chapter 4 reviews the results in comparison to current practices and evaluates how the method can fit into existing Integrity Management programs. Chapter 5 presents conclusions drawn from the work and recommendations for future work.

Chapter 2

Literature Review

This literature review builds upon the context established in the Introduction, focusing on recent efforts to apply statistical techniques to improve the validation process.

2.1 Background

Given the potential benefits that can be realized by accurately calibrating the in-line inspection results and the advances in readily available computing power, recent emphasis has been placed on statistical techniques to improve the validation process. In general, the existing literature can be grouped into concentrated efforts in the following areas:

1. Identifying when to accept or reject an ILI run
2. Determining the minimum number of excavations required to fully validate a tool run
3. Improving the Corrosion Growth Rate assumption between runs
4. How to properly attribute measurement errors between field and ILI tools

These are all interrelated efforts towards the ultimate goal of reducing the measurement uncertainty surrounding the tool reported anomalies. Measurement uncertainty can be grouped into epistemic (systemic) and aleatory (random) categories shown in Figure 2-1. Sources of epistemic uncertainties include model error, human

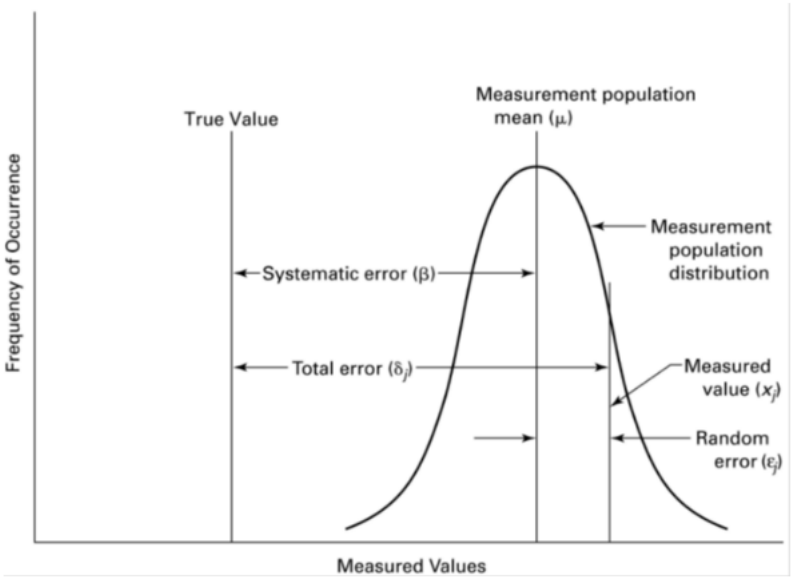


Figure 2-1: Distribution of repeated measured observations showing how systematic and random errors describe the relationship to the true value. [Tomar et al., 2009]

factor application variances, and tool biases. Epistemic uncertainty is not random and can be reduced with a more complete understanding of the system. By contrast, aleatory uncertainties may never be eliminated because of their nature, and examples include "variation in pipe manufacturing processes, signal fluctuations in sensor data collected... localized pipeline operating pressure variations and precise voltage levels in cathodic protection (CP) systems." [Cheng et al., 2016] These uncertainties are limited by the kinds of tools used, and as inspection technologies mature, the errors associated with these categories are reduced.

Tool calibration is a process to reduce and correct for epistemic uncertainty while providing bounds on the aleatory uncertainty. However, the process is made complex by the fact that actual run conditions, flaw morphologies, and sensor performance can vary from the ideal laboratory conditions under which a tool's sensors are qualified. Additionally, the true depth of a feature is never known unless a destructive test is done as in-field tools introduce uncertainties as well.

Stated mathematically, the following two equations capture the relationship between the actual feature depths and the depths observed by the field measurement

and ILI tools for a given anomaly (i):

$$\begin{aligned} dm_i &= \alpha + \beta da_i + \epsilon_i \\ df_i &= da_i + \varphi_i \end{aligned}$$

Where:

dm_i is the ILI measured anomaly depth

da_i is the actual anomaly depth

df_i is the field reported depth

α is the constant bias

β is the non-constant bias

ϵ is the ILI scattering error

φ is the field scattering error

Given perfect knowledge of all the components of the system, one would expect to be able to determine the α and β that describe how a tool trends features, removing all epistemic uncertainty. And if an ILI tool and in-field NDE were able to get repeated observations of the same feature under ideal conditions, one would expect ϵ and φ to be driven down to zero. The proposed techniques summarized below attempt to achieve these aims using statistical techniques.

2.2 Efforts to Improve Calibration

Joshi [Joshi, 2011] claims that a basic assumption in API 1163 is that uncertainties in the tool measurement error follow a Gaussian distribution and the performance verification method is based on a theory of binomial distribution for accepting/rejecting a hypothesis. His work proposes that error distributions may be asymmetric around a non-zero mean and infers a confidence level by fitting both the ILI and the field population to two different Gamma Distributions and taking the absolute value of the difference. While it would seem that this approach trades one set of assumptions about the measurement error for another, the useful insight is that there needs to be

flexibility in the assumed population distributions used by the validation model and that normality cannot be taken for granted.

Desjardins et al. [Desjardins et al., 2007] also present slight modifications to the standard API 1163 process in which hypothesis testing is used to determine whether to accept or reject the stated performance of the tool. Their additional contribution is a proposal to calibrate the result by removing the tool bias, which is determined to be the average difference between the ILI and field measurements. Their case study shows that doing so lowers the expected tolerance of the ILI tool from $\pm 10\%$ to $\pm 5\%$. Key assumptions made during their analysis was that the distribution error of the ILI tool is normal (consistent with API 1163 but not with Joshi), there is no bias in the field measurement, and the field tool's accuracy is much greater than the ILI tool.

2.3 Attribution of Measurement Error

Another approach to improving the accuracy of the ILI tool performance revisits the assumptions of the ILI tool error normality and the proper accounting of field measurement error. The former is important to help identify what may be outlier data that are included with the general population and can skew the tool performance. The latter is significant because if field measurement error is large, the ILI can be unnecessarily penalized for with a higher standard deviation that does not accurately represent the performance. Both are discussed in detail in the following sections.

2.3.1 ILI Measurement Error

Several works have investigated obstacles to the normality assumption in the measurement error of the ILI tool and the issues fall into two broad categories: depth-based and classification-based biases.

A case study in applying API 1163 guidance conducted by Smart and Haines [Smart and Haines, 2014] found that taking the entire corrosion population of an ILI run into account led to a lack of a normal distribution. Segmenting the corrosion anomalies into different classifications (slotting, pitting, etc.) did not reveal a dis-

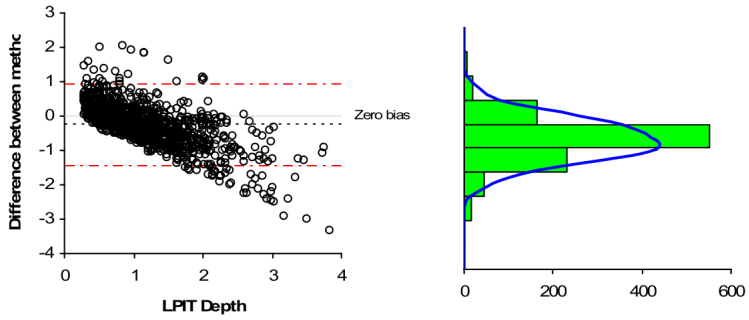


Figure 2-2: Example of the bias for pit depth [Tomar et al., 2009]
As the corrosion features get deeper along the depth percentiles in the left graph, the bias trends more negative. This corresponds with the ILI tool underestimating the size of larger defects.

cernible pattern. Further investigation by plotting the depths by mileage (to isolate soil conditions or other localized factors) showed that the measurement error gets larger as the true depth gets deeper. By grouping features into two populations (above and below 50% wall loss), each population was closer to a normal distribution. For the 50% and greater depth population, the lowest depth error that was determined to be ±20%, in contrast with the common vendor provided specification of ±10%. A 2012 study by Salama et al. [Salama et al., 2012] also notes that MFL tools over-estimated small defects and underestimated large defects, but concluded that MFL performance was between 12-21%wt. Dann and Huyse [Dann and Huyse, 2018] also provide evidence to support this depth-bias, noting that for errors in the top percentiles of measured features are highly affected by an over-sizing bias in a manner similar to Figure 2-2. They also note the effect such errors can have when propagated through to corrosion growth rates.

As discussed in Section 1.1.4, API 1163 provides a graphical tool for categorizing corrosion anomalies into various classifications that accounts for the physical limitations of the MFL tool in sensing anomalies that are oriented perpendicular to the sensor. (Figure 1-10) Ludlow [Ludlow, 2012] notes that when considering classification-based biases, sampling theory plays a significant role. This is because the Field verification program can only address a small percentage of the total population of ILI features. If each classification population must be treated differently, as

Table 2.1: Case Study using API 1163 Classifications

Geometry	Count	Mean Error	Standard Deviation of the error	Corresponding Tolerance @ 80% certainty (random only)
Axial Grooving	1	NA	NA	NA
Axial Slotting	6	-1.67	5.54	7.09
Circumferential Grooving	11	-6.15	7.44	9.52
Circumferential Slotting	14	-6.77	4.57	5.85
General	16	-6.18	4.83	6.18
Pinhole	86	-1.13	4.34	5.55
Pitting	31	-8.24	6.46	8.27

The tolerance is wider for challenging classifications like Circumferential Grooving features than the base tolerance established for General Corrosion. Reproduced from [Haines and Tomar, 2013]

shown by Haines and Tomar [Haines and Tomar, 2013] in Table 2.1, then it is likely that classification with a smaller number of constituent anomalies may not have a statistically significant amount of field verifications to allow conclusions to be drawn.

2.3.2 Field Measurement Error

McNealy et al. [McNealy et al., 2010] surveyed the state of the art of common field measurement tools, which are shown in Table 2.2. Of note, both the measurement error and tool bias of the pit gauge are much more dependent on operator training and application, hence the large range in variability. In the process used by Enbridge, Li et al. [Li et al., 2016] noted that ultrasonic thickness gauges have expected tool tolerances of ± 0.31mm and pit gauges have tool tolerances of ± 0.50mm.

Table 2.2: Performance ranges of Field Measurement tools

Tool	Error	Certainty
Pit Gauge	$\pm 1.3\% - 10\%$	80%
Ultrasonic Pen Probe	$\pm 2\%$	80%
Laser profilometry	$\pm 1.5\%$	80%

Summarized from [McNealy et al., 2010]

Ellinger et al. [Ellinger et al., 2016] used 3,009 matched pairs of ILI and field depth measurement and noted that as actual field depth increases, MFL ILI tool accuracy decreases. This occurred in both the ILI reported measurements as well as the field-called depths. Anomaly length was not found to have a meaningful correlation between reported and field-measured, but very short or very long features were reported to have lower performance.

2.3.3 Statistical Tools to Improve Error Attribution

Two established statistical tools for decomposing the scatter between two measurement systems with non-zero measurement errors looking at the same feature include the Grubbs method, developed in 1948 [Grubbs, 1948], and the more recent Jaech method developed in 1981. The Grubbs method assigns components of constant biases and measurement errors to both tools. Then, using the variances and covariances of the individual tools, the variance of the measurement error estimator $\hat{\sigma}_j^2$ assigned to the j-th tool is given by:

$$V(\hat{\sigma}_j^2) = \frac{2\sigma_j^4}{n-1} + \frac{\sigma_\phi^2\sigma_1^2 + \sigma_\phi^2\sigma_2^2 + \sigma_1^2\sigma_2^2}{n-1}$$

Where:

$V(\hat{\sigma}_j^2)$ is the measurement error estimator of the j-th tool
σ_j is the standard deviation of the j-th tool
σ_1 is the standard deviation of the 1st tool
σ_2 is the standard deviation of the 2nd tool
σ_ϕ is the covariance between the two tools

The Grubbs estimator provides no value in the case of a "negative measurement error," which can occur when the ratio between the variances of the two tools becomes larger than three. To correct for this, the Jaech approach provides an alternative by using a constrained expected likelihood formulation that integrates a sharing function and the total scatter. [Morrison et al., 2000] However, the Jaech formulation becomes complicated in the three tool case and is largely intractable for the four or more tool case.

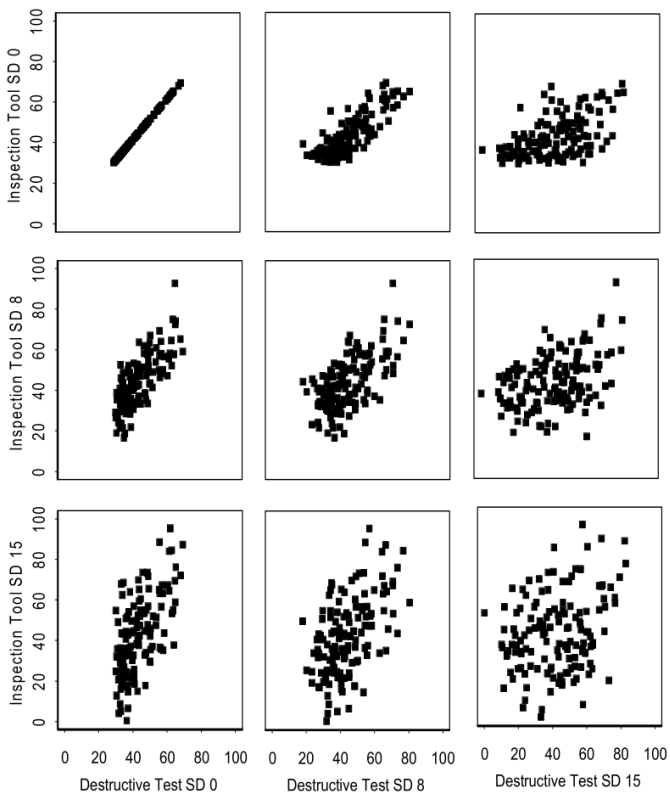


Figure 2-3: Illustration of relative scatter between two tools [Worthingham et al., 2002] The correlation signal can be lost in the noise as the scatter on tool one (X axis, left to right) or tool two (Y axis, increasing top to bottom) increases.

2.4 Models for Improving Validation

A Bayesian approach is proposed by Worthingham et al. [Worthingham et al., 2002], in which the measurement error for the two tools can be described using a multivariate normal distribution and the attribution of the error between tools is sampled from a posterior variance-covariance Inverse-Wishart matrix. This Bayesian approach is one of two broad approaches to stepping beyond simply deciding whether or not to accept the vendor-stated ILI performance specification and towards a statistical calibration of the ILI data. The other approach falls under the umbrella of the Error-in-Variables Model proposed by Caleyo et al. [Caleyo et al., 2004] and subsequently refined and reused by multiple works.

2.4.1 Error-in-Variables Model

The classical Error-in-Variables (EIV) procedure shown in Figure 2-4 can be used to determine the slope of the best-fit line when both the ILI and field instruments have

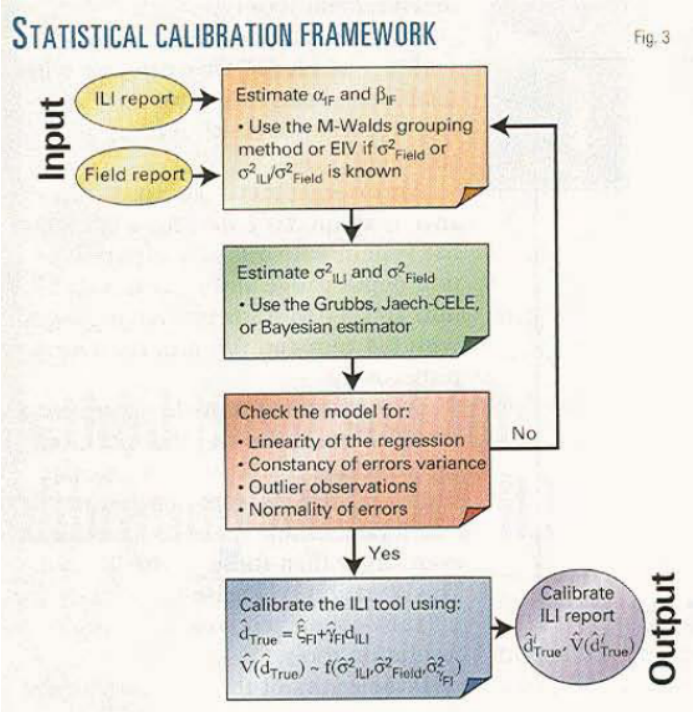


Figure 2-4: Calibration of ILI data using the Errors-in-Variables Model [Caleyo et al., 2004]

known errors. What complicates the application of the procedure is when the errors (especially systemic) are still unknown, which can be the case when the first batch of field verifications for an unproven ILI run are received. In this case, the fit of an Ordinary Least Squares model may or may not warrant introducing a constant bias to the ILI tool.

Caleyo et al. [Caleyo et al., 2005] address the uncertainty in the parameters of the calibration model line by using a variation of the Wald grouping method, where the independent variable is partitioned into two groups and a straight line is passed through the mean points of the groups. From this, values of β and α (which describe the linear correlation between field and ILI measurements, i.e. the Unity Plot line from API 1163) can be calculated. In comparing the Wald method to the traditional EIV model, an alternative formulation of β was proposed to address the case where measurement errors for both tools were large and similar.

2.4.2 Bayesian Efforts

Worthingham et al.'s initial use of the Bayesian estimator was primarily to compare the Grubbs and Jaech estimators for multiple tool cases. While they identified that confidence intervals for measurement errors using the Bayesian method may be tighter than a frequentist methodology, that question was outside the scope of their study.

Subsequent works in applying the Bayesian method in both the Al-Amin et al. [Al-Amin et al., 2012] and Dann and Huyse papers focus on improving the understanding of the relationship between in-line inspection years. Dann and Huyse introduce a combination of deterministic and probabilistic models to account for sizing bias in the corrosion growth process, but the application of the Bayesian model was focused on estimating the corrosion growth rate, not on the error sizing process.

The Al-Amin et al. work aimed calibrate the parameter of multiple ILI tools simultaneously and then compare the tool performance in relation to one another. The improvements over the Grubbs and Jaech estimators it the ability to account for systematic bias in one or both tools. In contrast with the EIV model, the bias parameters α and β are sampled from the product of the observations and prior distributions informed by previous experience using a Markov Chain Monte Carlo method. Like Worthingham et al., the Wishart distribution is used as a prior for the multivariate generalization of the variance-covariance matrix. Stated succinctly:

$$\begin{aligned} dm_{ij} &\overset{ind}{\sim} MVN(\mu_i, \Sigma_\epsilon) && \text{i=1,2,..,m} \\ \mu_i &= \alpha + \beta df_i \end{aligned}$$

Where:

dm_{ij} is a m x n matrix of the i-th defect inspected in the j-th year
$\overset{ind}{\sim} MVN$ indicates the assignment of a multivariate normal probability distribution to a random variable, with independency
μ_i is the mean matrix of estimated calibrated field data for the j-th year
Σ_ϵ is a n x n covariance matrix of scattering error between paired years
α is a n x 1 vector of constant biases determine for each j-th inspection year
β is a n x 1 vector of non-constant biases determined for each j-th inspection year
df_i is a m x 1 vector of field data for the i-th defect

By applying this Bayesian model to a case study for a single pipeline with multiple successive inspections, the authors were able to make relative statements about the performance of the ILI tool based on the expected parameters. The parameters were used as the foundation for a hierarchical Bayesian corrosion growth model.

Table 2.3: Calibration line parameters results from Bayesian model proposed by [Al-Amin et al., 2012]

ILI 2004 (Vendor A)			ILI 2007 (Vendor B)			ILI 2009 (Vendor B)			ILI 2011 (Vendor A)		
α_1 (%wt)	β_1	σ_1 (%wt)	α_2 (%wt)	β_2	σ_2 (%wt)	α_3 (%wt)	β_3	σ_3 (%wt)	α_4 (%wt)	β_4	σ_4 (%wt)
2.04	0.97	5.97	-15.28	1.40	9.05	-10.38	1.13	7.62	4.84	0.84	5.94

	ILI 2007 (Vendor B)	ILI 2009 (Vendor B)	ILI 2011 (Vendor A)
ILI 2004 (Vendor A)	$\rho_{12} = 0.70$	$\rho_{13} = 0.72$	$\rho_{14} = 0.82$
ILI 2007 (Vendor B)	-	$\rho_{23} = 0.78$	$\rho_{24} = 0.71$
ILI 2009 (Vendor B)	-	-	$\rho_{34} = 0.74$

In this case study, the Bayesian approach was simultaneously applied to four tools run on the same line to determine the Unity plot parameters. The first table indicates that the 2004 run is the most accurate as α_1 and β_1 are close to zero and unity respectively and σ_1 is lower in comparison to 2007 and 2009. The table below also shows the correlation coefficients ρ_{kl} between all four of the tools. In this case study, the correlation coefficients between the 2004 and 2011 tools and the 2007 and 2009 tools were higher, which is representative of the fact that they were from the same tool vendor.

The present work intends to explore a similar model to the one developed by Al-

Amin et al. However, instead of comparing across multiple years of ILI data, the focus will be on refining the tolerance used within a single campaign while field verifications are on-going in comparison with the API 1163 procedure. Additionally, the effect of anomalies subject to a degraded (wider) tolerance will be explored.

Chapter 3

Bayesian Modeling Approach

In this chapter, the proposed Bayesian Inference model is described, including a brief background into the statistical techniques for handling the stochastic variables that describe the uncertainty. The priors for the model are discussed, as well as the relevant assumptions required by the model. Additionally, the process of creating the two simulated data sets is detailed.

3.1 Motivating Question

The Bayesian Inference model proposed below seeks to examine whether or not the tool tolerance (represented by the standard deviation) can be lowered after performing several field trials. The associated parameters that define the Unity Plot can also provide further insight into the tool's bias. Proper use of these data can have wide-ranging effects since, as discussed in Section 1.2, tolerance can have significant effects on the execution of an integrity management program.

Consistent with API 1163, the underlying data required for this process are still the same: historical dig documentation, calibration test piece, or a subset of features from the current ILI population that are field-verified. The paired ILI-to-verified data will then be used to establish the performance for the rest of the population.

Additionally, the same concerns outlined in Section 2.3 around feature classification and measurement error as a function of depth are still relevant and do not

invalidate this process. For example in Table 2.1, the ILI identified a majority of features with "Pinhole" geometry, for which the tool has a stated tolerance of ±10% but the model has established a tolerance of ±5.55% from 86 data points. The same ILI has identified a smaller amount of "Axial Grooving" features for which the tool vendor has a stated tolerance of ±15% but no verification digs have been performed yet. The pipeline operators can continue to use the vendor stated tolerance as a prior and update their predicted tolerance as more data become available, or they can try to establish another prior either by leveraging a covariance assumption for tool performance on pitting and slotting features established on a prior run.

3.2 Bayes Rule

The Bayesian approach to inference is a method of formally updating a prior knowledge (expressed as a probability) based on observed evidence. In contrast with a frequentist perspective, it allows for unknown parameters of a physical process to be treated as distributed variables rather than deterministic values.

Stated mathematically, the Theorem states:

$$p(\theta|X) = \frac{p(\theta) \times p(X|\theta)}{p(X)}$$

In the context of the present work:

$p(\theta|X)$ represents the posterior, or the final conditional distribution of the random value of interest $(\alpha, \beta, \sigma$, etc...) given the observed data
$p(\theta)$ represents the prior distribution of a variable, based on tool vendor specifications or knowledge of past runs in the line
$p(X|\theta)$ is the likelihood, or the conditional distribution of the observed data given the influence of the random value in question
$p(X)$ is the marginal likelihood in the denominator that serves as a normalizing constant to ensure the final result satisfies the probability theorem

When applied to the validation problem, this process is used to determine the

variables that describe the unity plot line. As discussed in Section 2.1, of interest is the estimated value of sigma, which is used in the calculation for the amount of tool tolerance to add on to the depths of undug features as a conservative estimate of the aleatory uncertainties that exist around the measurement error. For example, a corrosion pit called as 60% deep by a standard MFL tool may be treated as a 70% deep when making decisions about where to excavate unless a smaller tolerance can be justified.

3.3 Markov Chain Monte Carlo

Given the complexity of many applications, the computation of the components of the Bayes Rules may not be readily achievable. Techniques to overcome the lack of closed-form solutions evolved as computational power became more readily available, and the Markov Chain Monte Carlo (MCMC) method is one simulation technique. In this technique, random samples of the uncertain parameters can be generated to form a Markov process whose stationary distribution is the joint posterior distribution of the parameters.

In contrast with conventional memory-less Monte Carlo simulation methods, sample states (which can represent a system with multiple stochastic parameters) generated by each sequence of a Markov Chain are dependent on the results of the prior sequence. More specifically, if a randomly generated sample state has a higher probability than the last, it is added to the chain to be evaluated next. If it does not have a higher probability, it can still be accepted based on the selection of an Acceptance-Rejection algorithm. This process is repeated until the sample states converge.

The present work uses the PyMC3 implementation of the MCMC method. The default acceptance-rejection sampler selected is the No-U-Turn Sampler (NUTS), which is an extension of the Hamiltonian Monte Carlo algorithm.

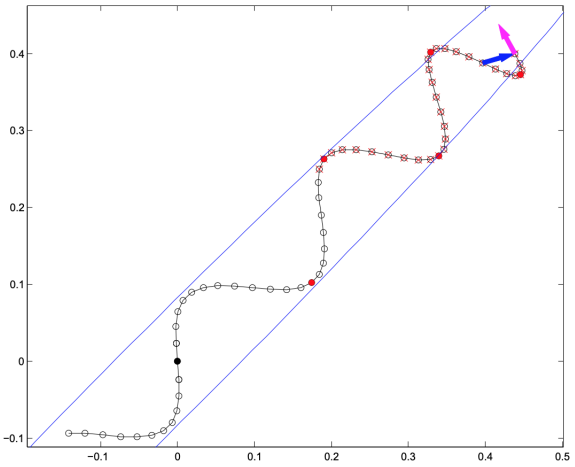


Figure 3-1: Markov Chain Monte Carlo Illustration [Hoffman and Gelman, 2014] NUTS MCMC state evaluation during one iteration. The blue ellipse is a contour of the target distribution, the black solid circle is the starting position, the black open circles are the potential next states, the red solid circles are positions associated with states that must be excluded from the set of possible next samples because their joint probability is below acceptance criteria, and the positions with a red "x" through them correspond to additional states that must be excluded per the NUTS algorithm.

3.4 Model Description

Recall from the discussion in Chapter 2 that the ILI-measured and field verified reported depths can be defined as follows, incorporating various biases and measurement errors:

$$dm_i = \alpha + \beta da_i + \epsilon_i \tag{3.1}$$

$$df_i = da_i \tag{3.2}$$

Where:

dm_i is the ILI measured anomaly depth

da_i is the actual anomaly depth

df_i is the field reported depth

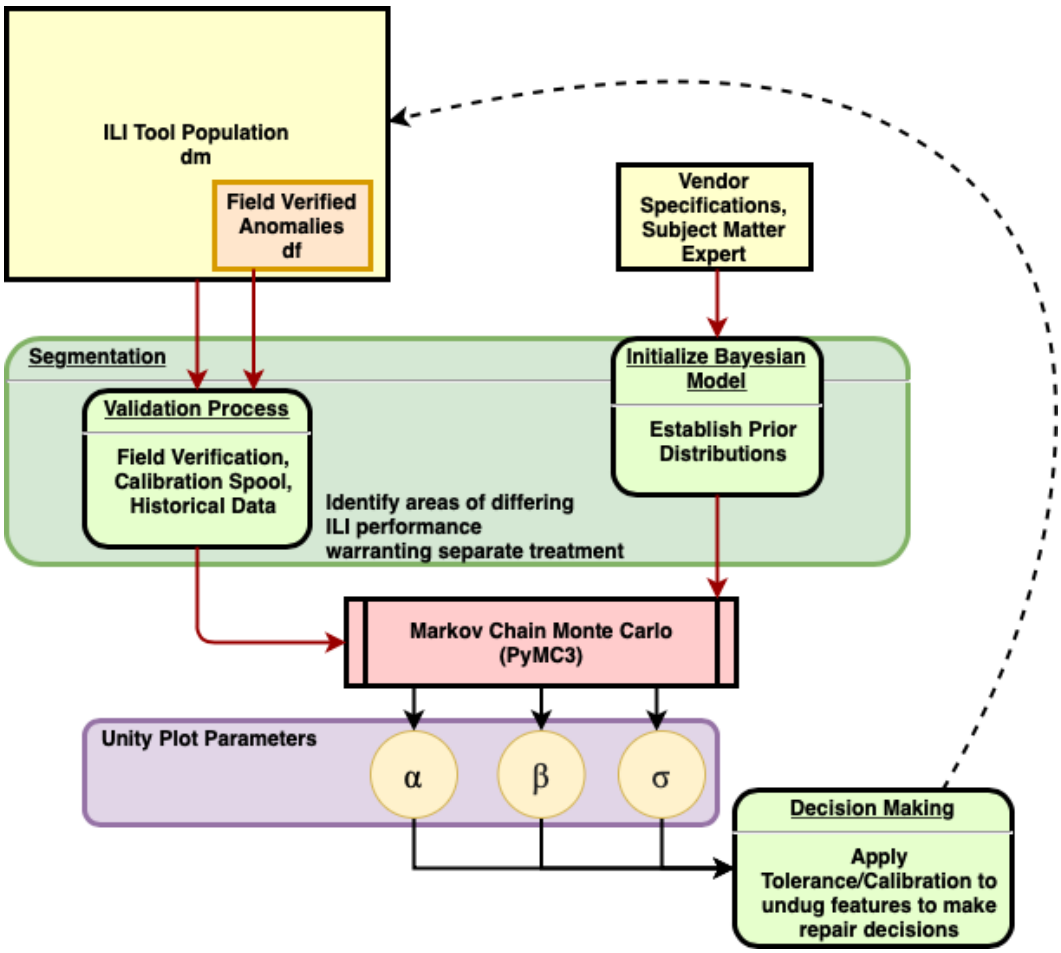
α is the constant bias

β is the non-constant bias

ϵ is the ILI scattering error

$\epsilon_i < 2\sigma$ to be within conventional ILI tolerance intervals of two standard deviations to represent the equivalent of 95% of a normally distributed population

Provided a set of dm_i and df_i, the goal is to determine the best fitting parameters for α, β that minimize ϵ and φ. Below is a directed acyclic graph which summarizes the outline of the proposed model.



1. A subset of the ILI tool population is matched with field data

2. If warranted, populations are segmented according to depth or classification if varying tool performance is expected

3. Priors for the Unity plot line parameters α, β, and σ are established for each population

4. An MCMC is run to determine the posterior distributions for each parameter of the Unity plot line

5. Using the Unity plot described by α, β, and σ, identify the confidence interval that can be ascribed to the tolerance interval of two standard deviations (percentage of data points that fall within the bounds)... i.e. "80% of observed data points fall within ±6.5% of the Calibration line"

6. If satisfactory, use this tolerance for the rest of the undug features in making repair decisions, or identify outliers and re-segment

3.4.1 Assumptions in Model

For this model, field random error was modeled with a much smaller scatter than the ILI tool. The Bayesian model does not attempt to remove the scatter (as in the Grubbs or Jaech method) and instead uses the field depth as an approximation of the true feature depth (i.e. $df_i = da_i$).

In actual scenario with field provided data, the ratio between field and ILI measurement error can be determined against the true value of a manufactured defect in a test spool to establish the appropriate values or a segment that is destructively tested to confirm the true depth. However, on many pipeline projects, true depth data are not available and the field data are established using past metrics of operator qualifications using a particular tool.

3.4.2 Selection of Priors

Recalling Equation 2.1, the Bayesian model requires that priors be specified for each of the stochastic variables. Initial proposals for prior distributions are as follows:

The prior for α is Gaussian distribution with a mean of 0 and a standard deviation of 0.1. This is informed by industry data that the constant bias is often minimized

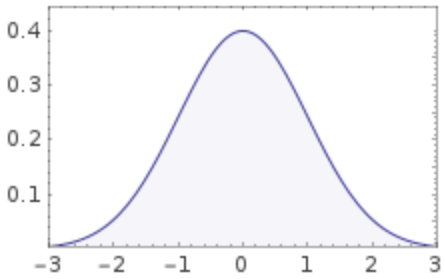


Figure 3-2: Prior Distribution for α

to zero but there may be minor positive or negative deviations. Large values are possible, hence there is no hard limit, but large biases would provide support for rejecting the tool run.

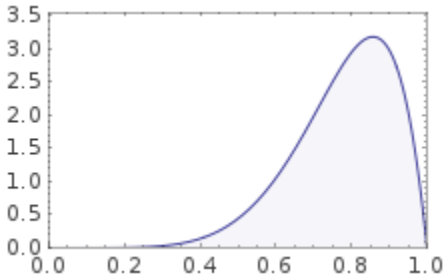


Figure 3-3: Prior Distribution for β

β follows a Beta distribution with shape parameters a=7 and b=2. This leads to a left-skewed distribution weighted closer to values around 1. Ideal ILI performance would lead to a 1:1 bias. From the discussion in Section 2.3.1, literature has shown a strong tendency for ILI tools to undercall deeper features, which are captured by a value of β under 1. Values much smaller than 1 are unlikely.

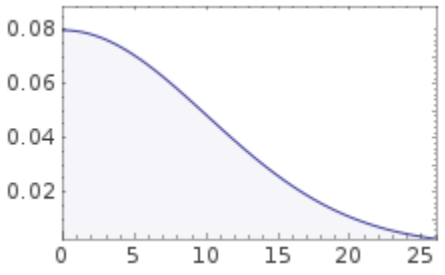


Figure 3-4: Prior Distribution for σ

σ is a Half-Normal Distribution with a standard deviation of 10. This results in

a wide, positively constrained distribution that reflects the uncertainty in the tool tolerance, which can vary from 5-20% depending on the tool vendor specification.

3.5 Generating Test Data

Data are randomly generated according to the following rules: 1000 sample features are generated with actual depth following an Inverse Gamma Distribution. Consistent with the literature review, this reflects that most corrosion features are relatively low-grade, but a small portion can quickly grow due to external factors. [Joshi, 2011] The upper range of features was cut off at 75% to capture the fact that features deeper than 75% are likely already repaired or will have reached actionable criteria to make them eligible for repair regardless of any further analysis.

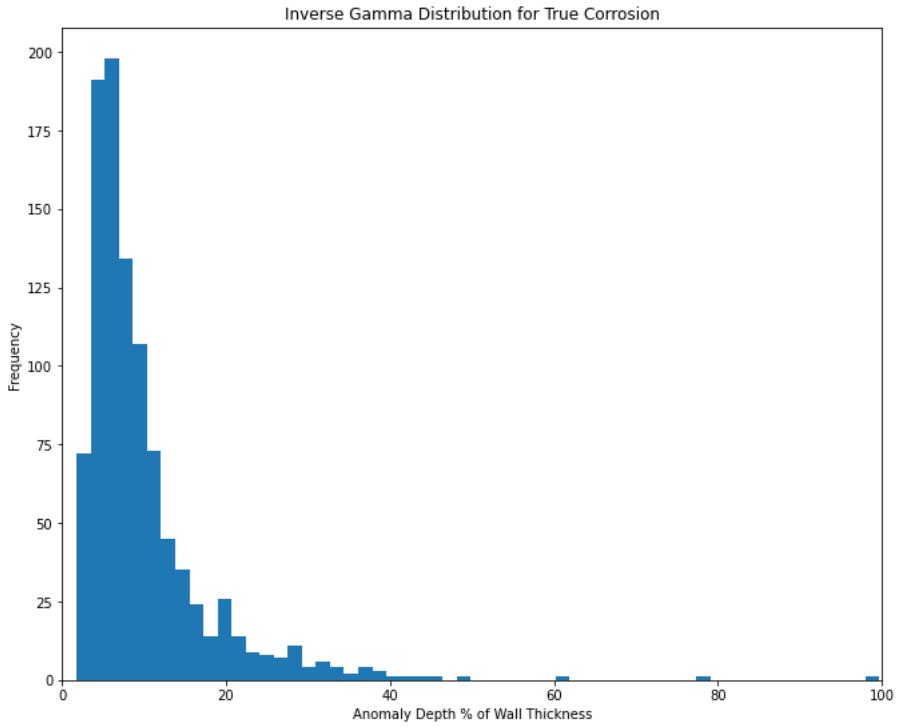


Figure 3-5: Generated population distribution for the corrosion depths

For Base features, a normally distributed scatter with 10% tolerance was used to introduce ILI scatter. This is consistent with the average of current ILI tool performance surveyed in literature. [Salama et al., 2012] For Challenging that represent

sub-optimal MFL conditions, the normally distributed scatter was increased to 20% to reflect the challenging conditions for the operation of the MFL sensor as covered by Section 1.1.2, including small or narrow features or areas where there is sensor lift-off. These features will be referred to as the "Challenged" set. The resulting generated data are shown below.

Performance Specifications

	General	Pitting	Axial grooving	Circumferential grooving
Depth at POD = 90 %	0.10 t	0.10 t	0.10 t	0.10 t
Depth sizing accuracy at 80 % certainty	±0.10 t	±0.10 t	±0.15 t	±0.10 t
Length sizing accuracy at 80 % certainty	±15 mm (0.59")	±10 mm (0.39")	±10 mm (0.39")	±10 mm (0.39")
Width sizing accuracy at 80 % certainty	±15 mm (0.59")	±12 mm (0.47")	±12 mm (0.47")	±12 mm (0.47")

Abbreviations: POD = Probability of Detection; t = wall thickness

Figure 3-6: Example of selective tool performance [Rosen Group, 2013]
In this example, Axial Grooving would be treated as the Challenged feature set while General, Pitting, and Circumferential Grooving would all be included in the Base feature set.

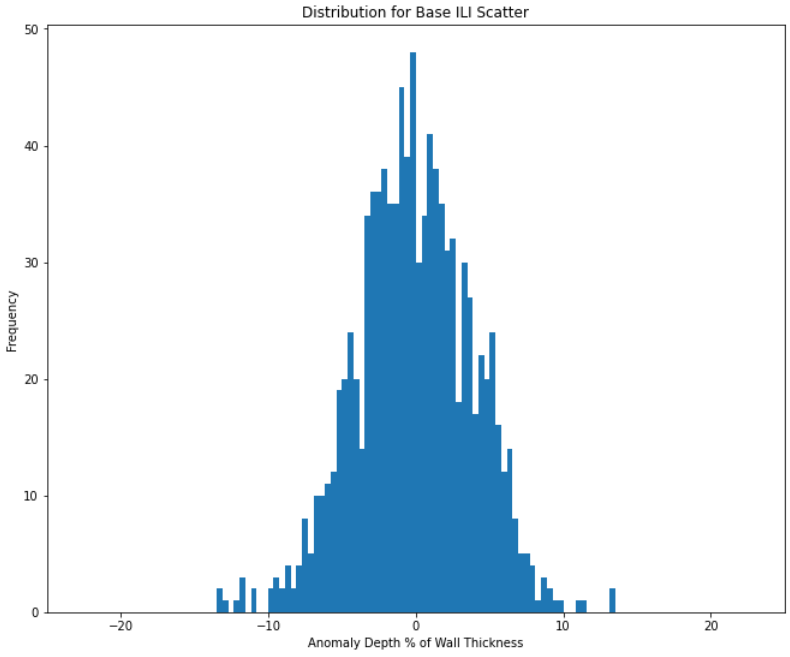


Figure 3-7: Base tolerance with approximately ±10% scatter

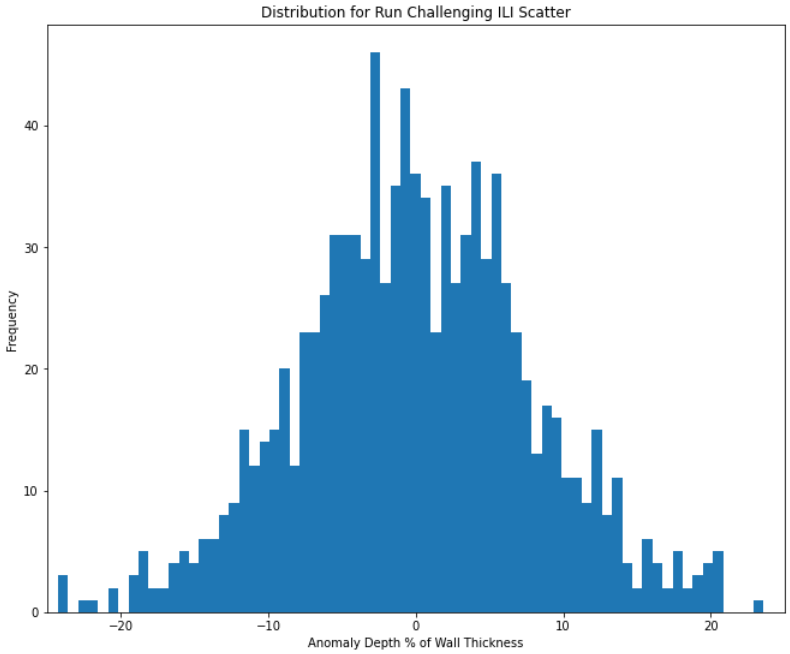


Figure 3-8: Challenged tolerance distribution with approximately ±20% scatter

Unnamed: 0	df	dm	dm2	dtrue
0	9.93966	12.6636	11.1429	11.1429
1	7.69109	11.1755	14.8136	8.73314
2	3.80154	6.5279	4.94826	4.28874
3	3.9366	8.61165	4.28852	4.28852
4	2.51423	6.61991	0.810554	3.28423
5	14.5743	17.7071	9.85316	15.8704

Figure 3-9: First 5 rows of the simulated data
dtrue on the right side is the original dataset generated according to the Inverse Gamma distribution. df represents the field measurement of the true depth with a small amount of scatter added (±3%). df is the reference value used in the Unity Plot. To represent ILI relative performance, scatter was added to the elements of dtrue with the respective tolerance: dm reflects the Base data with ±10% scatter added and dm2 captures the Challenged data with ±20% scatter added.

Chapter 4

Results and Discussion

This chapter discusses the results of the Bayesian approach and compares it to the standard practices outlined in API 1163, namely the Unity Plot. As introduced in Chapter 3, the model was run on two different data sets, one representing features subject to a standard level of scatter representative of Base ILI performance while another represents Challenging features with a larger amount of scatter. The benefits and trade-offs of the model are reviewed against performance specifications. Finally, potential changes required for further applications are discussed.

4.1 Simulation Results

Base Features

Table 4.1: Bayesian Results for Base Tolerance

Variable	Mean	Std Dev
alpha (%WT)	0.241	0.088
beta	0.878	0.010
sigma (%WT)	3.613	0.080

Figure 4.1 shows the posterior distributions for the variables of interest and the trace plots for the MCMC runs for the population representing the population subject to base scattering errors of ±10%. The estimated values for the parameters are

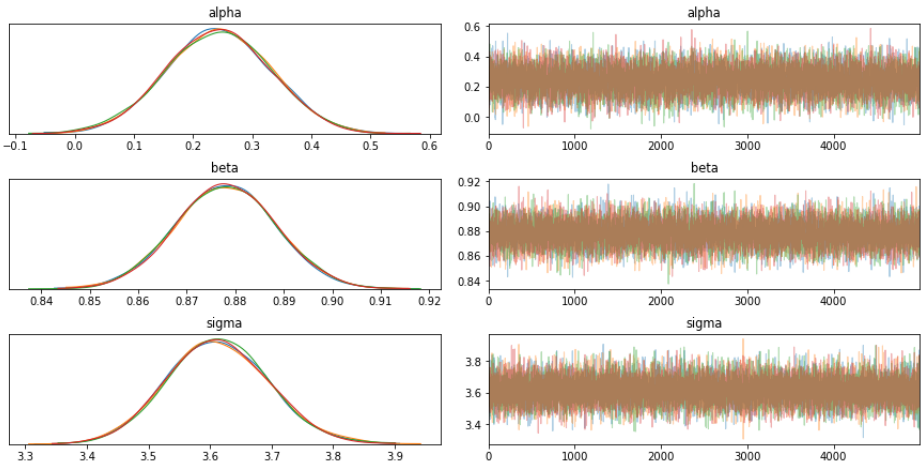


Figure 4-1: Posterior results for the three stochastic variables in the Unity Plot line for the Base case
The left column consists of a smoothed histogram of the marginal posteriors of each stochastic random variable while the right column contains the samples of the Markov chain plotted in sequential order.

summarized in Table 4.1. As expected for a nominally performing tool, there is little systematic bias (α is close to zero) and the non-constant bias β is closer to one. Implementing the common practice of using two standard deviations as tolerance bands to describe 95% of a normally distributed population results in an equivalent tool performance of 7.2%.

Challenging Features

Table 4.2: Bayesian Results for Challenging Tolerance

Variable	Mean	Std Dev
alpha (%WT)	0.338	0.093
beta	0.729	0.012
sigma (%WT)	4.963	0.110

For the population representing challenging features with an expected tool tolerance of ±15%, the equivalent results are summarized in Figure 4.1 and Table 4.1. Note that β has a lower value as compared with the base population performance, indicative of some over-calling of features by the ILI tool (reported to be larger than

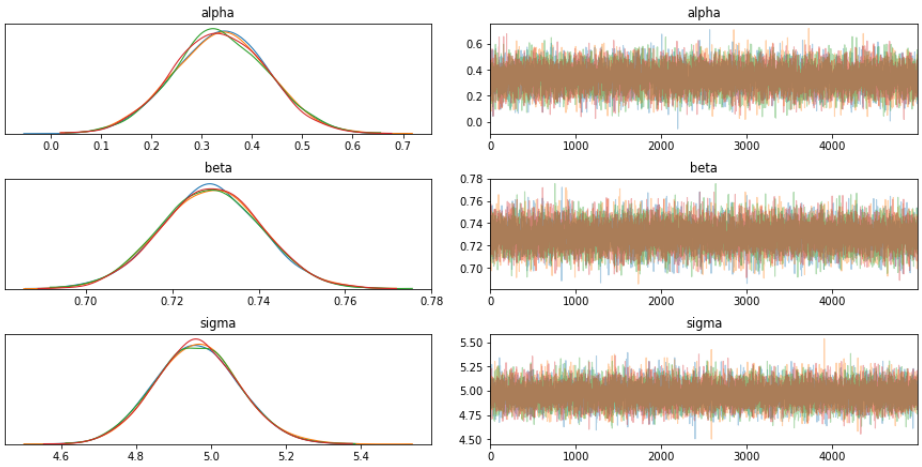


Figure 4-2: Posterior results for the three stochastic variables in the Unity Plot line for the Challenging case

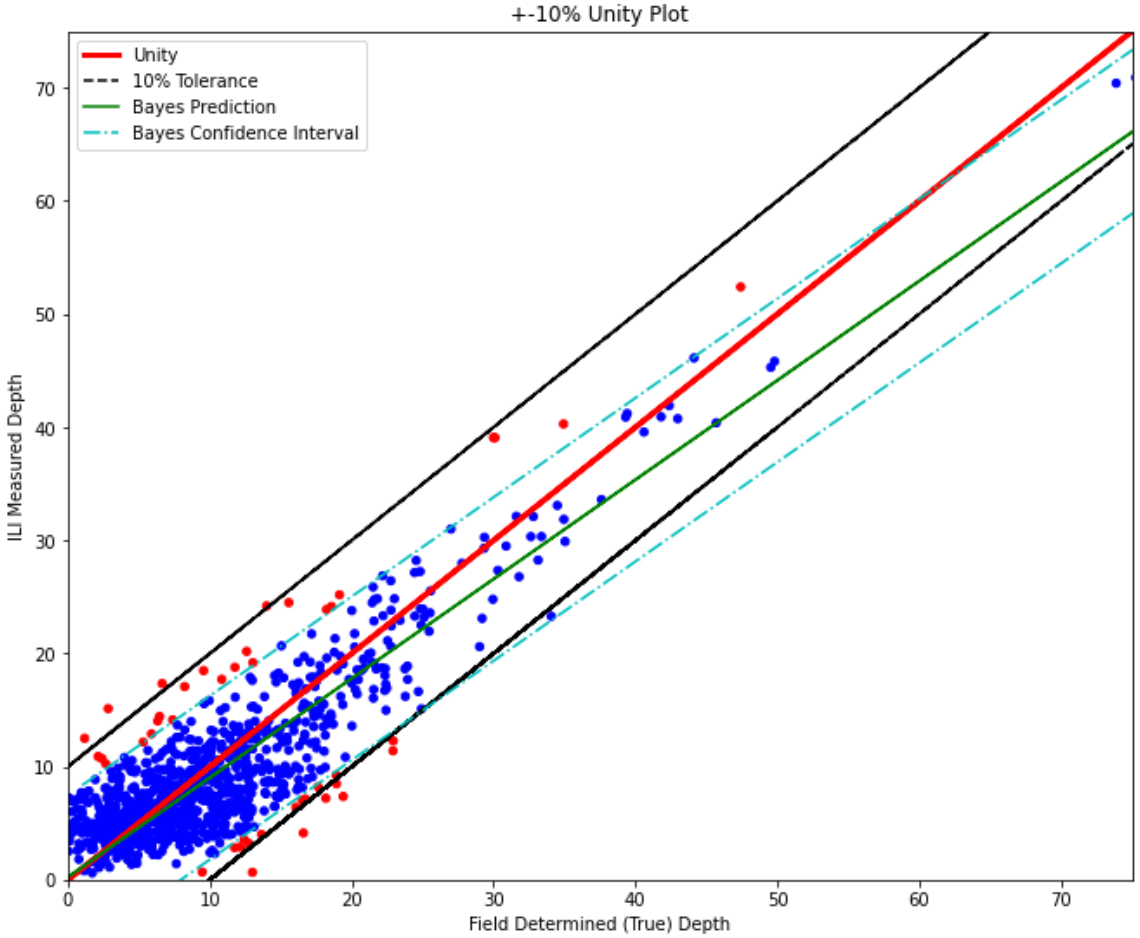
found in the field). Additionally, twice the standard deviation provides for an equivalent tool performance of 9.8%.

4.2 Unity Plot of Base Features

To better visualize the results of the model, the equivalent line determined by the simulated parameters was plotted against a traditional Unity Plot (Figure 4.2). For the Base scenario, the Bayesian model shows slight deviation from the Unity Plot only for larger features, reflecting the weighting that the over-called 70% anomaly at the upper-right corner had on the trend. Interestingly, the tighter tolerance suggested by the Bayesian model does aggressively cull out some lower grade features as potential outliers.

Figures 4.2 and 4.2 show the residual error plots for both the traditional Unity Plot and the Bayesian model, respectively.

Figure 4-3: Unity Plot for Base Features



The Unity Plot showing the 1:1 line (red) with standard tolerance offsets in black. The Bayesian calibration process was run on the plotted features which results in the green calibration line with a tighter tolerance in the dashed cyan lines. A majority of features fall within the tolerance bands of both methods, but there are more over-called (ILI reports the depths greater than they are) features that land outside of the Bayesian calibration line, shown as points plotted in red. This is a conservative result since it results in the ILI overestimating features that may trigger excavations, but an excessive tendency towards overcalling can lead to unnecessary field work.

Figure 4-4: Plot of Residuals for Base Case: Traditional

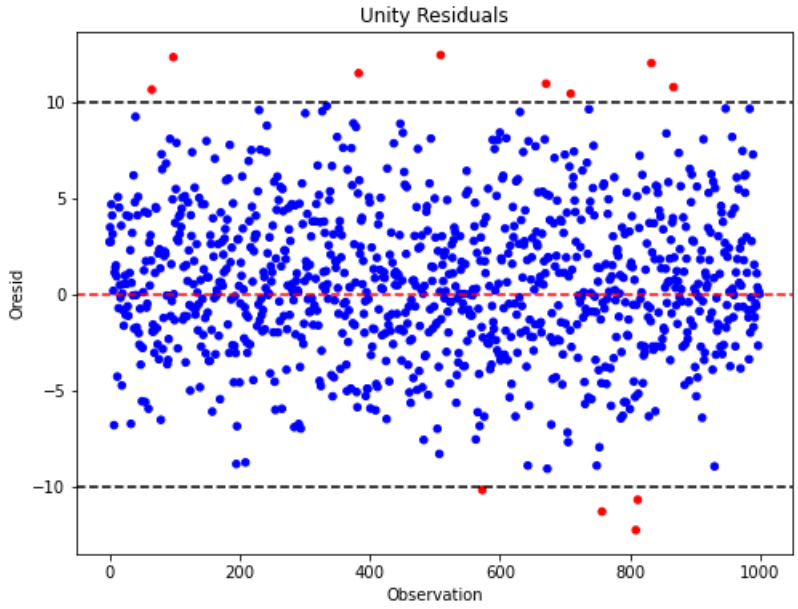
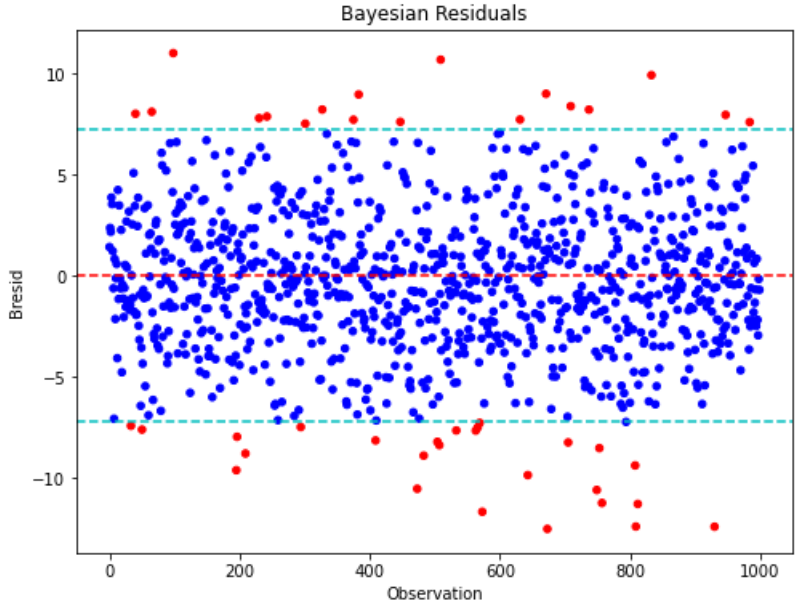


Figure 4-5: Plot of Residuals for Base Case: Bayes

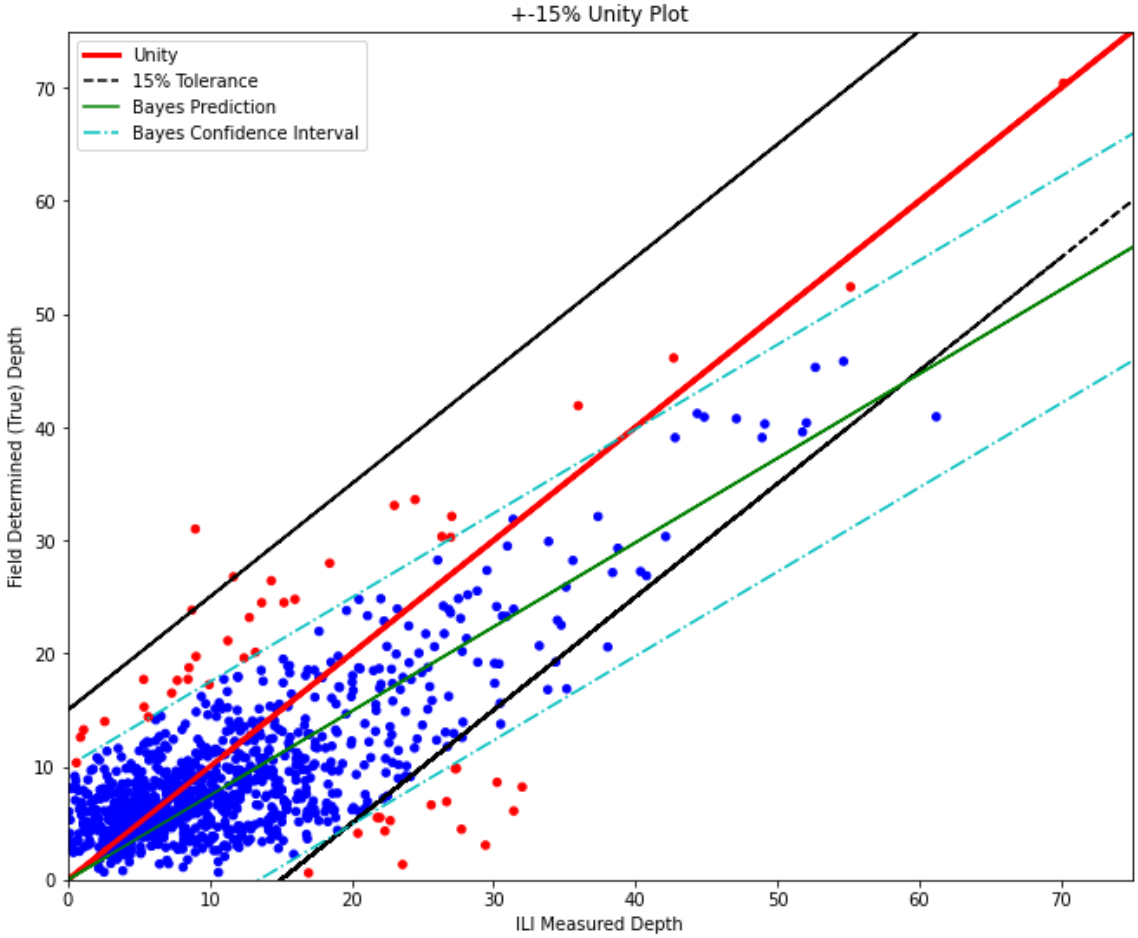


Residual error plots for both the traditional and Bayesian methods. The y-axis shows the error between the actual feature depth versus the expected value from the calibration line. Features that fall outside of the tolerance are colored red. As seen on the Unity Plot, the tighter tolerance on the Bayesian residuals does result in more features being over-called (more observations fall below the -6% tolerance band). However, the number of features above the upper tolerance band is only slightly larger than the traditional method.

4.3 Unity Plot Challenged Features

The more interesting scenario is the representation of the challenging population. Here we see the tendencies observed in the base case magnified (Figure 4.3). The value of the model is still showing that the expected tool tolerance of ±15% can still be tightened with little loss in confidence of that the tool tolerance will encompass the measurement errors around deeper features. (Figures 4.3 and 4.3).

Figure 4-6: Unity Plot for the Challenged Features



Similar to Figure 4.2, the red 1:1 line and corresponding black tolerance bands of ±15% (vendor specification) are shown. The Bayesian calibration line (green) deviates from this line more than in the Base feature-set, conservatively reflecting the trend observed in the literature that ILI-measured features have increased measurement errors as the feature depth increases. While there are a small portion of features (colored red) that appear to fall outside of the corresponding Bayesian tolerance bands (dashed cyan), these are mostly lower depth pits that do not pose as much of a risk. The following residual plots show a clearer comparison when the scaling is accounted for.

Figure 4-7: Plot of Residuals for Challenged Case: Traditional

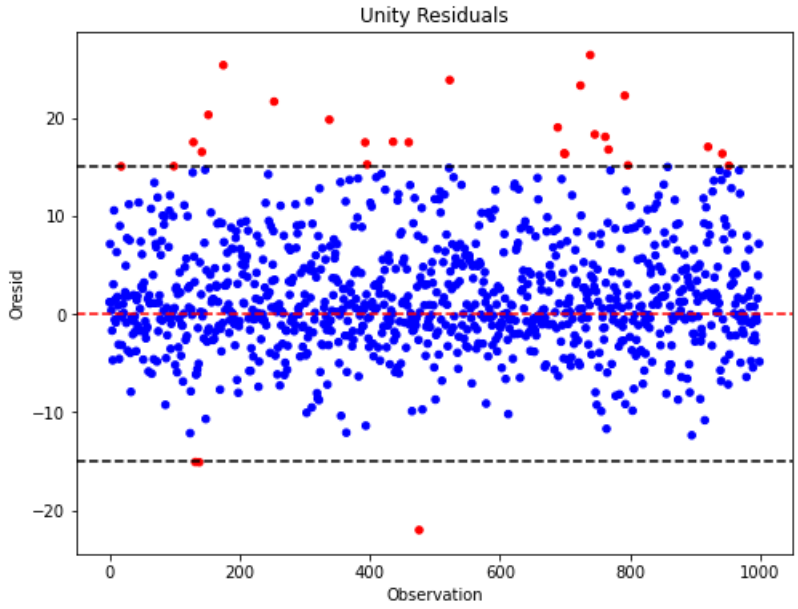
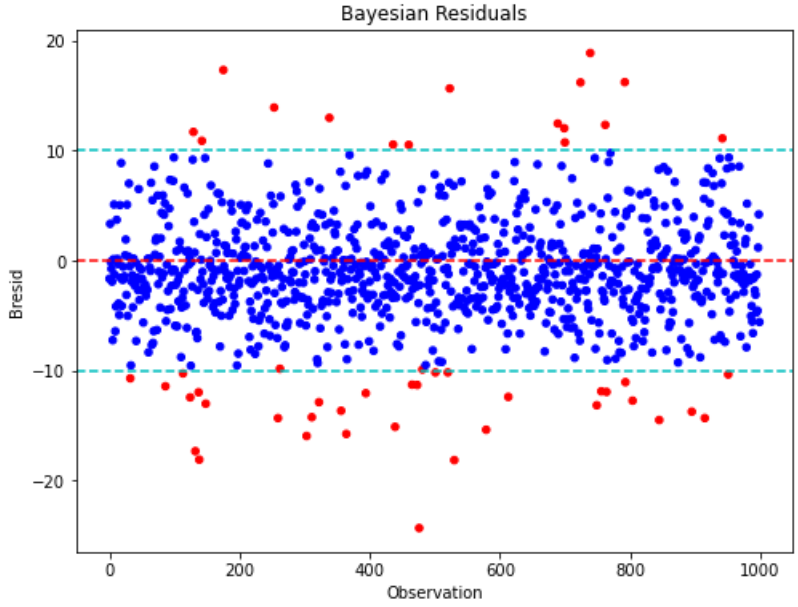


Figure 4-8: Plot of Residuals for Challenged Case: Bayes



Residual error plots for both the traditional and Bayesian methods. The y-axis shows the error between the actual feature depth versus the expected value from the calibration line. As was seen in the Base case, the Bayesian model allows for a tighter tolerance band while limiting the number of features that are over-called (errors greater than +10%).

4.4 Performance Metrics

To summarize the performance of the tools, the Mean Absolute Error, Mean Square Error, and Root Mean Square Error are reported below for the two approaches. The Mean Absolute Error corresponds well to the information that was conveyed in the Residual Plot. However, because of the safety-based nature of pipeline integrity means that larger errors carry more weight, the Root Mean Square error may prove more indicative of the actual economic value of the relative performance improvements. For reference, an Ordinary Least Squares model was included as well. The confidence levels of anomalies within vendor reported specifications and two standard deviations of the model calibrated line are also reported below in Tables 4.4 and 4.4).

Table 4.3: Comparison of Unity and Bayesian: Base

Technique	MAE	MSE	RMSE	Conf Level
Unity	13.7479	2.8734	1.6951	0.991
OLS	21.8284	3.3244	1.8233	NA
Bayesian	11.6901	2.6466	1.6268	0.991

Table 4.4: Comparison of Unity and Bayesian: Challenged

Technique	MAE	MSE	RMSE	Conf Level
Unity	38.4370	4.3893	2.0951	0.892
OLS	21.8284	3.3244	1.8233	NA
Bayesian	24.2332	3.6868	1.9201	0.953

Here we can see that the Bayesian approach does perform well when compared against the traditional Unity plot for thee Base features (both have a 99% confidence level) and is superior in the Challenged set (95% as compared with 89% provided by the traditional method).

4.5 Thoughts on Application

The main benefit of employing the Bayesian model is that with enough samples, the pipeline operator can confidently use a more accurate alternative to the vendor

provided tolerance. This process can be done iteratively such that the parameters can be recalculated quickly based on successive measurements. In contrast to traditional methods in which operator judgment is required to select the specific level of tolerance to apply, this elegantly captures the process of updating knowledge based on observed performance. At the end of a repair campaign, the calibrated parameters can serve as priors for the next round of inspections.

However, this does not completely remove the need for engineering judgement. There are two areas where operator experience and pattern recognition can greatly influence the results. The first is in the selection of the distribution of the priors: the nature of a Bayesian process is that when there are fewer observed samples, more weight is placed on the knowledge represented by the priors. If this is not representative of reality, the output is of little value.

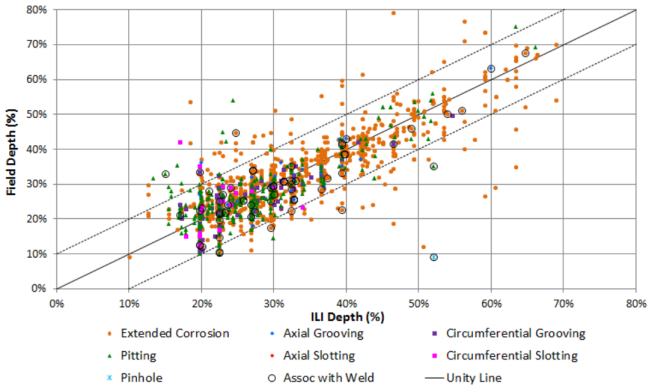


Figure 4-9: Unity Plot with Different Classification Groups [Li et al., 2016] Mixing features types can unnecessarily penalize the tolerance for the entire population based on the performance of a few outliers.

A corollary to this, and a result of sampling theory, is that judicious selection of the appropriate segmentation of the populations can also have a large effect. For example, mixing the two (base and challenged) populations would result in applying a lower tolerance to features we do not have as much confidence in while penalizing the performance of the tool on base features. For example, the plot in Figure 4-9 shows features (circled) that could be excluded in a separate Unity Plot to improve the tolerance applied to the remaining general population.

Additionally, as seen in works like Worthingham et al. covered in Section 2.4.2,

the Bayesian Inference framework is flexible enough to assign measurement errors between the ILI tool and the field tool through the use of a covariance matrix and associated prior distribution. Evaluation of such a model's performance in the base and challenged scenarios is outside the scope of the present work but may be considered in the future.

Finally, it should be noted that the ideal process is for the operator to work closely with the tool vendor to provide the dig results for re-calibration based on the raw MFL signals. This, however, does not minimize the results here since the pipeline operator is ultimately responsible for stewarding the integrity of the line based on their understanding of the tool tolerance. Additionally, building a model of the field measurement error rests largely on the operator, with the tool vendor providing only a minor support role in that decision-making process.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 5

Conclusions and Future Work

Although MFL-based ILI tools have become more accurate at detecting and sizing corrosion features in pipelines, any integrity management program must take into consideration the potential for measurement errors to effect reported results. Since this information is the basis for risk reduction decision making about where to prioritize repairs, it is vitally important that pipeline operators apply the appropriate tolerance level on the ILI-reported features.

5.1 Summary of Results

The Bayesian model proposed in this work was found to be a capable alternative to the primary API 1163 validation process, with lower mean average error and higher tolerance. It allows the pipeline operator to move beyond simply accepting or rejecting the tool vendor's performance specification and instead allows for an alternative tolerance to be stated with the appropriate confidence interval. Like the more recent works identified in the current literature, it offers the flexibility of assigning measurement errors between the field and ILI tools and for incorporating prior knowledge about constant and non-constant biases in the selection of the priors. Another strength of the Bayesian model is that it allows the prior to dominate when there is a lack of information about the performance for a particular feature classification until more information is provided to sufficiently overrule the prior.

The main limitation of the Bayesian model is that it has yet to be applied to actual inspection data. Further testing and implementation would allow for identification of any discrepancies. Additionally, pipeline operators will need properly document the assumptions behind taking a less conservative approach to applying the tool tolerance compared with the specifications provided by the vendor in accordance with API 1163. Ultimately, the greatest value of the Bayesian model may be as a supplemental approach for working with the ILI tool vendor for noting when tolerances can be adjusted.

5.2 Recommendations for Future Work

Several works have begun discussing using Hierarchical Bayesian models to incorporate the stochastic representation of measurement errors into a larger model that incorporates physical parameters like coating condition or soil acidity to describe the overall temporal degradation of the pipeline. Hierarchical models may also prove valuable in describing the physical parameters that influence the MFL signal before the anomalies are reported and can capture the constituent uncertainties that result in the final tolerance. Examples of such factors are the tool speed or sensor stand-off. By dissecting the contributing factors, vendors and pipeline operators can determine what inspection variables need to be modified to improve tool accuracy.

While outliers are partially mitigated in a Bayesian model with a proper selection of the prior, spurious data can still influence the results. The art of outlier detection and segmentation of the anomaly population (i.e. deep versus shallow feature performance) still requires engineering judgment, but methods of intelligently identifying potential outliers can aid in this decision making.

Likewise, concepts from the transfer learning class of problems can assist with the sampling of classification problems. If a pipeline operator has many features that fall under the "pitting corrosion" category that are field verified, how much confidence should they have in features that fall under the "slotting" category that have not been field verified yet? A modification of the variance-covariance priors may be applied

where the same MFL tool is treated as multiple tools with some level of correlation.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix A

PyMC3 Code

```
import pymc3 as pm
import pandas as pd
import math as math

# use pandas to load generated DataFrame
count_data = pd.read_excel("ilidatagenerated.xlsx")
n_count_data = len(count_data.index)
X = count_data["dm2"]
y = count_data["df"]

# variables for pymc inputs
basic_model = pm.Model()

with basic_model:
    # Priors for unknown model parameters
    alpha = pm.Normal('alpha', mu=0, sigma=.1)
    beta = pm.Normal('beta', mu=1,sigma=.1)
    sigma = pm.HalfNormal('sigma', sigma=10)
```

```
    # Expected value of outcome
    mu = alpha + beta*X

    # Likelihood (sampling distribution) of observations
    #Y_obs = pm.MvNormal('Y_obs', mu=mu, chol=chol, observed=y)
    Y_obs = pm.Normal('Y_obs', mu=mu, sigma=sigma, observed=y)

map_estimate = pm.find_MAP(model=basic_model)
map_estimate

with basic_model:
    # draw 500 posterior samples
    trace = pm.sample(500)

trace['alpha'][-5:]

with basic_model:
    # instantiate sampler
    step = pm.Slice()

    # draw 5000 posterior samples
    trace = pm.sample(5000, step=step)

pm.traceplot(trace);
pm.summary(trace).round(4)
pm.plot_posterior(trace);
```

Bibliography

[Al-Amin et al., 2012] Al-Amin, M., Zhou, W., Zhang, S., Kariyawasam, S., and Wang, H. (2012). Bayesian model for calibration of ILI tools. In *Proc. Bienn. Int. Pipeline Conf. IPC*, volume 2, pages 201–208.

[Amaya-Gómez et al., 2019] Amaya-Gómez, R., Sánchez-Silva, M., Bastidas-Arteaga, E., Schoefs, F., and Muñoz, F. (2019). Modeling of pipeline corrosion degradation mechanism with a Lévy Process based on ILI (In-Line) inspections. *Chem. Eng. Trans.*, 77:823–828.

[American Petroleum Institute, 2013] American Petroleum Institute (2013). In-Line Inspection Systems Qualification Standard.

[Baker, 2008] Baker, M. J. (2008). Pipeline Corrosion: Final Report. *U.S. Dep. Transoprtation Pipeline Hazard. Mater. Saf. Adm. Off. Pipeline Saf.*

[Caleyo et al., 2007] Caleyo, F., Alfonso, L., Espina-Hernández, J. H., and Hallen, J. M. (2007). Criteria for performance assessment and calibration of in-line inspections of oil and gas pipelines. *Meas. Sci. Technol.*, 18(7):1787–1799.

[Caleyo et al., 2004] Caleyo, F., Alfonso, L., Hallen, J. M., Gonzalez, J., and Pere (2004). Method proposed for calibrating MFL, UT ILI tools. In *Artic. Oil Gas J.*

[Caleyo et al., 2005] Caleyo, F., Alfonso, L., Hallen, J. M., and Perez-Baruch, E. (2005). Statistical method for the calibration of pipeline in-line inspection data. In *Pipeline Pigging Integr. Manag. Conf. 2005.*

[CEPA, 2016] CEPA (2016). Metal Loss Inline Inspection Tool Validation Guidance Document.

[Cheng et al., 2016] Cheng, K., Abdolrazaghi, M., Hassanien, S., and Watt, C. (2016). Effect of calibration of measurements on integrity reliability analysis. In *Proc. Bienn. Int. Pipeline Conf. IPC*, volume 2.

[Dann and Huyse, 2018] Dann, M. R. and Huyse, L. (2018). The effect of inspection sizing uncertainty on the maximum corrosion growth in pipelines. *Struct. Saf.*, 70:71–81.

[Desjardins et al., 2007] Desjardins, G., Reed, M., and Nickle, R. (2007). ILI performance verification and assessment using statistical hypothesis testing. In *Proc. Bienn. Int. Pipeline Conf. IPC*, volume 2, pages 485–491.

[Dillon, 1982] Dillon, C. (1982). Forms of Corrosion Recognition and Prevention: NACE handbook 1 volume 1. *Corros. Underst. Basics*, (81):99–143.

[Dutta and Ghorbel, 2008] Dutta, S. M. and Ghorbel, F. H. (2008). Magnetic flux leakage sensing: Current practices and mathematical analysis. In *ASME Int. Mech. Eng. Congr. Expo. Proc.*, volume 9 PART B, pages 981–989.

[Ellinger et al., 2016] Ellinger, M. A., Bubenik, T. A., and Moreno, P. J. (2016). ILI-to-field data comparisons - What accuracy can you expect? In *Proc. Bienn. Int. Pipeline Conf. IPC*, volume 1.

[Grubbs, 1948] Grubbs, F. E. (1948). On Estimating Precision of Measuring Instruments and Product Variability. *J. Am. Stat. Assoc.*, 43(242):243–264.

[Haines and Tomar, 2013] Haines, H. H. and Tomar, M. (2013). ILI Tool Calibration Based on In-ditch Measurement with Related Uncertainty. *Contract PR-218-103608*.

[Hoffman and Gelman, 2014] Hoffman, M. D. and Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, 15:1593–1623.

[Joshi, 2011] Joshi, A. V. (2011). Statistical analysis of in-line inspection performance with Gamma distribution. In *NACE - Int. Corros. Conf. Ser.*

[Li et al., 2016] Li, Y., Fredine, G., Hubert, Y., and Hassanien, S. (2016). Making integrity decisions using metal loss ILI validation process. In *Proc. Bienn. Int. Pipeline Conf. IPC*, volume 1.

[Ludlow, 2012] Ludlow, J. (2012). Statistical tolerance intervals: A better approach to in-line inspection performance assessment. In *NACE - Int. Corros. Conf. Ser.*, volume 3, pages 2423–2437.

[McNealy et al., 2010] McNealy, R., McCann, R., Van Hook, M., Stiff, A., and Kania, R. (2010). In-line inspection performance III, Effect of In-ditch Errors in Determining ILI Performance. In *Proc. Bienn. Int. Pipeline Conf. IPC*, volume 4, pages 469–473.

[Mora et al., 2004] Mora, R. G., Powell, D., and Harper, W. V. (2004). Methodology for statistically designing and assessing metal loss in-line inspection validation programs. Technical report.

[Morrison et al., 2000] Morrison, T., Mangat, N., Desjardins, G., and Bhatia, A. (2000). Validation of an in-line inspection metal loss tool. In *Proc. Bienn. Int. Pipeline Conf. IPC*, volume 2, pages 839–844.

[Offshore Mag., 2012] Offshore Mag. (2012). TDW launches new diameter inline inspection tool. *Offshore Mag.*

[Peng, 2020] Peng, X. (2020). Analysis of Magnetic-Flux Leakage (MFL) Data. 56(6).

[PHMSA, 2018] PHMSA (2018). PHMSA: Stakeholder Communications - Corrosion.

[Pipe Operators Forum, 2012] Pipe Operators Forum (2012). Guidance on Field Verification Procedures for In-Line-Inspection.

[Rosen Group, 2013] Rosen Group (2013). Rocorr - MFL-A Specifications.

[Salama et al., 2012] Salama, M. M., Nestleroth, B. J., Maes, M. A., Rodriguez, C., and Blumer, D. (2012). Characterization of the accuracy of the MFL pipeline inspection tools. In *Proc. Int. Conf. Offshore Mech. Arct. Eng. - OMAE*, volume 6, pages 247–251. American Society of Mechanical Engineers Digital Collection.

[Shi et al., 2015] Shi, Y., Zhang, C., Li, R., Cai, M., and Jia, G. (2015). Theory and application of magnetic flux leakage pipeline detection.

[Smart and Haines, 2014] Smart, L. and Haines, H. (2014). Validating ILI Accuracy Using API 1163.

[Timashev and Bushinskaya, 2009] Timashev, S. A. and Bushinskaya, A. V. (2009). Statistical analysis of real ILI data: Implications, inferences and lessons learned. In *Rio Pipeline Conf. Expo. Tech. Pap.*, volume 2009-Septe.

[Tomar et al., 2009] Tomar, M. S., Fingerhut, M., and Yu, D. (2009). Qualification of ILI performance in accordance with API 1163 and the potential impact for management of pipeline integrity. In *Proc. Bienn. Int. Pipeline Conf. IPC*, volume 2, pages 693–703.

[Ulrich, 1996] Ulrich, L. W. (1996). DOT's perspective on in-line inspection.

[Worthingham et al., 2002] Worthingham, R., Morrison, T., Mangat, N. S., and Desjardins, G. (2002). Bayesian estimates of measurement error for in-line inspection and field tools. Technical report.