Phase I Report

Investigate Fundamentals and Performance Improvements Of Current In-Line Inspection Technologies For Mechanical Damage Detection

Prepared for Pipeline Research Council International, Inc (PRCI)

<table>
<tr><td rowspan="4">BLADE ENERGY PARTNERS
2600 Network Blvd.
Suite 550
Frisco, TX 75034
Telephone: (972) 712-8407
Fax: (972) 712-8408

www.blade-energy.com</td><td colspan="2">Job Number:
PRCI-U206-002</td></tr>
<tr><td>Version:
001</td><td>Date:
May 2008</td></tr>
<tr><td colspan="2">Purpose:
This Phase I report provides a comprehensive and in-depth review of the current status of in-line-inspection technologies, including, but not limited to, Magnetic (Axial MFL, Circumferential MFL), Ultrasonic (UT), and Geometrical (Caliper) methods, in terms of their capabilities, limitations and potentials in detection, discrimination and characterization of various forms of pipeline mechanical damage, such as dents, dents with corrosion, and dents with cracks, gouges and dents with gouges.</td></tr>
<tr><td colspan="2">PRCI Award:
PRCI Contract PR-328-063502
DTPH56-06-T-000016 ECT #204</td></tr>
</table>

Version	Date	Issued As	Author By	Checked By	Project
000	April 2007	Phase I Report Draft	Richard McNealy Ming Gao Roger McCann Pablo Cazenave Ravi Krishnamurthy	RMN RMC	RMK
001	May 2008	Phase I Review Comments and Final	RMc/MG	RMc	RMK

TABLE OF CONTENTS

APPENDICES:

List of tables

List of Figures

1.0 EXECUTIVE SUMMARY

The Pipeline Research Council International, Inc. (PRCI) engaged Blade Energy Partners Ltd to conduct a study evaluating the capability of current in-line inspection (ILI) technologies used to detect and characterize mechanical damage (MD) anomalies.

The primary objective of this study is to assist the pipeline industry in selecting ILI technologies that are best suited for detecting and sizing the types of mechanical damage that may pose integrity concerns, and/or are required to be addressed by the existing Regulatory Rules. The practical need is driven by both the recent changes in Regulatory requirements, vis-a-vis mechanical damage, and the latest developments of ILI technologies aiming to detect and size such damage.

Based on the information provided by the six participating ILI vendors, and taking full advantage of extensive previous work, this report presents the results of Phase I of the Project, including updated capabilities and deficiencies of the current MD ILI technologies, performance claims, and supporting validation data. In addition, this Phase 1 report details the fundamental approaches embodied within each technology then analyzes validation data and derives performance conclusions from available data. Finally, this report identifies further testing to be conducted within Phase II of the Project.

The performance of the current MD ILI technologies was evaluated mainly on dent depth sizing accuracy, but also on gouge and metal loss within dents. The methods and procedures specified in API 1163, namely, Binomial Distribution Analysis, Confidence Interval Analysis and Least Square Linear Regression Analysis, were used to analyze the validation data provided by both the ILI Vendors and by Pipeline Operators. Phase I also assessed performance specifications beyond those considered in API 1163, e.g., Probability of Detection, Probability of False Calls, and Probability of Identification.

Six participating ILI vendors provided information about current technologies they use to detect and size plain dents, and mechanically induced gouges without dents, as well as coincident damage caused by corrosion, gouges and cracks located within dents. However, only five Vendor Dent Sizing Technologies were evaluated. The results from validation data sets containing fewer than 9 points (Vendor Technology F) were not included in the overall critical comparisons.

For plain dents, four Vendor Technologies, whose sample size exceeded this minimum, were analyzed. Table I summarizes the analysis results. The Technologies referenced in Table 1 include Direct Arm Measurement Caliper with Electromagnetic sensors (DAMC-EM) and Triaxial MFL with Eddy current sensors.

In addition, Table 1 also includes the results of analyses made by pooling data from the same type of technologies A-G and A-C-G in order to provide an overall insight into the level of the current sizing performance of these types.

Data Analyzed	Technology Type	Tolerance for Certainty = 0.8 at 95% Confidence Level			
		Sample Size for Validation	Limits of Detection (%OD)	Binomial Distribution Analysis	Clopper-Pearson Certainty Interval Method
Technology A	DAMC (EM)	130[a]	0.5	±1.10%	±1.22%
Technology G	DAMC (EM)	20[b]	0.5	±0.74%	±0.74%
Technology C	DAMC (EM)	15[c]	0.5	±0.51%	±0.78%
Technology E	{Long Field} MFL [Hall-3][ID/OD EM]	273[d]	2.0	±0.78%	±0.80%
Technology A and G	DAMC (EM)	150	0.5	±0.93%	±0.93%
Technology A, C, and G	DAMC (EM)	165	0.5	±0.88%	±0.88%

Table 1: *Summary of Dent Sizing Performance*

a) Direct examination observations of MD from NPS 16 pipeline
b) Direct examination observations of MD from multiple ILI
c) Laboratory Pull Test for 20 plain dents in NPS 8 pipe
d) Dent depth predicted by Technology E and validated against multiple tool size ILI performed by Caliper [DAMC]-Technology D. The present evaluation of the most statistically significant sample, available for Technology E, assumed the dent depth sizing performance of the referenced Technology D at 85% certainty of +/- 0.60 % OD. (A level consistent with the validation results observed for DAMC (EM) Calipers).

From Table 1, it is seen that most of the plain dents that require repair or evaluation by the current PHMSA/OPS Integrity Rules can be detected and sized by commercially available ILI tools specialized to detect either deformation, i.e., Technologies A-C, or metal loss, i.e., Technology E. It is noted that MFL with 3 axis sensors, as demonstrated by Technology E, may predict dent depth as accurately as the current caliper technologies. Mechanically induced gouges without dents (data not included in the table) can also be reliably detected and sized by some of the MFL tools. The following are the main findings of the capability of the current ILI technologies for plain dent assessment:

- On average, a dent depth measurement tolerance of ±0.77% OD for a certainty of 0.8 at 95% confidence level (sample size n = 438 with 360 success). Measurement tolerances for individual technologies vary from ±0.51% to ±1.10% OD based on Binomial Distribution Analysis or ±0.74% OD to ±1.22% based on Clopper-Pearson Certainty Interval method.
- Dent length and width are measured, but scatter in the measurements places a large uncertainty on the accuracy of any individual measurement.
- Regression analyses suggest that pull tests provide a better correlation between dent dimensions predicted by in-line tools and direct validation measurements than validation tests using field excavation data. This result is expected given the potential for re-

rounding and re-bounding of pipeline dents and the challenges surrounding the physical measurement of mechanical damage in field excavations.
- Regression analyses show that the depth sizing performance of one MFL technology, Technology E, may be comparable to that of caliper technology on mechanical damage assessment (±0.78%OD, 80% of time). However, a full confirmation of the depth sizing performance is required to reliably characterize the performance of this technology.

In case of Coincident damage, Vendor data suggested that next to plain dents (PD), dents with metal loss (DML) represented the most widely inspected condition. Consequently, the large data sample available for dents with metal loss (DML) provided the basis for a comprehensive evaluation of Vendor Technologies in characterizing Coincident damage. However, none of the participating Vendors made performance claims relating to coincident feature sizing (metal loss, corrosion or gouge within dents) or probability of detection and minimum threshold for detection and reporting for current MD Technologies. To assist in clarifying these issues, performance of ILI Technologies was evaluated using the coincident damage date sets to determine the Proportion of Detection, Proportion of False Calls and Proportion of Identification. These values approximate the Probability of Detection (POD), Probability of False Calls (POFC) and Probability of Identification (POI), which were estimated by statistical methods.

Six technologies were evaluated for dent with metal loss capability using validation data. Four of them are MFL plus caliper technologies identified by the participating Vendors and two represent data for two technologies provided by three operators using similar in-line inspection technologies. All of these Technologies consisted of the integrated analysis of caliper based deformation data and magnetic flux leakage (MFL) signals. In addition, case study and literature data were reviewed and considered in the overall analysis. Table 2 gives a summary of the evaluation results.

The key conclusions regarding the current MD technologies for co-incident damage characterization are as follows:

- All MFL technologies demonstrated capability to detect metal loss within dents with either a strict application of the Subject Matter Expert Analysis (SMEA) or combined assessment of MFL and Caliper data.
- Two technologies (C and J), combining MFL and Caliper data, demonstrated capability for detecting Dent-Coincident-With-Metal-Loss with proportion of detection and proportion of identification greater than 94% (Table 2). This corresponds to POD and POI of approximately 89% at a 95% confidence level.
- Even though Technologies H and I have a high proportion of detection (87%), their POD interval at 95% confidence level is rather wide (from 70% to 96%) due to relatively small sample size.
- Metal loss data indicates the MFL Technologies have success in detecting metal loss that is less than 10% wall thickness, coincident with plain dents that range in size from 2% to 6% OD. However, the available data is insufficient to quantify a detection performance for all Technologies.
- Using the same Performance Specification (80% Certainty at 95% Confidence), corrosion sizing performance in dents was estimated to be:

– ± 15% WT depth tolerance for Technology C,
– ± 6.4% WT depth tolerance for the combined data of Technologies H and I,

- One Technology (E) qualitatively demonstrated capability to detect cracks within dents, within limitations of orientation, but insufficient data was available to quantify a performance,
- The discrimination of gouges from corrosion within dents was claimed by a limited number of Technologies. Validation data was not sufficient to confirm such claims.

Data Analyzed		Technology Type	Data Used to Calculate Proportion		Proportion (x/n)		
			Total Excavation Investigations (a)	DML Excavations (based on ILI report) (b)	Detection	False Calls	Identification
Vendors	J	MFL+ Caliper Combo	61	58	94.50%	10.30%	94.20%
	A+B	MFL+ Caliper	138	82	60.00%	67.10%	100.00%
	(H, I)+ (F,G)	MFL+ Caliper	26	23	87.00%	13.00%	80.00%
	C	MFL+ Caliper Combo	34	26	100.00%	3.80%	100.00%
Operators	K+N(Operator G1)	MFL + Caliper	27	8	66.70%	25.00%	100.00%
	K+N (Operator G2)	MFL + Caliper	114	37	86.10%	16.20%	96.80%
	K+L (Operator L2)	Transverse MFL + Caliper	63	56	78.80%	53.60%	84.60%

Table 2: Summaries of the evaluation results

a) Total number of excavations including excavations for DML, corrosion and others
b) The number of excavations for DML only

Phase II will focus on utilizing operator excavation data for the various technologies studied in Phase I. A uniform field MD assessment protocol will be developed and utilized by all the participating operators during field assessment of mechanical damage. The reduction of field MD measurement errors will provide further insight into current MD technology with a particular focus on coincident damage and detail dent shape characterization. Additional pull through testing may also be considered in Phase II, however it will be dependent on schedule and cost.

2.0 BACKGROUND

2.1 Motivation

Mechanical damage is normally divided into two categories, dents and gouges, which are deformations in the pipe wall that serve as failure initiation sites. Dents typically result from a purely radial deformation. A pipe impinging on a rock may result in a dent. If the pipe also slides on the rock, a dent with a gouge may result. Third-party mechanical damage, caused during construction and excavation, is a common cause of gouges. A gouge normally results in a highly deformed, work hardened surface layer and may involve metal removal. Undetected mechanical damage to pipelines from outside forces can lead to leaks or ruptures. Mechanical damage can result in either immediate or delayed failure.

A majority of the anomalies caused by outside forces do not have dire consequences. However, a few prominent pipeline failures have been attributed to mechanical damage[1]. While dents are common, failures from dents alone (i.e., dents without additional surface mechanical damage such as scratches and gouges) are relatively rare. Dents with additional surface mechanical damage result in immediate failure approximately 80 percent of the time[2]. In the remainder of mechanical damage events, damage is not severe enough to cause immediate failure. However it may lead to delayed failure if the internal pressure is raised sufficiently, if corrosion or cracking develops in the damaged material, or if there is pressure-cycle fatigue.

The pipeline industry has multiple in-line inspection methods to inspect for mechanical damage. The most commonly used methods include in-line deformation (caliper) tools, which measure the bore diameter, and magnetic flux leakage technology (MFL) that detects metal loss. As shown in Figure 1, the immediate incidents are higher in number as compared to delayed. The inspection technology can have a large impact on the delayed incidents and some impact on the immediate incidents.

The basic in-line inspection tools to assess mechanical damage are deformation, or caliper, tools that assess the bore diameter of a pipe. These tools are often used to assess dent depth and length at the inner surface of a pipe. Dents with depths greater than those allowed by regulations are identified for excavation and/or analysis. The issue with deformation tools is that they cannot determine other co-incident damage such as corrosion, removed metal, cold working, or cracking at the dent location. Without knowing these conditions, many benign dents are needlessly identified for excavation and repair, while other critical dents that eventually fail may pass the inspection criteria.

---

[1] Kiefner, John, Kolovich, Carolyn, Mechanical Damage Technical Workshop, Feb 28, 2006 Houston Tx.
[2] Rosenfeld, M.J. Proposed new guidelines for ASME B31.8 on assessment of dents and mechanical damage. GRI, May 2001.

	Total number of reportable incidents from all causes 1985 through 2003	Number of immediate incidents from mechanical damage	Number of delayed incidents from mechanical damage	Ratio of immediate to delayed
300,000 miles of natural gas transmission and gathering pipelines	1583	440 (28% of total)	49 (4% of total)	9 to 1
160,000 miles of liquid petroleum pipelines	3366	724 (21% of total)	153 (5% of total)	5 to 1

Figure 1: Analysis of Reported Mechanical Damage Incidents in the USA[1]

The utilization of MFL tools along with caliper/deformation tool allows operators to identify co-incident damage. In addition, there are newer MFL tools (utilizing all components of magnetic field data) that exhibit potential to characterize dents without caliper data, and also identify co-incident crack like defects. However, there is a lack of a consistent view of the capabilities of current ILI technologies to discriminate and quantify mechanical damage. Hence the issues with the understanding of ILI technologies are:

- How well can caliper, MFL and other tools and/or a combination of tools,
    - Characterize the shape and size of dents and other mechanical damage features?
    - Discriminate the presence of corrosion/gouge/crack in dents?
    - Detect and characterize residual stresses and metallurgical changes by mechanical damage process
- Are there validation data that validates the expected performance of ILI technologies?

The purpose of this project was to address these issues with currently available data. Consequently, research conducted within this project was the result of collaborative effort between the vendor community and operators to integrate field data with the data provided by caliper, MFL and other MD tools. In order to partially address these issues the current project

- Provides insight into the current capabilities of ILI technologies for mechanical damage.
- Provides data to validate assessment capabilities of in-line inspection tools.
- Identifies the strengths and limitations of current technology utilizing quantifiable performance measures.
- Establishes focus for immediate and future research.

2.2 Prior Developmental Work

Prior published research was surveyed and found to address the following topics in mechanical damage relevant to the current Project:

- Definitions of mechanical damage
- Inspection technologies applicable for mechanical damage

Multiple qualitative definitions for the conditions arising from mechanical damage can be found in the literature. However, there is limited guidance or agreement for quantitative descriptions. The conditions arising from mechanical damage are normally divided into two categories: dents and gouges.

ASME B31.8S defines mechanical damage as a type of metal damage caused by the application of an external force, possibly resulting in denting, coating removal, metal removal, moved metal, cold working of the underlying metal, and/or introduction of residual stresses.

The definitions for dents and deformations vary among many industry standards. Some of the more accepted definitions for dents include:

- API 1160[3] - Plain dents are a local change in surface contour but not accompanied by a stress concentrator

- API 1163[4] - Dent: A local change in piping surface contour caused by an external force such as mechanical impact or rock impact; ovality; out of roundness, i.e. egg shaped or broadly elliptical.

- API 1156[5] - Dent: "… when a perceptible, longitudinally and circumferentially local deviation from cylindrical curvature exists."

- PDAM[6] - Dent: "a depression which produces a gross disturbance in the curvature of the pipeline, caused by contact with a foreign body, resulting in plastic deformation of the pipe wall.

The most significant difference between sources cited lies in quantifying the extent of the deformations. These criteria vary from "perceptible" deformation to definitions based on the minimum limits of detection for inspection technology. Figure 2 depicts a generalized view of dents in pipe illustrating some of this issues involved in delineating the extent (length and width) of deformations.

---

[3] American Petroleum Institute. API Standard 1160, Managing System Integrity for Hazardous Liquid Pipelines. November 2001.
[4] American Petroleum Institute. API Standard 1163, In-Line Inspection Systems Qualification Standard. June 2004.
[5] American Petroleum Institute. API Publication 1156, Efects of Smooth and Rock Dents on Liquid petroleum Pipelines. November 1997.
[6] A. Cosham and P. Hopkins. The Pipeline Defect Assessment Manual (PDAM), a Report to the PDAM Joint Industry Project. May 2003.

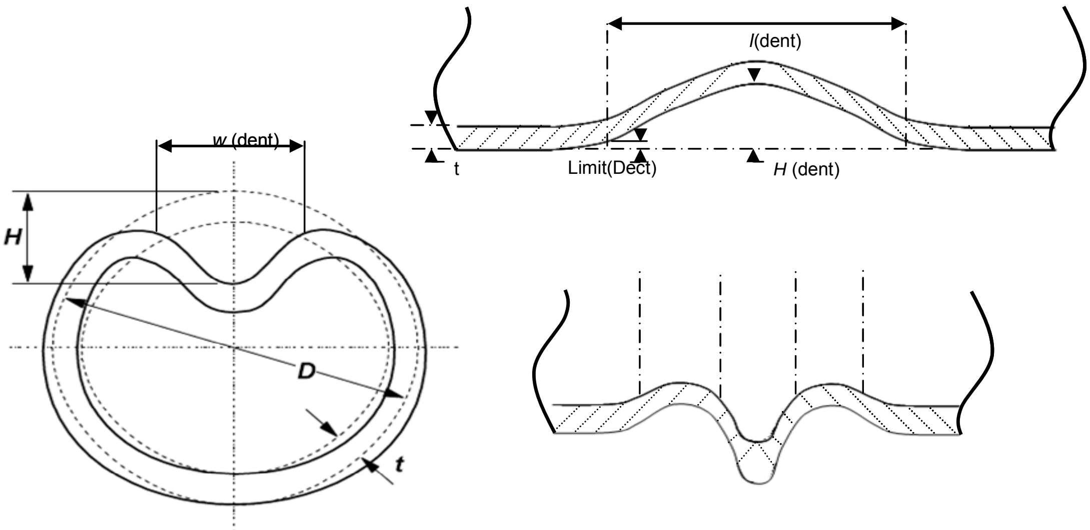


Figure 2: Typical Configurations for Dents

Inspection technologies have been employed to manage mechanical damage threats. Despite, prior work recognized significant technology gaps in the detection and discrimination of mechanical damage, and the need to develop new technologies[7,8,9], industry has realized substantial success in detection and discrimination of mechanical damage.

Technologies and associated ILI tools currently used for mechanical damage detection and discrimination can be categorized as one of three types: dimensional (Calipers), electromagnetic (MFL), ultrasonic (UT, EMAT)[10].

Dimensional measurement technology (mainly, Calipers) directly measures the deviation from circular form of the pipe wall and has been used for detecting, locating and sizing dents, wrinkles, cold bends, etc. This technology is generally regarded as providing the most accurate results for sizing dents and wrinkles at specified detection thresholds (for example, 2% OD for dents), but it is not capable of detecting other defects associated with dents such as corrosion, cracks, and gouges. Therefore, dimensional technology in conjunction with other technologies, such as electromagnetic technology, is utilized to reveal the severity of mechanical damage defects.

---

7 Teitsma, A., "Technology Assessment for Delivery Reliability for Natural Gas-Inspection Technologies: REFC", GTI, USDOE-NETL DE-FC26-02NT41647, Nov 2004.

8 Panetta, P.D. et al, "Mechanical Damage Characterization in Pipelines", Pacific Northwest National Laboratory, October 2001, DOE DE-AC06-76RLO1830.

9 Bubenik, T.A., et al, In-Line inspection Technologies for Mechanical Damage and SCC in Pipelines- Final Report, DOT DTRS56-96-C-0010 June 2000.

10 Davis, R. J. and Nestleroth, J. B., "Pipeline Mechanical Damage Characterization by Multiple Magnetization Level Decoupling," *Review of Progress in Quantitative Nondestructive Evaluation*, Volume 18, Plenum New York, 1999.

MFL (Magnetic Flux Leakage), both axial and circumferential fields, has been shown to be sensitive to geometric and magnetic changes due to mechanical damage and is capable of detecting some mechanical damage [11,9]. Figure 3 describes the conditions comprising mechanical damage that result in changes to MFL signals. The mechanical damage signal may be driven by geometric changes (local metal loss or moved metal and global ovality and denting). Other parts of the signal are mainly associated with changes in magnetic properties resulting from stresses, strains, and metallurgical changes of line pipe steels.

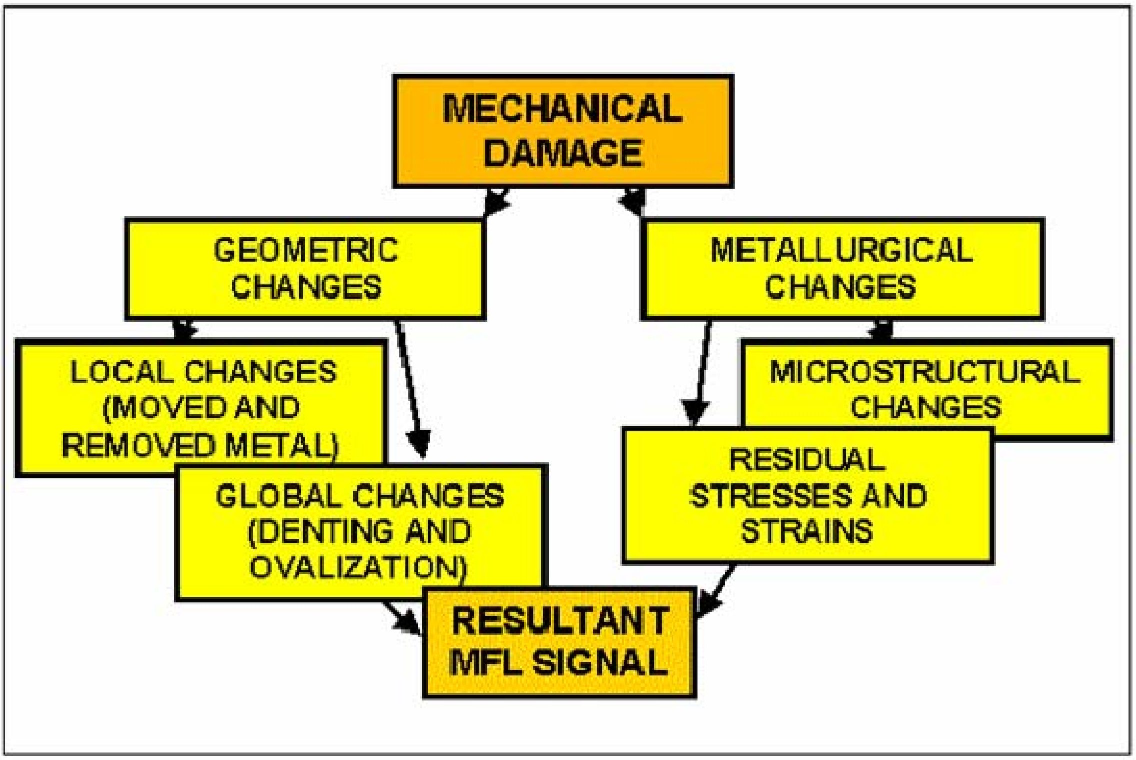


Figure 3: MFL signals at Mechanical Damage

The sensitivity and effectiveness of MFL in detecting mechanical damage strongly depend on the sensor and magnetic field configuration. The magnetic flux leakage produced by mechanical damage defects is a three dimensional vector quantity, axial, radial and circumferential. The most commonly used MFL tools are capable of detecting deformations, but limited to record flux leakage `data only in either an axial, or combined axial & radial axes.

The other two magnetic vector components (radial and circumferential) of flux leakage produced by the mechanical damage defects contain additional information that can be useful for discrimination and sizing[7,8]. Therefore, a 3-axis sensor has been developed to record flux leakage in all axes. However, the differences in detection and discrimination of mechanical damage between the various MFL technologies have not been quantified and assessed in the open literature. This quantification and assessment is essential to decision making for pipeline integrity.

[11] Panetta, P.D., et al, "Mechanical Damage Characterization in Pipelines", PNNL-SA-35467, Oct 2001

A multiple magnetization technology has also been developed in parallel to the 3-axis sensor technology [12,10]. This technology utilizes a low field magnetization to distinguish plastic strain (cold work/residual stress) within dents. Commercial tools are now available for better detection, discrimination and sizing of mechanical damage defects[9]. For example, one vendor appears to use a high/low field decoupled tool with the capability of discriminating and sizing cold work, corrosion and gouge.

Ultrasonic technologies have been used to detect dents and cracks. Identification of the dent by UT is feasible, however dent sizing is inconsistent. Despite high degree of capability in detection of cracks in plain pipe, the probability of crack detection within dents will decrease with the size of the dent. Sizing of cracks within dents isn't feasible with any degree of reliability. Additionally, application of UT requires a liquid medium, consequently practical operational considerations exist for application in gas pipelines. To overcome this limitation, Electromagnetic Acoustic Transducer (EMAT) technology is being developed for gas pipelines. However, the EMAT commercialization is recently initiated for cracks in plain pipe. There is little quantitative understanding regarding EMAT tools capability to detect and discriminate mechanical damage.

It was recognized by PRCI that significant data and experience from the implementation of existing technologies for the detection and characterization of mechanical damage has been made available within the Industry. The intent of the current work is to leverage existing data around all of these technologies to define a comprehensive view of the strengths and limitations of the existing ILI technologies.

2.3 Project Objectives and Scope

This project was conducted for the Pipeline Research Council by Blade Energy Partners Ltd under contract PR-328-063502 and PHMSA transaction agreement DTPH56-06-000016 Project #204. It is Phase I of a two-phase project that evaluates current technologies for detecting mechanical damage of pipelines. The overall objectives of this Project include:

- Evaluate capabilities of deformation and MFL based inspection tools that may detect and possibly discriminate mechanical damage.
- Identify current capabilities of mechanical damage inspection technologies used in the pipeline industry.
- Provide data to validate assessment capability of in-line inspection tools while tying these results back to fundamentals and performance characteristics.

Phase I provides a comprehensive and in-depth review of the current status of in-line-inspection technologies in terms of their capabilities, limitations and potentials for detection, discrimination and characterization of various forms of pipeline mechanical damage, such as dents with corrosion, dents with cracks (and other secondary features), gouges and dents combined with gouges. These technologies include, but are not limited to, Geometrical (Caliper) methods, Magnetic (Axial MFL, Circumferential MFL), Ultrasonic (UT), and Electromagnetic Acoustic Transducers (EMATs). Phase II for this research follows with supplemental testing to further quantify the performance of the most promising technologies and establish their applicability limits.

The detection, characterization and measurement of Mechanical Damage (MD) in pipelines can involve assessment of the following conditions:

Pipeline Dents or Deformations

- Circumferential position (top or bottom) of pipe
- 3 dimensional size and shape (depth, axial length, circumferential width, contour)
- Strain (elastic and plastic)

Stress Concentrations within Dents

- Metal loss due to corrosion
- Metal loss due to gouging
- Cracks, fatigue or environmentally assisted
- 3 dimensional size and orientation
- Effect of deformations on ILI tool performance (Limits of Detection, Probability of Identification and Sizing Tolerances)

In order to develop a comprehensive assessment of current ILI technologies, this project's approach was as follows:

- Survey the current ILI technology and methods.
- Gather data and results of their applications in MD assessments or laboratory tests.
- Provide an in depth statistical assessment of the gathered data.
- Characterize current performance capabilities of the technologies and methods supported by the data.
- Relate the results to their impact on MD Integrity Strategies.
- Identify promising current technologies and methods that lack sufficient validation data to quantify their performance capability.

3.0 CURRENT MECHANICAL DAMAGE TECHNOLOGIES AND CLAIMED CAPABILITIES

Technologies and associated in-line inspection (ILI) tools currently used for mechanical damage detection and discrimination can be categorized as one of three types:

- Dimensional (Caliper and High Resolution Caliper)
- Electromagnetic (MFL)
- Ultrasonic (UT, EMAT) measurements.

The following In-Line Inspection technology vendors agreed to participate in the current project and provided validation data for cases in which these technologies were employed for detection and discrimination of mechanical damage:

- BJ Pipeline Services
- Baker Hughes Pipeline Management Group
- Enduro Pipeline Services
- GE Oil & Gas Pipeline Inspection
- Rosen Inspection Technologies
- T.D. Williamson Inc

Each ILI vendor provided responses to a technology questionnaire prepared as part of this Project. The data returned in these questionnaires described the general features of technologies developed and/or employed by that ILI vendor and indicated whether validation data (laboratory, complimentary ILI tool data or direct field examination measurements) were available. The vendors were asked to identify only current commercial technologies and to exclude developmental research efforts.

After assessing the technology questionnaires, individual meetings were conducted between Blade Energy Partners personnel and participating ILI vendors. The purpose of theses meetings was to gather additional understanding of the inspection technologies and the actual validation data. A thorough understanding of the details of each technology served to provide generic technical descriptions for this report, thereby avoiding the use of trade-names. Each ILI tool available has been provided a generic technology label. Details on each of the technology are provided in Appendix A.

Most of the vendors identified multiple technologies employed to detect mechanical damage, for which validation was available. All of the vendors reported that current Mechanical Damage (MD) assessment has four components:

- Sensing Technology
  - > Deformation Sensing
  - > Co-incident Sensing

- Mechanical Design

- Data Analysis
- Measurement Reporting Capability

Data describing the characteristics and capabilities within each of the essential components was gathered. The differentiating features of the technologies within each component are discussed in detail in the following sections.

3.1 Sensing Technology

The types of deformation and feature sensors employed provide the first level of differentiation between technologies. The participating vendors reported using multiple types of calipers, MFL type tools and/or ultrasonic sensing technologies for mechanical damage. None reported using EMAT transducers in applications where validation data in the form of direct measurements would be available during the current Project. In the majority of cases the detection and discrimination of mechanical damage relies on the integration and analysis of multiple data streams. Often this data originates from multiple in-line inspection tools or from tool technologies that have been mechanically combined into a single tool.

3.1.1 Deformation Sensing

Vendors identified three types of geometry sensing technologies that are used to detect and characterize global deformations such as ovality and plain dents in the pipe wall:

- Direct Arm Measurement
- Indirect Electromagnetic Measurement
- Magnetic Flux Leakage (MFL)

Table 3 summarizes these technologies in terms of generic category descriptors used in this report.

Direct Arm Measurement sensing technologies are generally employed in tools known as calipers. Calipers have been historically grouped into two categories: single-channel tools and multi-channel tools. Single-channel tools only give distance traveled and the minimum pipeline diameter along a pipeline. They are useful on new construction projects or on line segments that have never been pigged. No participating vendor cited a single-channel tool as being currently employed to assess mechanical damage.

All Direct Arm Measurement Calipers (DAMC) considered in this project employ multiple mechanical arms (fingers) that measure the interior of the pipe geometry by contacting the inner surface of the pipe. Each arm is equipped with a sensor that measures the angle of the arm (using Hall Effect transducers). Radial movement of the sensor is converted into deviations of the pipe wall from circular form. Multi-channel DAMC tools provide data, such as deformation depth, deformation orientation, length and width of the deformation, and the ability to make longitudinal and circumferential strain calculations based on the rate of movement in each geometry sensor. Various proprietary mechanical designs are incorporated in DAMC's to address issues associated with sampling intervals, inspection speed excursions, sensor bounce, vibration and sensor coverage of the internal pipe surface.

Technology	Sensor	Number of Sensor Rings or Planes	Generic Technology Key
A	Direct Arm Measurement with Electromagnetic Sensor	1	DAMC (EM)
C	Direct Arm Measurement with Electromagnetic Sensor Combo with MFL	1	DAMC(EM)
D	Direct Arm Measurement	1	DAMC
E	Longitudinal Field MFL-3 Axis	1	{Long Field} MFL [Hall-3] [ID/OD EM]
F	Indirect Electromagnetic Caliper	1	IEMC
G	Direct Arm Measurement with Electromagnetic Sensor	2	DAMC (EM)
J	Direct Arm Measurement Combo with MFL	1	DAMC
K	Direct Arm Measurement	1	DAMC

Table 3: Categorization of Deformation Assessment Technologies

For DAMC(EM) technologies the ends of the mechanical arms in contact with the pipe wall are equipped with direct contact points (hardened steel tips, pads or rollers) and, depending on the particular technology, may be augmented with electromagnetic (EM) sensors on the tips of the arms (Hall Effect or Eddy Current). The electromagnetic sensors provide an additional data stream for analyses that address potential sensor liftoff, which occurs due to tool speed/sensor inertia, deformation geometry, interior cleanliness of the pipe and other conditions.

Indirect Electromagnetic Calipers (IEMC) was also identified for mechanical damage assessment. This type of deformation sensor does not rely on direct contact with the pipe wall. A ring of electromagnetic sensors (Eddy Current) mounted on a fixed diameter ring, with a diameter sufficient to allow for passage of the tool through a maximum expected bore restriction, gauge the distance between the ring and pipe wall.

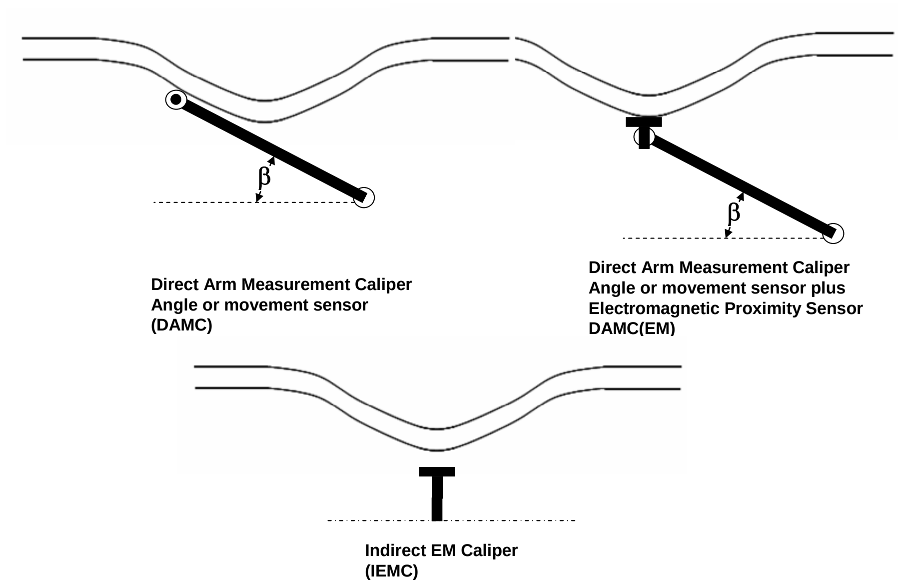


Figure 4: Illustrations of the three types of caliper sensor technology reported for mechanical damage technologies.

The probability of detection (POD) and sizing performance for deformations are greatly influenced by the circumferential resolution and the coverage of the sensing area of the caliper sensors[12]. Transducer resolution, data sampling rate and contact area of sensors also affect performance.

Design and spacing of the sensing arms determines the circumferential resolution of a caliper tool. Narrow arms provide greater resolution of the contour. Conventional resolution direct arm calipers are typically designed with the same number of arms as the nominal pipe size (NPS), such as 12 arms or 20 arms for NPS 12 or 20 pipelines, respectively. The circumferential spacing between the sensing arms is approximately 3.14 inches when the number of sensors equals the NPS[13]. Higher resolution caliper tools of the types identified by the participating ILI vendors for mechanical damage are designed with more narrow spacing, typically 1 to 3 inches. All of the direct arm type calipers considered in this project, except Technology G, consisted of a single plane of deformation sensors. For single plane direct arm tools the sensor spacing decreases with bore restriction. Therefore, the maximum coverage of the internal pipe surface at nominal NPS with a single plane of direct arm sensors is close to the specified bore reduction (typically 75%).

[12] Beuker, T, Rahe, F.,"High Quality Smart Pig Inspection of Dents, Compliant with the US Code of Federal Regulations", Oil and Gas Processing Review 2005.
[13] Michael Baker Jr, Inc.,"Dent Study-Final Report", TTO Number 10 DTRS56-02-D-70036, page 18, Nov 2004

Higher coverage is achieved with technologies employing dual offset, direct arm sensor rings (Technology G) or indirect electromagnetic measurement, IEMC (Technology F).

Accuracy in measurement of deformation depth is also affected by the sensor coverage or direct arm spacing. If the maximum depth of a deformation is not hit by a sensor, there is a potential for under-sizing.

Linear resolution of a deformation contour is related to the sampling rate. Sampling rate may be in terms of either elapsed time or distance traveled. When the sampling rate is by elapsed time, the resolution can be affected by variations in velocity as the tool traverses the pipeline. When the sampling rate is by distance traveled (as determined by an odometer), rapid acceleration or deceleration can cause odometer slippage and introduce errors in measurements. Linear sampling at intervals of 0.25 inch and less is characterized as high linear resolution, while sampling at intervals of 1 inch and greater is characterized as low linear resolution[3]. All MFL tools can detect presence of dent; however characterization is compromised by the influence of stress. The stress distribution around a mechanically damaged region is very complex, consisting of plastic deformation and residual (elastic) stresses. Consequently in a traditional MFL signal it is difficult to differentiate the geometric element from the stress factor.

Technology E is the only technology that does not use a caliper data stream to characterize deformations. This technology utilizes 3-axis Hall Effect sensors together with a secondary ID/OD sensor and tool inertial data to detect and size dents. In addition, Technology E uses magnetic models to incorporate both stress and geometry effects into interpreting MFL signals from dents. Stress and geometry radial MFL signals from circular dents have distinct characteristics[14]:

1. Geometry signal: the geometry signal is characterized by a set of primary central peaks and secondary central peaks. The secondary central peak lies between, and often overlaps signals associated with the shoulders of dents.
2. Stress signal: In circular dents, there are stress peaks at the dent base and also in the rim region. The peak associated with the dent base stresses lies 'underneath' the primary central geometry peak. The stress rim peaks form a partial halo that combines with the geometry peak to cause an extensive halo around the dent.

Generally, dents are not considered to be reliably identified and sized by MFL data especially purely axial MFL signal. This is mostly due to the insensitivity of axially orientated magnetic field sensors. The flux leakage is spherical in nature and its axial, radial and circumferential vector components can be identified and measured by applying a cylindrical frame of reference coinciding with the pipeline axis. The magnetic signature of a deformation is unique and is not generated due to an absence of material, but rather the signal is determined by the shape of the dent. Strong neodymium iron boron magnets typically saturate the steel to approximately 300 Oersteds. At this level of magnetic saturation, the effects of plastic deformation, cold working, or residual stresses can be eliminated or at least minimized.

---

[14] Lynann Clapham*, Vijay Babbar, Alex Rubinshteyn, "Understanding Magnetic Flux Leakage Signals from Dents", Queen's University, Kingston, Ontario, Canada ,IPC2006-10043

Typically, it is the radial and circumferential components of a MFL signal that give the most information about a deformation, not the axial component. The radial component of a deformation is characterized by a central region of the signal pattern that roughly corresponds to the shape (length and width) of the deformation, when viewed in the vendor's analysis software. A halo that is recorded on the edge of the deformation is in part due to the geometry of the dent, but it is primarily due to the way in which the tool traverses the deformed surface. As the tool travels through a deformation, the sensor ride, tool center, and magnetic pole position all play roles in the creation of the halo. The circumferential component of a deformation is characterized by two equal and opposite peaks. The amplitude and shape of these peaks are very important when assessing a deformation[15]. Sizing algorithms relating the peak depth of dents to the amplitude and shape of these peaks have been incorporated in Technology E. Other technologies utilize data from multiple magnetic data streams for analysis in the differentiation of dents but did not report sizing capability based on MFL alone.

3.1.2 Coincident Damage Sensing

Magnetic Flux Leakage (MFL) and Ultrasonic Tools (UT) in combination with deformation tools are utilized to detect and discriminate localized mechanical damage in the form of gouges and cracks, corrosion or gouges within dents. These types of damage were broadly categorized as localized damage due to moved or removed metal by Bubenik et al[16]. The sensitivity to deformation detection is based on the type of sensor assembly, the size of the sensors, and the type of sensors. MFL tools may not produce a signal response due to deformations that result in ovality or pipe that is gradually egged shaped with no sharp contour changes. However, MFL tools do respond to any sharp portions of the deformation by the sensors "lifting off" the inside pipe surface. This creates a magnetic field change, causing a signal response. The proper analysis of the MFL raw data is key to gaining information about the deformation anomalies. In all cases (except for Technology E), the mechanical damage processes reported include integrated analysis of caliper based deformation data in conjunction with MFL data. The longitudinal length of a deformation may be determined by measuring the length of pipe where there was sensor disturbance. The circumferential width may be determined by the number of channels or sensors affected. In many cases, there is a relationship between the depth of a symmetrical deformation and the circumferential width.

The signal shape for a typical MFL signal from a dent is fundamentally different than that seen from metal loss. The signal is due to two effects that occur at the same time. First, the sensor orientation relative to the local pipe wall changes. The sensor still records the axial field, but the pipe wall is no longer parallel to the sensor. Since the flux field is a vector quantity, the resultant measurement changes. Second, residual stresses and strains change the local magnetic properties of the pipe. Dent signals show characteristic peaks near the start and finish of the dent with a relatively low signal through the defect.

---

[15] Scott Ironside, Characterization of Mechanical Damage Through Use of The Tri-Axial Magnetic Flux Leakage Technology, IPC 2006-10454.

[16] Bubenik, T.A., et al, In-Line inspection Technologies for Mechanical Damage and SCC in Pipelines- Final Report, DOT DTRS56-96-C-0010 June 2000.

Metal loss creates a different change in the magnetic flux field and thus different signals than deformation "lift off." Deformations that contain metal loss may be distinguished by their combined metal loss and deformation lift-off signals However, certain deformations that have a small radius of curvature and thus a steep slope may cause significant lift-off of the MFL sensors, thereby reducing the ability to detect metal loss.[17] The integration and analysis of other data streams, such as tool acceleration, data from inertial sensors and secondary ID/OD discrimination data, are utilized in the discrimination of co-incident damage in dents.

The typical MFL signal from a cold worked region will generally exhibit a signal shape that is fundamentally different from that of both metal loss and a dent because flux in the region immediately below the cold worked area decreases. This change occurs because the cold worked region, which is on the side opposite the sensor, carries more flux, thereby reducing the flux in the rest of the pipe. In addition, there is a slight increase in signal at either end of the feature. These signal features are characteristic of mechanical damage. Current mechanical damage process employs signal processing, unique for each vendor, enabling detection and discrimination of these characteristic features for mechanical damage.

A pipeline will tend to re-round due to internal pressure after impact by an indenter. Because of the denting and re-rounding process, residual stresses and plastic deformation arise at the outer edge of the maximum dent length and width. These stresses give rise to a small amount of magnetic deformation in the re-rounded area. Viewing of the magnetic signals using vendor propriety computer software provides a "pseudo-color data display"[18] generated by algorithms equating signal strength, or rate of change, to color representations. This "color display" provides a method for visually interpreting the MFL signal behavior (Figure 5 and Figure 6). A halo signal can be observed in such displays and is caused by a ring of magnetic flux change surrounding a defect that has been re-rounded from internal pipe pressure. This halo is typically largest at the maximum dent length. The halo signature is visible at high magnetization levels and, while some detail is not apparent at a high magnetization level, there are characteristics that have been used by ILI analysts for detection and discrimination of mechanical damage. The halo length and the maximum dent depth are related, and this relationship is being used to estimate dent depth. Similar relationships have been exploited by Technology E to predict dent depth and qualitative observations regarding halo signals are utilized by all the identified MFL based technologies to identify mechanical damage.

The discrimination between mechanical damage consisting of dents with metal loss and non-mechanical damage metal loss features, or between corrosion and gouges are aided by observations of the halo signals.

[17] Michael Baker Jr., "Dent Study, Final Report", PHMSA TTO Number 10 DTRS56-02-D-70026, page 17.
[18] Nestleroth, J.B., Bubenick, T.A.,"Magnetic Flux Leakage (MFL) Technology For Natural Gas Pipeline Inspection", GRI Report, Feb 1999.

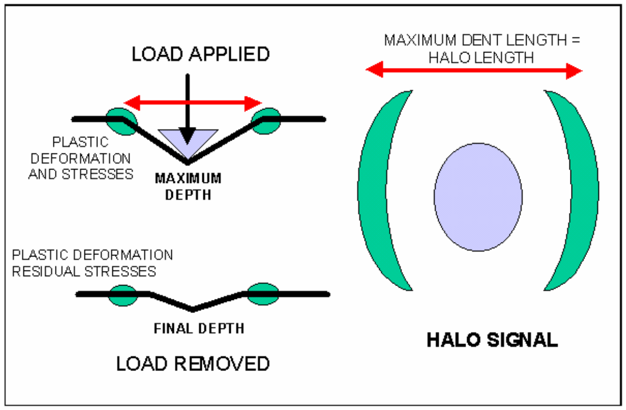


Figure 5: Descriptions of Halo representations in MFL analysis and suggestion of relationship of halo dimensions to dent depth from Bubenik.[16]

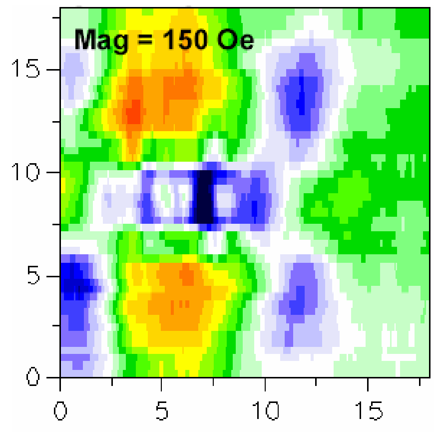


Figure 6: Color scan representation of mechanical damage from high field strength MFL from Tietsma illustrating halo signatures.

ILI Vendors reported that Hall type magnetic flux leakage sensors, with various physical and magnetic vehicle configurations, or ultrasonic sensors were the principal technologies employed for detection and discrimination of coincident features within dents such as metal loss, gouges, cracks or gouges on the body of pipe without deformation.

The physical configurations of the current mechanical damage technologies for coincident mechanical damage features were characterized according to the following categories:

Technology Type for Coincident Features
- MFL or UT

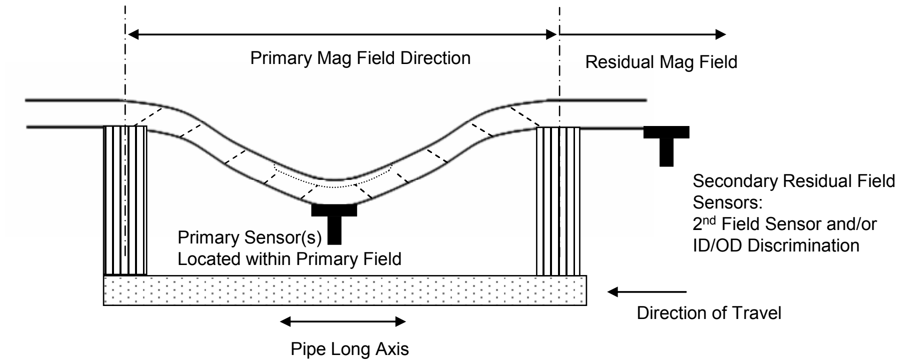
Primary Sensor Field Directional
- Magnetizer Field Direction for MFL
- Directional Bias for UT

Primary Sensor Type, Technologies or combination within primary sensor element
- Coil type or Hall Effect Sensor, # of Axes
- ID/OD Discrimination, Eddy Current or Hall Effect Sensors
- Ultrasonic Sensor Type if utilized

Secondary Sensors

- Sensors located outside of the principal sensor element or magnetizer vehicle

Figure 7 illustrates the technology categorization coding utilized for coincident damage technologies.



{Primary Magnetic Field Direction}	Type	[Primary Sensor Type]	(Secondary Residual Field Sensors)
{Long Field} {Circ Field} {UT Direction Bias}	MFL UT	[Coil] [Hall-#Axis] [ID/OD EM] [Long Wave] [Shear Wave]	(ID/OD EM) (Hall-#Axis)

Figure 7: Technology Characterization Key for Coincident Damage Sensing

Table 4 is a summary the tool technologies evaluated in this study for detection and discrimination of mechanical damage. The definition of discrimination used in this study is borrowed from Bubenik: to discern and identify dents, gouges, dents with coincident metal loss (corrosion or gouges) or cracks. Figure 8 summarizes the type of mechanical damage these various technologies are expected to detect and discriminate.

Technology	Generic Technology Key	Category
A	DAMC (EM)	Deformation
B	{Long Field} MFL [Hall-1] [ID/OD EM]	Deformation/Local Changes Coincident Damage
C	{Long Field} MFL [Hall-1] [ID/OD EM] (Hall-1) DAMC(EM) Combo	Local Changes Coincident Damage
D	DAMC	Deformation
E	{Long Field} MFL [Hall-3] [ID/OD EM]	Deformation/Local Changes Coincident Damage
F	IEMC	Deformation
G	DAMC (EM)	Deformation
H	{Long Field} MFL [Hall-1] [ID/OD EM]	Local Changes Coincident Damage
I	{Circ Field} MFL [Hall-2] [ID/OD EM]	Local Changes Coincident Damage
J	{Long Field} MFL [Hall-2] [ID/OD EM] DAMC Combo	Local Changes Coincident Damage
K	DAMC	Deformation
L	{Circ Field} MFL [Hall-1] [ID/OD EM]	Local Changes Coincident Damage
M	{Long Bias}UT[Shear]	Local Changes Coincident Damage
N	{Long Field} MFL [Hall-1] [ID/OD EM]	Local Changes- Operator Data

Table 4: Technology for Mechanical Damage Presented by Participating ILI Vendors and Operators

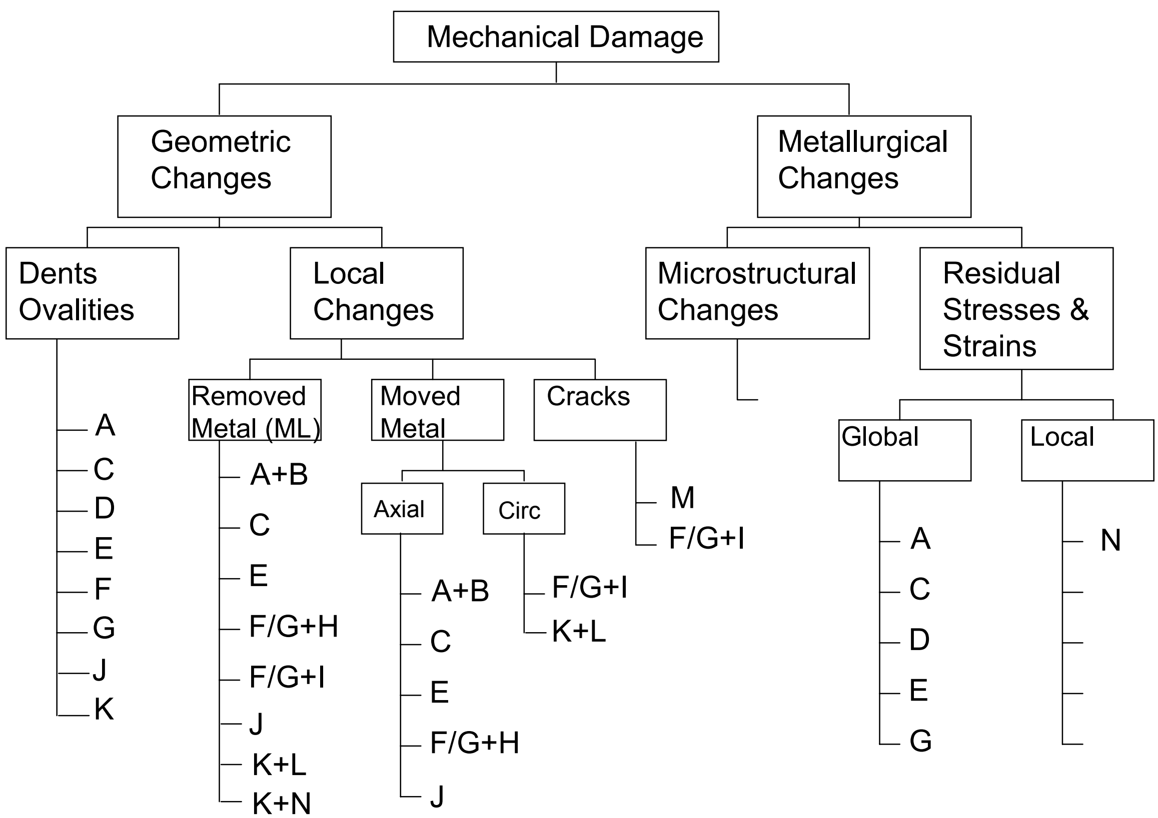


Figure 8: Mechanical Damage Sensing Technology Generic Description Key

3.2 Mechanical Design

The tools identified by the participating vendors represent either independent technologies for measuring deformations and coincident damage (moved or removed metal), or combined technologies within a single vehicle. Detailed descriptions are given in Appendix A.

The ability of sensors to remain in contact with the pipe wall when riding over deformations has been recognized by many vendors as being critical to resolution and sensitivity of dent measurements, and various vendors have incorporated proprietary design features. Several vendors also identified keeping the tool core on the centerline of the pipeline as another proprietary design consideration. This issue is of particular importance for deformation tools as they traverse across dents and through bends. When considering a current technology for assessment of mechanical damage the user should remain aware of the availability of such features within various technologies and evaluate the applicability of these features to the conditions of the pipelines under consideration. Given the proprietary nature of these mechanical designs, no specifics are discussed in this report.

Caliper is a direct measurement device, which is calibrated. All of the vendors described calibration procedures with varying details depending on the individual designs for the tools. Tools for the detection and discrimination of coincident damage represent more complex systems and, as such, are more difficult to qualify than calipers. All of the vendors reported detailed qualification procedures for components of these tools. Given the proprietary nature of these tools no specifics of these procedures are discussed in this report.

3.3 Data Analysis

According to US IMP Rules[19] plain dents up to 6% OD are generally allowed to remain in a pipeline, except for pipeline segments affecting High Consequence Areas, where more stringent requirements apply, such as top of pipe dents where 2% OD limits apply. The Rules also identify conditions such as features that require further investigation, dents (without specified severity) with coincident metal loss, cracking or stress risers. As a result, the participating vendors reported that pipeline operators and inspection vendors have had to agree on limits for threshold reporting of dent depths and coincident metal loss, cracking or stress risers. Vendors reported threshold limits for dents ranging from 2% down to, in some instances, the lowest detection limits of the technology that was used. Vendors also reported threshold limits for coincident damage depth, such as metal loss (corrosion or gouges) or cracking, to vary from the lowest detection limits for the applied technology (any depth > 0%) to threshold depths as high as 20% depending on the customer.

A key feature in MD analyses for all technologies is alignment of the multiple data streams from independent vehicles.

Vendors reported that the analysis of inspection data for mechanical damage from all technologies is highly dependent on interpretation by trained subject matter experts (SME). All

[19] 49 CFR 195.452, 49 CFR 192 Subpart O.

of the vendors reported similar overall qualifications for Subject Matter Experts evaluating MD data. ASNT ILI-PQ-2005[20] Level II individuals are utilized for classification of mechanical damage features with Level III qualifications required for final sizing and discrimination based on proprietary processes developed by the vendors for their individual technologies.

3.4 Measurement Reporting Capability

Participating vendors are not making performance claims related to coincident feature sizing (metal loss, corrosion or gouge within dents), probability of detection, or minimum thresholds for detection and reporting for any current MD technology. All mechanical damage performance beyond plain dents is offered on a best endeavor basis with reporting thresholds on agreement with the customer.

Validation data in the form of laboratory pull tests or direct examination measurements from excavations was supplied by the vendors for the purpose of demonstrating the claimed capabilities. An analysis of the validation data is presented in Appendix B of this report. Additional validation data was provided by three pipeline Operators (one liquid pipeline and two gas pipeline operators) who utilized some of the technologies identified in this research and this data.

Descriptions of the technologies based on the data provided by the participating ILI vendors are detailed in Appendix A. Individual technologies cited represent in-line tools that can be run as an independent inspection vehicle. It is noted in the category descriptions where a technology requires a combination of data from another separate technology and the associated vehicle.

[20] American Society for Nondestructive Testing ANSI/ASNT. Standard ILI-PQ-2005, In-line Inspection Personnel Qualification and Certification. August 2005

4.0 PERFORMANCE EVALUATION

4.1 Validation Data

Evaluation of the performance of the current technologies was conducted using the validation data provided by both participating vendors and operators. The participating ILI vendor's provided performance data for all of their current technologies to detect, discriminate and size mechanical damage. Additionally, one liquid pipeline operator and two gas pipeline operators volunteered results from recent direct examinations following an in-line inspection of MD.

The validation data represents samples from populations of all detections and measurements reported by the individual technologies. The analysis performed within the current project considers characteristics of the samples and statistical methods to determine performance characteristics.

Performance of in-line tools sizing and characterizing deformations reflects:

- Errors from the ILI tool itself
- Errors in in-ditch validation measurements
- Variability inherent in the pipe, most notably, variability in shape and size of deformations due to changes in internal pressure (re-rounding) and changes in external confining forces (re-bounding).

For present purposes of this report

- Rebounding causes a decrease in dent depth due to elastic unloading that occurs when the indenter is removed during excavation.
- Re-rounding causes an increase in dent depth due to decreased internal pressure during excavation.

Errors affecting ILI system sizing performance for dent depth are illustrated schematically in Figure 9.

The effect of re-rounding has been analyzed and specific guidelines made for sizing its effect[21]. However, data from mechanical damage excavations is very unlikely to contain the operative pressure at the time of intervention/excavation, because this information is not consistently recorded.

[21] A. Le Bastard. *Influence of internal pressure for depth measurement on a dent.* Paper #10103, International Pipeline Conference 2006. Calgary, Alberta, Canada, September 25th to 29th, 2006.

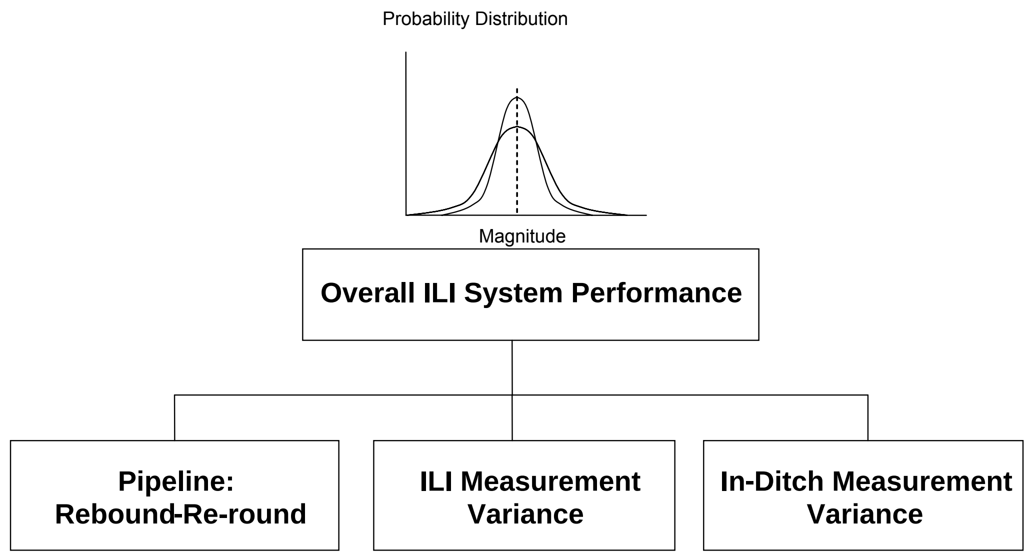


Figure 9: Sources of errors in dent measurements

The effects of the soil bed removal during excavations are less well understood and are difficult to evaluate in field. The possibility of direct dent measurement in the field before excavation (in addition to ILI tool measurement) is very difficult, if not impossible, to achieve.

4.2 Performance Based on Samples

The approach used to understand the capabilities of current MD technologies was to evaluate performance parameters based on the available validation data. It is standard statistical procedure to use samples taken from a population, such as validation measurements from a population of ILI predictions, to estimate properties of the population, such as error in ILI predictions. We used binomial distributions, confidence intervals, and linear regression to assess various properties of data obtained for this project. Specifics of these techniques are discussed Appendix B.

Binomial Distribution Analysis – This method extends a method outlined in API 1163[26]. The method assumes that the validation measurements are independent of each other and the "successful" measurements have a Binomial distribution. The analysis estimates the probability (or certainty) of finding "x" out of "n" measurements within the desired tolerance, or estimates the probabilities of detection and identification for a given confidence level for a relatively small sample.

Confidence Intervals (CI): Confidence intervals for the certainty are the recommended procedure in API 1163[28]. Confidence intervals can also be used to estimate a tool's certainty for a given tolerance and confidence level, or to estimate the probabilities of detection and identification for a given confidence level from a relatively small sample. Many confidence intervals for binomial variables appear in the literature.[22,23,24] We have chosen one, the Clopper-Pearson confidence interval, that always gives conservative results[25].

[22] Brown, L. D., Cai, T. T., DasGupta, A., 2001. Interval Estimation for a binomial proportion (with discussion), Statist. Sci. 16, 101-133.

Regression Analysis- Regression Analysis examines the relation of a dependent variable (e.g., ILI measure dent depth) to specified independent variables (e.g., field measured dent depth). A detailed discussion of regression analysis is found in Appendix B. Part of this discussion shows that errors associated with approximating data by a regression model must satisfy certain properties in order for the model to be acceptable. In particular, the use of regression equations to calibrate tools, or otherwise adjust data to account for biases in the tools, should be avoided if possible, whenever these properties are not approximately satisfied.

Both binomial distribution analysis and confidence interval techniques were used for probability of detection, probability of identification and probability of false call analysis. All of the above three methods were employed to evaluate validation data to predict sizing tolerances. The results and findings are briefly summarized in this section. The detailed results of this performance analysis are presented in Appendix B.

4.3 Measures for Technology (ILI Tool) Performance

ILI Tool (technology) performance, in general, is evaluated by the following five performance measures[26,27]

- Accuracy of Sizing
- Probability of Detection (POD)
- Probability of False Calls (POFC)
- Probability of Identification (POI)
- Locating

Accuracy of Sizing is a measure of an ILI technology's ability to accurately predict an anomaly's dimensions, typically, depth, length, and width. API 1163 indicates that the sizing accuracy shall include the following three parameters

- Tolerance, e.g., ± 10% WT in depth for corrosion
- Certainty, e.g., 80% of time
- Confidence level, e.g., 95%

Two types of statistical methods, namely binomial distribution analysis[27] and confidence interval analysis[28] were used to assess both Vendor and Operator data. This statistical approach takes into

---

[23] Brown, L. D., Cai, T. T., DasGupta, A., 2002. Confidence Intervals for a binomial proportion and asymptotic expansions. Ann. Statist. 30, 160-201.
[24] Piegorsh, W. W., 2004. Sample sizes for improved binomial confidence intervals, Comp. Statist. & Data Anal, 46, 309-316.
[25] McCann, R., McNealy, R., and Gao, Ming, In-Line Inspection Performance, II: Validation Sampling, NACE, Corrosion 2008, Paper No. 1177
[26] API 1163:"In-line Inspection Systems Qualification Standard", First Edition, August 2005.
[27] Desjardins, G. ,Reed, M, Nickle, R., ILI Performance Verification and Assessment Using Statistical Hypothesis Testing. IPC 2006, paper no: 10329, 2006.
[28] McCann R., McNealy R., Gao M.: "In-Line Inspection Performance Verification", NACE Corrosion 2007, #0713, Nashville, TN, 2007.

account the use of small samples to validate a tool's sizing performance. A detailed discussion of these statistical methods is given in Appendix B.

In the pipeline industry, it traditionally uses a unity graph plot of ILI and field data along with the least square regression analysis to determine the error bands (i.e. tolerance) for a given set of data at the desired confidence level resulting in the following two parameters:

- Tolerance, e.g., ± 10% WT for corrosion
- Confidence level, e.g., 80%

The advantage of this two parameter method is that the unit graph plot explicitly shows the relationships between ILI and excavation data (slope and coefficient of determination, R^2), error bands (standard deviations), as well as an estimate of the tool bias (intercept). This provides an opportunity to investigate ILI's systematic errors, possible sources of errors, and appropriate adjustment or calibration. However, this method does not provide information on how good or bad it could be to use a small size of field excavation samples (n) to represent the large population of anomalies (N) on the pipeline to validate the tool sizing performance. The assessment results (e.g., error bands and a linear equation) are valid only within the data set used for regression. Predictions should not be extended beyond the range of the data set used.

Therefore, these two statistical approaches (i.e. binomial/confidence interval and least squares linear regression analyses) are considered to be both complementary and supplementary. For the present study both approaches are used to evaluate the technologies' sizing performance and compare the results as appropriate. The detailed procedure and assumptions for analysis is given in Appendix B.

Probability of Detection (POD) is the standard measure of the ILI technology's ability to find and report pipeline features and anomalies that existed on the pipeline. POD is defined as the number of anomalies correctly detected and reported by the ILI tool divided by the total number of anomalies on the pipeline[26]:

POD = [# times detected]/[# of anomalies] x 100% per anomaly/feature type/size

where, the "# times detected" is the number of features (per type) detected by the ILI tool (technology) and the "# of anomalies" is the number of anomalies (per type) on the pipeline.

Probability of False Call (POFC) A closely associated measure against "Detection" is the frequency that the tool falsely reports an anomaly where no anomaly exists. The probability of false call (POFC) is defined as the probability of a non-existing feature being reported as a feature[27]:

POFC = [# of false calls]/ [# of anomalies + # of false calls] x 100% per feature type

The ability of detection and the number of false calls are often related. Increasing the ability of detection might result in an increase in false calls. This is in part due to ILI signal interpreters and analyzers being afraid of missing critical defects when faced with the choice of whether or not to report ambiguous or ill-defined anomalies.

Probability of Identification (POI) is the standard measure of the technology's ability to properly discriminate an anomaly from other types of anomalies that exhibit similar signals. The probability of identification is defined as the probability that an anomaly or other feature, once detected, is correctly classified[27]:

POI = [# times correctly identified]/[total # of detected anomalies] x 100% per feature type

It is important to note that all these probabilities are defined in terms of the total population of anomalies. Knowledge of the total population is very rare, because the entire pipeline would need to be exposed to direct examination. This is practically unfeasible. Only a portion of the pipeline is exposed and assessed for tool performance evaluation. Therefore, traditionally the ratio of "success of detection", or the rate of "success of identification", (x) to the total number of samples (n), i.e., x/n, is reported as POD, or POI, respectively[27]. Similarly, the ratio of false calls (also denoted as x) to the number of samples (n), x/n is used to represent POFC. Obviously, POD, POFC and POI, obtained in this way are NOT the true probabilities, BUT the proportions of detection, false calls and identification. To avoid confusion, in this report we use **Proportion of Detection, Proportion of False Calls and Proportion of Identification** to indicate the values that are calculated from the ratio of "x/n", and **Probability of Detection (POD), Probability of False Calls (POFC) and Probability of Identification (POI)** to indicate the values that are estimated from the statistical methods. It is important to understand the difference between a probability and a proportion that approximates it.

Locating is a measure of a technology's ability to correctly report the location of a reported pipeline anomaly. Since locating is not a concern for the current ILI technology for MD inspection, only the first four measures were used in detail for the present study.

4.4 Dent Depth Sizing Performance (Tool Tolerance)

4.4.1 Binomial Distribution and Confidence Interval Analysis

Assessment of tool performance began with depth measurements because depth measurements were the most widely available data. Evaluation of depth sizing accuracy with an assumption that no errors or insignificant errors were involved in the validation measurements.

Both binomial distribution analysis and confidence interval methods were used. There are several binomial confidence intervals in the literature[22,23,24]. The choice of confidence intervals is partially dependent on whether a guaranteed nominal confidence level is required. The Clopper-Pearson confidence interval was selected for use in this report because the Clopper-Pearson interval guarantees that the actual confidence is at least the nominal confidence level. The lower (left) and upper (right) limits of confidence intervals were calculated using the one tail approach.

Table 5 shows that the two methods provided comparable results. The tool tolerance determined from the binomial distribution analysis is either the same or slightly smaller than those obtained from the Clopper-Pearson confidence interval method. The reason for this is probably related to the nature of the Clopper-Pearson confidence intervals, which is conservative and guarantees that the actual coverage probability is always equal to or above the nominal confidence interval.

Analyses were performed on data from six Vendor Technologies, four of the technologies showed depth tolerance within ± 1.0% OD at a certainty of 0.80 and a 95% Confidence level. Combining data from Technologies A & G and A & C & G showed a tolerance of ± 0.93% while Technology A showed about 10% larger than ± 1.0% OD for a given certainty of 80% and confidence level of 95%. Technology C showed the smallest tolerance of ±0.51% OD from binomial distribution analysis but showed ±0.78% OD (that is comparable to Technology G and E) from the Clopper–Pearson confidence interval analysis. It is noted that the sample size for Technology C is also smaller than the others. Therefore, the difference in tool tolerance between Technologies G, C, and E may be insignificant.

Liquid and Gas pipeline operators provided data to supplement availability from vendors. Every effort was made to match operator data to the definitions of current mechanical damage technologies defined by the vendors. Operator the data is identified by the subscript “o”. It must be noted that for operator data, assessment of mechanical coincident damage, such as dents with metal loss, was not made using the full current process identified by the vendors in all instances. Data sets with fewer than 9 points (Technology F) will not be included in discussions regarding results and critical comparisons. Technologies represented by Vendor and Operator data are discussed in detail and compared Table 5.

Data from the operators, showed a much larger depth tolerance, ranging from ± 1.6%OD to ± 4.48%OD. Reasons for the larger tolerance are most likely associated with re-rounding and rebounding of dents and the accuracy of field measurements.

Data Analyzed	Technology Type	Tolerance for Certainty = 0.8 at 95% Confidence Level			
		Sample Size for Validation	Limits of Detection (%OD)	Binomial Distribution Analysis	Clopper-Pearson Certainty Interval Method
Technology A	DAMC (EM)	130[a]	0.5	±1.10%	±1.22%
Technology G	DAMC (EM)	20[b]	0.5	±0.74%	±0.74%
Technology C	DAMC (EM)	15[c]	0.5	±0.51%	±0.78%
Technology E	{Long Field} MFL [Hall-3][ID/OD EM]	273[d]	2.0	±0.78%	±0.80%
Technology A and G	DAMC (EM)	150	0.5	±0.93%	±0.93%
Technology A, C, and G	DAMC (EM)	165	0.5	±0.88%	±0.88%

a) Direct examination observations of MD from NPS 16 pipeline
b) Direct examination observations of MD from multiple ILI
c) Laboratory Pull Test for 20 plain dents in NPS 8 pipe
d)Dent depth predicted by Technology E and validated against multiple tool size ILI performed by Caliper [DAMC]-Technology D. The present evaluation of the most statistically significant sample, available for Technology E, assumed the dent depth sizing performance of the referenced Technology D at 85% certainty of +/- 0.60 % OD. (A level consistent with the validation results observed for DAMC (EM) Calipers).

(a)

Technology	Tolerance for Certainty = 0.8 at 95% Confidence Level			
	Total (n)	Successes (x)	Binomial Distribution Analysis	Clopper-Pearson Certainty Interval Method
Technology Ao	28	24	±1.60%	±1.60%
Technology Co	58	18	±4.48%	±4.65%
Technology Jo	17	9	±1.63%	±1.63%
Technology Ko	166	95	±2.37%	±2.37%

(b)

Table 5: Tool Depth Sizing Tolerance evaluated using binomial and confidence interval techniques for a certainty of 80% and confidence level of 95% (a) vendors' data and (b) operators' data

Confidence intervals for certainty at a given tolerance of ± 1.0% OD and given confidence level of 95% were also determined. Table 6 gives the results, showing that the binomial distribution analysis provides consistent, but slightly higher than the lower bound values of certainty interval estimated by the Clopper-Pearson confidence interval technique. From Table 6, it is seen that all of the vendors' technologies, except for Technology A, exceed a 80% certainty with 95% confidence level. For Technology A, its lower bound certainty is 77.6%, while it's true certainty is not known, but lies between 77.6% and 88.9%.

No vendor-supplied data was available for Technology K (DAMC type caliper). Pipeline operators made dig validation data available from inspections using a technology that matches sensing technology consistent with Technology K and is marked as Technology Ko in this report.

The data provided by the operators showed much lower certainties, between 21.1% and 70.2%. Again, these low certainty values most likely resulted from re-rounding, re-bounding and the accuracy of field measurements, suggesting that a consistent protocol may be required for field validation sizing.

Technology	Certainty for Tolerance = ±1.0% OD					
				Binomial, 95% Confidence for Given x, n	Clopper-Pearson	
	Total (n)	Successes (x)	Proportion (x/n)		Lower Bound	Upper Bound
Technology A	130	109	0.838	0.784	0.776	0.889
Technology G	20	20	1.000	0.867	0.861	1.000
Technology C	15	15	1.000	0.829	0.819	1.000
Technology E	273	251	0.919	0.890	0.887	0.945
Technology A and G	150	129	0.860	0.812	0.805	0.904
Technology A, C, and G	165	144	0.873	0.828	0.812	0.919

(a)

Technology	Certainty for Tolerance = ±1.0% OD					
				Binomial, 95% Confidence for Given x, n	Clopper-Pearson	
	Total (n)	Successes (x)	Proportion (x/n)		Lower Bound	Upper Bound
Technology Ao	28	24	0.857	0.745	0.702	0.950
Technology Co	58	18	0.310	0.226	0.211	0.425
Technology Jo	17	9	0.529	0.364	0.311	0.740
Technology Ko	166	95	0.572	0.512	0.506	0.637

(b)

Table 6: Certainty for depth tolerance of ±1.0 OD at 95% confidence level: (a) vendors' technologies and (b) operators' data

4.4.2 Depth Sizing Accuracy – Linear Regression Analysis

Tool depth sizing performance was further analyzed using linear regression analysis. The equations of regression lines (intercept and slope), the Coefficient of Determination (R^2), the Standard Error (σ), and the Normality of Error Distribution (as measured by the value of the Tukey lambda) of the Vendor and Operator data were investigated in great detail. Detailed information of the analysis is given in Appendix B.

4.4.2.1 *Linear Regression Analysis of Vendors' Dent Depth Data*

The analysis started with vendor data to show what is the "best" that can be expected from the current mechanical damage assessment technologies. Operator data was then examined to augment technologies for which vendor data was insufficient, or not available. It is recognized that operator data may not necessarily represent the complete, or best, mechanical damage assessment technology available from vendors.

Figure 10 summarizes the ILI and validation data provided by vendors along with linear regression lines and lines with ±2σ (standard error) from the regression lines. From Figure 10, it is seen:

1. **Technology G:** Among the six technologies investigated, Technology G exhibits the most desirable behavior on the following:

   - All ILI measurements are within ± 1% OD error as compared to the corresponding validation (field) measurements.
   - Errors and regression residuals are normally distributed with mean = 0 and constant standard deviation.
   - The regression line fits the data well, with the coefficient of determination $R^2 = 0.9$.

   However, the regression equation of Technology G is $y = 0.7816x + 0.3639$, which is significantly different from $y = x$. A regression line significantly deviated from $y = x$ reflects systematic errors and a bias of the tool, if errors in validation data are insignificant as compared with the ILI measurement:

   - The tool over predicts dent depth for shallow dents with depth < 1.67% OD
   - The tool under predicts dent depth for dents deeper than 1.67% OD
   - Based on one available literature source[29], the systematic errors are more likely from the field measurement (validation data) due to the best fit line having a slope less than one. However, this may not be the case for Technology G, because validation was performed against the pull test for which errors of the field measurement may have been minimized.

The results from this analysis suggest the pull-through test is an effective and accurate way for validation of tool performance, because errors from re-rounding, re-bounding are eliminated. Other errors from validation measurements can be minimized. The observed errors, both systematic and random, should be attributed to the technology itself as well as the execution process of the pull test.

.

2. **Technology C:** The results of linear regression also call attention to Technology C:

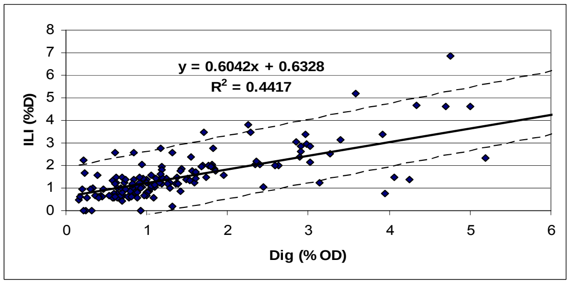
   - The highest coefficient of determination $R^2 = 0.92$
   - The regression equation is $y = 1.096x + 0.088$, which is reasonably close to the ideal equation $y = x$.

[29] G. Desjardins: "Assessment of ILI Tool Performance", Corrosion 2005, paper# 05164, NACE, 2005,

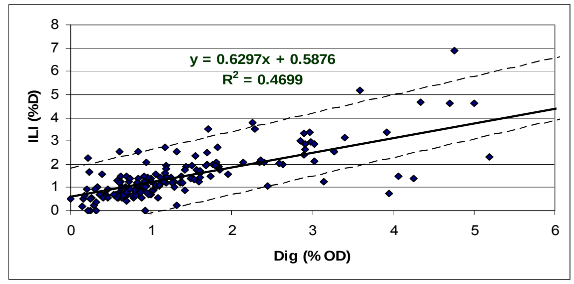
- The errors and residuals for Technology C are unlikely to be normally distributed. This means conclusions should not be based on the regression analysis.
- Technology C only contains 15 data sets within the range of 0 and 3% OD, with most data for depth less than 1%OD. Therefore, the established linear regression relationship is limited to shallow dents.
- More data with wider depth coverage is needed to support the analysis and to make a definitive conclusion even though the available data is promising.

3. **Technology E**: provided 275 validation data points, the largest data set for evaluation among the six technologies. This is about 25 and 13 times more data than Technology C and Technology G, respectively. Furthermore, the depth coverage is also the largest, from 1.5%OD up to 8%OD, with the most data in the range between 2 and 6%OD. Therefore, it is expected that Technology E provided the most representative and statistically meaningful data for evaluation.

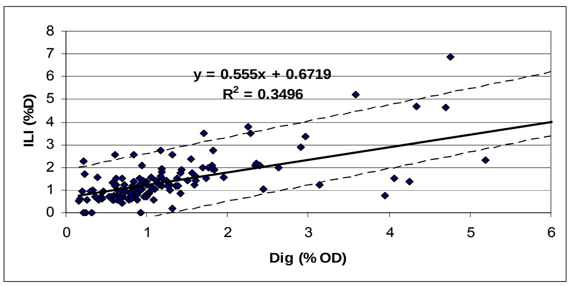
   - The regression line fits the data well, with a coefficient of determination $R^2 = 0.8$ (i.e., $R \approx 0.9$), lower than those of Technology C and G but significantly better than other technologies with their R^2 in the range from 0.35 to 0.47
   - The linear regression line equation shows $y = 0.42 + 0.914x$, which is close to $y = x$ (slope =1) with a systematic bias of 0.42
   - The errors and residuals for Technology E are likely to be approximately normally distributed. This means that conclusions can be based on the regression analysis.



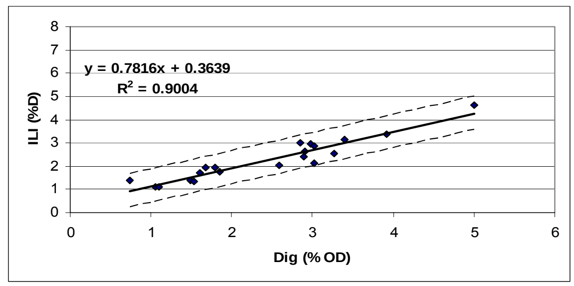
(1) Technology A,G –DAMC(EM)



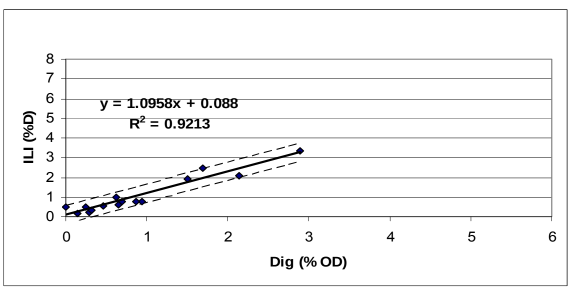
(2) Technology A,C,G- DAMC (EM)



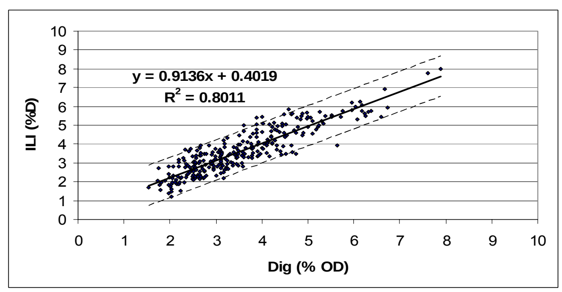
(3) Technology A- DAMC(EM)



(4) Technology G- DAMC(EM)



(5) Technology C- DAMC(EM)



(6) Technology E- {LongField}MFL[Hall-3][ID/OD EM]

Figure 10: Dig and ILI measurement of vendors' dent data. Solid Line – regression best fit line; Dashed Line- regression line $\pm 2\sigma$ (standard error).

More importantly, it is known that the validations for Technology E were conducted against Technology D, a DAMC type caliper, not the actual field (dig) measurements. Since errors associated with Technology D measurement are assumed to be significant as compared to those of Technology E, the first assumption required for linear regression, namely, "X values have no error or insignificant error compared to errors in Y values", may have been violated and the assessment results were questioned. Therefore, a detailed analysis was performed to account for the errors in the validation measurements***Error! Bookmark not defined.***. The analysis shows:

- Based on the 273 data points and an assumption that Technology D caliper measurement has an error of at most 0.6% OD at 80% of time, a linear relationship between

Technology E predictions and Technology D validation measurements can be established and have the form:

$$y = 0.0261 + 1.0802x + (\varepsilon - 1.0802\delta),$$

Where: y = Technology E measurement with error
ε = error in Technology E measurement
x = Technology D measurement with error
δ = error in Technology D measurement

- By using the estimated variances $\sigma_\varepsilon^2 = 0.1291$ and $\sigma_\delta^2 = 0.2190$ from the 273 data points and an assumption that both ε and δ are normally distributed with mean = zero, the error term $(\varepsilon - 1.0802\delta)$ in the equation $y = 0.0261 + 1.0802x$ has a standard deviation of $\sigma = 0.5632$.

- With the estimated standard deviation and the established equation $y = 0.0261 + 1.0802x$, the accuracy of depth sizing for Technology E is calculated, *that is*, ± 0.72% OD at a confidence level of 80%. In addition;

  - Technology E slightly over-predicts the caliper measurement due to the slope of 1.082 that is > 1 and the intercept = 0.026 that is > 0.
  - Practically, the equation $y = 0.0261 + 1.0802x$ is very close to the ideal equation of $y = x$. This implies that the two technologies (E and D) give essentially the same results, allowing for errors inherent in each technology.

The significance of the analysis demonstrates that a full conformation of the depth sizing performance for validating Technology D is required to reliably characterize the performance of Technology E. With the currently assumed Caliper tolerance (0.6%OD and 80% of time), the depth sizing performance of Technology E, i.e., the multi-vector MFL technology, is comparable to that of the caliper technology on mechanical damage assessment.

Similar analysis can be readily preformed to the technologies as long as the sizing error of validation measurement is given.

4. **Technology A**: Evaluation of Technology shows

- Significantly lower coefficient of determination R^2 (0.35) than those obtained for Technologies G, C and E
- Larger standard deviations
- Larger skewness;
- The errors and residuals for these technologies are unlikely to be normally distributed

The reason for Technology A having the lowest correlation between ILI and validation measurements cannot be determined whether this is due to the technology itself, or field measurements, or re-rounding and rebounding, or a combination, but they are certainly the possible causes.

Finally, the pooled data of Technologies A&G and A&C&G exhibits similar low correlations as Technology A. This is because the sample size (or, weight) of technology A is much larger than that of Technology C and G. The correlation trend is dominated by Technology A.

From the above analysis, it does illustrate the value of above ground pull tests for understanding the actual performance of mechanical damage assessment tools. It is also demonstrated that the linear regression method provides an opportunity to investigate sizing performance in terms of systematic errors, error bands, possible error sources if the errors in validation measurements are relatively small or are known (not small but accurate and reliable). With regard to the sizing accuracy, the linear regression method often provides tool tolerance that is represented by the error bands on the unity graph plot at various confidence levels[29]. As discussed previously, because the linear regression method analyzes error bands within a given data set, it is incapable of assessing the certainty and confidence level, i.e., how good or bad using such a small size of field excavation sample (n) to represent the large population of anomalies (N) on the pipeline. Therefore, the value reported by linear regression can be either consistent or significantly different from that reported by binomial distribution analysis, depending on the sample size and the magnitude of scattering of the data.

Technology	Linear Regression (Unit Graph) Method			
	R^2	R	Standard Deviation	Tolerance at 80% Confidence Level
Technology A	0.350	0.592	0.834	±1.07%
Technology G	0.900	0.949	0.289	±0.37%
Technology C	0.921	0.960	0.273	±0.35%
Technology E	0.801	0.895	0.563	±0.72%
Technology A and G	0.441	0.664	0.789	±1.01%
Technology A, C, and G	0.470	0.686	0.765	±0.98%

Table 7: Depth Tolerance, standard error and correlation coefficient estimated from the linear regression analysis for Vendors' data

Table 7 shows the estimated tool tolerance at 80% confidence level (i.e., 1.282 standard deviation) from the linear regression method for the vendor technologies, which ranges from ±0.35% OD to ± 1.07% OD.

4.4.2.2 Analysis of Operators' Dent Depth Data

Data supplied by the operators were placed in six groups according to the ILI technologies used to obtain the data. Only four technologies, A_o, C_o, J_o, K_o have sufficient data for meaningful analysis. Here, subscript "o" denotes operator data from possible previous generations of caliper technology. Conclusions and/or critical comparisons based on operator supplied data are limited to illustrate statistical significance of sample size or to provide data relating to a technology where none was available through the vendor. Figure 11 summarizes the ILI and validation data provided by the vendors along with linear regression lines and lines indication $\pm2\sigma$ (standard error) from the regression lines. From Figure 11, it is seen that

- All operator data show low Coefficient of Determination, R^2 (0.278 to 0.479) except for Technology A_o (0.819). This reflects large scatter about the regression lines in Figure 11.
- The operator data also show significantly larger standard deviations, except for Technology A_o, than the vendor data.
- The normality study shows that it is reasonable to assume that residuals and errors are normally distributed for Technology C_o but not for others. Therefore, data from other technologies have deviated from normal distribution and may not result in meaningful regression results.
- Technology A_o shows a standard deviation of 0.817 and Coefficient of Determination 0.819, which is comparable to the vendor data. Furthermore, the regression equation for Technology A_o is $y = 0.848x + 0.409$, indicating no large differences between ILI predictions and dig measurements. This suggests that re-rounding and rebounding are most likely not presented in the dig data. This may explain why data from technology A_o behaved more like vendor data than the other operators' data
- The regression lines for other technologies appear to significantly deviate from the ideal relationship $y = x$. The large intercepts and small slopes suggest the technologies over-predict when dent depth is small. The over-prediction of the dent depth is most likely associated with errors from re-rounding and rebounding when the field measurements were performed.
- An improved and consistent protocol for field validation measurement and analysis for mechanical damage assessment to minimize errors, in particular, errors from re-rounding and re-bounding, is needed.

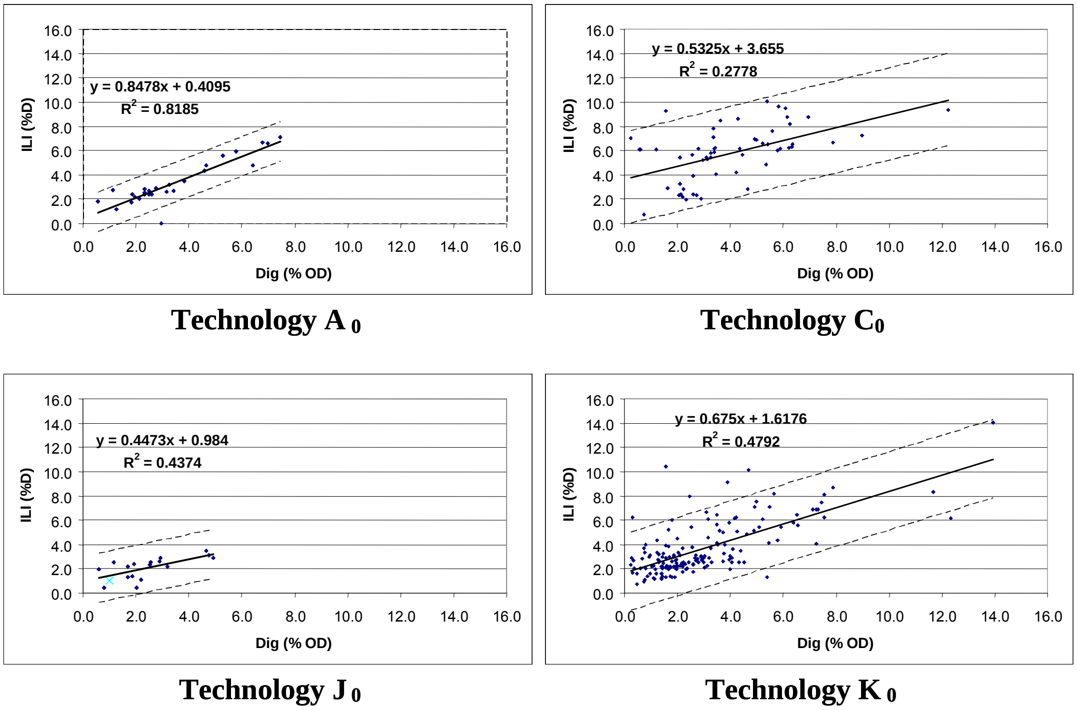


Figure 11: Dig and ILI measurements for technologies supplied by operators (dashed lines indicate regression line ±2 (Standard Error)

Comparison of Sizing Accuracy between Linear Regression and Binomial Distribution Analysis for operators' data is shown in Table 8.

Technology	Linear Regression (Unit Graph) Method				Binomial Distribution Analysis
	R^2	R	Standard Deviation	Error Band +/-1.28 Standard Deviation (80% Confidence)	Tolerance for Certainty=0.8 and 95% Confidence Level
Technology A_o	0.819	0.905	0.817	±1.04%	±1.60%
Technology C_o	0.278	0.527	1.906	±2.44%	±4.48%
Technology J_o	0.437	0.661	1.008	±1.29%	±1.63%
Technology K_o	0.479	0.692	1.561	±2.00%	±2.37%

Table 8: Depth Tolerance, standard error and correlation coefficient estimated from the linear regression analysis for operators' data

Table 8 shows the estimated tool tolerance at 80% confidence level (i.e., 1.282 σ) estimated by the linear regression method for the operator data, which are in the range from ±1.04% OD to ±

2.44%OD. For comparison, the tool tolerances estimated by the binomial distribution analysis for a given certainty of 80% at 95% confidence level are also included. Because of possibly re-rounding and rebounding errors, care must be taken if tolerance data is used for integrity assessment.

A complete analysis of data using the linear regression method is provided in the Appendix B. Appendix B also provides the analysis of length, width, and ML within the dents.

4.5 Proportion and Probability of Detection, False Calls and Identification for Dent with Metal Loss (DML)

The performance of current ILI technologies for characterizing mechanical damage is further evaluated in terms of their Proportions and Probabilities of Detection, False Calls, and Identification. The performance evaluation is focused on the ability of the tools to characterize dents with metal loss (DML), the condition with the most data.

Gouges are differentiated from metal loss due to corrosion by notch like profiles; localized cold working associated with moved metal and altered mechanical properties. Five technologies demonstrated capabilities to discriminate gouges. These technologies combine multiple sensor data streams, including deformation and magnetic signals such as low field magnetic permeability measurements or multiple magnetic field vectors provide a qualitative discrimination for gouges.

For present purposes, a definition of technology **reliability** is need. **Reliability** is the consistency of a set of measurements or measuring instruments. A technology is reliable if it yields consistent results of the same measure. It is unreliable if repeated measurements give significantly different results. However, the vendor and operator data did not include multiple inspections of the same pipeline segment or repeated pull tests of the same test piece.

The accepted standard for quantifying inspection reliability[30] is Probability of Detection (POD) which is interpreted graphically in Figure 12.

[30] Georgiou, George. Probability or Detection (PoD) Curves, Derivation, applications and limitations, Jacobi Consulting Limited for HSE 2006, Research Report 454, Crown Publishing.

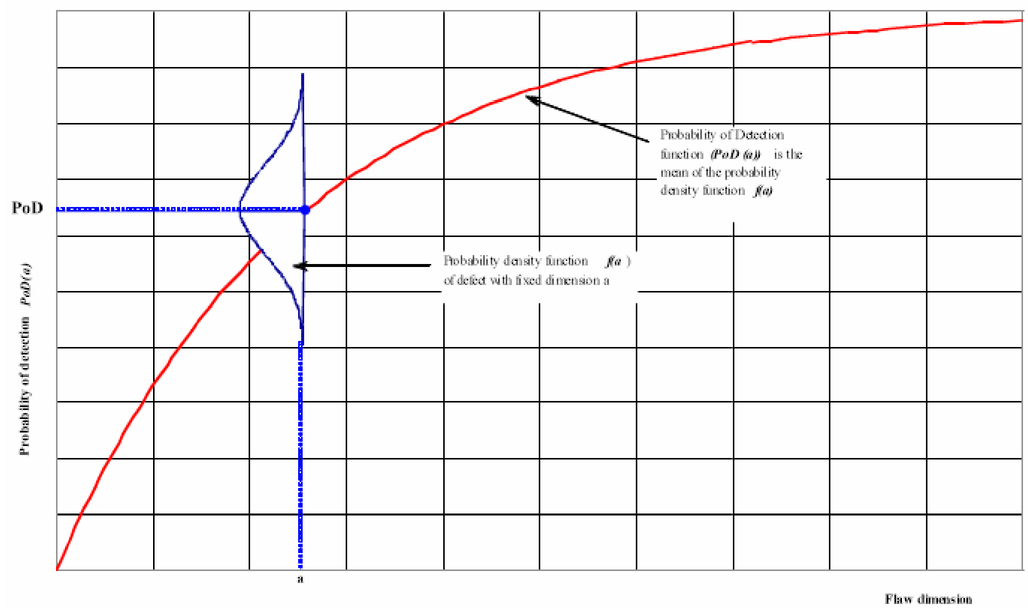


Figure 12: Schematic of the POD for flaws of fixed dimension for 'hit/miss' data[31]

The demonstrated ability to detect coincident damage within dents, such as corrosion, gouges and cracks, varied among the current technologies. All of the MFL technologies demonstrated capability to detect metal loss within dents. MFL technologies, utilizing radial and circumferential magnetic vector data, claimed capability to axially detect cracks and gouges by assessing the orientation of a magnetic field. Discrimination between corrosion and gouges was claimed for MFL technologies using single and multi-axis Hall sensors with Subject Matter Expert Analysis. Five (5) vendors provided verification data and, of these, two referenced published technical papers for their data. The papers reported their ILI mechanical damage tool predictions (dents or dents w/ metal loss or corrosion) and the results of excavation direct examinations. From this data the probabilities of detection of mechanical damage, probability of false identification of a condition and probability of correct identification were estimated, but the small data sets and differences in pipeline attributes made critical comparisons impossible.

4.5.1 Available Validation Data (DML)

Validation data from six technologies, four (J, A+B, H+I, and C) of them from the participating vendors and two (K+N, K+L) from three operators, indicated that they are capable of detecting Dents with Metal Loss. All of these technologies rely of the integrated analysis of deformation data and magnetic flux leakage signals. The six technologies and the data used to evaluate the Proportion and Probability of Detection, False Calls and Identification of DML are listed in Table 9.

[31] Georgiou, George. Probability or Detection (PoD) Curves, Derivation, applications and limitations, Jacobi Consulting Limited for HSE 2006, Research Report 454, Crown Publishing.

Technology		Data Used to Calculate Proportion and Probability of Detection, False Calls, and Identification					
		Total Investigations	DML Excavations (based on ILI report)	Correct Calls (True and Quasi)	True Calls (True DML)	False Calls	Missed Calls
Vendors	J (Combo)	61	58	52	49	6	3
	A+B	138	82	27	27	55	18
	H, I +(F,G)	26	23	20	16	3	3
	C (Combo)	34	26	25	25	1	0
Operators	K+N (Operator G1)	27	8	6	6	2	3
	K+N (Operator G2)	114	37	31	30	6	5
	K+L (Operator L2)	63	56	26	22	30	7

Table 9: Technologies and data used for the Proportion and Probability of Detection, False Calls and Identification calculations

In the table, the headings of each of the data columns are detailed as follows:

- Total Investigations
  - The total number of excavations including excavations for DML, corrosion, and others
- DML Excavations
  - The number of excavations for DML only
- DML Detected (True and Quasi True DML)
  - Dent w/ Metal Loss (True)
  - Dent w/ corrosion (True)
  - Dent w/ gouge (True)
  - Dent w/ crack (Quasi True)
  - Other features with signals similar to DML, (for example, ID restriction or mismatch – anything that results from a sensor lift-off)
- True Calls
  - The number of true calls as defined above
- False Calls
  - The number of non-existing features being reported as features
- Missed Calls
  - Plain Dent call, but Field found DML (missed)
  - ML/Corrosion call, but Field found DML (missed)
  - Nothing called, but Field found DML (missed)

4.5.2 PROPORTION OF DETECTION, FALSE CALLS AND IDENTIFICATION

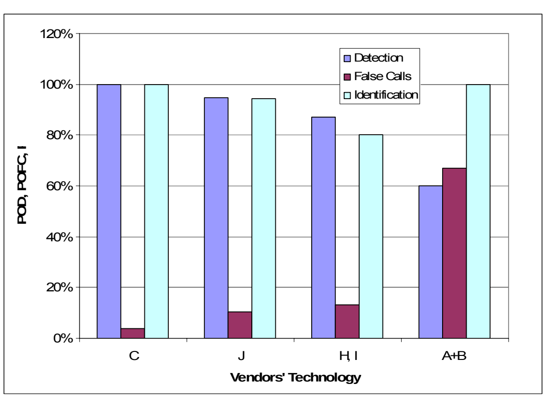
The Proportions (rates) of Detection, False Calls and Identification are given in Table 10. It is seen that Technology C has the highest Proportion of Detection (100%), the highest Proportion of Identification (100%) and the lowest Proportion of False Calls (3.8%) for DML prediction.

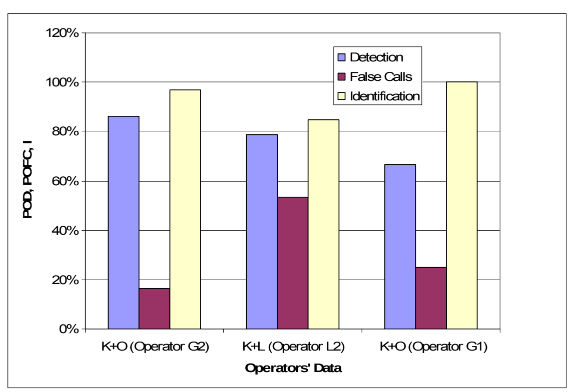
Technology J is the second with Proportion of Detection = 94.5%, Proportion of Identification = 94.2% and Proportion of False calls = 10.3%. However, the sample size of Technology J is about twice that of Technology C. Therefore, the difference in the respective Probabilities may not be significant and will be further analyzed later. The Proportion of Detection of the current technologies ranges between 60% and 100%. Figure 13 contains bar plots that compare the Proportions of Detection, False Calls and Identification of these technologies.

Data from the operators showed that Technology K+N from Gas Operator 2 have the highest Detection (86.1%), the lowest False Calls (16.2%) and relatively high Identification (96.8%) performance. The same technology from Gas Operator 1 showed a lower Proportion of Detection and higher Proportion of False Calls, along with the highest Proportion of Identification. Further, the sample size provided by Gas Operator 2 is 37 DML, about 4.5 times the size provided by Gas Operator 1 (8 DML). Therefore, the performance of Technology K+N provided by Gas Operator 2 could be statistically better than that provided by Gas Operator 1.

Technology		Proportion (x/n)		
		Detection	False Calls	Identification
Vendors	J Combo	94.5%	10.3%	94.2%
	A+B	60.0%	67.1%	100.0%
	H, I +(F,G)	87.0%	13.0%	80.0%
	C Combo	100.0%	3.8%	100.0%
Operators	K+N (Operator G1)	66.7%	25.0%	100.0%
	K+N (Operator G2)	86.1%	16.2%	96.8%
	K+L (Operator L2)	78.8%	53.6%	84.6%

Table 10: Proportions of Detection, False calls and Identification Calculated from Vendors and Operators Data.





(a) **(b)**

Figure 13: Proportion Plots (a) Vendors' Technologies and (b) Operators Data, showing Technology C has the highest Proportion of detection and Identification and the Lowest Proportion of False Calls.

4.5.3 *PROBABILITIES OF DETECTION, FALSE CALLS AND IDENTIFICATION*

Proportions of Detection, False Calls and Identification were calculated, along with estimates of the Probabilities of Detection, False Calls and Identification using both the Binomial Method and confidence intervals. The results are given in Table 11.

Technology		POD				POFC				POI			
		Proportion	Binomial, 95% Confidence for Given x,n	Clopper-Pearson Confidence Interval α=0.05		Proportion	Binomial, 95% Confidence for Given x,n	Clopper-Pearson Confidence Interval α=0.05		Proportion	Binomial, 95% Confidence for Given x,n	Clopper-Pearson Confidence Interval α=0.05	
				Lower	Upper			Lower	Upper			Lower	Upper
Vendors	J	94.5%	89.0%	86.5%	98.5%	10.3%	19.4%	4.6%	19.4%	94.2%	88.4%	85.8%	98.4%
	H,I +(F,G)	87.0%	75.1%	69.6%	96.3%	13.0%	30.4%	3.7%	30.4%	80.0%	65.6%	59.9%	92.9%
	A+B	60.0%	48.9%	46.7%	72.3%	67.1%	75.6%	57.6%	75.6%	100.0%	89.9%	89.5%	100.0%
	C	100.0%	89.1%	88.7%	100.0%	3.8%	17.0%	0.2%	17.0%	100.0%	89.1%	88.7%	100.0%
Operators	K+N (Operator G1)	66.7%	45.0%	34.5%	90.2%	25.0%	60.0%	4.6%	60.0%	100.0%	65.1%	60.7%	100.0%
	K+N (Operator G2)	86.1%	76.4%	73.0%	94.4%	16.2%	29.5%	7.3%	29.5%	96.8%	90.8%	85.6%	99.8%
	K+L (Operator L2)	78.8%	67.2%	63.8%	89.6%	53.6%	65.1%	41.8%	65.1%	84.6%	72.7%	68.2%	94.6%

Table 11: A comparison the calculated Proportions and Probabilities of Detection, False Calls and Identification

Table 11 shows the estimated POD and POI from the binomial method are comparable to the lower bound of POD and POI estimated from the confidence interval method while the estimated POFC from the binomial method are consistent with the upper bound of POFC estimated from the confidence interval method. Both values of POD and POI are lower while the POFC values are higher than the respective proportions of detection, identification and false calls. These lower values of POD and POI and higher values of POFC derived from statistical methods reflect the sample size effect. It should be noted, the conditions for POD and POI analyses here are not completely consistent with those required by the methods (binomial and confidence interval analysis) because the samples are not random. However, the results are likely to be indicative of the actual probabilities if there is no systematic bias in the way the data was collected, e.g., "obviously" erroneous measurements were not omitted.

4.6 Coincident Damage Sizing (DML)

Data for metal loss sizing coincident with a dent occurred in only two relatively small data sets. The standard deviation for depth sizing was ±6.5%WT that corresponds favorably with expected body of pipe metal loss sizing performance from MFL in-line inspections. However, sizing of features within deformation is offered on a best effort basis by the inspection vendors for current mechanical damage technologies.

Using confidence intervals, the validation sample for Technology C was found to be consistent with a performance of 80% certainty, ±15% WT depth tolerance and 95% confidence. The performance analysis of the combined data of Technologies H and I indicate a tolerance of about ±6.4% WT (or ±5.3% WT, if outliers are ignored) for the same performance specification. These differences should not be emphasized due to the small data sets used in the analysis. The metal loss data from Technologies C, H and I together with individual case study examples from Technologies E and J indicate the MFL based technologies have success in detecting metal loss less than 10% wall thickness coincident with plain dents in the range of 2% to 6% OD. However, the data set size is insufficient to fully quantify a detection performance or provide for critical comparisons. The details of the validation data and analysis are provided in Appendix B

5.0 INTEGRITY APPLICATIONS

5.1 Mechanical Damage Assessment Criteria and Models- Importance of In-line Inspection Measurements

Engineering assessment of mechanical damage for the purpose of fitness for service following an appropriate in-line inspection is an essential step to determine and prioritize repair and remediation of pipeline integrity. There are two levels of assessments in accordance with the guidance provided by US regulations and industry standards:

1) Depth based assessment
2) Strain based assessment

Depth based assessment requires knowledge of ILI tool accuracy in depth sizing and feature characterization. This knowledge is particularly important for features under "immediate" conditions because mitigative action is based on ILI reported data without excavation validation[32]. Strain-based assessment requires a high-resolution geometry profile as well as sizing accuracy because bending and membrane strain analysis requires data on dent geometry and size. Therefore, procedures for ILI performance validation, referenced in API 1163[33] and other publications[34,35], should be used to validate the accuracy of the geometry data.

The purpose of this section is to develop an understanding, through a brief but comprehensive review, (details provided in Appendix C), of

- Assessment guidance provided by US regulations and industry standards
- Parameters (depth and strains) that are used for integrity assessment of mechanical damage
- How these parameters are linked to in-line inspection

This review identifies the basic and special requirements for deformation tool performance in anomaly detection, defect discrimination, and defect sizing.

5.2 Applications: Integrity Assessment and ILI Inspection

The PHMSA/OPS integrity management regulations in North America define mechanical damage conditions requiring repair or evaluation as:

[32] Anon; "Managing System Integrity for Hazardous Liquid Pipelines", API 1160, 2001.
[33] Anon: In-Line Inspection Systems Qualification Standards", API 1163, June 2004.
[34] Desjardins, G. ,Reed, M, Nickle, R., ILI Performance Verification and Assessment Using Statistical Hypothesis Testing. IPC 2006, paper no: 10329, 2006.
[35] McCann, R., McNealy, R. and Gao M. *"In-Line Inspection Performance Verification"*. NACE Corrosion 2007, Paper 07132, Nashville, TN, March 2007.

- Dents with any indication of metal loss, cracking or stress riser (response time orientation dependent)
- Dents with depths greater than 2% depending on orientation and location
- Dents with depths greater than 6% depending on orientation and location
- Gouges greater than 12.5% of nominal wall

Effective integrity management depends on understanding the capabilities of in-line tools in order to determine and apply appropriate technologies in response to potential pipeline threats. The Phase I research provides an indication of the capabilities of current technologies to detect and discriminate limited mechanical damage conditions. The current integrity evaluation methods for mechanical damage conditions require specific data produced by in-line inspection technology. The sensitivity of integrity conclusions to inspection data is discussed in Appendix C of this report. The amount and quality of data available within Phase I indicated that lab pull testing could further improve and quantify the applicability of specific technologies to mechanical damage threats.

The mechanical damage technologies that were identified by the participating ILI vendors and analyzed in Phase I are summarized in

Table 12 together with symbols indicating the level of validation concluded within Phase I. The validation levels are:

- Validation data was sufficient to quantify detection or discrimination
- Limited validation data in the form of anecdotal case studies or examples
- Claim of commercial capability, but no validation data available

The current mechanical damage assessments are employed as caliper type deformation technologies in combination with metal loss and/or crack technologies. Implementation may reflect data combined from multiple inspection tools run within relatively close time intervals or as a single inspection run with multiple sensor technologies combined within a single tool.

The selection of an appropriate current technology for mechanical damage should consider the frequency, severity and types of conditions encountered.

● Quantitative Validation Data

◐ Qualitative Case Study Data

○ Claimed Capability- No Data

Generic Technology Key		Mechanical Damage Condition Capabilities																	
		Pipeline Deformation									Localized Changes or Coincident Damage								
		Detection- Position & Orientation	Maximum Depth	Length	Width	Shape	Nominal Wall Thickness	Global Strain	Local Elastic Strain	Local Plastic Strain	Metal Loss	Corrosion/Gouge Discrimination	Cracks	Feature Through Wall Depth	Feature Length	Feature Width	Metallurgical Changes	Orientation
A	DAMC (EM)	●	●	●	●	●		◐										
B	{Long Field}MFL[Hall-1][ID/OD EM]	●		●	●						●			●	●			●
C	{Long Field}MFL[Hall-1](ID/OD EM)(Hall-1) DAMC(EM)	●	●	●	●	●		◐		○	●	◐		●	●			●
D	DAMC	●	●	●	●	●		◐										
E	{Long Field}MFL[Hall-3][ID/OD EM]	●	●	●	●	●		◐			●	◐	◐	●	●			●
F	IEMC	●	●	●	●	●												
G	DAMC (EM)	●	●	●	●	●		◐										
H	{Long Field}MFL[Hall-1][ID/OD EM]	●		●	●						●	◐		●	●			●
I	{Circ Field}MFL[Hall-2][ID/OD EM]	●		●	●						●	◐		●	●			●
J	{Long Field}MFL[Hall-2](ID/OD EM)(DAMC)	●		●	●	●		◐			●	◐		●	●			●
K	DAMC	●	●	●	●	●		◐										○
L	{Circ Field}MFL[Hall-1](ID/OD EM)	●		●	●						○		◐	○	○			
M	{Long Bias}UT[Shear]																	○
N	{Long Field}MFL[Hall-1](ID/OD EM)	●		●	●						○				○	○		○

Table 12: Current mechanical damage capabilities and level of validation

The PHMSA/OPS Integrity Rule for liquid pipelines addresses assessment using tools capable of detecting both deformation and corrosion anomalies. However, this guidance from PHMSA/OPS recognizes the ability of MFL type tools to identify dents, but not their reliability for sizing.[36] This same guidance suggests the capability for detection by MFL could be considered in circumstances where all indications of dents are excavated, examined and repaired if required. The ability to detect smooth plain dents was claimed by all the current MFL technologies. These claims were substantiated by qualitative and quantitative validation data that was obtained as part of this project. Validation data demonstrating capability of MFL based dent detection and characterization was obtained for only Technology E, which is axial field MFL that uses three-axis sensors coupled with subject matter expert analysis. Technology E also claimed the capability of predicting dent depths. Data obtained for Technology E indicated a strong correlation between dent predictions and caliper data results. Further testing is recommended for Phase II of this research to understand the reliability of such measurements.

With the increased frequency of mechanical damage in pipelines, the accuracy of dent depth measurements has become important with respect to the identification of conditions requiring an immediate response. The accuracy of wall thickness, depth, length and shape data is important, because this data is required to perform assessment methods identified in Appendix C. Data demonstrating the detection of plain dents less than 0.5%OD was obtained for direct arm calipers augmented with electromagnetic proximity sensors. Data demonstrating the detection of plain dents less than 2.0%OD was obtained for only one MFL technology, Technology E, but further reliability testing is suggested. Sizing performance reflects a total system performance, including errors associated with in-ditch measurement and re-rounding of pipe. The data from the Phase I research provides some level of critical comparison for deformation (dent) technologies:

- **DAMC (EM) Calipers:**
  - A,C,G
  - Limit of Detection 0.5%, POD .75 to 1.0 depending on vendor and pipe size.
  - 80% Certainty Depth Tolerance: ± 0.51%OD to ± 1.10%OD
- **DAMC Calipers:**
  - Jo, Ko
  - Limit of Detection 0.5%
  - 80% Certainty Depth Tolerance: ± 1.63%Dia to ± 2.37%OD (Operator Non-current MD technology)
- **MFL[Hall-3]:**
  - E
  - Limit of Detection 2.0%
  - 80% Certainty Depth Tolerance: ± 0.78%Dia (Phase II Reliability Testing Recommended)
  - No profile capability

All of the current caliper type technologies can determine dent profiles for use in strain based assessment. The sensor density, or pipe wall coverage, affects the accuracy of calculated strains. Selection of deformation tools intended to provide data for strain based assessment should consider the sensor spacing. The current technologies, in general, provide sensor densities (when

[36] PHMSA/OPS, PRIMIS Implementing Integrity Management for Liquid Operators FAQ 6.17

coupled with appropriate interpolation algorithms), that well represent dent geometries, but sensor spacing less than 1.0 inch provide the most accurate representation for use in evaluations, such as the finite element method.

The current MFL type technologies demonstrated capability to detect metal loss with dents. Validation data was made available by vendors for individual technologies. Most of the vendors indicated that their technologies incorporate proprietary mechanical designs that minimize sensor lift-off when traversing plain dents less than 6%OD in depth. All of the vendors indicated that metal loss sizing within dents was offered on a best endeavor basis. The validation data indicated success in the detection of metal loss less than 10% wall thickness depending on the orientation and aspect ratio of metal loss. Sizing is offered within the context of current mechanical damage assessment. However, it remains the responsibility of the operator to specify it, and the responsibility of the ILI vendor to agree to reporting limits for metal loss coincident within deformations.

The detection of gouges represents an important mechanical damage condition. It is differentiated from metal loss due to corrosion, by notch-like profiles, localized cold working associated with moved metal and altered mechanical properties compared with removed metal caused by corrosion. The PHSMA/OPS liquid rule identifies gouges or notches greater than 12.5% wall thickness as a condition requiring evaluation. This requirement represents a challenge for the current mechanical damage technologies. Discrimination relies, largely, on a pseudo-deductive process recognizing the increased likelihood for gouging based on the circumferential orientation of features and coincidence with local deformations. The analysis of multiple sensor data streams, including deformation and magnetic signals, such as low field magnetic permeability measurements or multiple magnetic field vectors, provide a further qualitative discrimination for gouges. However, the discrimination of gouges at or near 12.5% wall thickness without associated deformation of the pipe wall (dent) remains difficult with the current technology depending on the secondary manual signal interpretation employed by the ILI vendor.

It is emphasized that current mechanical damage technologies involve full secondary analysis by trained subject matter experts. Therefore, the discovery timeline requirements from the regulations must be considered.

6.0 PHASE I SUMMARY AND CONCLUSIONS

6.1 Technology Summary

Each current technology consists of a process using one or more sensor technologies, integration of multiple inspection data streams, specialty data analysis and subject matter expert interpretation. Participating ILI vendors provided data for seven Caliper, MFL and UT technologies with claimed capabilities for detecting mechanical damage. Results of an analysis of this data can be briefly summarized as follows:

- **All caliper technologies defined dent depth, length, width and shape.**
  - DAMC(EM) calipers provided the best validated performance
- **All MFL technologies demonstrated ability to detect coincident damage in the form of Dents with Metal Loss.**
  - All but one technology requires caliper data stream
  - Subject Matter Expert Analysis Required
  - Validation data available for confidence interval analysis (POD, POI, POFC)
- **One technology (E) qualitatively demonstrated capability to detect cracks within dents, within limitations of orientation, but insufficient data was available to quantify performance.**

Data supplied by Vendors clearly indicates a higher performance level for ILI technologies than does the Operator data. The causes for this discrepancy are debatable, but are likely to include

1. Vendor data was gathered under more controlled conditions
   a. Current technology was used.
   b. Data represents the best effort for the technologies.
   c. Tool speed was likely to have been carefully controlled.
   d. Pipe was likely not to have significant corrosion, wax build-up, or other inner surface abnormalities.
   e. More accurate validation measurement technology, compared to routine field measurements, may have been used.
   f. Vendor data was not collected within the context of a "blind" study.
2. Operator data came from routine tool runs and did not have the benefit of special analysis by the vendor. Therefore, comparisons with current mechanical damage technology should be limited to sensing technology recognizing the following considerations:
   a. Older sensing technologies were used to obtain some of the data.
   b. Tools were used as provided by the vendor. No special data analysis for mechanical damage was done.
   c. Effects of re-rounding and rebounding appear to be present in some data.
   d. Tool speed (no data supplied) was whatever occurred during the run.
   e. Inner surface of pipes had unknown conditions.
   f. Unknown validation measurement technology was used.

None of the concerns detailed above could be addressed in this study due to absence of data. The following concerns standout as needing further investigation:

- Re-rounding
- Rebounding
- Tool speed
- Accuracy of validation measurements

The validation of current technologies for the detection and discrimination of metallurgical changes caused localized residual stress and strain were not offered by any of the vendors.

6.2 Performance Analysis and Critical Capability Comparisons

The capabilities of current technologies are evaluated in terms of their sizing and probabilities of detection, identification and false call (POD, POI and POFC) based on the available data. Two approaches are used for the evaluation: (1) binomial distribution and confidence interval method for sizing, POD, POI and POFC analyses and (2) linear regression analysis for determining correlations between ILI predictions and field measurement and standard errors in ILI sizing. This section summarizes the findings and conclusions from the data analysis, which is presented in detail in Appendix B.

Current ILI mechanical damage technologies, as indicated by validation data from Vendors, can

- Accurately predict dent depths.
- On average accurately predict dent length and width, but scatter in the measurements places a large uncertainty on the accuracy of any individual measurement.
- On average predict reasonably accurately metal loss coincident with a dent, but scatter in the measurements places a large uncertainty on the accuracy of any individual measurement.

Summaries of the results from the Phase I research relating to plain dents and dents with metal loss follow.

6.2.1 PLAIN DENTS

Sizing performance for plain dents was evaluated to compare capabilities of current technologies. Insufficient data was available for plain dents to calculate POD, POI and POFC. The current technologies, based on vendor data, accurately predict dent depths. On average, a measurement tolerance of ±0.77% OD for a certainty of 0.8 at 95% confidence level (sample size n = 438 with 360 successes). Measurement tolerances for individual technologies evaluated varied from ±0.51% to ±1.10% OD (Technology F (IMEC) excluded as discussed below) as shown in Table 13. From the table it seems that five of the analyses showed depth tolerance within ± 1.0% OD at a certainty of 0.80 and a 95% Confidence level. Among them, Technologies C, G (DMAC-EM) and E (multi-axial MFL) exhibited tolerances of ±0.51%OD, ±0.74%OD and ±78%OD, respectively. Because of the small sample size (15 data points) as compared to technology E (273 data points), the difference between Technologies C, G and E may not be significant. Combining data from Technologies A & G and A & C & G showed tolerances of ± 0.93% and ± 0.88% respectively, while Technology A showed a tolerance about 10% larger than ± 1.0% OD for an assumed certainty of 80% and confidence level of 95%. The evaluation of tool tolerance is based on an assumption that the error in validation is negligible (small) or insignificant as compared to that of the ILI prediction. Unfortunately, this may not be a valid assumption. Actual tool performance could be better assessed if validation measurements errors were determined and included in the calculations.

Technology	Tolerance for Certainty = 0.8 at 95% Confidence Level			
	Total (n)	Successes (x)	Binomial Distribution Analysis	Clopper-Pearson Certainty Interval Method
Technology A	130	109	±1.10%	±1.22%
Technology G	20	20	±0.74%	±0.74%
Technology C	15	15	±0.51%	±0.78%
Technology E	273	251	±0.78%	±0.80%
Technology A and G	150	129	±0.93%	±0.93%
Technology A, C, and G	165	144	±0.88%	±0.88%

Table 13: Dent Sizing Performance data for Vendors

Data from the operators summarized in Table 14 showed a much larger depth tolerance, ranging from ± 1.6%OD to ± 4.48%OD, than data from the vendors. The discrepancy in tolerances is most likely due to re-rounding and rebounding of dents, and the accuracy of field measurements.

Technology	Tolerance for Certainty = 0.8 at 95% Confidence Level			
	Total (n)	Successes (x)	Binomial Distribution Analysis	Clopper-Pearson Certainty Interval Method
Technology Ao	28	24	±1.60%	±1.60%
Technology Co	58	18	±4.48%	±4.65%
Technology Jo	17	9	±1.63%	±1.63%
Technology Ko	166	95	±2.37%	±2.37%

Table 14: Dent Sizing Performance data from Operator Data

Vendor and operator data were combined and re-grouped according to technologies' sensor type, detection limits, POD, and sizing performance. Results are given in the Table 15.

Sensor Type	Limits of Detection (%OD)	POD	80% Certainy Depth Tolerance @95% Confidence	Shape Reporting (Length and Depth)
DAMC (EM) Calipers	0.5	0.75 to 1.0	±0.51%OD to ±1.10%OD	Yes
DAMC Calipers	0.5	Not Determined	±1.63%OD to ±2.37%OD	Yes
IEMC Calipers	2.0	Not Determined	±3.0%OD[+]	Yes
Mfl [Hall-3]	2.0	Not Determined	±0.78%OD**	No

Table 15: Dent Performance data for Operators and Vendors

On average, the current technologies based on the vendor data accurately predict dent length and width, of plain dents but have a high standard deviation.

Above ground pull through tests suggests that Technology G- DMAC(EM) provides one of the better correlations between dent dimensions predicted by in-line tools and direct validation measurements, (80% proportion within ±0.51%OD). This result is expected given the potential for re-rounding and rebound of pipeline dents and the challenges surrounding the physical measurement of mechanical damage in field excavations. However, even under the well controlled conditions of a pull through test, data for Technology G indicated a bias in the relationship between depth prediction and validation depth, Figure 16. Data from Technology G provided the best statistical fit and depth sizing. This suggests that pull-through tests provide an effective and accurate procedure for tool performance validation because measurement errors, such as those due to re-rounding and re-bounding, are minimized. The observed errors, both systematic and random, can then be attributed to the technology itself and/or the execution process of the pull test. Pull-through tests provide a sound basis for comparison of the tool performance between technologies.

The validation data for Technology F (IMEC) indicated a depth sizing performance with a certainty of 0.8 at 95% confidence of +/- 3.0 %OD. The number of data validation data points was small and as a result this technology was removed from the critical comparisons.

Data provided by the operators showed a much larger depth tolerance, ranging from ± 1.6%OD to ±4.48%OD. Reasons for the larger tolerance are most likely associated with errors in field measurements and re-rounding and rebounding of dents. Efforts were made to correlate the operator with vendor specified current mechanical damage technologies. Care must be taken in comparisons of Vendor and Operator data to avoid conflicts between previous and current generations of mechanical damage technology.

Data showed Technology E, multi-axial MFL, has essentially a comparable sizing accuracy as Caliper type, DAMC technology. When potential errors in validation measurement are taken into account, the ILI measurements almost match the validation measurements. Because Technology E is capable of detecting metal loss and potentially cracks in dents, further characterization of this technology for mechanical damage assessment is recommended.

Dent length and width measurements were obtained for one technology. These consisted of measurements on 20 dents using Technology G (DAMC-EM). The ILI measured lengths and widths had means of 95.9% and 99.3%, respectively (as percentages of the validation lengths and widths). Unfortunately, the standard deviations were very large, 19.4% and 26.9%, respectively. Thus, on average, the measurements were accurate, but individually they may be highly inaccurate.

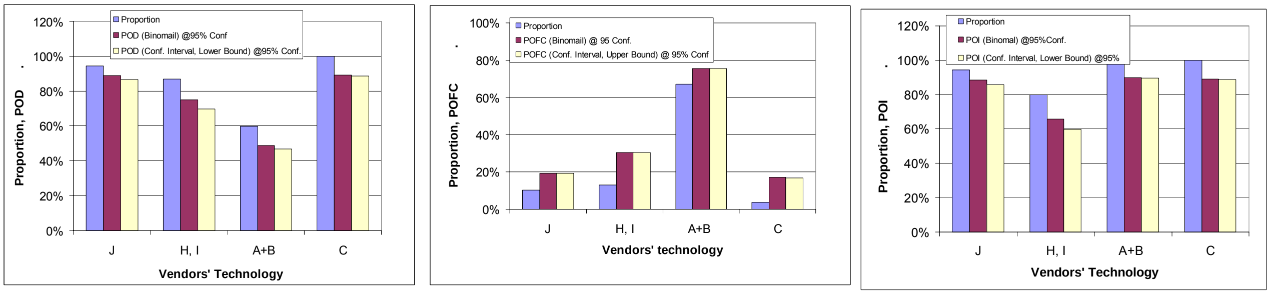
Only 43 (14%) of the 301 dents in the Operator supplied data had both ILI and validation measurements for the length of the dent. The data were lumped without regard for technology because there were too few data for any single technology to do a reasonable analysis. The resulting errors in ILI predicted lengths had a mean of 70% of the validation measurement length with a standard deviation of 33%. In short, limited conclusions can be drawn from this data.

6.2.2 Dents with Metal Loss

From limited data available in Phase I, it was observed that the MFL technologies evaluated had demonstrated capability to detect metal loss within dents. Technologies J and C, combined MFL and Caliper technologies, were capable of detecting Dents coincident with metal loss approximately 89% of time at 95% confidence level.

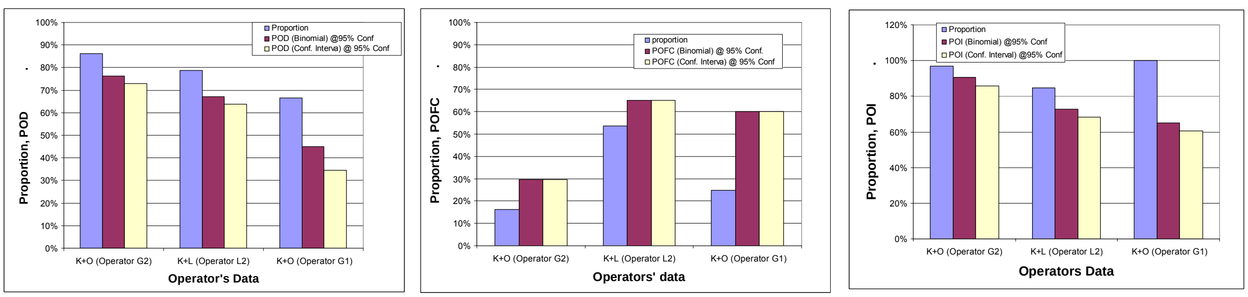
Most of the vendors employed proprietary mechanical design features to minimize sensor lift off. Using confidence intervals, the validation sample for Technology C is consistent with a performance of 80% certainty, ±15% WT depth tolerance and 95% confidence. The performance analysis of the combined data of Technologies H and I indicate a tolerance of about ±6.4% WT (or ±5.3% WT, if outliers are ignored) for the same performance specification. These differences should not be emphasized due to the small data sets used in the analysis. The metal loss data from Technologies C, H and I together with individual case study examples from Technologies E and J indicate the MFL based technologies have success in detecting metal loss less than 10% wall thickness coincident with plain dents in the range of 2% to 6% OD. However, the data is insufficient to fully quantify a detection performance.

Technology E, utilizing radial and circumferential magnetic vector data, claimed capability to detect axial cracks and gouges depending on magnetic field orientation. However there is not any quantitative data of statistical significance to draw concrete conclusions. Discrimination between corrosion and gouges was claimed for MFL technologies using single and multi-axis Hall sensors with Subject Matter Expert Analysis but insufficient data was made available to quantify discrimination.



(a) (b) (c)

Figure 14: A comparison between "Proportion" and Probability of Detection, False Calls and Identification using the binomial and Confidence Interval methods (Vendor Technology: (a) Detection, (b) False calls and (c) Identification



(a) (b) (c)

Figure 15: A comparison between "Proportion" and Probability of Detection, False Calls and Identification using the binomial and Confidence Interval methods (Operator Data) (a) Detection, (b)False Calls and (c) Identification

The findings of POD, POFC, and POI for technologies investigated are summarized as follows, based on the data presented in Table 11, Figure 14 and Figure 15: (all results are at a 95% confidence level)

- Technology H+I has a POD = 75%, but high POFC (30.4%) and low POI (65.6%)
- Technology A+B shows a lowest POD = 48.9% among the Vendor data and highest POFC (75.6%), but better POI = 89.9%
- Operator Technology K+O (G2) shows a comparable POD = 76.4% and POFC (29.5%) as Vendor Technology H+I, but better POI (90.8%).
- Other two operators' data Technologies K+L, (L2) and K+N, (G1) shows low POD (67% and 45%), high POFC (65% and 60%) and low POI (73% and 65%), respectively.
- Vendor Technologies J and C show to have POD (at least 87% and 86%, respectively. Also these technologies exhibit low POFC, at most 17% and 19%, respectively 19% and high POI at least 89% and 86%, respectively.

6.2.3 *PHASE I LIMITATIONS*

The data provided from both vendors and operators did not indicate multiple inspections of the same pipeline segments or repeated pull tests of the same test piece. Therefore, a complete determination of reliability measures, such as POD and POI for both plain dents and dents with metal loss, was not possible. An understanding of the population of mechanical damage conditions or features by a technology along with multiple opportunities to detect and measure the population is essential to provide full understanding of reliability.

The evaluation of validation data also provided insight into the potential for errors in validation measurements themselves. The direct physical measurement of depth, length and width of dents from the outside the pipes for comparison with ILI measurements is complicated by the ILI data measured from the inside of the pipe. There are analogous considerations for validation measurements of metal loss, corrosion or gouges within deformations. Complete understanding of the reliability and performance of current technologies should include an understanding of the performance of the externally applied validation measurement technologies.

In summary, a complete understanding of reliability requires a common population of mechanical damage features and conditions that can be inspected multiple times with multiple technologies. In addition the potential for errors arising from changes in dent dimensions due to re-rounding, rebounding and validation measurements must be controlled or fully understood.

The considerations outlined above helped to develop the recommendations for further research and testing in Phase II of this project.

6.3 Phase II Recommendations

Based on Phase I study results, the objectives of Phase II study are two fold:

- Increased understanding of capabilities
  - Using performance criterion and lessons learned from Phase I data
- Establish a common basis for a valid comparison between technologies in plain dent sizing and for detection and discrimination of dents with metal loss.

Presentation of Phase I results to the MD1-2 team members during a project team meeting resulted in the following priorities from liquid and gas pipeline operators for increased understanding of MD capabilities from current technologies:

- Discrimination of coincident damage is the most important
  - DML (Gouges vs Corrosion)
  - Dents with Cracks
- 6 Technologies identified with some capability for DML
  - Tech C {Long Field}MFL[Hall-1](ID/OD EM)(Hall-1)
  - Tech E {Long Field}MFL[Hall-3](ID/OD EM)

- Tech H {Long Field}MFL[Hall-1](ID/OD EM)
- Tech I {Circ Field}MFL[Hall-2](ID/OD EM)
- Tech J {Long Field}MFL[Hall-2](ID/OD EM)
- Tech L {Circ Field}MFL[Hall-1](ID/OD EM)

The following approaches are recommended to address the limitations of the Phase I data:

- Validate protocol(s) for direct assessment of dents and dents with metal loss
- Obtain significant population of MD features using the identified technologies for direct examination using validated protocols.

The approaches address the following specific limitations from the Phase I results:

- Plain Dents
  - Validation limited by errors in data
    - Re-Rounding
    - Unknown Error in Direct Exam Validation

- Plain Dents with Coincident Damage
  - Validation limited by sample size, DML only
  - Unknown error in Direct Exam Validation

These recommendations embody the following detailed scope:

Obtain a single 30 inch diameter line pipe test specimen with approximately 10 MD features (dents and dents with metal loss or gouges). This test piece will be subjected to detailed direct examination using typical protocols normally employed by Operators to measure dent dimensions and metal loss within dents. These protocols will be validated against laser based profile measurements.

Upon completion of the field direct measurement protocol Phase II proposed to gather ILI and validation data from pipeline Operators. Gas and Liquid pipeline operator Team Members volunteered to pool data from their 2008 MD digs from the 6 current technologies. Detailed validation measurements using a validated protocol are recommended to be employed by these operators.

The advantages of validating pooled operator MD data are:

- Best coverage of pipe dimensions, dent shape, size, and dent with ML/cracks with large defect population
- Results representing the actual conditions for both ILI runs and field measurement

Direct measurements of dents and dents with metal loss obtained from the pooled Operator data will be compared against the predictions from the in-line inspection technologies using the same processes employed in the Phase I research. The use of validated protocols for in-ditch examination will allow for consideration of direct examination error in the data evaluation.

The combination of in-ditch validation protocol and significant data sample size provides the best opportunity to fully understand the capabilities of the identified technologies by limiting the number of factors contributing to validation error.

6.4 Acknowledgements

The authors would like to acknowledge the continuing assistance provided by participating ILI vendors and the pipeline operators in providing their technology and validation data.

This page was left intentionally blank

Appendices

Phase I Report Investigate Fundamentals and Performance Improvements of Current In-Line Inspection Technologies for Mechanical Damage

TABLE OF CONTENTS:

List of FIGURES:

List of TABLES:

A.0 APPENDIX A- CURRENT MECHANICAL DAMAGE TECHNOLOGIES

Six (6) ILI inspection vendors agreed to participate in this research by providing technical data regarding their technologies and methods. They were (in alphabetic order):

- Baker Hughes (C-Pig)
- BJ Pipeline Services
- Enduro
- GE PII Pipeline Solutions
- Rosen
- TD Williamson (Magpie Systems)

Vendor responses to a preliminary questionnaire confirmed that they have developed and implemented various technologies to assess mechanical damage. Blade Energy personnel conducted a review of all currently applied mechanical damage technologies in order to identify all essential and important input parameters, whose criticality and sensitivity are quantitatively established in this report. The review included detailed interviews and data gathering at the ILI vendor locations to obtain a fundamental understanding of their respective technologies. Data was collected from all of the participating ILI vendors and also from four relevant pipeline operators to obtain a comprehensive view of MD threat data. Blade Energy considers this information to be confidential and makes reference to technologies by generic type and not by trade names or source.

Data provided by the participating ILI vendors and pipeline operators included:

1) Documentation of contractual performance tolerances and confidence.
2) Identification of "non-contractual" capability.
3) Document process maps for implementation of vendors' MD methods
    a. Drive applications and limits (Liquid/Gas, Differential Pressures)
    b. Pre-inspection gauging and cleaning
    c. Tracking requirements
    d. Understanding of the role and extent of SME analysis, in addition to the software based signal processing and pattern recognition (R&R factor)
    e. Client level software demonstration of information available and reported.
4) Identification of vendor based laboratory validation tests and results
5) Identification of pipeline operator results data
6) Identification of pipeline operator excavation data

All data from the participating vendors and pipeline operators are confidential and are treated in accordance with terms and conditions in agreements between the operators and PRCI.

A.1.1 TECHNOLOGY A

Category : Deformation
Technology Type: Direct Arm Measurement Caliper with Electromagnetic Sensors
Claimed Capability: Detection and measurement of geometric changes; dents and ovalities
General Description:

- Direct Read Arms, Hall Effect measurements at tips
- Calibrated with each build; Detailed calibration process
- Base Deformation Resolution Limit: 0.015" to 0.080" depending on Diameter
- Normal Specified Dent Depth Resolution: 0.125", limit of Detection 1.0% OD or .125"
- Arm Spacing approximately 1" for sizes 6 to 36 NPS

Analysis and Reporting

- Deformations include ovalities, dents and wrinkles
- Dents can be accurately sized and identified using Caliper data
    - Depth, Length, Width and Shape

A.1.2 TECHNOLOGY B

Category: Coincident Damage (used in conjunction with Technology A)
Technology Type: Longitudinal Field MFL
{Long Field}MFL[Hall-1][ID/OD EM] + DAMC (EM)
Claimed Capability: Detection and measurement of localized, moved and removed metal reported as metal loss.
General Description:

- Longitudinal orientation high field saturation magnetization
- Hall Effect, Single Axis Primary Sensor, circumferential spacing approx 0.25" to 0.3125" for sizes 8 to 36 NPS located with the primary magnetic field
- Eddy Current Sensor for ID/OD discrimination within the primary magnetic field

Analysis:

- Subject Matter Expert Data Analysis required
- MD analysis requires three independent parameters, Hall data, ID/OD data, caliper and inertial data.
- Software overlays caliper data on top of MFL data for analysis

Reporting

- Metal Loss within dents only, no discrimination for gouge, corrosion or cracks
- No standards for reporting limits for ML within dents (i.e., 0, 10%, 20%, etc)
    - Reporting limits based on agreement with customer
- Dents can be accurately sized and identified using MFL plus caliper data
    - Depth, Length, Width plus shape
- Dents with ML > 10% can be identified with reasonable accuracy, but no sizing performance can be specified
- Dents with shallow corrosion (< 5%) may have large number of false calls
- May miss metal loss in severe dents (> 6% OD)
- For small NPS (4-6") identification of dents can look like ML therefore caliper data is recommended.

Validation Data
The vendor supplied 138 direct examination observations regarding mechanical damage from a 16 inch pipeline.
Mechanical Damage predictions:

Deformations: Plain Dents and Wrinkles
Coincident Damage: Deformations with Metal Loss

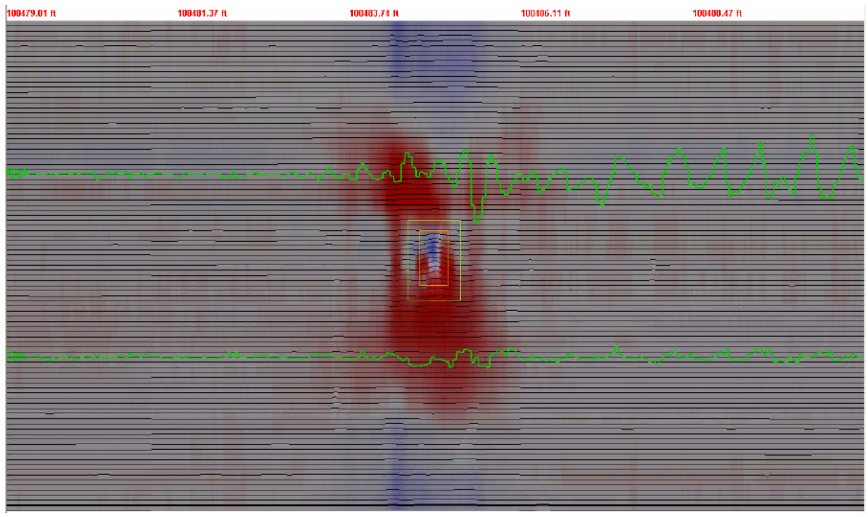


Figure 1: Typical Analysis for Technology A+B; Overlay of Primary MFL sensor, ID/OD data and gyro data for a plain dent.

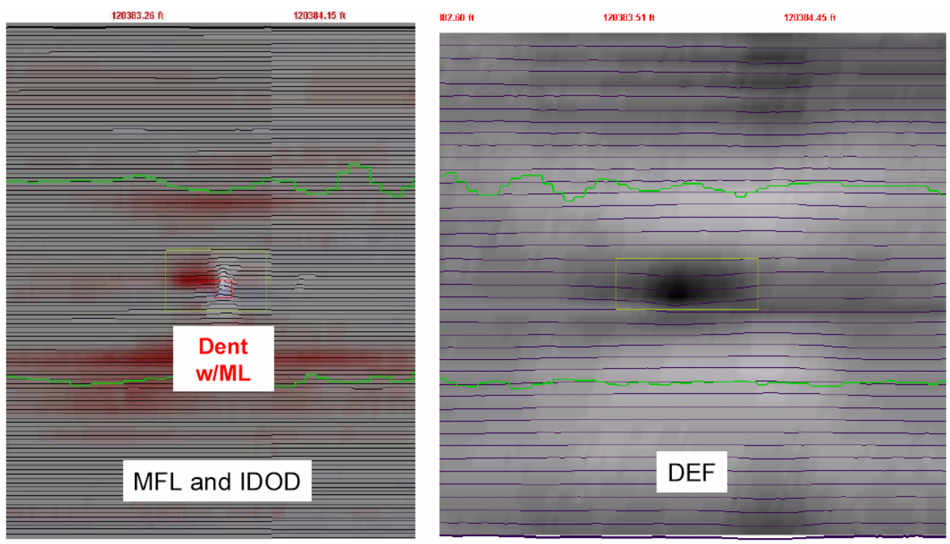


Figure 2: Typical Analysis for Technology A+B; Overlay of Primary MFL sensor, ID/OD data and gyro data for a dent with metal loss showing integrated evaluation of dent data with metal loss signal discriminates metal loss within dent.

A.1.3 TECHNOLOGY C

Category: Coincident Damage + Deformation (combination tool)
Technology Type: Longitudinal Field MFL
{Long Field}MFL[Hall-1](ID/OD EM)(Hall-1) + DAMC(EM)

Claimed Capability: Detection and measurement of deformations (ovalities, dents and wrinkles); localized moved and removed metal reported as metal loss.
General Description:

- Longitudinal orientation high field saturation magnetization
- Hall Effect, Single Axis Primary Sensor, circumferential spacing approx 0.497" to 0.142" for sizes 8 to 24 NPS located with the primary magnetic field
- Hall Effect sensors trailing outside of primary magnetic field
- Hall Sensors for ID/OD discrimination within the primary magnetic field
- Direct Measurement Arm Calipers with Hall (EM) sensors at tips of arms
    - 0.001 inch arm measurement sensitivity
    - 0.5% Limit of Detection Normal Specification with lower reporting limits upon agreement
    - Arm spacing 1.0 inch max for NPS 6 through 24

Analysis:

- Subject Matter Expert Data Analysis required
- Combination of magnetic flux leakage, deformation, inertial and internal/external data sets for classification. These are further enhanced by the addition of the residual magnetic field data set for discrimination of MD
- Software overlays caliper data on top of MFL data for analysis.

Reporting

- Metal loss within dents only, limited conditional discrimination possible for gouges, corrosion or cracks. "Potential" gouges or cracks may be identified.
- No standards for reporting limits for ML within dents (i.e., 0, 10%, 20% etc)
    - Reporting limits based on agreement with customer
- Dents can be accurately sized and identified using MFL plus Caliper data
    - Depth, Length and Width plus shape
    - Claim POD=1.0
- Dents with ML > 10% can be identified with reasonable accuracy, but no sizing performance specified
- May miss metal loss in severe dents (> 6% OD) depending on complexity of shape
- For small NPS (4-6") identification of dents can look like ML, caliper data is always incorporated in analysis.

Validation Data
The vendor supplied 14 case studies from direct examination observations regarding mechanical damage from multiple in-line inspections; tool sizes ranging from 4 to 24 NPS.
Mechanical Damage predictions:

Deformations: Plain Dents, Wrinkles

Coincident Damage: Deformations with Metal Loss, conditional discrimination for gouges

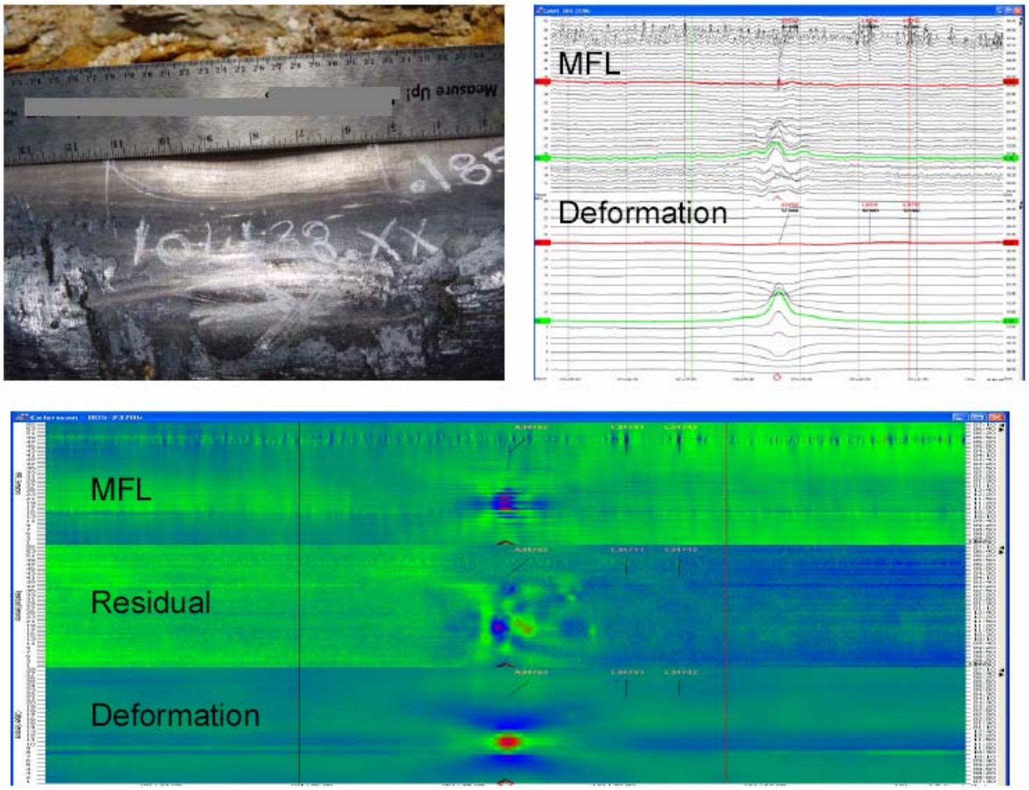


Figure 3: Technology C, 2.2 % OD depth plain dent in 8 inch pipe with no metal loss, Actual depth 2.3 % OD.

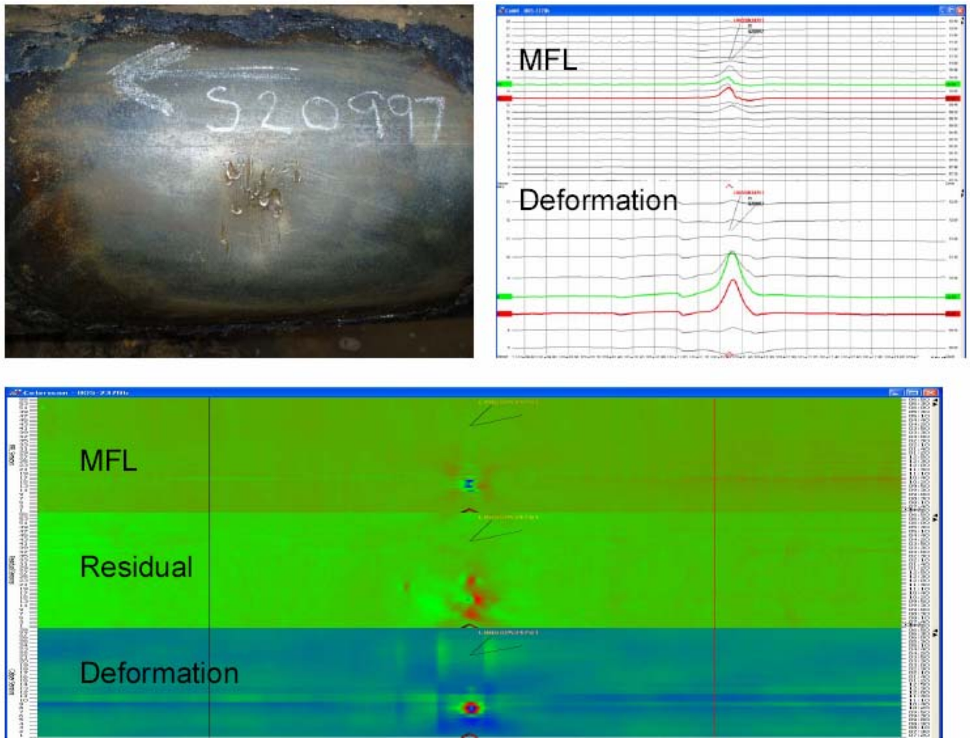


Figure 4: Technology C, 3.6% OD depth plain dent in 8 inch pipe with 20% metal loss predicted, actual depth 3.1 % OD with 12% metal loss.

Figure 5 shows an example from Technology C with multiple gouge features and the deformation profiles. Figure 6 shows the associated remnant (or low field) and saturated high field MFL data sets indicating the flux leakage patterns at each location shown in Figure 5. Table 1 lists for Technology C, the predicted metal loss depths and deformations along with results of analysis from the remnant low field data, providing a qualitative confirmation of residual stresses associated with the deformed areas.

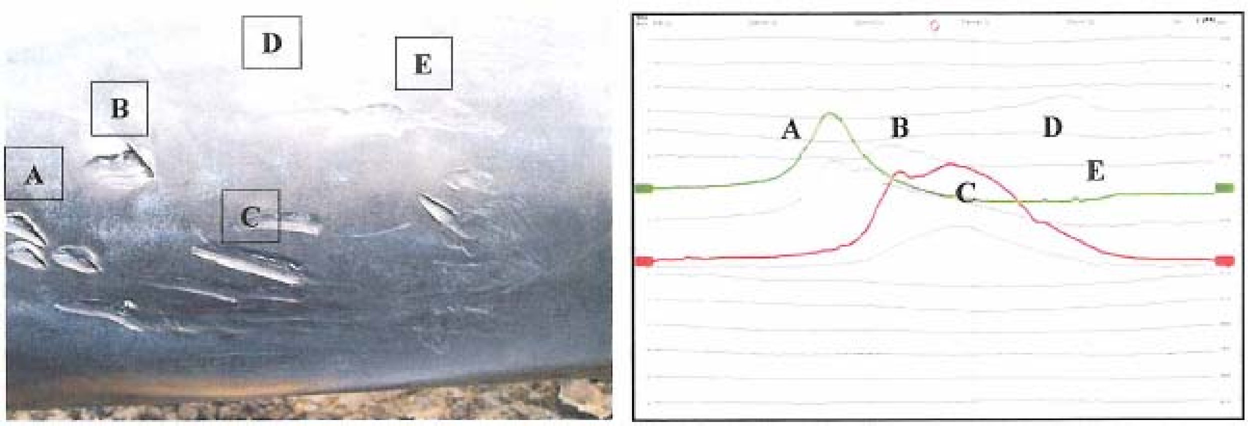
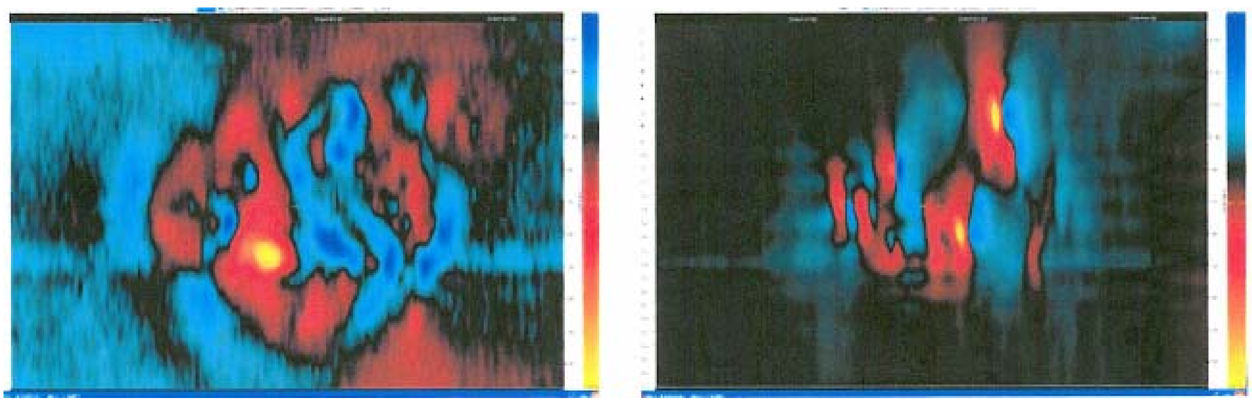


Figure 5 Technology C case study, 5 gouge features with deformation data, note no deformation detected for Feature (E)



(a) **(b)**

Figure 6 Remnant of Low Field MFL signals and (b) Saturated Field MFL signals for 5 gouge features from Technology C

Zone	Metal Loss (%)	Deformation (in)	Remnant Field signal
A	14.10%	0.119	Yes
B	20.70%	0.027	Yes
C	20.20%	0.164	Yes
D	22.00%	0.013	Yes
E	14.80%	0.000	No

Table 1: Technology C data streams for gouge area shown in Figure 5

A.1.4 TECHNOLOGY D

Category: Deformation
Technology Type: Direct Arm Measurement Caliper, DAMC
Claimed Capability: Detection and measurement of geometric changes; dents and ovalities
General Description:

- Direct Read Arms, roller tips
- Detailed calibration process
- Normal Spec Resolution, dent depth: ± 0.100”
- Arm Spacing approximately 1”-1.5” for sizes 8 to 56 NPS

Analysis and Reporting

- Deformations include ovalities, dents and wrinkles
- Global strain with inertial measurement unit
- Dents can be accurately sized and identified using caliper data
  - Depth, Length, Width plus shape

Validation Data
No validation data was made available for this technology.

A.1.5 TECHNOLOGY E

Category: Coincident Damage + Deformation (used in conjunction with Technology D)
Technology Type: Longitudinal Field MFL
{Long Field}MFL[Hall-3][ID/OD EM]

Claimed Capability: Detection and measurement of localized dents, moved and removed metal reported as metal loss, corrosion, gouges and cracks.
General Description:

- Longitudinal orientation high field saturation magnetization
- Hall Effect, 3 Axis Primary Sensor, circumferential spacing approx 0.25” to 0.40” for sizes 8 to 48 NPS located within the primary magnetic field
- Eddy current sensors for ID/OD discrimination within the primary magnetic field

Analysis:

- Subject Matter Expert Data Analysis required, manual pattern recognition using “pseudo-color data display”
- Axial vector provides confirmatory discrimination data
- Radial vector defines location and extent of physical deformation and discrimination of ML interaction within the deformation defined by the extent of the halo pattern.
- Circumferential vector important for discrimination between metal loss and plain dents through recognition of halo “lobe” pattern

Reporting

- Reports dents, corrosion within dents, gouges within dents or possible cracks within dents
- No standards for reporting limits for ML within dents (i.e., 0, 10%, 20% etc)
  - Reporting limits based on agreement with customer
- Dents can be sized from MFL data stream alone.
  - Depth, Length, Width
  - Limited shape data

Validation Data
Case study data provided for discrimination. 273 Measurements of predictions of dent depth predicted by Technology E validated against Technology D DAMC in-line inspection. 83 data points extracted for analysis. Case study examples provided for circumferential and radial vectors utilized in analysis of MD.

Mechanical Damage predictions:

- Deformations; Plain Dents, Wrinkles
- Coincident Damage: Deformations with metal loss, corrosion and gouges with conditional detection for cracking (orientation dependent)

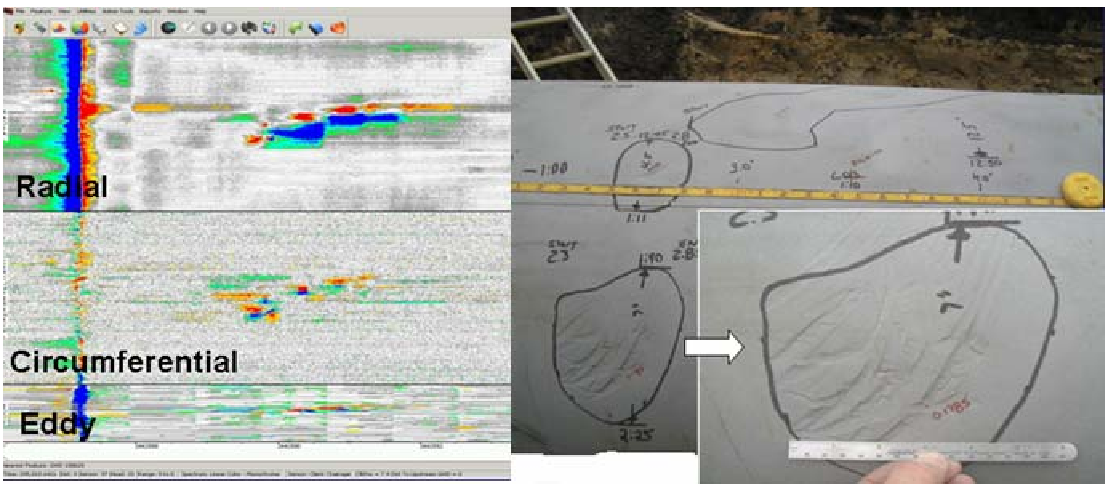


Figure 7: Technology E analysis and case study verification of gouging.

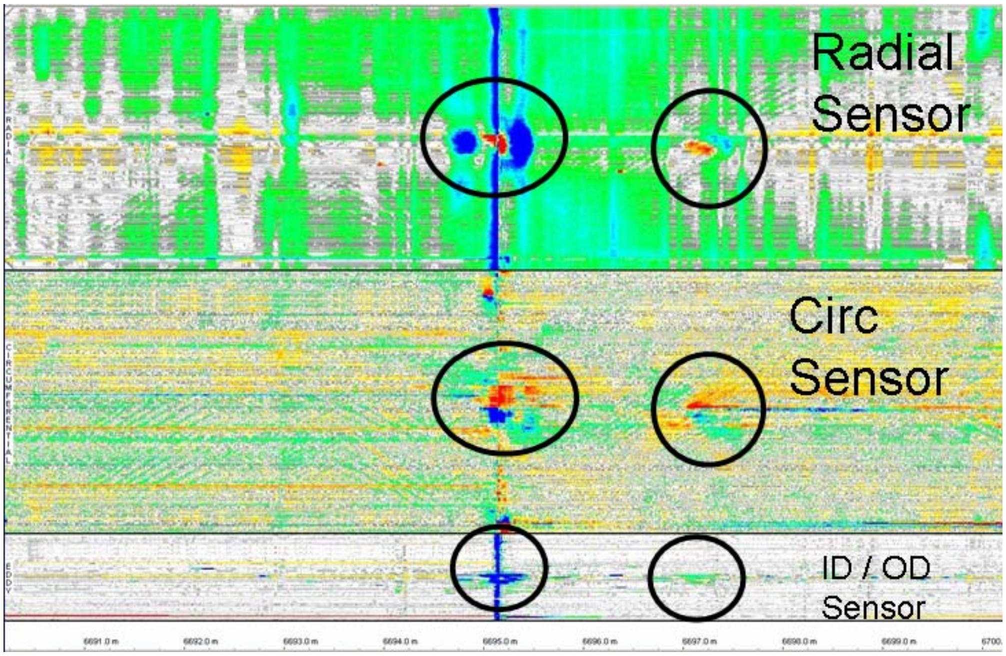


Figure 8: Technology E analysis and case study for plain dent identification.

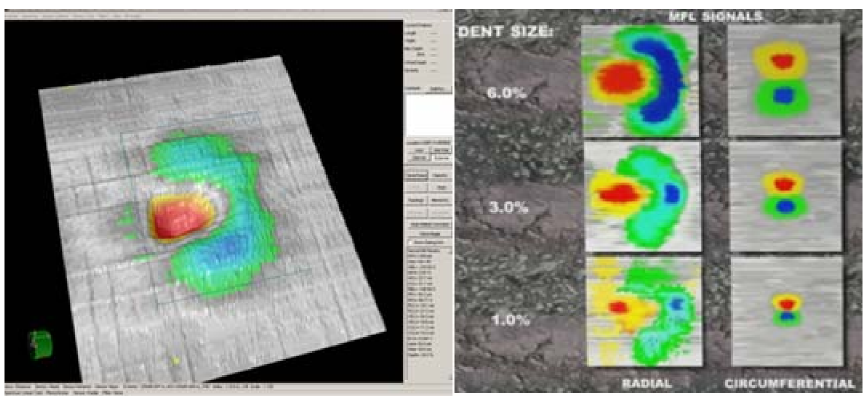


Figure 9: Technology E "pseudo color" representation of radial vector used to detect halo patterns characteristic of dents. Shape of halo pattern is key to differentiating dents from gouges. Also shown is the generalized relationship for dent sizing based on radial and circumferential vector signal analysis.

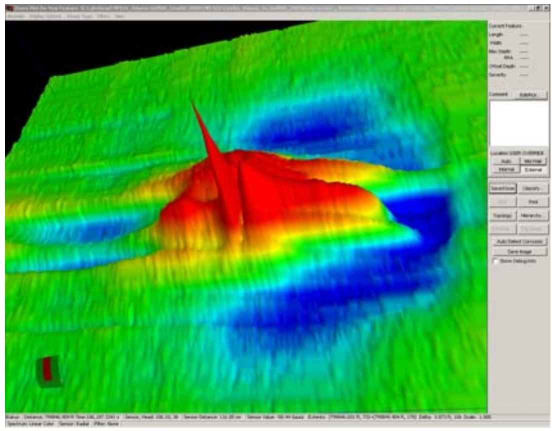


Figure 10: Technology E radial vector analysis. Halo pattern is typical for a dent with signal spike characteristic of a circumferentially oriented crack. Compare with plain dent radial representation. Case study confirmed in field as a crack.

A.1.6 Technology F

Category: Deformation
Technology Type: Indirect Electromagnetic Caliper
Claimed Capability: Detection and measurement of geometric changes; dents and ovalities
General Description:

- Electromagnetic, Eddy current ring, 8 sensors fixed mounted measure standoff from pipe wall
- Detailed calibration process
- Limit of Detection 1.5 % ID at 90% probability of detection

Analysis and Reporting:

- Deformations include ovalities, dents and wrinkles
- Global strain with inertial measurement unit
- Dents can be accurately sized and identified using caliper data
    - Depth, Length, Width plus shape

Validation:
The vendor supplied 5 case studies from direct examination observations regarding dents from multiple in-line inspections; tool sizes ranging between 8 and 26 NPS.

Mechanical Damage Predictions
Deformations: Plain Dents

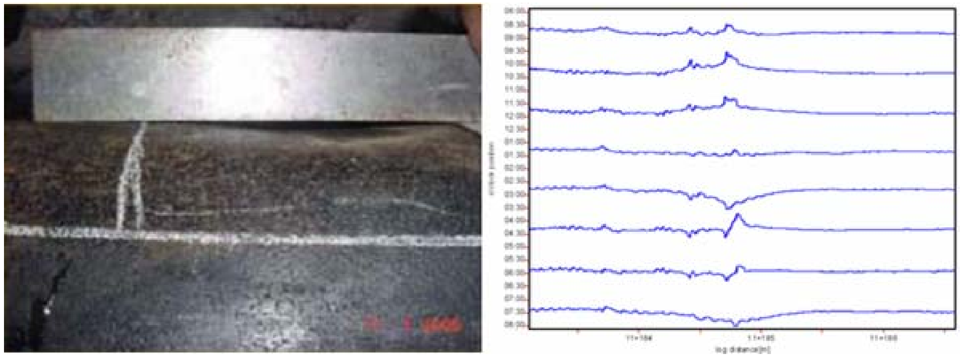


Figure 11: Technology F case study, IEMC eight channel Eddy current inspection of 8 inch pipe, shallow dent predicted < 2.0% OD, validated by direct examination at 0.3% OD.

A.1.7 TECHNOLOGY G

Category: Deformation
Technology Type: Direct Arm Measurement Caliper with Electromagnetic Sensors
Claimed Capability: Detection and measurement of geometric changes; dents and ovalities
General Description:

- Direct Read Arms, Eddy current sensors at tips
- Calibrated with each build, detailed calibration process
- Deformation Resolution Limit: 0.100" depending
- Normal Spec Dent Depth Resolution: 0.125", limit of detection 1.0% OD at 90% POD
- Full circumference pipe wall coverage; two, tandem, in line sensor rings

Analysis and Reporting

- Deformations include ovalities, dents and wrinkles
- Dents can be accurately sized and identified using caliper data
    - Depth, Length, Width plus shape

Validation

The vendor supplied 3 case studies from direct examination observations regarding dents from multiple in-line inspections; tool sizes ranging between 8 and 16 NPS. Laboratory pull test data provided for 20 plain dent features in 8 inch pipe.

Mechanical Damage Predictions
Deformations: Plain Dents

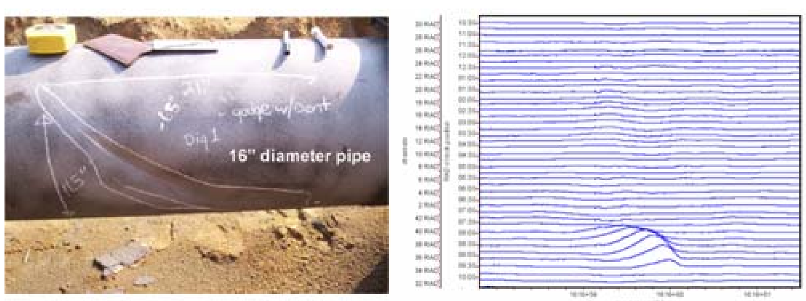


Figure 12: Technology G case study 16 inch pipe, 3.8% OD prediction, 4.0% OD validated

A.1.8 TECHNOLOGY H

Category: Coincident Damage (used in conjunction with Technology F or Technology G)
Technology Type: Longitudinal Field MFL
Claimed Capability: Detection and measurement of localized, moved and removed metal reported as metal loss or gouges.
General Description:

- Longitudinal orientation high field saturation magnetization
- Hall Effect, Two Axis Primary Sensor, axial and radial vectors circumferential, located with the primary magnetic field
- Eddy Current Sensor for ID/OD discrimination within the primary magnetic field

Analysis:

- Subject Matter Expert Data Analysis required, manual pattern recognition using "pseudo-color data display"
- MD analysis requires multiple independent parameters, Hall data, ID/OD data, caliper and inertial data.
- Software overlays caliper data on top of MFL data for analysis

Reporting

- Metal Loss within dents only, no discrimination for gouges, corrosion or cracks
- No standards for reporting limits for ML within dents (i.e., 0, 10%, 20% etc)
    - o Reporting limits based on agreement with customer
- Dents can be accurately sized and identified using MFL plus Caliper (Technology F or Technology G) data
    - o Depth, Length, Width plus shape
- Dents with ML > 10% can be identified with reasonable accuracy, but no sizing performance specified
- Discrimination of gouges is possible, subject to angular orientation of linear features with respect to magnetization field.

Validation
The vendor supplied 6 case studies from direct examination observations regarding dents from multiple in-line inspections; tool sizes ranges were 8, 16 and 26 NPS

Mechanical Damage Predictions
Deformations; Plain Dents
Coincident Damage: Deformations with metal loss, conditional discrimination for gouges

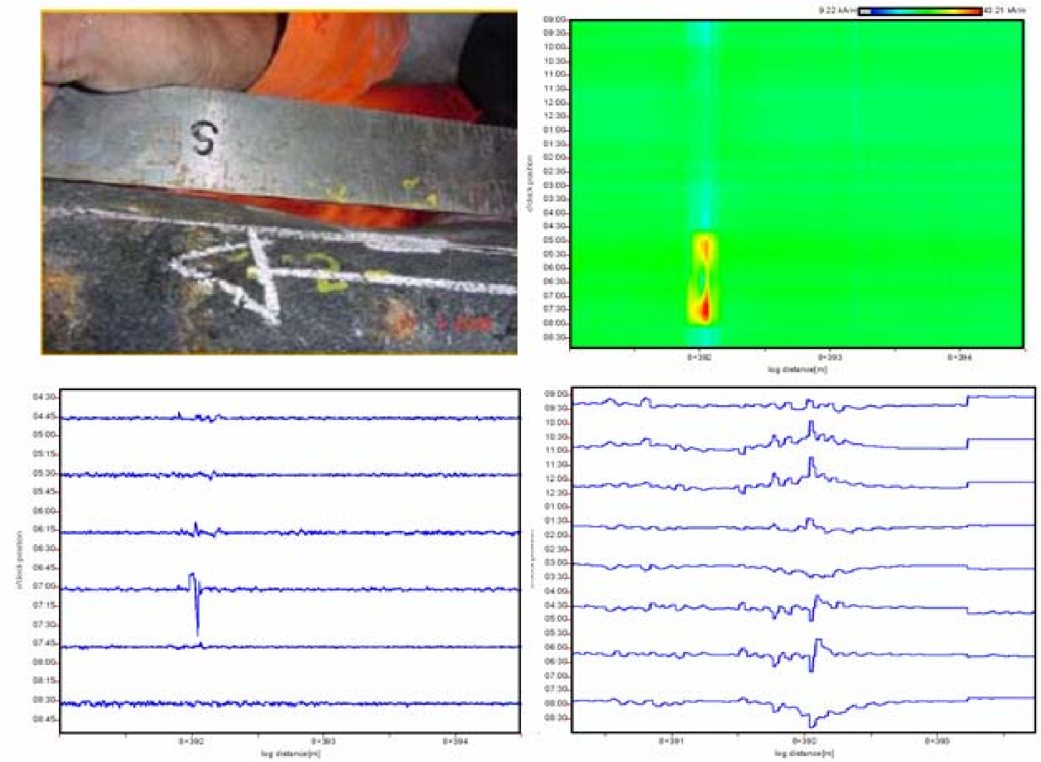


Figure 13: Technology H case study, 3.6% plain dent predicted, 4% plain dent validated. Analysis based on integrated evaluation of longitudinal field MFL plus ID/OD eddy current and deformation from Technology F.

A.1.9 TECHNOLOGY I

Category: Coincident Damage (used in conjunction with Technology F or Technology G)
Technology Type: Circumferential Field MFL
Claimed Capability: Detection and measurement of localized, moved and removed metal reported as metal loss, gouges.
General Description:

- Circumferential orientation high field saturation magnetization
- Hall Effect, Two Axis Primary Sensor, axial and radial vectors circumferential calculated, located with the primary magnetic field
- Eddy Current Sensor for ID/OD discrimination within the primary magnetic field

Analysis:

- Subject Matter Expert Data Analysis required, manual pattern recognition using “pseudo-color data display”
- MD analysis requires multiple independent parameters, Hall data, ID/OD data, caliper and inertial data.
- Software overlays caliper data on top of MFL data for analysis

Reporting

- Metal loss within dents only, no discrimination for gouges, corrosion or cracks
- No standards for reporting limits for ML within dents (i.e., 0, 10%, 20% etc)
    - o Reporting limits based on agreement with customer

- Dents can be accurately sized and identified using MFL plus Caliper (Technology F or Technology G) data
  - Depth, Length, Width plus shape
- Dents with ML > 10% can be identified with reasonable accuracy, but no sizing performance specified
- Discrimination of gouge is possible, subject to angular orientation of linear features with respect to magnetization field.

Validation

The vendor supplied 3 case studies from direct examination observations regarding dents from multiple in-line inspections; tool sizes ranging were 16 NPS. The vendor provided 26 predictions of mechanical damage with direct examination validations for Technology I.

Mechanical Damage Predictions

Deformations; Plain Dents

Coincident Damage: Deformations with Metal Loss, conditional discrimination for gouges

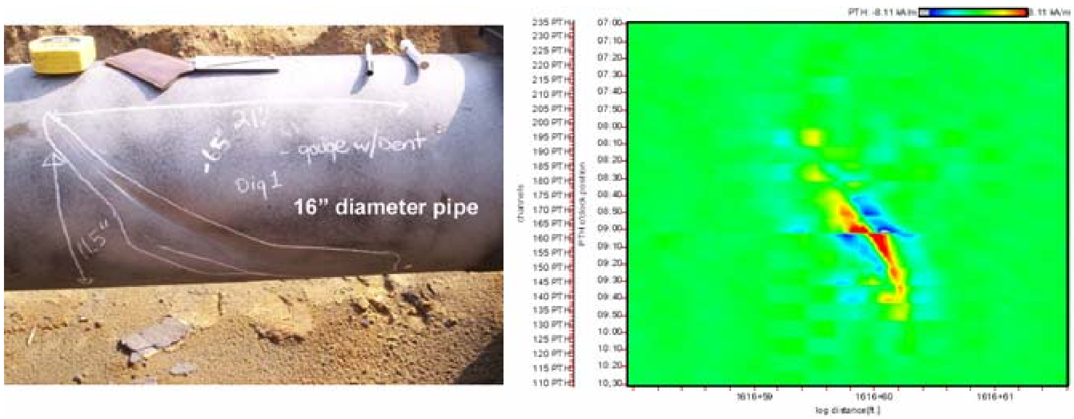


Figure 14: Technology I, dent with metal loss predicted as 3.8% OD deformation in conjunction with Technology G. Validated as dent with gouge, 4.0% OD.

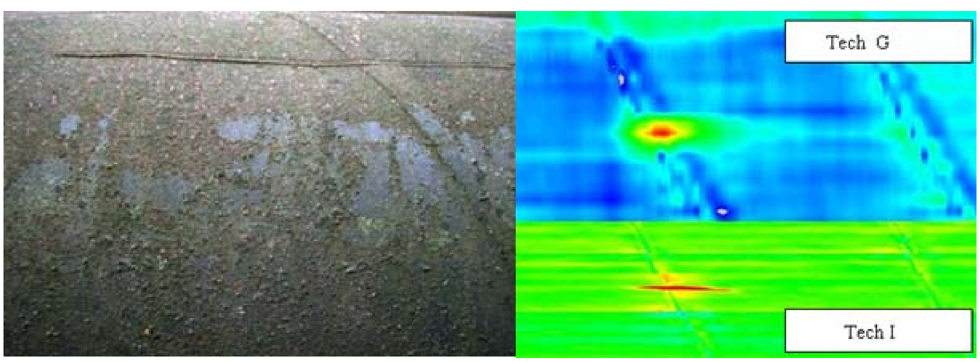


Figure 15: Technology I with Technology G, predicted as dent (0.9% OD) with metal loss, validated as 0.7% dent with gouge.

A.1.10 Technology J

Category: Coincident Damage + Deformation (combination tool)

Technology Type: Longitudinal Field MFL

Claimed Capability: Detection and measurement of deformations; ovalities, dents and wrinkles. Localized, moved and removed metal reported as metal loss.
General Description:

- Longitudinal orientation high field saturation magnetization
- Hall Effect, Single Axis Primary Sensor, circumferential spacing approx 0.250" for sizes 4 to 42 NPS located with the primary magnetic field
- Hall Effect sensors trailing outside of primary magnetic field
- Hall Sensors for ID/OD discrimination within the primary magnetic field
- Direct Measurement Arm Calipers
  - with Hall (EM) sensors at tips of arms for NPS > 16 inch
  - 0.080" inch arm measurement sensitivity
  - 0.5% Limit of Detection
  - Arm spacing 1.0 inch max for NPS 4 through 42

Analysis:

- Subject Matter Expert Data Analysis required
- Combination of magnetic flux leakage, deformation, inertial and internal/external data sets for classification, these are further enhanced by the addition of the residual magnetic field data set for discrimination of MD.
- Software overlays caliper data on top of MFL data for analysis

Reporting

- Metal loss within dents only. "Potential" gouges may be identified.
- No standards for reporting limits for ML within dents (i.e., 0, 10%, 20% etc)
  - Reporting limits based on agreement with customer
- Dents can be accurately sized and identified using MFL plus Caliper data
  - Depth, Length, Width plus shape
- Dents with ML > 10% can be identified with reasonable accuracy, but no sizing performance specified

Validation
The vendor supplied 3 case studies from direct examination observations regarding dents from multiple in-line inspections; tool sizes ranging were 16 NPS.

Mechanical Damage Predictions
Deformations; Plain Dents
Coincident Damage: Deformations with metal loss, conditional discrimination for gouges

A.1.11 Technology K

Category: Deformation
Technology Type: Direct Arm Measurement Caliper
Claimed Capability: Detection and measurement of geometric changes; dents and ovalities
General Description:

- Direct Read Arms, wheels at tips
- Dent Depth Accuracy (85% Conf) ± 1.0% Dia to ± 0.4% Dia for sizes 4 to 42 NPS
- Arm Spacing approximately 2.0 inch to 6.5 inch for sizes 4 to 42 NPS

Analysis and Reporting

- Deformations include ovalities, dents, wrinkles

- Dents can be accurately sized and identified using Caliper data
    - Depth, Length, Width plus shape

A.1.12 TECHNOLOGY L

Category: Coincident Damage (used in conjunction with Technology K)
Technology Type: Circumferential Field MFL
Claimed Capability: Detection of deformations; ovalities, dents and wrinkles. Localized moved and removed metal reported as metal loss.
General Description:

- Circumferential orientation high field saturation magnetization
- Hall Effect, Single Axis Primary Sensor, for sizes 12 to 36 NPS located within the primary magnetic field
- Hall Sensors for ID/OD discrimination trailing the primary magnetic field

Analysis:

- Subject Matter Expert Data Analysis required
- Combination of magnetic flux leakage, deformation, inertial and internal/external data sets for classification, these are further enhanced by the addition of the residual magnetic field data set for discrimination of MD
- Software overlays caliper data on top of MFL data for analysis

Reporting

- Metal loss within dents can be discriminated as corrosion or gouge.
- No standards for reporting limits for ML within dents (i.e., 0, 10%, 20% etc)
    - Reporting limits based on agreement with customer
- Dents can be accurately sized and identified using MFL plus Caliper data
    - Depth, Length, Width plus shape
- Dents with ML > 10% can be identified with reasonable accuracy but no sizing performance specified

A.1.13 TECHNOLOGY M

Category: Coincident Damage (used in conjunction with Technology K)
Technology Type: Ultrasonic (UT)
Claimed Capability: Detection and measurement of cracks within dents.
General Description:

- Longitudinal orientation feature bias due to UT sensor orientation
- 45 degree pitch/catch ultrasonic transducers for sizes 10 to 34 NPS

Analysis:

- Real time data processing algorithm
- Subject Matter Expert Data Analysis required for MD
- Combination with caliper data stream

Reporting

- Reports identification, position and orientation of plain dents but no dent sizing
- Planar (crack-like) features within dents.
- Reporting limits for crack length and depth based on agreement with customer
- Dents with cracks > 1mm can be identified.

- May miss cracks in severe dents (> 6% OD) due to sensor lift off

A case study was referenced for ultrasonic Technology M in which a reference was made to a series of loop tests. The ultrasonic technology is designed to detect axially oriented cracks, but is sensitive to lift-off and signal loss when traversing dents. A case study (flow loop test) was reported to involve 16 inch pipe with dents of un-reported dimensions and coincident cracks. The case study reported successful detection of the cracks within dents but no performance data was available.

A.1.14 TECHNOLOGY N

Category: Local Residual Stress (hard-spots) and metal loss
Technology Type: Longitudinal Field MFL
Claimed Capability: Detection and measurement of hard-spots due to thermal treatment or cold work.
General Description:

- Multiple tool technique, variation of low level remnant field technology
- One inspection vehicle with saturated magnetic field followed by separate vehicle with magnets removed

Analysis:

- Integration of data and comparison of two tool runs
- Subject Matter Expert Data Analysis required for hard-spots
- Combination of caliper data stream for MD analysis

Reporting

- Metal loss
- Hard-spots longer than 100 mm and harder than 350 HB

Technology N represents a technique to detect metallurgical changes (hard-spots) employing two consecutive in-line inspection runs with and without saturated field magnets. The data is integrated and comparisons made between the two data streams to identify hard-spots. No data or case studies were available. This technology was not identified by the vendor as a current technology for dents with metal loss but was employed by Operators G1,G2,L2

B.0 APPENDIX B - PERFORMANCE ANALYSIS

B.1 Performance Assessment- Vendor and Operator Data

A typical sizing Performance Specification for wall loss measured by ILI tools includes

1. A tolerance, e.g., within 10% of nominal wall thickness. This determines whether an ILI measurement is a "success" or "failure".
2. A certainty (p), i.e., a lower bound for the proportion of reported depths that are within the specified tolerance, e.g., 80% ($p \geq 0.8$).
3. A confidence level $(1 - \alpha)$ 100% indicating the confidence, e.g., 95% with $\alpha = 0.05$, with which the certainty is satisfied. This is a measure of the reproducibility of results by another set of measurements.

A Performance Specification analogous to that for wall loss can be stated for dent measurements by requiring a tolerance as a given percentage of OD, a certainty, and a confidence level. A confidence level of 95% ($\alpha = 0.05$) is common for engineering application and will be used in the following discussion, but should not be considered as established (see Reference 6 page 26).

It is not possible to determine the certainty of any inspection tool. The very best that can be hoped for is a reasonably small interval that contains it. This is due to the simple fact that a set of measurements, no matter how carefully or randomly obtained, has a possibility of not being completely representative of all measurements made by, or that could be made by, the tool. A facetious example may illustrate this point. Suppose you want to estimate the proportion of students at a university that weigh more than 230 pounds. In order to accomplish this you select 50 students at random and determine the proportion that weighs more than 230 pounds. It is possible, exceedingly unlikely, but still possible, that all of the 50 students are linemen on the football team. In such a case, you would conclude that all of the students weigh more than 230 pounds. This is clearly an erroneous result and illustrates that care must be taken with statistical analyses. There is a high probability that a second sample of 50 students would yield entirely different proportion. The confidence level is a measure of the reproducibility of results.

The correspondence between tolerance and confidence level is counterintuitive to some. A large confidence level means there is a high likelihood the tool meets the tolerance requirement. This may imply the tolerance requirement is not very demanding (relatively speaking), i.e., the percentage of wall thickness is relatively large. Thus, a large confidence (small α) may, but not necessarily, correspond to a relatively large tolerance. For an extreme example, the confidence level is 100% if the tolerance is ±100% of wall thickness. However, an exceptionally accurate tool will have a high confidence level if the tolerance is significantly larger than the accuracy of the tool. For example, an ultrasonic tool that measures wall thickness to within 0.001 in. will have nearly a 100% confidence level if the tolerance is ±10% of wall thickness. In short, the confidence level must be considered together with the tolerance.

The previous paragraph can be summarized by a desired tolerance or confidence level may force constraints on the remaining parameter. Changing the tolerance changes the definition of success. Increasing the tolerance tends to increase the number of successes and consequently, the likelihood that the tool satisfies its Performance Specification. A change in tolerance necessitates re-determining whether the tool satisfies its certainty and tolerance requirements.

B.1.1 REVIEW OF MATHEMATICAL METHODS

An ILI dent measurement will be called a **success** if it is within a pre-assigned tolerance, measured as % OD, of the dig measurement. Again we assume the dig measurements are much more accurate than the ILI measurement. This is debatable, but necessary for the current assessment of the data provided by the vendors and operators.

Suppose exactly x measurements from a random sample of n measurements have a specific property, e.g., a dent was detected where there was a dent, the depth of a dent was measured within a given tolerance, or no dent was detected where there was no dent. The goal is to use this sample proportion x/n of measurements with the specific property to estimate the proportion of all measurements with that property.

The probability that there are exactly x successes in n measurements is given by the binomial distribution

$$P(x,n,p) = C_x^n p^x (1-p)^{n-x}$$

where

- n = the total number of measurements,
- x = the number of “successful” measurements,
- p = the certainty, e.g., 80% of time (p = 0.8), that a measured value is within tool tolerance of the actual value,
- $C_x^n = \dfrac{n!}{(n-x)!\,x!}$

P(n, x, p) can be calculated in Excel using a worksheet function:

P(x, n, p) = BINOMDIST(x, n, p, False)

The cumulative probability that there are x or fewer successes

$$\sum_{i=0}^{x} P(i,n,p) = \sum_{i=0}^{x} C_i^n p^i (1-p)^{n-i} \qquad (1)$$

can also be also easily calculated in Excel by BINOMDIST(x, n, p, True)

The inverse problem of estimating p, given exactly x successes in a random sample of n measurements, has been well studied with estimates often described in terms of confidence intervals[34,35]. A confidence interval for a certainty p is an interval in which we expect to find the true value of the certainty. The endpoints of the interval are calculated using information from a sample of tool measurements. A confidence interval has an associated confidence level, often written as $(1 - \alpha)100\%$. For example, $\alpha = 0.05$ gives 95% confidence. Suppose we construct all possible confidence intervals using a fixed procedure and random samples of the same size from a population. The confidence is the percentage of these intervals that contain the true value of the certainty. In particular, 95% confidence means that 95% of *all* possible confidence intervals for the certainty, determined by a given procedure and random samples (of the same size) of measurements by the ILI tool, actually contain the certainty.

As mentioned before, the endpoints of a $(1 - \alpha)100\%$ confidence interval (p_L, p_U) for the certainty are determined using sample data. The uncertainty caused by a sample not being the entire population is one reason there is uncertainty as to whether the interval actually contains the certainty and why there is a "confidence" associated with a confidence interval.

One of the first confidence intervals (p_L, p_U) for a proportion (certainty in the present case) was the Clopper-Pearson confidence interval. Given x successes in n measurements, the Clopper-Pearson $(1 - \alpha)100\%$ confidence interval is obtained by choosing p_U so that

$$\sum_{i=0}^{x} C_i^n p_U^i (1 - p_U)^{n-i} = \frac{\alpha}{2} \tag{2}$$

and by choosing p_L so that

$$\sum_{i=x}^{n} C_i^n p_L^i (1 - p_L)^{n-i} = \frac{\alpha}{2} \tag{3}$$

or, equivalently, so that

$$\sum_{i=0}^{x-1} C_i^n p_L^i (1 - p_L)^{n-i} = 1 - \frac{\alpha}{2} \tag{4}$$

Eqs. (2) and (4) can be written in Excel as

$$\text{BINOMDIST}(x, n, p_U, \text{True}) = \frac{\alpha}{2} \tag{5}$$

and

$$\text{BINOMDIST}(x - 1, n, p_L, \text{True}) = 1 - \frac{\alpha}{2} \tag{6}$$

respectively.

Thus, p_L and p_U for the Clopper-Pearson confidence interval can be easily found using solver in Excel, or by iteration. However, p_U and p_L can also be calculated directly in Excel using the worksheet function for the inverse beta distribution:

$$p_U = \mathrm{BETAINV}(1 - \alpha/2, x + 1, n - x) \quad (7)$$

$$p_L = \mathrm{BETAINV}(\alpha/2, x, n - x + 1) \quad (8)$$

Paradoxically many $(1 - \alpha)100\%$ confidence intervals are not truly $(1 - \alpha)100\%$ confidence intervals. They are only approximate $(1 - \alpha)100\%$ confidence intervals. Their actually coverage (true proportion of $(1 - \alpha)100\%$ confidence intervals that contain the certainty) varies depending on values of x, n, and p. Many papers, including Refs.[1,2,3], show that it is not uncommon for $(1 - \alpha)100\%$ confidence intervals to have coverage that is less than $(1 - \alpha)100\%$, depending on x, n, and p. Confidence intervals that have at least $(1 - \alpha)100\%$ coverage, such as the Clopper-Pearson interval, tend to be conservative in the sense that their actual coverage in many cases is significantly more than $(1 - \alpha)100\%$, thereby making them much larger than necessary. Many confidence intervals that have less than nominal coverage tend to have lengths that are shorter than confidence intervals with nominal coverage and coverage that, on average, is at least nominal. Thus, a choice of confidence intervals is partially dependent on whether guaranteed nominal coverage is required. Here the choice is for at least nominal coverage. In this report eqs. (7) and (8) are used to determine endpoints of all confidence interval for proportions.

The Clopper-Pearson confidence interval sometimes bears the title "exact" because it is based directly on the binomial distribution rather than any approximation to the binomial distribution. This interval is also described as conservative because of the discrete nature of the binomial distribution. [4]

A method, referred to here as the **Binomial Distributional Analysis**, similar to that used to calculate a Clopper-Pearson confidence interval also appears in the pipeline literature.[5] Values determined by both Binomial Distributional Analysis and confidence intervals appear in the discussion in the body of this report for comparative purposes.

The commonly used "textbook" $(1 - \alpha)100\%$ confidence interval for a population proportion p approximates a binomial distribution with a normal distribution to obtain the following endpoints:

---

[1] Brown, L. D., Cai, T. T., DasGupta, A., 2001. Interval Estimation for a binomial proportion (with discussion), Statist. Sci. 16, 101-133.
[2] Brown, L. D., Cai, T. T., DasGupta, A., 2002. Confidence Intervals for a binomial proportion and asymptotic expansions. Ann. Statist. 30, 160-201.
[3] Piegorsh, W. W., 2004. Sample sizes for improved binomial confidence intervals, Comp. Statist. & Data Anal, 46, 309-316.
[4] Bousma, A. "Confidence Intervals for a Binomial Proportion", Department of Statistics & Measurement Theory, December 16, 2005.
[5] G. Desjardins: "Assessment of ILI Tool Performance", Corrosion 2005, paper# 05164, NACE, 2005

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{9}$$

where $\hat{p}$ = x/n and $z_{\alpha/2}$ is the (1 – α/2)100-th percentile of the standard normal distribution. Values of $z_{\alpha/2}$ are found in most statistics textbooks. For example, $z_{0.025}$ = 1.96 for a 95% (α = 0.05) confidence interval. It appears that eq. (9) was used to obtain the entries in Table 9, page 37, of API 1163. It is sometimes overlooked that Eq. (9) requires that n is "sufficiently large". "Sufficiently large" is often interpreted as $n\hat{p} \geq 5$ and $n(1 - \hat{p}) \geq 5$, but other conditions leading to larger values of n have also been used. Eqs. (7) and (8) have no assumptions on the size of n.

B.1.2 Performance Assessment- Vendor Dent Data

Table 2 gives the number and proportion of ILI measurements that are within 1% OD of corresponding dig measurements. Table 3 gives 95% confidence intervals, as determined by Eq's. (7) & (8), for the certainties of the ILI tools that are implied by the numbers of ILI measurements within 1% OD of corresponding dig measurements. All of technologies, except for Technology F exceeded a 0.89 certainty with 95% confidence level. This is corroborated by the small Standard Deviations on the errors.

	Number of Successes	Total Number	Proportion
Technology A	109	130	0.84
Technology G	20	20	1.00
Technology C	15	15	1.00
Technology E	251	273	0.92
Technology F	4	7	0.57
Technology A,G	129	150	0.86
Technology A,C,G	144	165	0.87

Table 2: Number and proportion of successes, Tolerance = ± 1% OD

	Bounds on Certainty (α=0.05))	
	Lower	Upper
Technology A	0.78	0.88
Technology G	0.86	1.00
Technology C	0.82	1.00
Technology E	0.89	0.92
Technology F	0.23	0.87
Technology A,G	0.81	0.90
Technology A,C,G	0.82	0.91

Table 3: 95% Confidence interval for certainty, Tolerance = ± 1% OD

Let n_1, m_1, p_1 and n_2, m_2, p_2 denote the number of data, number of successes and proportion of successes in Table 2 for Technologies A and G, respectively. A standard hypothesis test found in many statistical textbooks allows us to conclude that (certainty for technology A) > (certainty for Technology G) at a 95% confidence level if

$$z = \frac{p_2 - p_1}{\sqrt{p(1-p) * \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} > z_{0.05} = 1.645$$

where $p = (m_1 + m_2)/(n_1 + n_2)$ and $z_{0.05}$ is the critical number that cuts a right tail with area 0.05 from a standard normal distribution. A straightforward calculation shows that z = 1.56. Consequently, the proportion of successes from Technology A cannot be distinguished from that of Technology G at a 95% confidence level. Similar calculations show that the proportions, except for that from Technology F, cannot be distinguished at a 95% confidence level.

Table 4 gives tolerances for the various technologies to be Not Inconsistent and Consistent with a certainty of 0.8 at a confidence level of 95%. "Not Inconsistent" means there is not enough evidence to make any decision about the certainty (p) of the tool. For example, the numbers in Table 4 for Technology A imply that, at a 95% confidence level,

1. Tool performance is not consistent with $0.8 \leq p$ if Tolerance < 0.59% OD.
2. There is insufficient evidence to make any decision about tool certainty if Tolerance is between 0.59% and 1.22% OD.
3. Tool performance is consistent with $0.8 \leq p$ if 1.22% OD ≤ Tolerance.

All the results are basically indistinguishable with the exception of Technology F. Any differences are of the magnitude expected from the variations in the data.

A detailed discussion of calculations leading to x_1 and x_2 can be found in a NACE Corrosion 2007 paper[6].

[6] R. McCann, R. McNealy, and M. Gao, In-Line Inspection Performance Verification, NACE Corrosion 2007 Conference and Expo, Paper 07132.

	Tolerance Levels for 0.8 ≤ p with 95% Confidence Level			
	Successes		Tool Performance	Tolerance (±% OD)
Technology A	x1	95	Not Inconsistent	0.59
	x2	113	Consistent	1.22
Technology G	x1	12	Not Inconsistent	0.26
	x2	19	Consistent	0.74
Technology C	x1	9	Not Inconsistent	0.16
	x2	15	Consistent	0.78
Technology E	x1	205	Not Inconsistent	0.64
	x2	231	Consistent	0.80
Technology F	x1	3	Not Inconsistent	0.70
	x2	7	Consistent	3.00
Technology A,G	x1	110	Not Inconsistent	0.57
	x2	129	Consistent	0.93
Technology A,C,G	x1	122	Not Inconsistent	0.54
	x2	142	Consistent	0.93

Table 4: Tolerance levels to be Not Inconsistent and Consistent with 0.8 ≤ p at 95% confidence level

B.1.3 ASSESSMENT OF ILI TOOL PERFORMANCE FROM OPERATOR DATA

The statistical methods that were used in Section B.1.2 to assess ILI Tool Performance for vendor data were also used to assess supplemental ILI Tool Performance data supplied by Liquid and Gas pipeline operators. Every effort was made to match operator data to the definitions of current mechanical damage technologies defined by the vendors with the data identified by the subscript "o". It must be noted that for operator data, any assessment of mechanical damage for coincident damage, such as dents with metal loss, were not made using the current process of data stream analysis identified by the vendors consistent with current technology.

Table 5 gives the number and proportion of successes (measurements within a given tolerance) when the tolerance is ±1% OD. Table 6 gives the 95% confidence interval for the proportion. For example, the certainty (p) of the ILI measurements provided by Technology J lies in the interval (0.30, 0.75) with confidence level 95%.

	Number of Successes	Total Number	Proportion
Technology A_0	24	28	0.86
Technology C_0	18	58	0.31
Technology J_0	9	17	0.53
Technology K_0	95	166	0.57

Table 5: Number and proportion of successes, Tolerance = ± 1% OD

	Bounds on Certainty (α = 0.5)	
	Lower	Upper
Technology A_0	0.70	0.95
Technology C_0	0.21	0.42
Technology J_0	0.31	0.74
Technology K_0	0.51	0.64

Table 6: 95% Confidence interval for certainty, Tolerance = ± 1% OD

Table 7 shows the number of successful measurements and associated tolerances for the performance of the ILI tool to be "Not Inconsistent" and "Consistent" with the certainty being greater than or equal to 0.8. "Not Inconsistent" means there is not enough evidence to make any decision about the certainty (p) of the tool. For example, the numbers in Table 7 for Technology J_0 imply that, at a 95% confidence level,

1. Tool performance is not consistent with $0.8 \le p$ if Tolerance < 1.06% OD.
2. There is insufficient evidence to make any decision about tool certainty if Tolerance is between 1.06% and 1.63% OD.
3. Tool performance is consistent with $0.8 \le p$ if 1.63% OD ≤ Tolerance.

	Tolerance Levels for 0.8 ≤ p with 95% Confidence Level			
	Successes		Tool Performance	Tolerance (±% OD)
Technology A_0	x1	18	Not Inconsistent	0.35
	x2	26	Consistent	1.60
Technology C_0	x1	40	Not Inconsistent	2.61
	x2	52	Consistent	4.65
Technology J_0	x1	10	Not Inconsistent	1.06
	x2	16	Consistent	1.63
Technology K_0	x1	122	Not Inconsistent	1.54
	x2	143	Consistent	2.37

Table 7: Tolerance levels to be Not Inconsistent and Consistent with 0.8 ≤ p at 95% confidence level

B.2 Sizing Accuracy- Linear Regression

A primary problem in assessing dig-ILI measurement data from both operators and vendors is the accuracy of the measurements to which the ILI measurements are compared. Presumably the measurements by vendors were made in more controlled conditions and, consequently, should be more accurate. The expectation was that data from vendors would have errors that were closer to zero on average with less spread (smaller standard deviation) than data from operators. This proved to be true, if we assume dig measurements have insignificant errors relative to ILI measurements. This assumption is debatable, but significant analysis cannot be done on the operator data without it. Quantitative knowledge of errors associated with dig measurements is essential for a thorough assessment of ILI tool performance. Section B.2.9 demonstrates one method for incorporating a known distribution of dig measurement errors into the assessment of

ILI measurements. Even without this knowledge, it was possible to glean information about ILI tool performance from data supplied by vendors and operators. This information comes from two approaches based on deviations from the ideal situation in which dig and ILI measurements are identical and functionally related by y = x:

1. Accuracy: What percentage of ILI measurements are within a given tolerance of the dig measurements? Accuracy was assessed by analyzing measurement errors directly.
2. Measurement trends: Is data consistent with the ideal relationship of y = x? Measurement trends were assessed with regression analyses.

These two approaches provide both complementary and supplementary information about the measurement of dents with ILI tools.

B.2.1 REVIEW OF LINEAR REGRESSION

Linear regression is the standard method used to obtain a linear function that "best-fits" the data. A regression equation relating X and Y can be obtained easily with a few clicks of a mouse using many software packages. This should be the first step of the regression analysis, but often is where the analysis ends. There are assumptions regarding the calculation of the regression coefficients that should be verified. A linear regression model has the form

$$Y = AX + B + \varepsilon$$

where ε is a random error. Linear regression analysis, using the method of least squares, finds estimators a and b for A and B so that

$$y = ax + b$$

is the line that "best-fits" the data. Four of the basic assumptions in linear regression analysis are:

- X values have no error (or insignificant error compared to errors in Y values).
- ε has zero mean.
- ε has constant standard deviation (rarely known), i.e., ε does not depend on X.
- ε is normally distributed.

It is impossible to verify whether ε has these properties, because ε is unknown. In practice, the distribution of ε is replaced by the distribution of residuals, $y_i - (a + bx_i)$, which is checked for the desired properties. How to check for these properties is described below. The important thing at the moment is that residuals, representing the error term ε, should be (approximately) normally distributed with zero mean and constant standard deviation. Notice that we made no assumptions about the distribution of errors while analyzing errors, but we make very specific assumptions about residuals (which are a counterpart of errors) in regression analysis. **A regression model should not be accepted if these assumptions are not approximately satisfied.** In particular, the use of regression equations to calibrate tools, or otherwise adjust data to account for biases in the tools, should be avoided, if possible, whenever the above conditions are not satisfied. A significant portion of the rest of this section investigates whether data provided by the operators and vendors is sufficient for the assumptions to be satisfied.

There are two standard ways to measure how well a regression fits the data:

- Coefficient of Determination (R^2): About $100(R^2)\%$ of the total variation in the sample y (ILI) values about their mean can be explained by the linear regression model. Roughly speaking, the better the regression line fits the data, the closer R^2 is to 1.
- Standard Error (s): The standard error is a measure of the amount of error in the prediction of y for an individual x. If the properties of an ideal regression analysis are present (See items 1- 3 below.), then about 95% of the observed y (ILI) values are within 2s of their respective values predicted by the regression equation.

Detailed discussions the Coefficient of Determination and Standard Error are found in most statistics text books.

A (regression) residual is the difference between an ILI measurement and the corresponding value predicted by the regression equation. That is,

$$\text{Residual} = d_{ILI} - (a(d_{dig}) + b)$$

where d_{ILI} denotes an ILI measurement, d_{dig} denotes the corresponding dig measurement, and $y = ax + b$ is the linear regression equation for ILI-dig data.

Similar to random error, ε, an ideal regression analysis will have

1. Mean value of residuals is 0. (This always occurs for a properly performed regression analysis)
2. Residual variance is independent of Field Measurements, i.e., residuals are uniformly distributed about the Mean, with no dependence on the Regression Approximation.
3. Residuals are normally distributed. (This implies residuals are symmetrically distributed.)

Significant variation from these conditions brings into question ofany conclusions based on the regression analysis. For this reason the latter two conditions will be checked as part of every regression analysis.

Distributions of measurement errors and residuals were assessed using three statistics: Mean, Standard Deviation and Skewness, where

- **Mean** is the average value of the residuals, which is always 0 for a properly performed linear regression.
- **Standard Deviation** is a measure of the "spread" of the distribution of the data. The larger the Standard Deviation, the more widely spread is the data.
- **Skewness** is a measure of the symmetry of the distribution of the data
    - Skewness > 0 indicates the right tail of the distribution is larger than the left tail.

- Skewness < 0 indicates the left tail of the distribution is larger than the right tail.
- Skewness = 0 indicates a symmetric distribution. In particular, Skewness = 0 for a normal distribution.
- The larger the magnitude of the skewness, the more pronounced is the lack of symmetry of the distribution of the data.

Two graphical methods were used to determine whether it is reasonable to assume a distribution is normal:

- Normal Probability plot- The normal probability plot is a graph, which if approximately linear, indicates it is reasonable to assume the data is normally distributed.
- Tukey Lambda Probability Plot Correlation Coefficient (PPCC) plot- A PPCC plot is based on a one-parameter (shape parameter) family of distributions. A PPCC plot is a graph of the shape parameter λ versus a correlation coefficient that measures how well the corresponding distribution fits the data. The larger the correlation coefficient (necessarily between -1 and 1), the better the distribution fits the data. The Tukey Lambda PPCC plot indicates whether it is reasonable to assume a symmetric distribution can be modeled by several common symmetric distributions. Specifically, if the maximum PPCC occurs at

    1. $\lambda = -1$, the distribution is approximately a Cauchy distribution.
    2. $\lambda = 0$, the distribution is a logistic distribution.
    3. $\lambda = 0.14$, the distribution is approximately a normal distribution.
    4. $\lambda = 0.5$, the distribution is U-shaped.
    5. $\lambda = 1$, the distribution is a uniform distribution.

    If the maximum PPCC occurs at approximately $\lambda = 0.14$, it is reasonable to conclude that a normal distribution is a good model for the data.

The analysis started with vendor data to show what is the "best" that can be expected from the current mechanical damage technologies. Operator data was then examined to augment technologies for which vendor data was insufficient or not available. It is recognized that operator data may not necessarily represent the complete, or best, mechanical damage assessment technology available from vendors.

B.2.2 REGRESSION ANALYSIS OF VENDOR DENT DATA

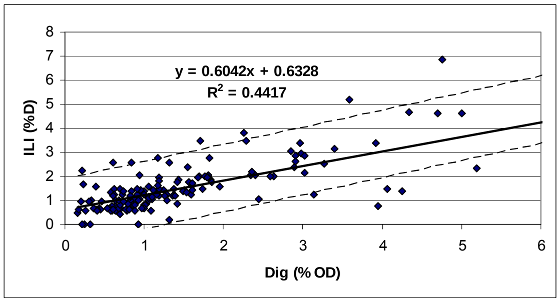
Dent depth sizing data provided by the participating vendors was analyzed as individual technologies by grouping data according to the technology used to obtain it. Table 8 gives the technology types, how technologies were grouped, and group names.

Technology Designation	Technology Type
A	DAMC(EM)
G	DAMC(EM)
C	DAMC(EM)
E	{Long Field}MFL[Hall-3][ID/OD EM]
F	IEMC
A and G	DAMC(EM)
A and C and G	DAMC(EM)

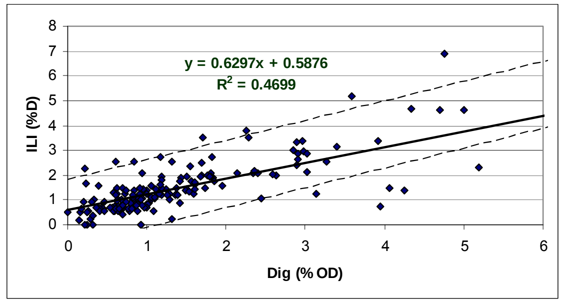
Table 8: Key for technology names for vendor data

No validation data was provided by the vendors for deformations using DAMC technologies. Calibration data was provided by multiple vendors who utilize mechanical contact caliper arms within their DAMC technologies. Such calibration data are useful in terms of expected sensitivity and limits of detection, but do not provide validation for overall systemic performance. As a result, data was requested from operators for the technologies lacking data but was not made available within Phase I .

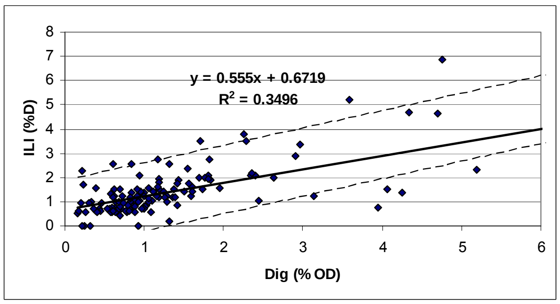
Figure 16 shows ILI and dig measurements supplied by the vendors for their technologies along with linear regression lines and lines indicating ± 2(Standard Deviations) from the regression line.



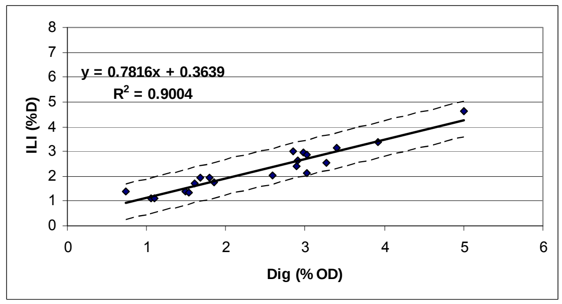
(1) Technology A,G –DAMC(EM)



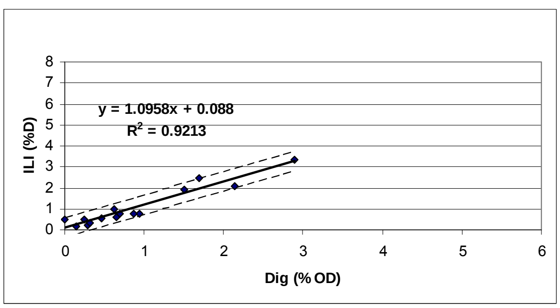
(2) Technology A,C,G- DAMC (EM)



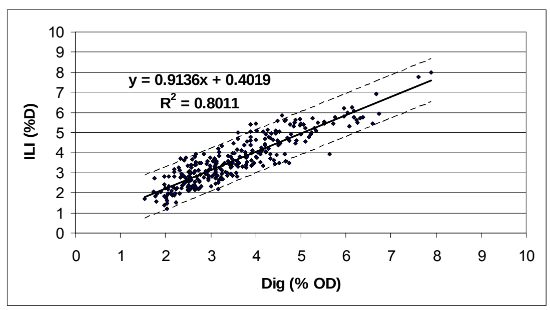
(3) Technology A- DAMC(EM)



(4) Technology G- DAMC(EM)



(5) Technology C- DAMC(EM)



(6) Technology E- {LongField}MFL[Hall-3][ID/OD EM]

Figure 16: Dig and ILI measurements of vendor dent data (dashed lines indicate regression line ±2(Standard Error)

Table 9 gives data counts along with the Mean, Standard Deviation, and Skewness of the residuals for the data associated with each technology data grouping. Technologies A and G, A and C and G have Standard Deviations and Skewness of roughly the same size. However, the Standard Deviations of Technologies G and C are statistically smaller than for the other technologies, as is shown in the next paragraph.

The greatest difference in Standard Deviations occurs between those of data from Technologies C and F. Let s_1 and s_2 denote the standard deviations given in Table 9 for data from Technologies C and F, respectively. Using a standard hypothesis test that can be found in many elementary statistics textbooks, we conclude that the Standard Deviation of data from Technology C is greater than that from Technology F at the 95% confidence level if $s_2^2/s_1^2 > F_{0.05}$, where $F_{0.05}$ is the critical number that cuts a tail with area 0.05 from an F distribution with $\nu_1 = 14 - 1$ and $\nu_2 = 5 - 1$ (14 and 5 are the number of data from Technologies C and F, respectively) degrees of freedom. A simple calculation shows $s_2^2/s_1^2 = 20.72$. According to a standard table, $F_{0.05}$ corresponding to the given degrees of freedom is less than 2.0. Consequently $s_2^2/s_1^2 > F_{0.05}$ and we conclude that the Standard Deviation of data from Technology F is greater than that from Technology C at a 95% confidence level. Analogous calculations show that:

- The Standard Deviation of data from Technology F is greater than those of the other Evaluations at a 95% confidence level.
- The Standard Deviations of data from Technologies G and C are smaller than those of the other Evaluations at a 95% confidence level.
- The Standard Deviations of data from Technologies G and C are not different at a 95% confidence level.
- The Standard Deviations of data from Technologies A and G, A and C and G, A only, and C only are not different at a 95% confidence level.

These facts should be treated with caution due to the relatively small data sets associated with these technologies.

	Count	Mean	Standard Deviation	Skewness
Technology A	130	0.000	0.831	0.441
Technology G	20	0.000	0.282	-0.223
Technology C	15	0.000	0.263	0.575
Technology E	273	0.000	0.543	0.205
Technology F	7	0.000	1.404	-0.475
Technology A,G	150	0.000	0.787	0.029
Technology A,C,G	165	0.000	0.763	-0.103

Table 9: Residual statistics (% OD) of vendor data

B.2.3 *RESIDUAL VARIANCE OF VENDOR DATA*

An acceptable regression analysis has:

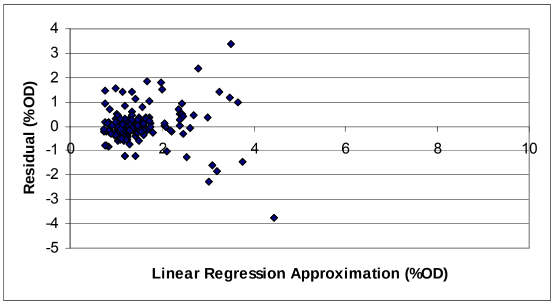
- Residual variance that is independent of the Regression Approximation.
- Residuals that are normally distributed.

In this section we investigate the residual variance for each technology. The distribution of residuals is investigated in the following section.

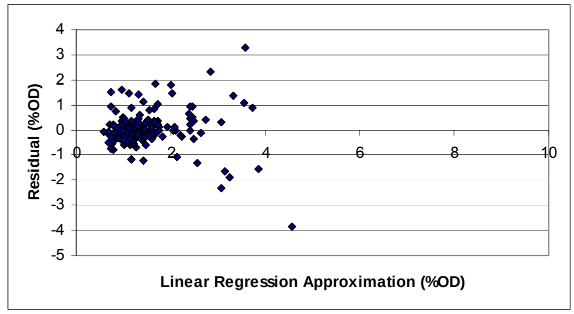
Figure 17 shows the residuals for the various linear regression analyses. For every technology, the distribution of residuals does not have a strong dependence on the linear regression approximation.

A positive (negative) residual means the ILI measurement over-predicts (under-predicts) the linear trend of the measurements. Ideally the residuals will be uniformly distributed horizontally and symmetrically distributed about 0 vertically when graphed against the linear regression approximation. That is, ideally, for each given value of the linear regression approximation (i.e., at each value of the dig depth) the residuals are symmetrically distributed (skewness = 0) with mean = 0 and have a common standard deviation.

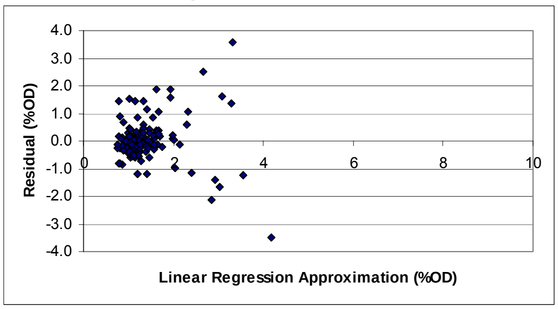
Since the mean of residuals is necessarily zero, a positive skewness implies there are more under-predictions of dent size by the ILI tool, but the over-predictions are larger in magnitude on average. This is readily apparent in the graph for technologies in Figure 17 (1). Conversely, a negative skewness implies there are more over-predictions, but the under-predictions are larger in magnitude on average. The larger the magnitude of skewness, the greater is the difference in the size of the tails of the distribution.



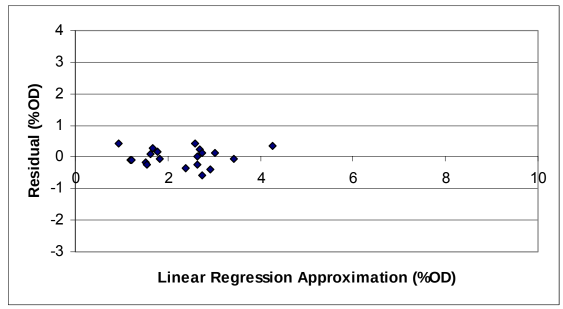
(1) Technology A,G –DAMC(EM)



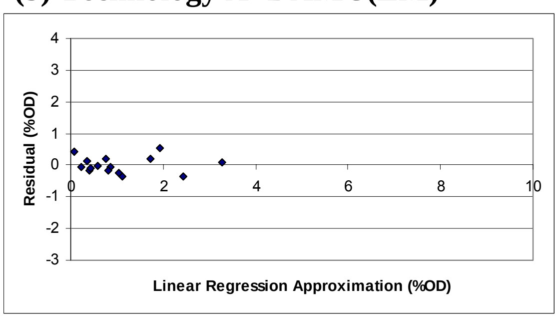
(2) Technology A,C,G- DAMC (EM)



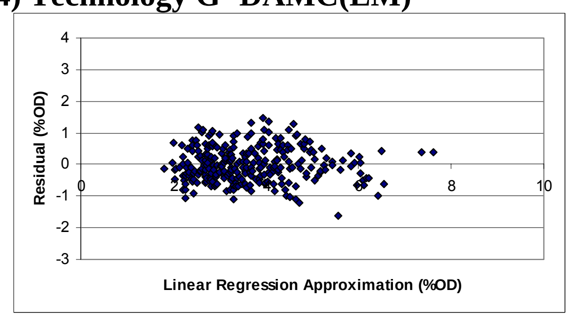
(3) Technology A- DAMC(EM)



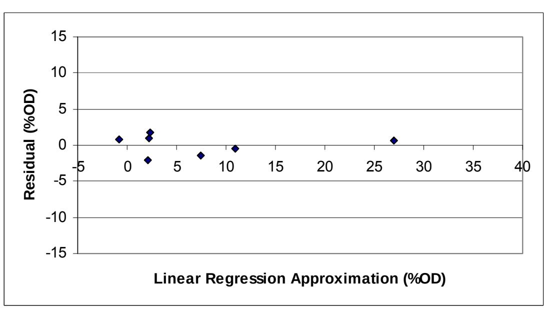
(4) Technology G- DAMC(EM)



(5) Technology C- DAMC(EM)



(6) Technology E- {LongField}MFL[Hall-3][ID/OD EM]



(7) Technology F- IEMC

Figure 17: Residuals from regression analyses for vendor dent data

B.2.4 *NORMALITY OF RESIDUAL DISTRIBUTIONS OF VENDOR DENT DATA*

Normal distributions are symmetric and, consequently, have skewness = 0. Even though none of the residuals have skewness = 0 (See Table 9), the skewnesses do not have sufficient magnitude to preclude the error terms (remember residuals are a surrogate for an error term) from being symmetrically distributed.

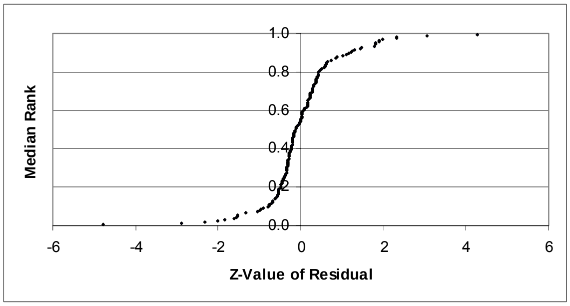
Figure 18 and Figure 19 show the Normal Probability and Tukey Lambda Probability Plot Correlation Coefficient plots. The Normal Probability plots can be grouped as follows:

1. Definite S-shape: Technology data groups A and G, A and C and G, A only, E only
2. "Bouncing around" a line: Technology C (if we ignore the extreme points on either end of the graph)
3. Roughly linear: Technology G
4. Unclear pattern due to insufficient data: Technology F

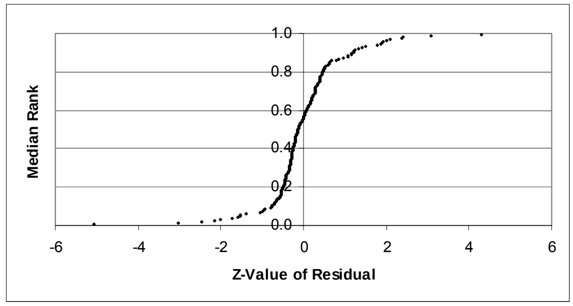
The Tukey Lambda Probability Plot Correlation Coefficient plots can be grouped as follows:

1. Maximum Correlation Coefficient clearly occurs at $\lambda < 0$: Technologies A and G, A and C and G, A only, C only
2. Correlation Coefficient at $\lambda = 0.14$ is not significantly different from the maximum Correlation Coefficient: Technologies: Technologies G only, E only.
3. Maximum Correlation Coefficient clearly occurs at $\lambda > 1$: Technology F, but there is too little data to be definitive.

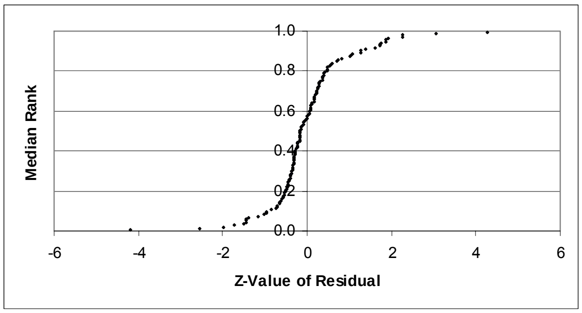
A linear Normal Probability plot and a maximum Correlation Coefficient at $\lambda = 0.14$ are indications that a distribution is normal. It follows that Technology G DAMC(EM) is the only technology for which it is reasonable to assume the residuals are normally distributed.



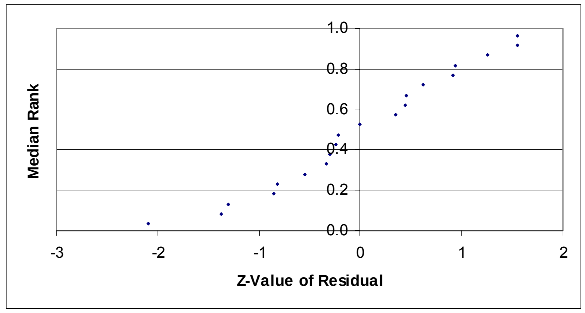
(1) Technology A,G –DAMC(EM)



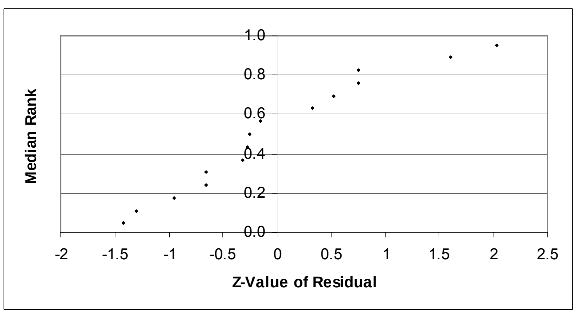
(2) Technology A,C,G- DAMC (EM)



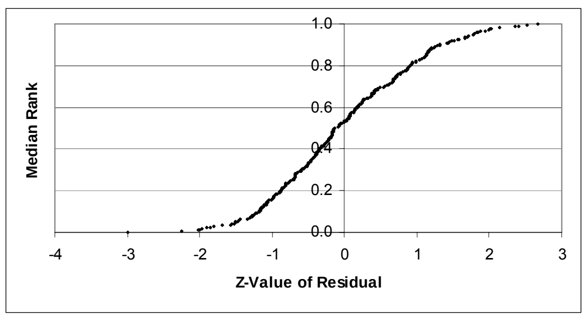
(3) Technology A- DAMC(EM)



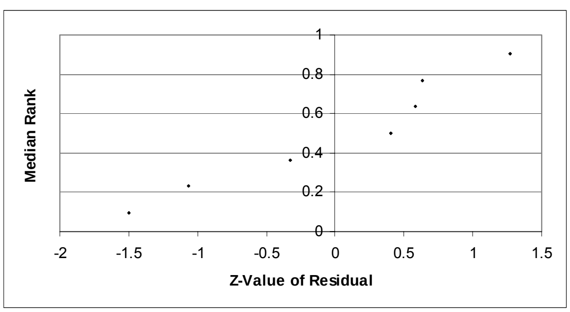
(4) Technology G- DAMC(EM)



(5) Technology C- DAMC(EM)

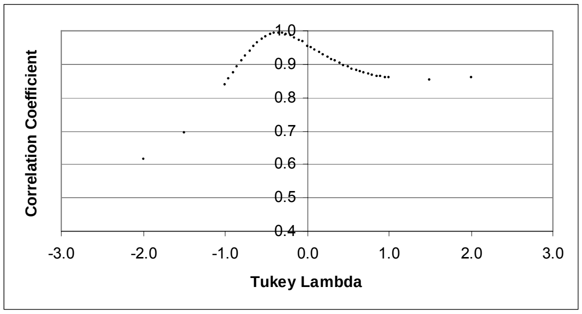


(6) Technology E- {LongField}MFL[Hall-3] [ID/OD EM]

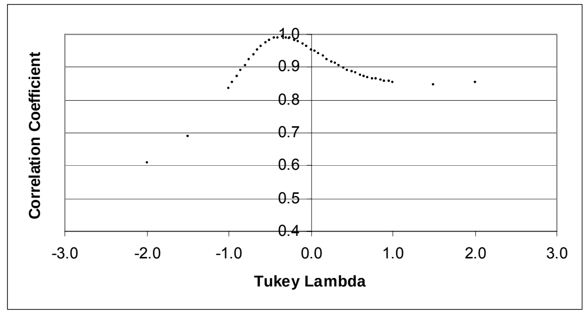


(7) Technology F- IEMC

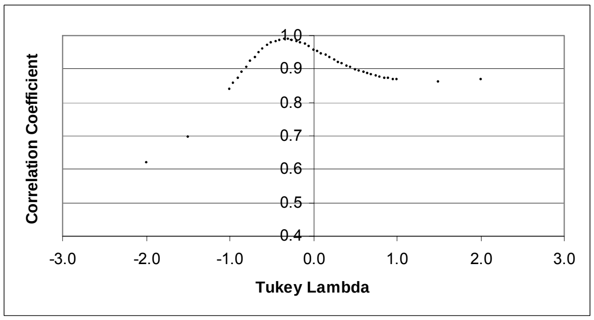
Figure 18: Normal Probability plot for residuals of vendor



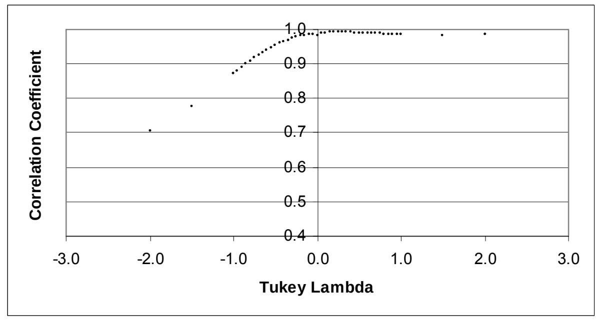
(1) Technology A,G –DAMC(EM)



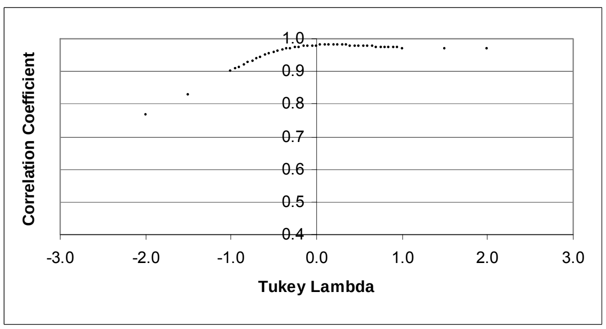
(2) Technology A,C,G- DAMC (EM)



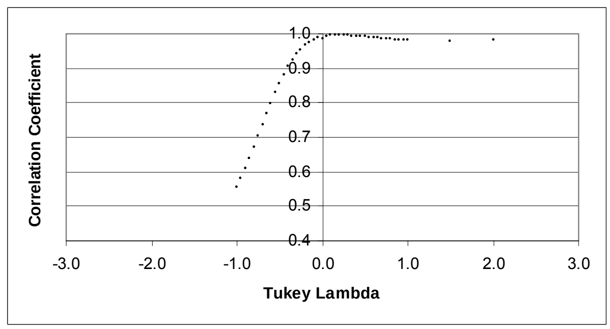
(3) Technology A- DAMC(EM)



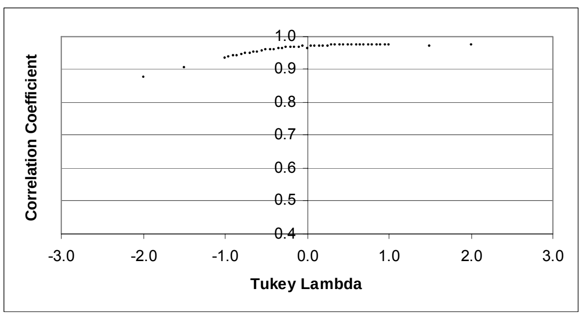
(4) Technology G- DAMC(EM)



(5) Technology C- DAMC(EM)



(6) Technology E- {LongField}MFL[Hall-3] [ID/OD EM]



(7) Technology F- IEMC

Figure 19: Tukey Lambda Probability Plot Correlation Coefficient plot for residuals

It is worth noting that a relationship between the data graphs in Figure 16 and the Normal Probability plots of residuals in Figure 18. Figure 18 shows that the Normal Probability plots for Technologies A and G, A and C and G, only A and only E have S-shaped Normal Probability plots. This implies that the tails of the distributions of these residuals are shorter than the tails of normal distributions. This means that bounds on residuals that are based on an assumption of a normal distribution, such as those indicated by dotted lines in Figure 16, actually contain a greater proportion than anticipated. In particular, the dotted lines in Figure 16 should bound more than 95% of the data points for Technologies A and G, A and C and G, only A and only E. This is seen to be true.

B.2.5 EQUATIONS OF REGRESSION LINES FOR VENDOR DATA

The Coefficient of Determination, R^2, is a measure of how well a regression line fits the data. About $100(R^2)\%$ of the total variation of the ILI measurements about their mean can be explained by the linear equation model determined by regression analysis. Table 10 gives the R^2 values for the regression analyses indicated in Figure 16. Only Technologies G, C and F DAMC(EM) have R^2 values greater than 0.90. No other technology or combination of technologies has an R^2 value greater than 0.83. This means that the relationship between ILI and dig measurements is much more linear, with less scatter, for the data associated with these technologies than for the other technologies or data groups. This is evident in Figure 16. However, as noted before, there is insufficient data to be definitive about Technology F. Table 10 also contains Standard Errors for the regression analyses. Notice that there is no apparent correlation between R^2 and Standard error. This is because they measure different variations from the regression line.

	R^2	Standard Error
Technology A,G	0.442	0.789
Technology A,C,G	0.470	0.765
Technology A	0.350	0.834
Technology G	0.900	0.289
Technology C	0.921	0.273
Technology E	0.801	0.544
Technology F	0.979	1.540
Technology A,G	0.442	0.789
Technology A,C,G	0.470	0.765

Table 10: R^2 values for graphs in Figure 16

The ideal regression line for ILI-Dig data is y = x. Consequently, we will determine whether data for each Technology supports y = x as a possible relationship between Dig and ILI data. Formulas for confidence intervals containing the coefficients of a regression line can be found in many statistical textbooks. In addition, these confidence intervals are part of the output from regression analysis in Excel. Table 11 gives 95% confidence intervals for the intercepts and slopes of regression lines for the Technologies indicated in Table 10. Only Technologies G, C and F have 0 in the confidence interval for the intercept and only Technologies C and F have 1 in the confidence interval for the slope. Thus, only Technology C allows the possibility of y = x being the true relationship between ILI and dig measurements with 95% confidence level.

It should be noted that residuals associated with Technology G behave very well, but the slope of the regression line (0.781) is small compared to the ideal value of 1.

			95% Confidence Interval	
		Coefficient	Left End	Right End
Technology A	Intercept	0.672	0.370	0.974
	Slope	0.555	0.414	0.696
Technology G	Intercept	0.364	-0.001	0.728
	Slope	0.782	0.625	0.938
Technology C	Intercept	0.088	-0.110	0.286
	Slope	1.096	0.928	1.264
Technology E	Intercept	0.402	0.157	0.647
	Slope	0.914	0.860	0.967
Technology F	Intercept	-1.139	-2.484	0.206
	Slope	1.080	0.912	1.248
Technology A,G	Intercept	0.633	0.349	0.917
	Slope	0.603	0.483	0.726
Technology A,C,G	Intercept	0.588	0.320	0.855
	Slope	0.630	0.517	0.742

Table 11: Confidence intervals for coefficients of regression lines for vendor data

B.2.6 CONCLUSIONS ABOUT REGRESSION ANALYSES FOR VENDOR DENT DEPTH DATA

None of the Technologies, or groups of Technologies, performed in an ideal manner. In particular, no technology, or data group of technologies, has all of the following desired properties:

a. Regression residuals that are normally distributed.
b. Slope of regression line close to 1.
c. Intercept of regression line close to 0.

B.2.7 ERROR DISTRIBUTIONS FOR VENDOR DENT DEPTH DATA

The error in an ILI measurement is the ILI measurement minus the verifying measurement. This assumes the verifying measurement is much more accurate than the ILI measurement. Table 12 gives statistics for the error in the ILI measurements. Errors from all deformation Technologies have small Means, which implies that their measurements are accurate on average. In fact, the Means are not statistically different according to a test that will be described in Section B.2.18 as part of the discussion of error statistics for operator data. Likewise, the Standard Deviations are relatively small, which implies there is relatively little scatter in the measured values. However, according to the test described in Section B.2.2:

- The Standard Deviation of data from Technology F is greater than those of the other Technologies at a 95% confidence level.
- The Standard Deviations of data from Technologies G and C are smaller than those of the other Technologies at a 95% confidence level.
- The Standard Deviations of data from Technologies G and C are not different at a 95% confidence level.
- The Standard Deviations of data from Technologies A and G, A and C and G, A only, and E only are not different at a 95% confidence level.

	Mean	Standard Deviation	Skewness
Technology A	0.091	0.964	-2.246
Technology G	-0.168	0.368	-0.003
Technology C	0.174	0.247	0.842
Technology E	0.095	0.553	0.137
Technology F	-0.51	1.570	-0.452
Technology A,G	0.056	0.910	-2.207
Technology A,C,G	0.067	0.872	-2.309

Table 12: Error statistics (% OD) for vendor dent data

On average the Means and Standard Deviations for the deformation vendor data give good results, with Technologies G and C doing slightly better than the others. However, this may be due to the data sets for these technologies being much smaller than for some of the other Technologies.

B.2.8 Normality of Error Distributions for Vendor Data

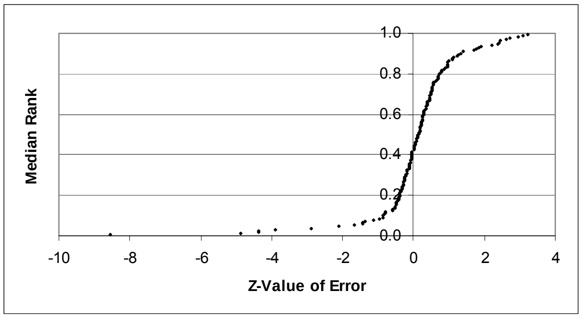
Figure 20 and Figure 21 show the Normal Probability and Tukey Lambda Probability Plot Correlation Coefficient plots. These graphs are similar to their counterparts for residuals in Figure 17. The Normal Probability plots can be grouped as follows:

1. Definite S-shape: Technology data groups A and G, A and C and G, A only, E only
2. "Bouncing around" a line: Technology C (if we ignore the extreme points on either end of the graph)
3. Roughly linear: Technologies G and F (if we ignore the extreme points on either end of the graph of Technology))

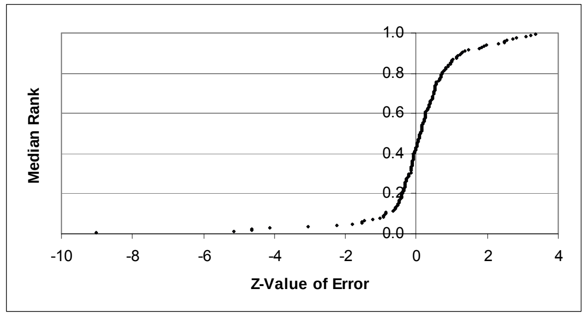
The Tukey Lambda Probability Plot Correlation Coefficient plots can be grouped as follows:

1. Maximum Correlation Coefficient clearly occurs at $\lambda < 0$: Technologies A and G, A and C and G, A only, and C only
2. Correlation Coefficient at $\lambda = 0.14$ is not significantly different from the maximum Correlation Coefficient: Technologies: Technologies G only, C only, E only, and F only.

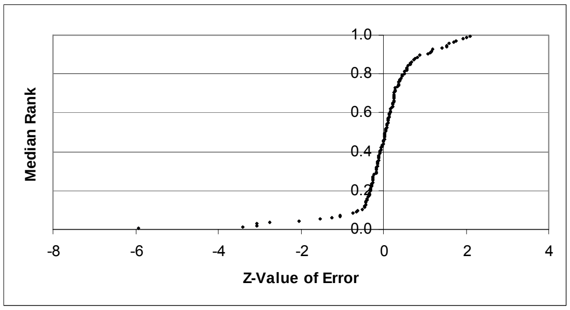
Since the linear Normal Probability plot and a maximum Correlation Coefficient at $\lambda = 0.14$ are indications that a distribution is normal, it follows that Technologies G and F (and possibly C) are the only technologies for which it is reasonable to assume the errors are normally distributed.



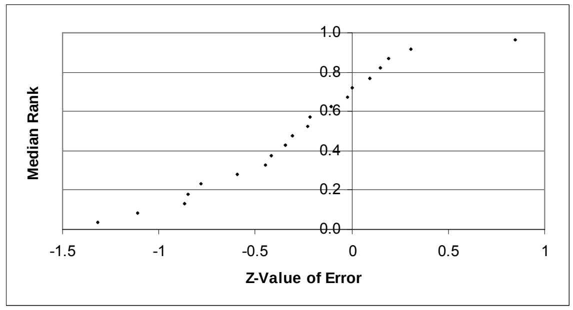
(1) Technology A,G –DAMC(EM)



(3) Technology A,C,G- DAMC (EM)

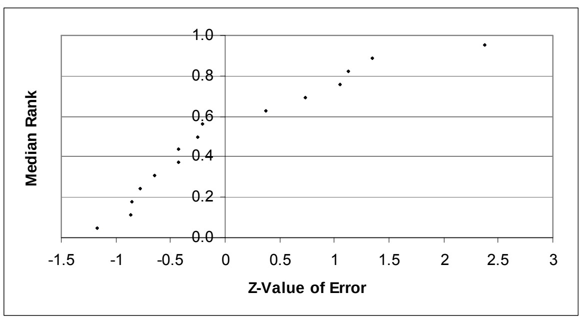


Technology A- DAMC(EM)

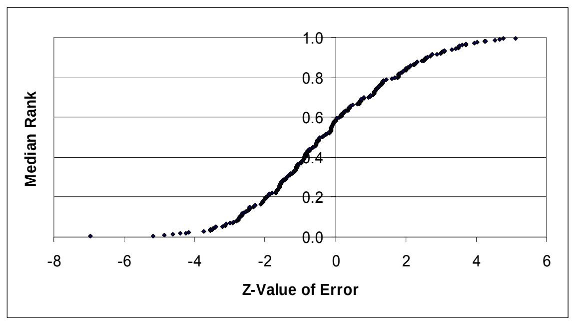


(3)

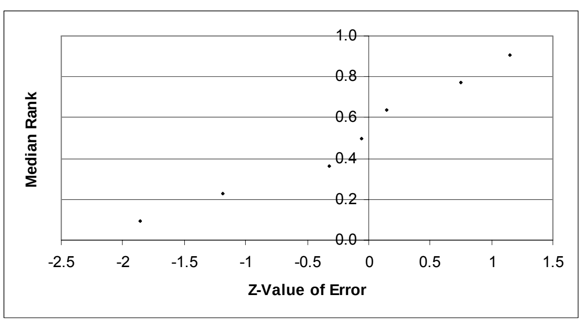
(4) Technology G- DAMC(EM)



(5) Technology C- DAMC(EM)

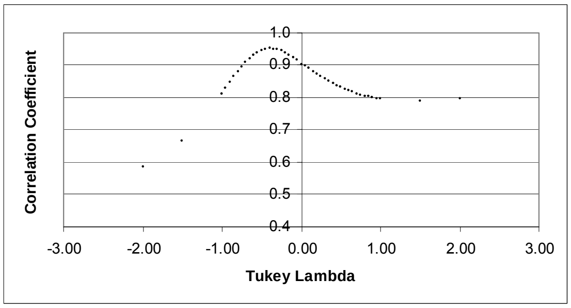


(6) Technology E- {LongField}MFL[Hall-3] [ID/OD EM]

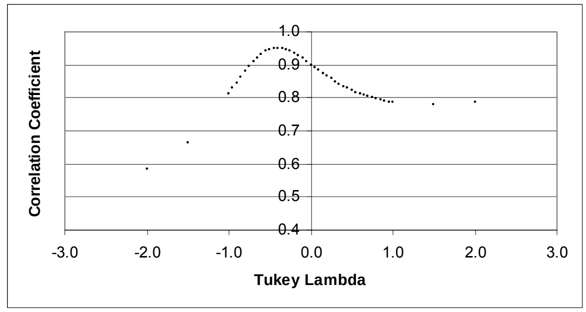


(7) Technology F- IEMC

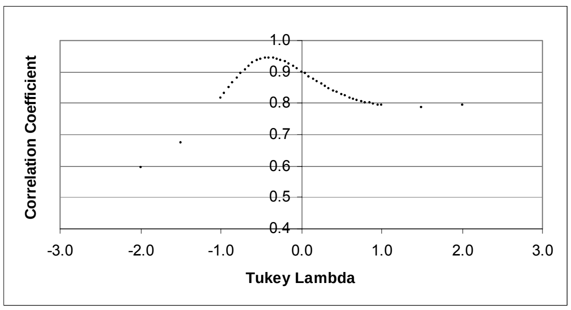
Figure 20: Normal Probability plot for errors of vendor dent data



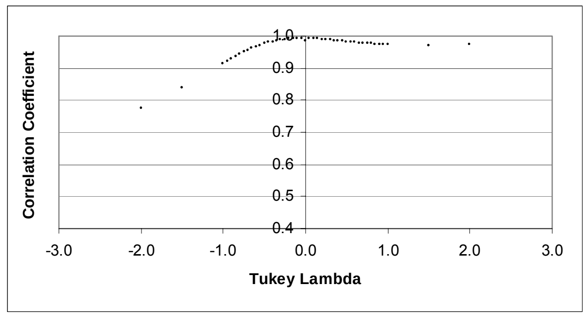
(1) Technology A,G –DAMC(EM)



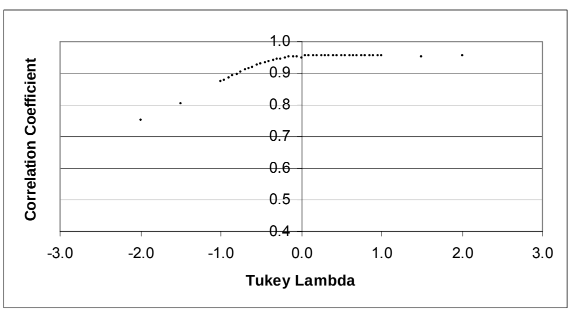
(3) Technology A,C,G- DAMC (EM)



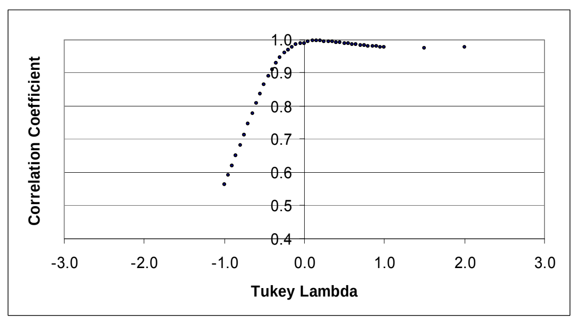
(3) Technology A- DAMC(EM)



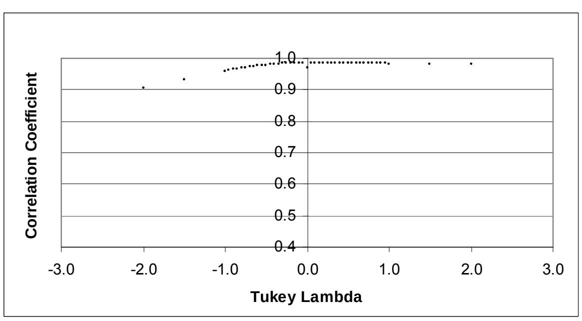
(4) Technology G- DAMC(EM)



(5) Technology C- DAMC(EM)



(6) Technology E- {LongField}MFL[Hall-3] [ID/OD EM]



(7) Technology F- IEMC

Figure 21: Tukey Lambda Probability Plot Correlation Coefficient plot for errors

B.2.9 *ACCOUNTING FOR KNOWN ERRORS IN DIG MEASUREMENTS*

No full performance specification (tolerance with associated confidence) was available for Technology D, a DAMC type caliper, which was used for validation measurements for Technology E, which measures peak dent depth, length and width from multi-vector MFL analysis. However, a performance specification for Technology K, which is similar to Technology D, claims that 85% of its caliper measurements have an error of at most 0.6% OD. We assumed the same performance specification held for Technology D and incorporated the resulting error distribution into our analysis. The conclusion of the analysis, described in detail below, is that the ideal (without errors) ILI measurements almost exactly match the ideal (without errors) caliper measurements. This implies that the two technologies give essentially the same results, allowing for inherent errors in each technology.

A complete development of methods mentioned in the following discussion can be found in Chapter 1 of the Footnote 7, whose notation is used for ease of comparison. The standard regression model is given by

$$y = \beta_0 + \beta_1\xi + \varepsilon$$

where the independent variable, ξ, is random and the error, ε, has mean zero and is uncorrelated with ξ. The unknown intercept, β_0, and unknown slope, β_1, are estimated (usually with a least-square technique) using a given a set of independent observations $(\xi_1, y_1), \ldots, (\xi_n, y_n)$. The corresponding measurement with error (ME) model assumes that the variables η and ξ are related by

$$\eta = \beta_0 + \beta_1\xi$$

but η and ξ can only be observed with errors. Instead of observing η and ξ directly, one observes

$$x = \xi + \delta \qquad \text{and} \qquad y = \eta + \varepsilon$$

where the errors δ and ε are uncorrelated. For the purposes of the present discussion

ξ = caliper measurement without error
x = caliper measurement with error
δ = error in caliper measurement
η = Technology E measurement without error
y = Technology E measurement with error
ε = error in Technology E measurement

---

[7] C-L Cheng and J. W. Van Ness, Statistical regression with Measurement error, Oxford University Press, New York, 1999.

For a sample of size n, the ME model can be formulated as follows. The unobservable "true" variables (ξ_i, η_i) satisfy

$$\eta_i = \beta_0 + \beta_1 \xi_i$$

for i = 1, … , n. However, the observed variables (x_i, y_i), which are the true variables plus errors (δ_i, ε_i).

$$x_i = \xi_i + \delta_i \qquad \text{and} \qquad y_i = \eta_i + \varepsilon_i$$

Ideally caliper and Technology E measurements are identical, so that $\beta_0 = 0$ and $\beta_1 = 1$. The following notation will be useful in determining whether Technology E performs close to this ideal.

$$\mu = \text{Mean}(\xi_i) \quad \text{and} \qquad \sigma^2 = \text{var}(\xi_i)$$

It is assumed that all δ_i have mean zero and common variance σ_δ^2 and that all ε_i have mean zero and common variance σ_ε^2. That is,

$$\text{Mean}(\delta_i) = 0,\ \ \text{var}(\delta_i) = \sigma_\delta^2,\ \ \text{Mean}(\varepsilon_i) = 0,\ \ \text{var}(\varepsilon_i) = \sigma_\varepsilon^2$$

for i = 1, … , n. It addition, it is assumed that all of the errors are independent.

A quantity with a hat "^" will denote an estimator for the quantity with the hat removed. For example, $\hat{\sigma}^2$ denotes an estimator for σ^2. An estimator for a statistic is often based on an analogous statistic obtained from sample data. In particular, the following estimators are obtained from the sample {(x_i, y_i): i = 1, … , n}:

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \hat{\mu}_y = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \hat{\sigma}_x^2 = s_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\hat{\sigma}_y^2 = s_{yy} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad \hat{\sigma}_{xy}^2 = s_{yy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Applying maximum likelihood estimation, if we can solve uniquely

$$\bar{x} = \hat{\mu} \quad (1)$$

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1\hat{\mu} \quad (2)$$

$$s_{xx} = \hat{\sigma}^2 + \hat{\sigma}_\delta^2 \quad (3)$$

(4)

$$s_{yy} = \hat{\beta}_1^2\hat{\sigma}^2 + \hat{\sigma}_\varepsilon^2 \quad (5)$$

$$s_{xy} = \hat{\beta}_1\hat{\sigma}^2 \quad (6)$$

for the six estimators $\hat{\mu}, \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2, \hat{\sigma}_\delta^2, \hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}^2, \hat{\sigma}_\delta^2, \hat{\sigma}_\varepsilon^2 \geq 0$, then we have parameter estimates for parameters in the ME model.

Assume that 85% of the caliper measurements have an error of at most 0.6% OD for Caliper Technology D (Taken from the claimed performance specification for DMAC Caliper Technology K in the absence of data for Technology D). A normally distributed random variable has 85% of measurements within 1.440 standard deviations of its mean. If we assume errors in caliper measurements are normally distributed, then $\hat{\sigma}_\delta$ = 0.6/1.440 = 0.4167 (for simplicity %OD will be dropped from all measurements), so that

$$\hat{\sigma}_\delta = \sigma_\delta^2 = 0.1736$$

Data for Technology E gives

$$\bar{x} = 3.5511, \quad \bar{y} = 3.6460, \quad s_{xx} = 1.4201, \quad s_{yy} = 1.4795, \; s_{xy} = 1.2974$$

Eq. (3) gives

$$\hat{\sigma}^2 = s_{xx} - \hat{\sigma}_\delta^2$$

From which $\hat{\sigma}^2$ = 1.2465 follows directly. Next we obtain $\hat{\beta}_1$ from Eq. (6), $\hat{\sigma}_\varepsilon^2$ from Eq. (5) and $\hat{\beta}_0$ from Eq. (2):

$$\hat{\beta}_0 = 0.1580, \quad \hat{\beta}_1 = 1.0408, \quad \sigma_\varepsilon^2 = 0.1291 \quad \hat{\sigma}^2 = 1.2465$$

Thus, the estimated slope ($\hat{\beta}_1$) and estimated intercept ($\hat{\beta}_0$) of the linear relationship between Technology E and caliper data determined by the data are close to those of the ideal result $\eta = \xi$.

If we approximate the true relationship $\eta = \beta_0 + \beta_1\xi$ by

$$\eta = \hat{\beta}_0 + \hat{\beta}_1\xi$$

then we have

$$y - \varepsilon = 0.1580 + 1.0408(x - \delta)$$

or, equivalently,

$$y = 0.1580 + 1.0408x + (\varepsilon - 1.0408\delta) \qquad (7)$$

If ε and δ are normally distributed with mean zero, then $(\varepsilon - 1.0408\delta)$ is normally distributed with mean zero and variance $\sigma_\varepsilon^2 + (1.0408)^2\sigma_\delta^2 \approx \hat{\sigma}_\varepsilon^2 + (1.0408)^2\hat{\sigma}_\delta^2 = 0.3172$. Thus, the error term has standard deviation 0.5632. A normal random variable has 80% of its values within 1.282 standard deviations of its mean. Consequently 80% of Technology E measurements are within 0.72 (= 1.282*0.5632)% OD of 0.1580 + 1.0408*(caliper measurement). That is, the measurements of Technology E tend to slightly over-predict the caliper measurements due to having slope $\hat{\beta}_1 > 1$ and intercept $\hat{\beta}_0 > 0$. Nonetheless, equation (7) is very close to the ideal equation y = x. This implies the two technologies (E and D) give essentially the same results, allowing for errors inherent in each technology.

If only 80% of the caliper measurements have an error of at most 0.6% OD, then the situation barely changes. A normally distributed random variable has 80% of measurements within 1.282 standard deviations of the mean, $\hat{\sigma}_\delta = 0.6/1.282 = 0.4680$ and

$$\hat{\sigma}_\delta = \sigma_\delta^2 = 0.2190$$

Proceeding as before

$$\hat{\beta}_0 = 0.0261, \quad \hat{\beta}_1 = 1.0802, \quad \sigma_\varepsilon^2 = 0.0780 \quad \hat{\sigma}^2 = 1.2011$$

The relationship between x and y now becomes

$$y - \varepsilon = 0.0261 + 1.0802(x - \delta)$$

or, equivalently,

$$y = 0.0261 + 1.0802x + (\varepsilon - 1.0802\delta) \qquad (8)$$

If ε and δ are normally distributed with mean zero, then $(\varepsilon - 10802\delta)$ is normally distributed with mean zero and variance estimator $\hat{\sigma}_\varepsilon^2 + (1.0802)^2\hat{\sigma}_\delta^2 = 0.3335$. Thus, the error term has standard deviation 0.5775. A normal random variable has 80% of its values within 1.282 standard deviations of its mean. Consequently 80% of Technology E measurements are within 0.74 (= 1.282*0.5775)% OD of (0.0261 + 1.0802*caliper measurement). This is almost the same tolerance as before.

Figure 22 shows that there is little significant difference between the graphs of Eqs. (7) and (8) without the error terms.

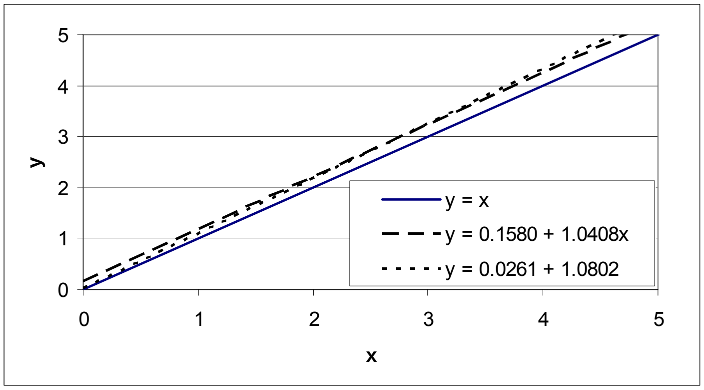


Figure 22: Illustration of small difference in regression with validation error removed

Since there is also no significant difference in the tolerances (0.71 and 0.74), we conclude that changing the certainty of the caliper error measurements from 85% to 80% makes no significant change in the results.

The significance of this analysis for Technology E is that validation errors account for almost all error in relating Technology E depth predictions to validation measurements based on the second in-line inspection tool (Caliper Technology D). Assuming the other possible errors in validation of dent depth inspection tools relate to pressure re-rounding or rebounding of pipe walls, this analysis demonstrates that in-line inspection using Technology E effectively eliminates these errors, thus only full conformation of the depth sizing performance for validating Technology D is required to reliably characterize the performance of Technology E.

B.2.10 Regression analysis of Pipeline Operator Data

Field measurements of deformation and metal loss were provided in MS Excel spreadsheets by four pipeline operators with pipelines in North and South America. Three of them operate gas pipelines and the other operates pipelines transporting liquids. In an effort to preserve anonymity, gas pipeline operators will be called Operator G1, Operator G2 and Operator G3, while the liquid pipeline operator will be called Operator L1.

Considerable time and effort was required to place the received data in a common format that allowed easy comparison of results among Operators. This required interpretation of some data and "review/verification" discussions with Operators. Some of the issues involved in processing the Operator data were to verify:

- A consistent set of measurement units used in both the reported and in-ditch measurements
- Actual nominal diameters were used to express dent depths as percentages of pipeline diameter. For instance, a 4" NPS pipeline has an actual nominal diameter of 4.5".

- Which columns in the databases correspond to dent data (depth, length and width) and which columns to metal loss data.
- Data units or formats are not changed by software or file format conversions.

As this report clearly shows, this unified database contains significant amounts of useful information. Unfortunately, not all of the data supplied by Operators allowed dent comparisons. The following describes the data from each Operator that was used in this study.

Operator G1

- Dent depth comparisons: 47
- Date gathered: November 2004 to November 2006
- Nominal diameter: 8, 10, 12, 16, 20, 24 and 30"
- Nominal thickness: 0.188, 0.203, 0.219, 0.250, 0.277, 0.292, 0.312, 0.317 and 0.375"
- Grades: A, A25, B, X40, X42, X52 and X60.
- Dent measurement technology: old versions of A and C, K

Operator G2

- Dent depth comparisons: 31
- Date gathered: August 2003 to May 2006
- Nominal diameter: 8, 16, 20, 26 and 30"
- Nominal wall thicknesses: 0.188, 0.250, 0.281 and 0.344"
- Grades: X42, X52, X56 and X60.
- Dent measurement technology: K.

Operator G3

- Dent depth comparisons: 7
- Date gathered: September 2003
- Nominal diameter: 20"
- Nominal wall thickness: 0.344"
- Grade: X70.
- Dent measurement technology: D

Operator L1

- Dent depth comparisons: 216
- Date gathered: August 2004 to March 2007
- Nominal diameter 6, 8, 10, 12, 16, 18 and 20"
- Nominal wall thickness: multiple
- Grade: multiple
- Dent measurement technology: old versions of A and C, J, K.

Operator L1 employed a longitudinal field MFL technology for metal loss without any special mechanical damage signal analysis. This MFL technology was categorized as Technology N but was not presented by the participating vendor as a current mechanical damage technology.

These Operators employed direct measuring arm calipers similar to the DMAC technologies presented in this research. As such they are referenced throughout this research with the subscript "o" modifier. The Operator data identified as L2 was taken from a publication where the operator utilized a Type K caliper and a transverse field MFL in-line inspection tool, technology L. [8]

Operator L1 provided an extensive database totaling 216 records on mechanical damage related features, all of which resulted in depth comparisons. Unfortunately, this database did not include diameter, wall thickness or SMYS information. However, most diameters could be extrapolated by comparing the absolute and relative depth of dents, and adjusting the results to the closest API 5L nominal diameter. The Operator later provided diameter information for those pipeline sections where it was not possible to extrapolate diameters.

B.2.11 Data from all 4 Operators combined

As described in the previous section, the data provided by all four Operators were compiled with one data format to simplify analysis. It was decided that it was not proper to analyze statistically this "lumped" data because of the vast number of differences in technologies, conditions under which it was gathered and ways in which it was gathered. However, it is interesting to consider some general trends in the "lumped" data.

Figure 23 compares the in-ditch and ILI measurements. Notice that data is scattered roughly about the line y = x (ideally the data would lie on y = x) with greater scatter above the line than below the line. The asymmetry of scatter may be due to rebounding being more prevalent than re-rounding.

[8] Eiber, Bob, Report on Overview Assessment of the 16 inch Diameter Olympic Pipeline Integrity to City-county Pipeline Safety Consortium of Washington, Robert Eiber Consultant Inc, Nov 2001

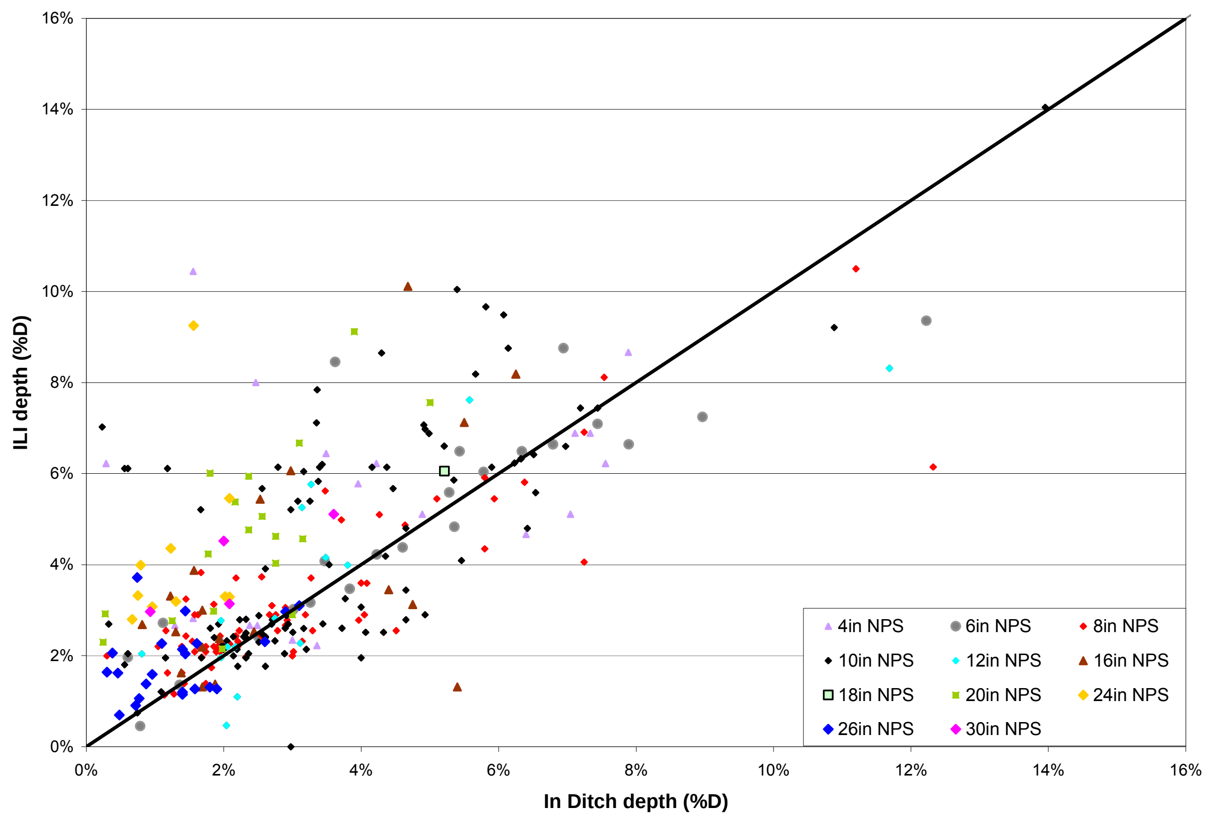


Figure 23: Dent depth comparison, data from all 4 Operators

Figure 24 shows that there is no strong relationship between the dent measurement errors (ILI depth minus In-Ditch depth) and either nominal pipe diameter or the ratio of nominal pipe diameter and wall thickness (D/t).

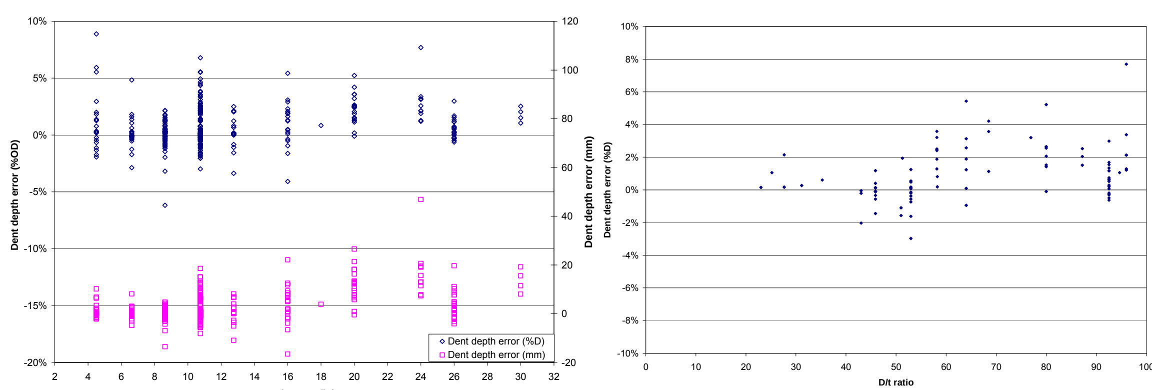


Figure 24: Dent depth error vs. nominal diameter and Dent depth error vs. D/t nominal diameter

Figure 25 shows that yield strength (SMYS) of the steel does not have a strong influence on ILI dent depth measurement error. There may be a slight increase in ILI measurement error and a slight decrease in the actual dent depth as SMYS increases, but the data is too scattered to make any firm conclusions.

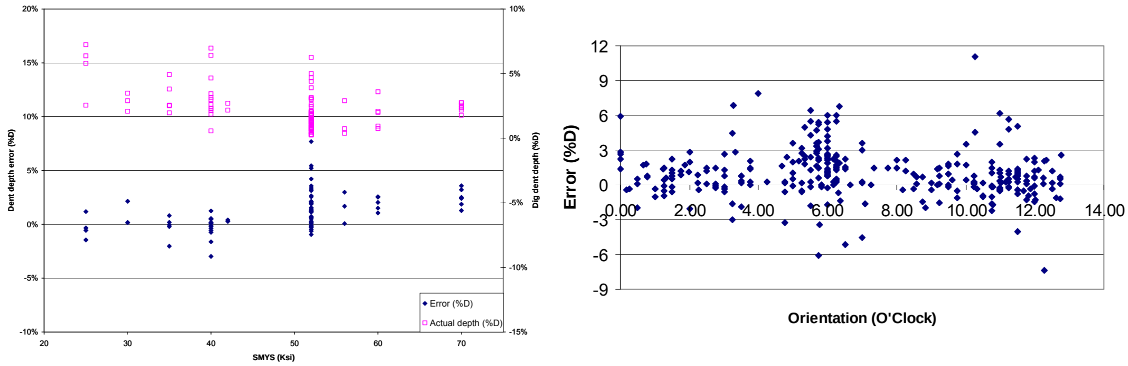


Figure 25: Dent depth error and Dent Dig Depth vs. Yield Strength, and Dent measurement error vs. dent orientation

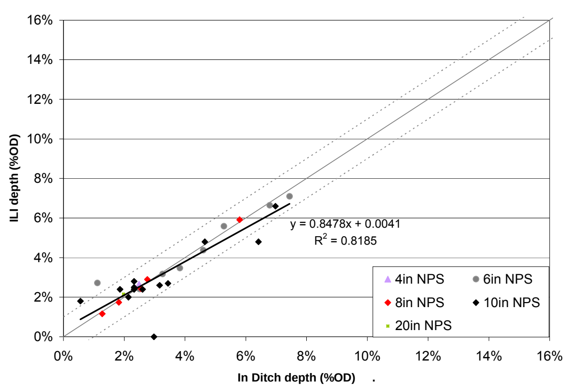
Figure 25 also shows errors in dent measurements with respect to orientation. The cluster of relatively large errors near 6 o'clock indicates that rebounding may influence dent measurements near the bottoms of the pipes more than elsewhere on the pipes.

B.2.12 Regression Analysis of Technologies represented in the Operator data

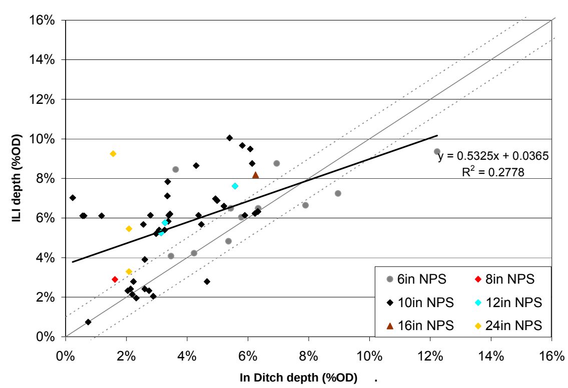
Data supplied by the Operators were grouped according to the ILI sensor technologies used to obtain the data. Technologies A, E, J, K and F are represented in the data. However, previous generations of caliper technologies A, C, E, J, F and K (denoted by A_0, C_0, etc to distinguish them from their descendents) are possibly present in the data. It is worth repeating here that the data obtained from the Pipeline Operators was obtained mainly over the period 2003-2006 and may not represent current mechanical damage assessment as characterized by the vendors.

Figure 26 shows Dig and ILI measurements for technologies supplied by operators, along with regression lines where appropriate. Only Technologies A_0, C_0, J_0, and K_0 have sufficient data for meaningful analyses, which are described in the next several sections.

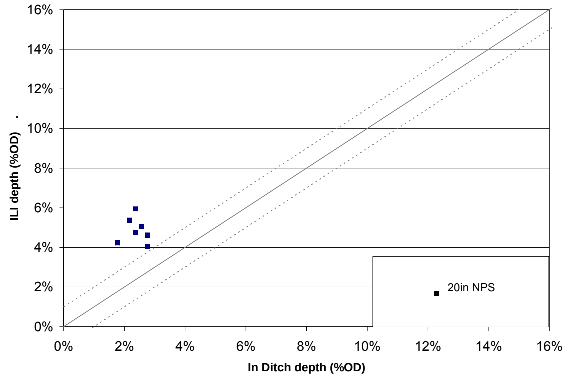
Figure 27 shows Dig and ILI measurements for these technologies along with the regression line and the regression line ±2 (Standard Error).



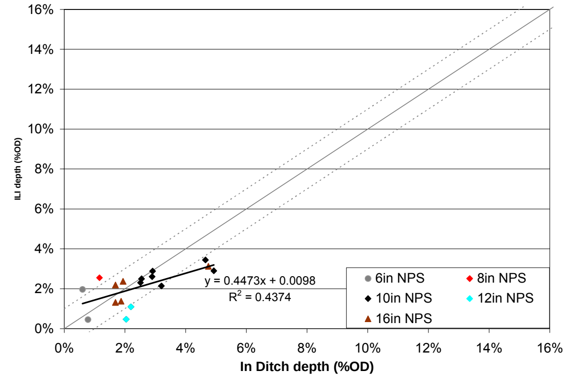
Technology A_0



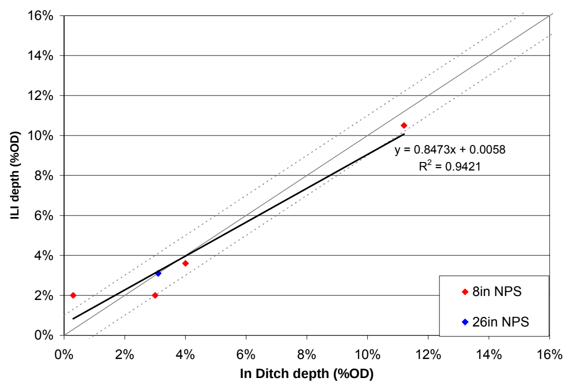
Technology C_0



Technology E_0

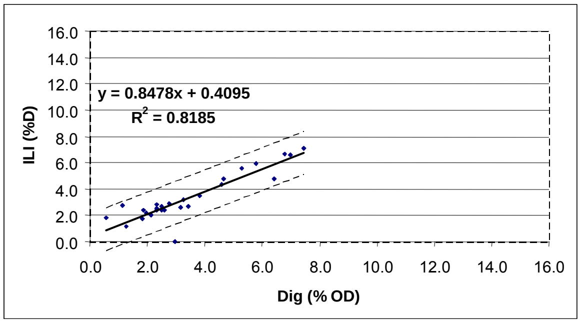


Technology J_0

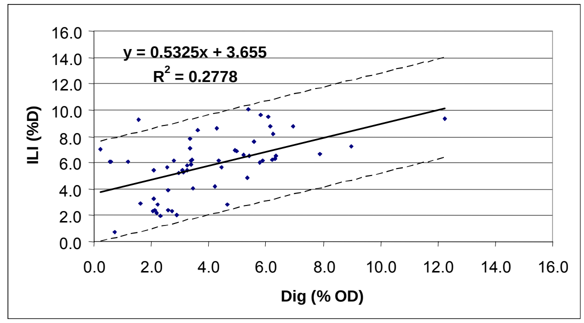


Technology F_0

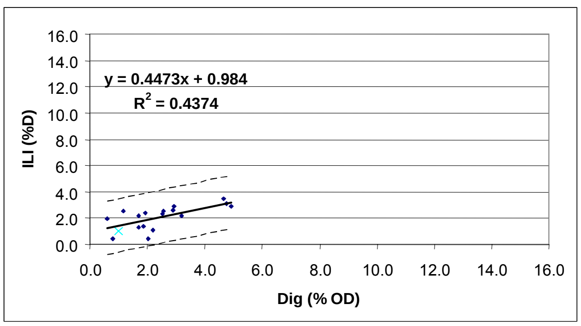
Figure 26: Dig and ILI measurements for technologies supplied by operators



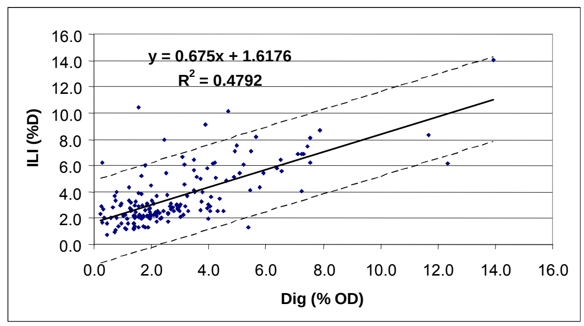
Technology A_0



Technology C_0



Technology J_0



Technology K_0

Figure 27: Dig and ILI measurements for technologies supplied by operators (dashed lines indicate regression line ±2 (Standard Error)

B.2.13 RESIDUAL VARIANCE OF OPERATOR DATA

An acceptable regression analysis has:

- Residual variance that is independent of the Regression Approximation.
- Residuals that are normally distributed.

In this section we investigated the residual variance for each Operator technology. The distribution of residuals is investigated in the following section.

Table 13 gives a data count along with the Mean, Standard Deviation, and Skewness of the residuals for Operator data. Technologies A_0 and J_0 have Standard Deviations that cannot be distinguished by the test described in Section B.2.2. The same is true for Technologies C_0 and K, except their Standard Deviations are much larger than for the other technologies. The larger standard deviations imply there is greater scatter about the regression lines for Technologies C_0 and K than those for Technologies A_0 and J. This is clearly shown in Figure 26.

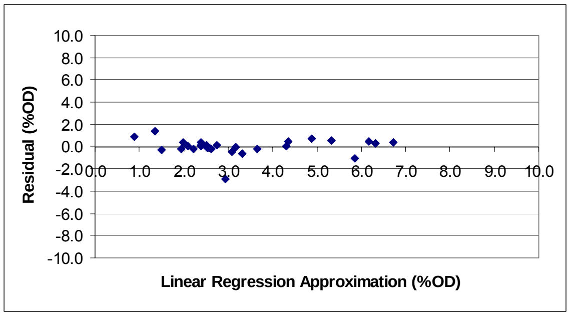
	Count	Mean	Standard Deviation	Skewness
Technology A_0	28	0.000	0.751	-2.123
Technology C_0	58	0.000	1.909	0.204
Technology Jo	17	0.000	0.660	-0.579
Technology Ko	166	0.000	1.557	1.502

Table 13: Residual statistics (% OD) for operator data

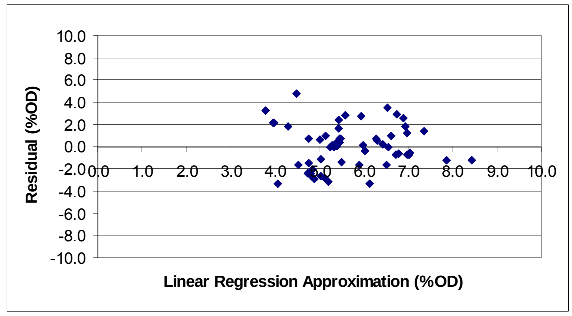
Figure 28 shows the graphs of residuals of the regression analyses for each Operator. Evidently the data is approximately uniformly distributed horizontally on the graph. This means the distributions of residuals are not overly dependent on the size of the dent. However, the vertical distribution of residuals is not always symmetric about 0. Residuals with large magnitude (absolute value) for Technology K_0 tend to be associated with positive residuals. These are "balanced" by a larger proportion of negative residuals with relatively small magnitudes. This means the distribution of these residuals are skewed to the right, which is precisely what is implied by the positive Skewness value in Table 13. This lack of symmetry in the residuals needs careful consideration.

Skewed to the right means the ILI measurements greatly over-predict Dig measurements more than they greatly under-predict Dig measurements. This could be due to rebounding. Rebounding makes a dent measured by the ILI tool smaller than the same dent measured after digging. This increases the "error" in the ILI measurement, which tends to increase the regression residual. Consequently, rebounding could be the cause of the large positive Skewness for Technology K_0. Analogously, re-rounding could be the cause of the large magnitude of the Skewness for Technology A_0.

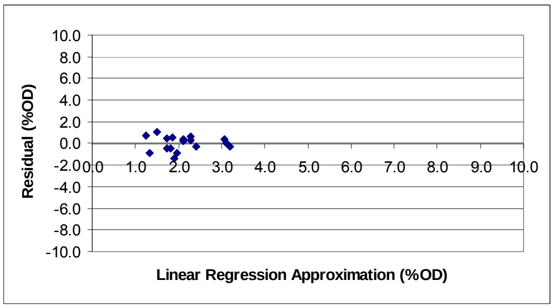
The lack of symmetry of the distribution of residuals is not sufficient to cause immediate rejection of conclusions based on the regression analyses, but such conclusions should be viewed with great skepticism, especially for Technology K_0.



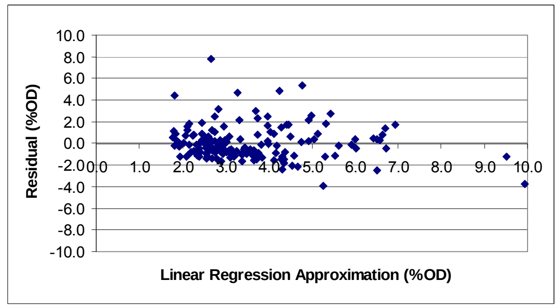
Technology A_0



Technology C_0



Technology Jo

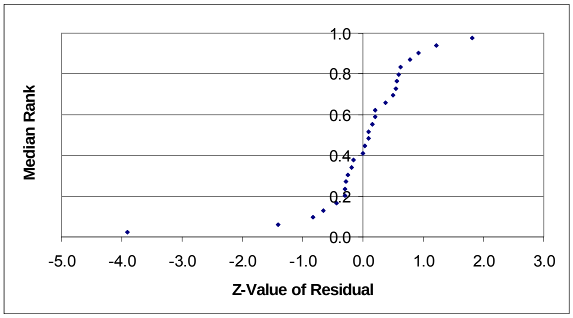


Technology Ko

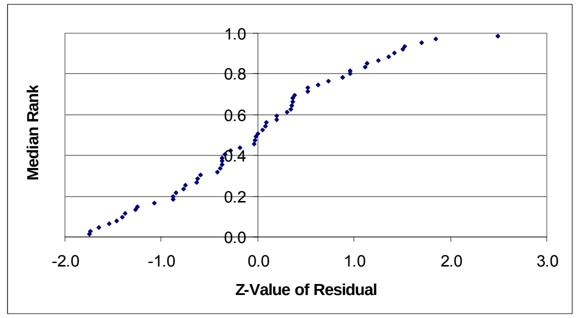
Figure 28: Residuals of operator data

B.2.14 Normality of Residual Distributions of Operator Data

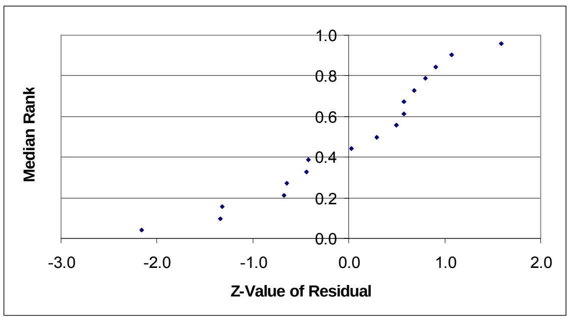
Figure 29 and Figure 30 show the Normal Probability and Tukey Lambda Probability Plot Correlation Coefficient plots, respectively, for the various Operators. The Normal Probability plots for Technologies A_0 and K_0 are definitely S-shaped, indicating it is not reasonable to assume the residuals are normally distributed. This is confirmed by maximum Correlation Coefficients occurring at λ's less than 0 in their Tukey Lambda Probability Plot Correlation Coefficient plots. The Normal Probability plot for Technology C_0 is nearly linear and the maximum Correlation Coefficient occurs at $\lambda \approx 0.14$. Therefore, it is reasonable to assume the residuals for Technology C_0 are normally distributed, but it is not reasonable to assume the residuals for Technologies A_0 and K are normally distributed. The distribution for residuals of Technology J_0 has properties between these two extremes, so the normality of the distribution is uncertain. Thus, a basic assumption for a valid regression analysis is not satisfied by the data for Technologies A_0, K_0, and possibly J_0.



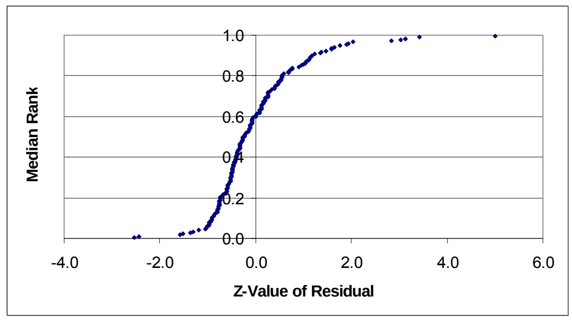
Technology A_0



Technology C_0

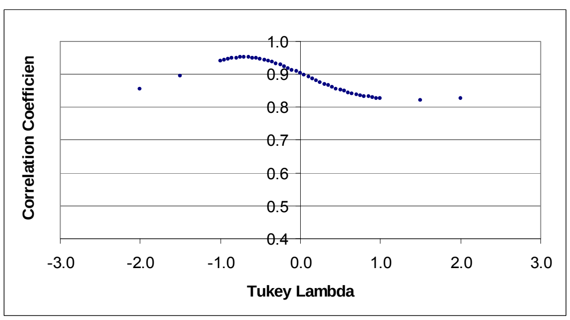


Technology Jo

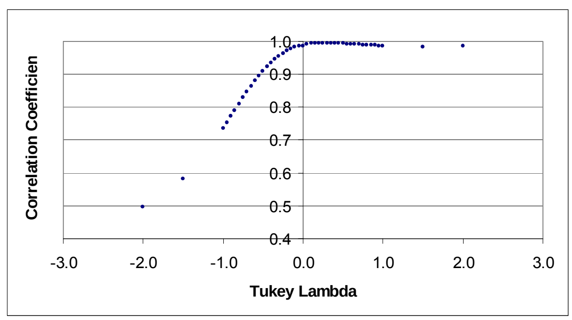


Technology Ko

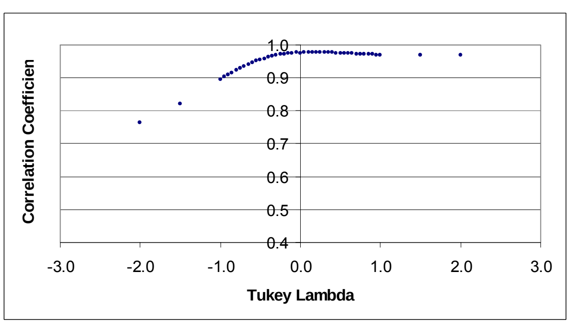
Figure 29: Normal Probability Plots for residuals of operator data



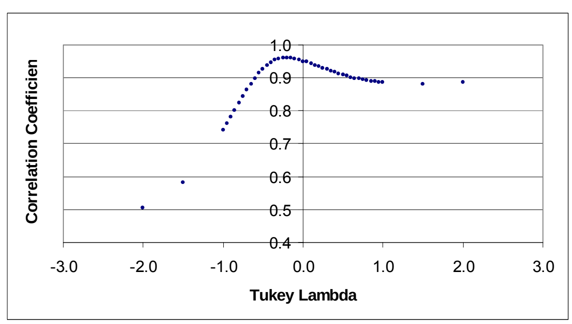
Technology A_0



Technology C_0



Technology Jo



Technology Ko

Figure 30: Tukey Lambda Probability Plot Correlation Coefficient plots for residuals of operator data

B.2.15 EQUATIONS OF REGRESSION LINES FOR OPERATOR DATA

The Coefficient of Determination, R^2, is a measure of how well the regression line fits the data. About $100(R^2)\%$ of the total variation of the ILI measurements about their mean can be explained by the linear equation model determined by regression analysis. Table 14 gives the R^2 values for the regression lines indicated in Figure 26. Clearly the regression results explain relatively little about the variation of the ILI measurements because the R^2 values are so small. This is reflected in the large scatter about the regression lines in Figure 26. It also reflects the relatively large values of the Standard Errors shown in Table 14.

	R^2	Standard Error
Technology A_0	0.819	0.817
Technology C_0	0.278	1.906
Technology Jo	0.437	1.008
Technology Ko	0.479	1.561

Table 14: R^2 values from Figure 26

Ideally, the ILI and dig measurements will be in perfect agreement, thereby making y = x the regression line. Using standard methods discussed in many statistics textbooks it is possible to construct 95% confidence intervals of the slope and intercept of each regression line just as we did for vendor data in Section B.2.5. Table 15 gives endpoints of these confidence intervals for the data provided by the operators. Important conclusions can be drawn from this table:

1. With 95% confidence level, the intercept for Technology C_0 is greater than 2.231, which is considerably greater than the ideal value of 0. Thus, values given by the regression equation over-predict when dent depth is relatively small. This is seen in the graph for Technology C_0 in Figure 26. The same is true, to lesser extents, for Technologies J and K.
2. With 95% confidence level, the slope for Technology C_0 is less than 0.760, which is less than the ideal slope of 1. This implies values from the regression equation under-predict when dents are relatively large. This is seen in the three rightmost points on the graph for Technology C_0 in Figure 26 Similar results hold for Technologies J and K.
3. With 95% confidence level, the intercept for Technology A_0 lies between -0.197 and 1.016. Consequently, it is not possible to reject the ideal intercept of 0 as a possibility. Likewise, with 95% confidence level the slope for Technology A_0 lies between 0.676 and 1.020. Consequently, it is not possible to reject the ideal slope of 1 as a possibility. In short, the data is not inconsistent with the ideal result of having y = x as the regression line. However, this is tempered by our earlier finding that conclusions based on the regression analysis may not be valid for Technology A_0 due to the behavior of its residuals.

			95% Confidence Interval	
		Coefficient	Left End	Right End
Technology A_0	Intercept	0.409	-0.197	1.016
	Slope	0.848	0.676	1.020
Technology C_0	Intercept	3.655	2.231	5.079
	Slope	0.533	0.305	0.760
Technology Jo	Intercept	0.984	0.095	1.873
	Slope	0.447	0.035	0.860
Technology Ko	Intercept	1.618	1.034	2.202
	Slope	0.675	0.564	0.786

Table 15: Confidence intervals for coefficients of regression lines

B.2.16 CONCLUSIONS ABOUT REGRESSION ANALYSES OF OPERATOR DATA

Data is of such quality that:

1. It is not possible to accept conclusions based on regression analysis for Technologies A_0, J_0, and K_0 with any reasonable level of confidence due to the behavior of residuals.
2. Even if regression analyses are accepted, as is possible in the case of Technology C_0, regression equations describe an unacceptably small percentage of the variation in the ILI data.

B.2.17 ERROR DISTRIBUTION OF OPERATOR DATA

The error in an ILI measurement is the difference between it and the corresponding Dig measurement. That is if d_{ILI} denotes an ILI measure and d_{dig} denotes the corresponding Dig measurement, then the error in the ILI measurement is

$$\text{Error} = d_{ILI} - d_{dig}$$

This assumes the Dig measurement is much more accurate than the ILI measurement. This assumption is open to question. There is no information on the accuracy of the Dig measurements, so without this assumption is it impossible to assess the data. When interpreting results this assumption should be kept in mind. Table 16 gives statistics for the errors related to each Technology.

	Mean	Standard Deviation	Skewness
Technology A_0	-0.109	0.804	-1.483
Technology C_0	1.772	2.174	0.467
Technology Jo	-0.393	0.976	0.257
Technology Ko	0.675	1.715	0.665

Table 16: Error statistics (% OD) for operator data

Part of the discussion of residual variance in Section B.2.13 described how re-rounding and rebounding increased "errors" in ILI measurements by decreasing the depth of the dent before the "dig measurement" is made. Re-rounding and rebounding could be part of the cause for relatively large magnitudes of Means, Standard Deviations and Skewnesses in Table 16.

B.2.18 Comparison of error Means of operator data

Four samples of error statistics for ILI measurements provided by Operators are given in Table 16. Let μ_1, μ_2, μ_3, and μ_4 denote the Mean errors of data from Technologies A_0, C_0, J_0 and K_0, respectively. The Tukey (or Tukey-Kramer) Multiple Comparison Procedure compares the Means of the populations simultaneously. This procedure is used to construct 95% confidence intervals of the difference of Mean errors of two technologies. Two Means are judged to be significantly different (at 95% confidence level) if the corresponding confidence interval does not contain 0. If a confidence interval contains 0, the conclusion is that there is no significant difference between the Means base on data used in the analysis. Table 17 gives endpoints of 95% confidence intervals for column Mean minus the row Mean. For example, a 95% confidence interval for

$$(\text{Mean error of Technology } C_0) - (\text{Mean error of Technology } J_0)$$

is (0.931, 3.399). Notice that 0 is not contained in any confidence intervals involving C_0. This means that the Mean error for Technology C_0 is statistically different from all the other mean errors at 95% confidence level. This is hardly surprising when you look at the values in Table 16. Possibly more surprising, is the fact that all the other confidence intervals contain 0, so that the other Mean errors are not statistically different at 95% confidence level.

There are standard tests in many statistics textbooks that compare the means of two populations. It is natural to question why to use the Tukey procedure instead of repeated applications of this procedure. The answer is that the Tukey procedure creates simultaneous confidence intervals at 95% confidence. That is, if the Tukey procedure is used a very large number of times, only about 5% of the time would an interval not contain the value of what it is estimating. If pair-wise 95% confidence intervals are created, the chance that at least one interval does not contain what it is estimating increases dramatically with the number of intervals calculated. This means that the confidence level for each confidence interval would have to be increased to assure an overall confidence level of 95%.

	Technology A_0		Technology C_0		Technology J	
	Left End	Right End	Left End	Right End	Left End	Right End
Technology C_0	-2.910	-0.851				
Technology Jo	-1.091	1.661	0.931	3.399		
Technology Ko	-1.697	0.131	0.415	1.780	-2.207	0.071

Table 17: Endpoints of 95% confidence intervals for difference of error means

B.2.19 Comparison of standard deviations of errors in operator data

The authors know of no simultaneous comparison procedure for Standard Deviations (or equivalently variance) analogous to the one just used for Means. Consequently, the Standard

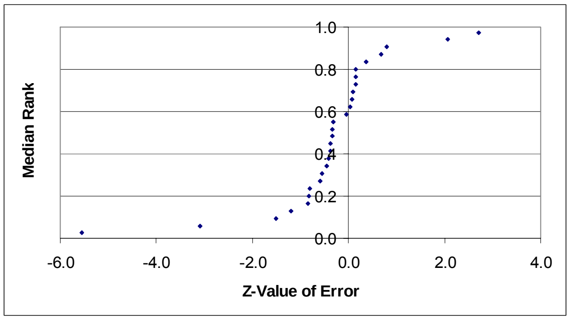
Deviations are compared pair-wise as in section 7.5.1. The greatest difference in Standard Deviations occurs between those of data from Technologies C_0 and A_0. Let s_1 and s_2 denote the Standard Deviations given in Table 16 for data from Technologies C_0 and A_0, respectively. Using a standard hypothesis test that can be found in many elementary statistics textbooks, we conclude that the Standard Deviation of data from Technology C_0 is greater than that from Technology A_0 at the 95% confidence level if $s_1^2/s_2^2 > F_{0.05}$, where $F_{0.05}$ is the critical number that cuts a tail with area 0.05 from an F distribution with $\nu_1 = 58 - 1$ and $\nu_2 = 28 - 1$ (58 and 28 are the numbers of data from Technologies C_0 and A_0, respectively) degrees of freedom. A simple calculation shows $s_1^2/s_2^2 = 7.31$. According to a standard table, $F_{0.05}$ corresponding to the given degrees of freedom is less than 2.0. Consequently $s_1^2/s_2^2 > F_{0.05}$ and we conclude that the Standard Deviation of data from Technology C_0 is greater than that from Technology A_0 at a 95% confidence level. Analogous calculations show that the Standard Deviation of data from Technology C_0 is greater than that of the other Technologies at a 95% confidence level. Similarly, the Standard Deviation of Technology K_0 is greater than those of Technologies A_0 and J_0, but the Standard Deviations of Technologies A_0 and J_0 are not different, at a 95% confidence level.

B.2.20 Summary of Comparisons of Means and Standard Deviations of Errors for Operator Data

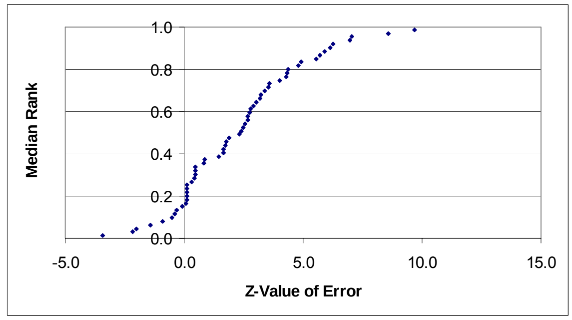
The previous two sections show that the Means and Standard Deviations of the errors for some technologies are statistically different from corresponding statistics of other technologies. In particular, the Mean and Standard Deviation of measurement errors from Technology C_0 are larger than the Means and Standard Deviations of the other technologies at a 95% confidence level.

B.2.21 Normality of Error Distributions of Operator Data

It is commonly assumed, often without justification, that measurements are normally distributed. While this assumption may be valid in some situations, the data provided by the operators does not support it being valid for ILI data measurements. Assessment of distributions of errors will be analogous to earlier assessment of distributions of residuals. The Normal Probability and Tukey Lambda Probability Plot Correlation Coefficient plots for the errors are shown in Figure 31 and Figure 32, respectively. Normal Probability plots for Technologies A_0, C_0 and K_0 have definite S-shapes, indicating the distributions are not normal. Only the Normal Probability plot for Technology J_0 resembles a straight line, and that is a weak resemblance at best. Consequently, it is not reasonable to assume that the error distributions for Technologies A_0, C_0 and K_0 are normal. The situation is unclear for Technology J_0.

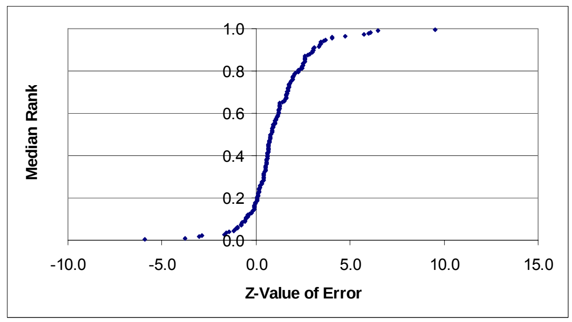


Technology A_0



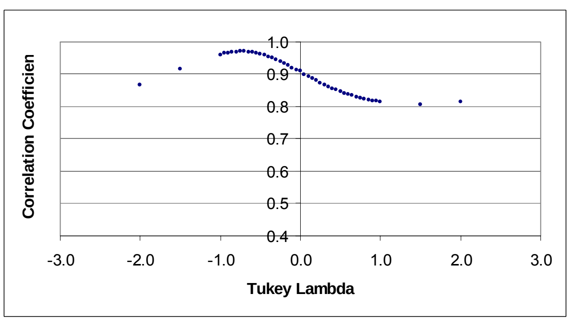
Technology C_0

Technology Jo

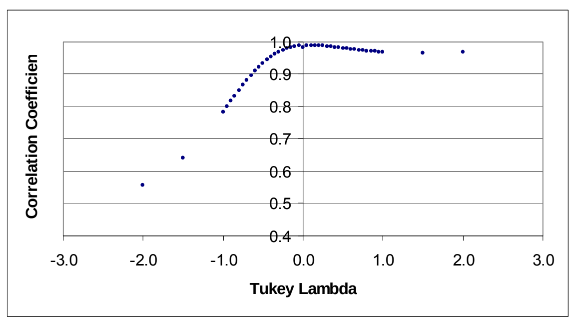


Technology Ko

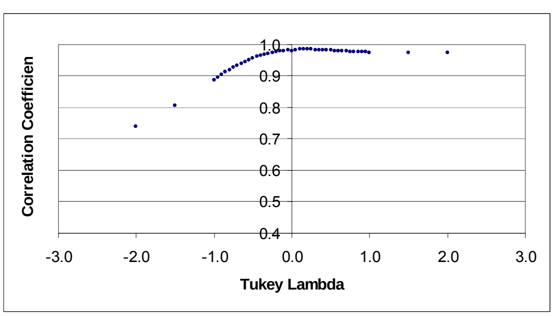
Figure 31: Normal Probability Plots for errors



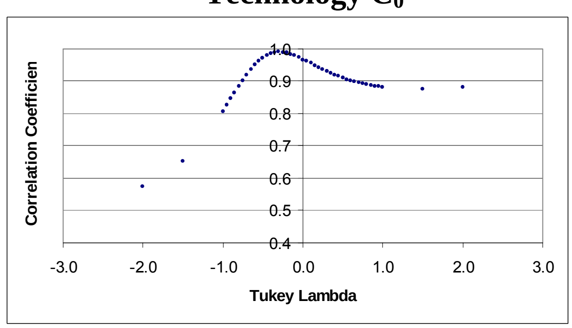
Technology A_0



Technology C_0



Technology Jo



Technology Ko

Figure 32: Tukey Lambda Probability Plot Correlation Coefficient plots for errors

B.3 Dent length and width measurements

Only 43 of the 301 dents with both Dig and ILI depth measurements had both Dig and ILI length measurements. The most measurements by one technology are 12. This is too small to do a reasonable analysis of individual performances of various technologies. Consequently, the data were lumped without regard for technology. Figure 33 shows the comparison of Dig and ILI measured lengths. It also shows ILI length as percent of Dig length. Each graph shows a clear tendency for ILI lengths to under-predict dig lengths. Table 18 shows that the mean ILI length is 70% of the Dig length with a Standard Deviation of 33%. The small Mean (ideally the Mean would be 100%) and large Standard Deviation indicate scattered data as evidenced in Figure 35. This discrepancy between ILI and dig lengths could be due to variation in inspection tool speed.

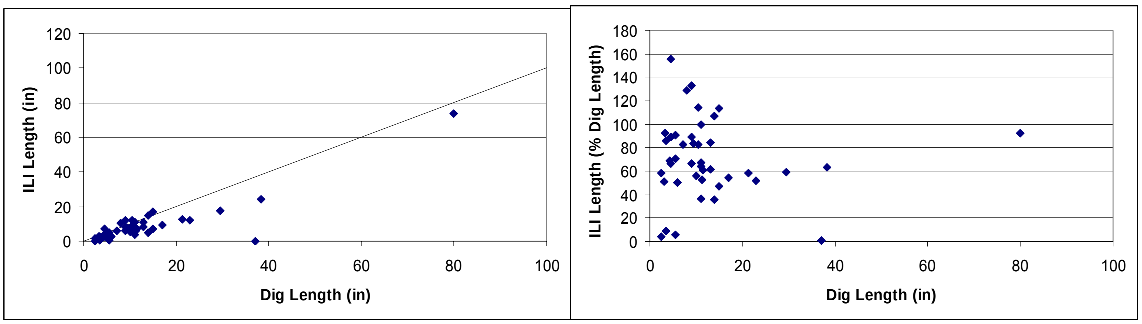
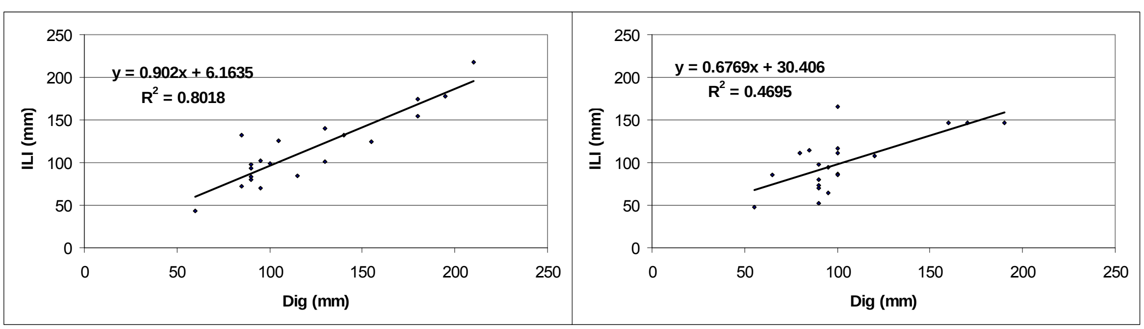


Figure 33: Dig and ILI dent comparison, and ILI length as % of Dig length

Mean	Standard Deviation
70.1%	33.1%

Table 18: Mean and Standard Deviation of ILI length as % of Dig Length

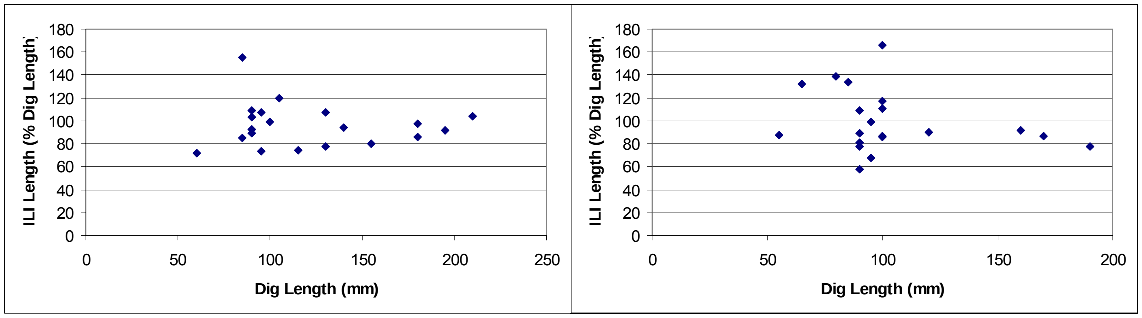
Only one Vendor provided data for length and width measurements. These consisted of measurements on 20 dents using Technology G. Figure 34 shows ILI and Dig measurements for dent lengths and widths, along with linear regression lines. It is clear that there is more scatter in the width data than in the length data.



(a) Dent Length **(b) Dent Width**

Figure 34: Dent lengths and widths measured by Technology G

Figure 35 shows the ILI lengths and widths as percentages of dig lengths and widths, respectively. There is insufficient data to make any conclusion, but it appears that the percentage error in length may decrease as length increases and the width predictions may increasingly under-predict dig measurements as length increases.



(a) Dent Length **(b) Dent Width**

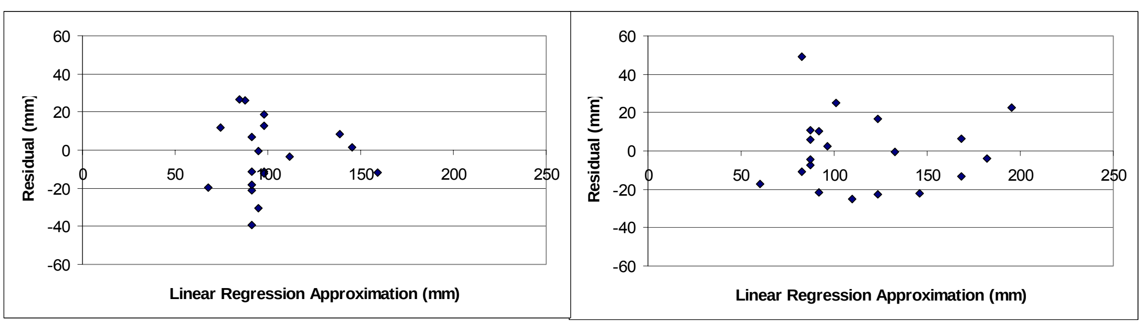
Figure 35: Dent lengths and widths measured by Technology G as % of dig measurement

Table 19 gives the Mean, Standard Deviation, and Skewness of the residuals for the lengths and widths. The large Standard Deviations reflect scatter seen in Figure 34.

	Count	Mean	Standard Deviation	Skewness
Length	20	0.000	19.049	0.814
Width	20	0.000	24.045	1.013

Table 19: Residual statistics (mm) for length and width data

Figure 36 shows the residuals for the linear regression analyses. In each case residuals are approximately uniformly distributed horizontally and vertically. However, the clumps of data near a linear regression approximation of 90 mm and the sparseness of data elsewhere make any conclusion uncertain.

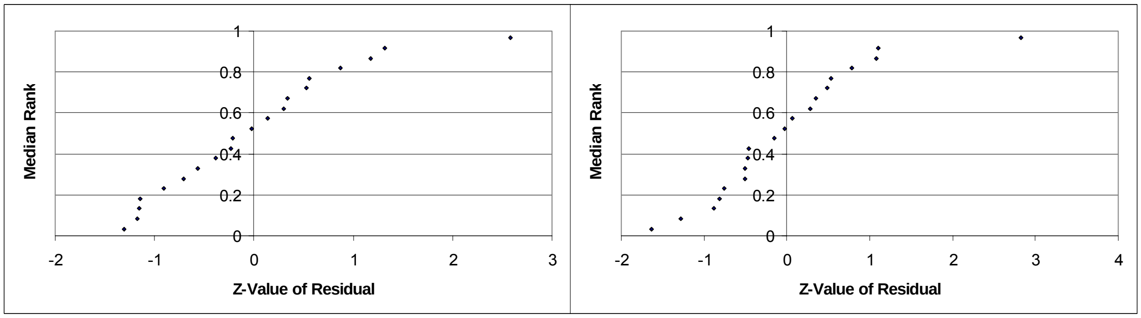


(a) Dent Length **(b) Dent Width**

Figure 36: Residuals from regression analyses for dent length and width data

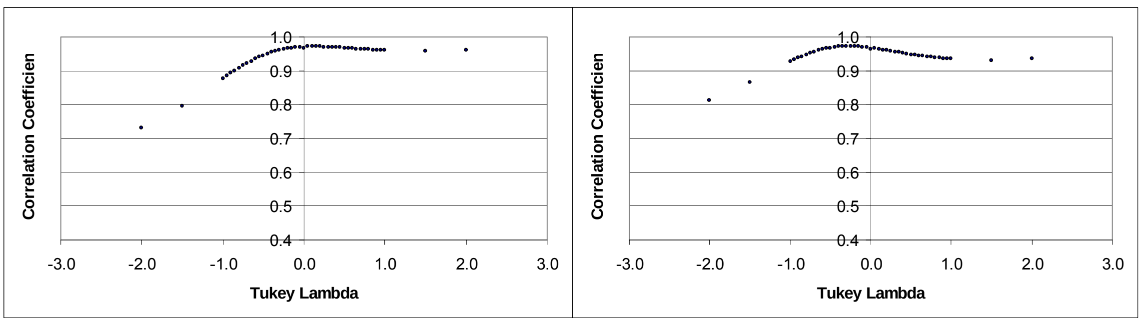
Figure 37 and Figure 38 show Normal Probability and Tukey Lambda Probability Plot Correlation Coefficient plots for the residues. Except for one point on the far right, the Normal Probability Plot for dent length residuals is roughly linear, while that for dent widths is S-shaped. These observations are corroborated by the Tukey Lambda Probability Plot Correlation Coefficient Plots which have the maximum Correlation Coefficient occurring at approximately 0.14 for length residuals and at a negative value for width residuals. In short,

- It is reasonable to assume residuals for length residuals are normally distributed.
- It is not reasonable to assume residuals for width residuals are normally distributed.



(a) Dent Length **(b) Dent Width**

Figure 37: Normal Probability plots for residuals of dent length and widths



(a) Dent Length **(b) Dent Width**

Figure 38: Tukey Lambda Probability Plot Correlation Coefficient plots for residuals of dent lengths and widths

Table 20 gives 95% confidence intervals for coefficients of regression lines shown in Figure 34. In each case 0 is in the confidence interval for the intercept and 1 is in the confidence interval for the slope. Thus, it is not possible to exclude y = x as being the true relationship between ILI and Dig measurements for both length and width with 95% confidence level.

			95% Confidence Interval	
		Coefficient	Left End	Right End
Length	Intercept	6.164	-19.888	32.215
	Slope	0.902	0.681	1.122
Width	Intercept	30.406	-6.480	67.291
	Slope	0.677	0.316	1.038

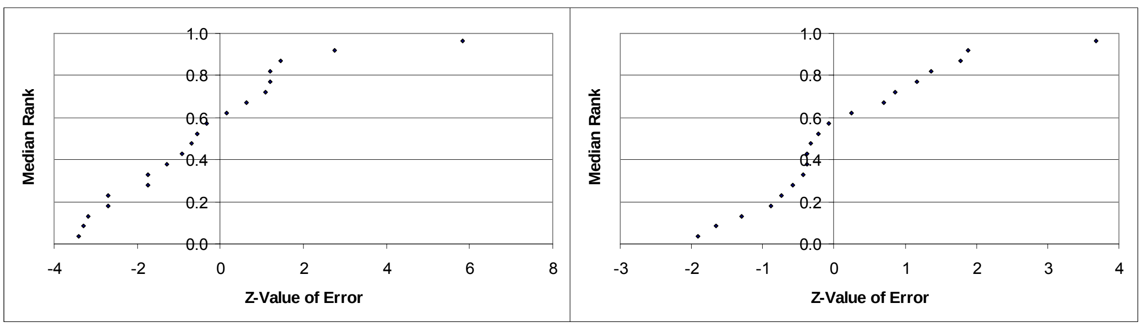
Table 20 Confidence intervals for coefficients of regression lines in Figure 34

Table 21 gives statistics for errors in the ILI measurements. On average the measurements are very good, but the Standard Deviations are quite large. This can be seen in Figure 35.

	Mean	Standard Deviation
Length	95.9%	19.4%
Width	99.3%	26.9%

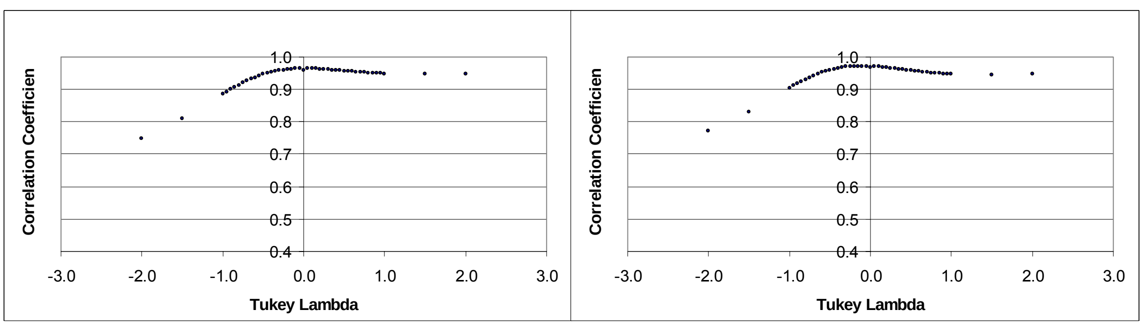
Table 21: Error statistics for vendor data % of dig measurement

Figure 39 and Figure 40 show the Normal Probability and Tukey Lambda Probability Plot Correlation Coefficient Plots for the length and width errors. From these graphs it appears to be reasonable (ignoring the data point on the far right in each graph in Figure 39) to assume both errors are normally distributed.



(a) Dent Length **(b) Dent Width**

Figure 39: Normal Probability plot for errors of length and width data



(a) Dent Length **(b) Dent Width**

Figure 40: Tukey Lambda Probability Plot Correlation Coefficient plot for errors of length and width data

Table 22 gives tolerances for the technology to be “Not Inconsistent” and “Consistent” with respect to length and width measurements for a certainty of 0.8 at a confidence level of 95%. Clearly the tolerances on width measurements are greater than on length measurements.

	Tolerance Levels for 0.8 ≤ p with 95% Confidence Level			
	Successes		Tool Performance	Tolerance (±% mm)
Length	x1	13	Not Inconsistent	17.0
	x2	19	Consistent	31.0
Width	x1	13	Not Inconsistent	21.0
	x2	19	Consistent	43.0

Table 22: Tolerance levels to be Not Inconsistent and Consistent with 0.8 ≤ p at 95% confidence level

There is no completely satisfactory way to compare the depth, length, and width measurements of Technology G, because length and width are measured in mm and depth is measured in %OD. Even if %OD is converted to mm, there is still a significant discrepancy in magnitude between dent measurements and the other measurements. Consequently, we must use indirect methods. We will use two methods that involve regression analysis:

- Compare slopes of regression lines. The closer to the ideal value of 1, the better the performance is. Intercepts could also be used if measurements are of the same magnitude.
- Compare R^2 values. The closer R^2 is to 1, the better the performance is.

Table 23 gives slope, intercept, and R^2 for depth, length, and width measurements made by Technology G. According to these values, Technology G performs poorest on width measurements. There is a trade off in the depth and length measurements. R^2 is larger for depth measurement, indicating less scatter in the measurements. However, the slope of the regression line for length measurements is closer to 1, indicating a better correspondence to the actual measurements. These should be taken only as indications of what might be, not as indications of what is. There is too little data and too much scatter to make any definitive conclusions.

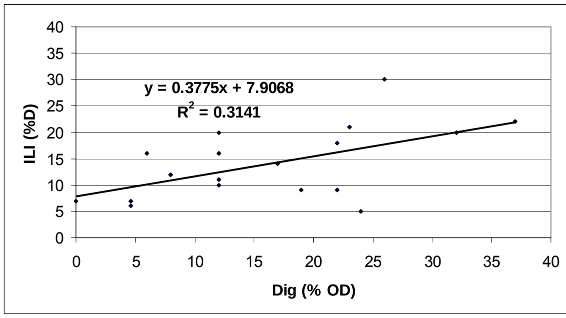
	Depth	Length	Width
R^2	0.90	0.80	0.47
Slope	0.78	0.90	0.68
Intercept	0.36	6.16	30.41

Table 23: R^2 for measurements made by Technology G

B.4 Sizing Performance for Coincident Damage

Sizing of metal loss (corrosion or gouge) features coincident with dents was generally reported to be offered on a best endeavor basis for the current mechanical damage technologies. Data comparing predicted depths and lengths of metal loss within dents from some current technologies was made available by the vendors. Figure 41 and Figure 42 show ILI metal loss depth and Dig measurements for Technology C and from data combined from Technologies H and I. Data pictured in Figure 41 contains two apparent outliers at Dig 7.43 and 31.25% WT. Figure 42 shows the data from Figure 41 with these two outliers removed. For ease of discussion Technology (H and I)’ will denote Technology H and I with the two outliers removed from consideration. A similar reduction in the data set for Technology C is not possible because there are no obvious outliers for Technology C, just scattered data.

Removing outliers needs some discussion. Removing outliers can be a very deceptive practice. Outliers may represent valid data points and indicate quirks in technology or inappropriate measurement procedures. We proceed on the assumption (and hope) that the outliers for Technology H and I are due to influences beyond the technology, and removing them gives a better assessment of Technology H and I. In any case, analysis results for Technology H and I, both with and without outliers, are presented for the reader's comparison.



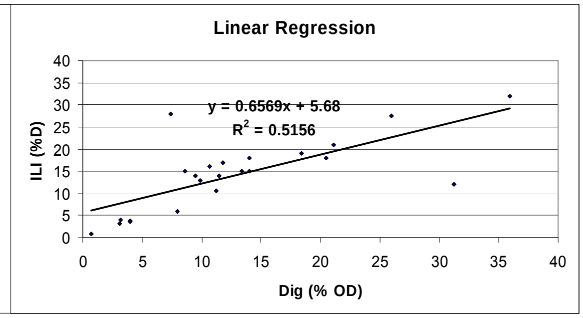


Figure 41: Metal Loss Depth ILI and Dig measurements, Technology C (left), Technologies H and I (right)

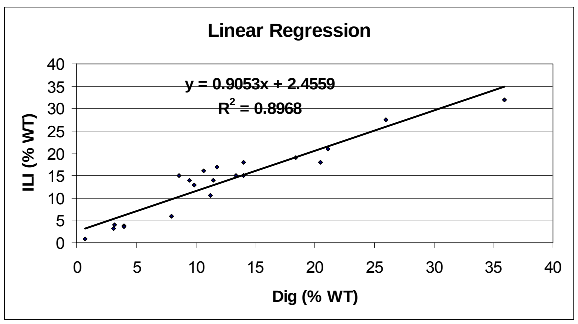


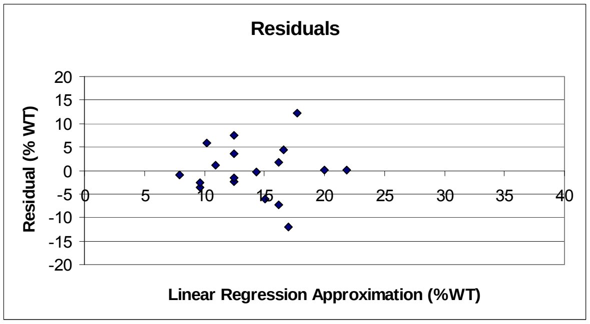
Figure 42: Metal Loss Depth ILI and Dig Measurements, Technologies (H and I)', two outliers removed from the previous Figure

Table 24 gives the Mean, Standard Deviation, and Skewness of the residuals for the metal losses indicated in Figure 41 and Figure 42. As expected, removing two outliers significantly reduced the Standard Deviation for Technology H and I.

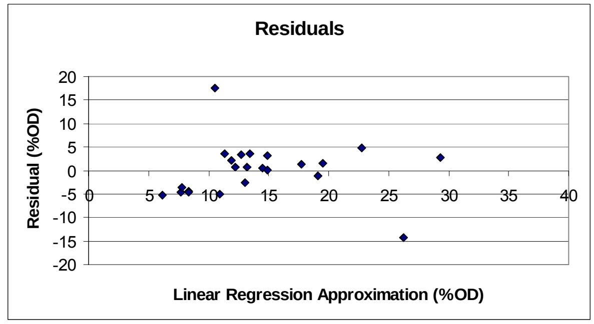
	Count	Mean	Standard Deviation	Skewness
Technology C	18	0.000	5.635	0.082
Technology H and I	23	0.000	5.757	0.551
Technology (H and I)'	21	0.000	2.594	0.376

Table 24: Residual statistics (mm) for length and width data

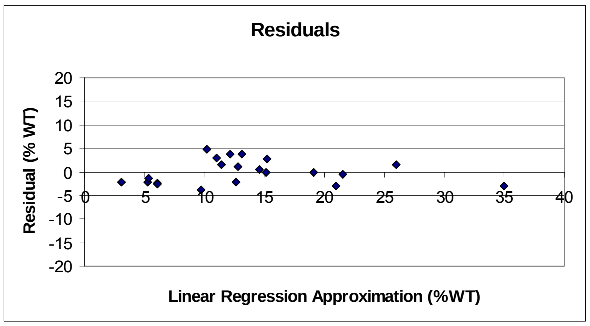
Figure 43 shows the residuals for the three regression analyses. In each case residuals are approximately uniformly distributed horizontally and vertically, except for outliers in Figure 43 (b).



(a) Technology C



(b) Technology H and I

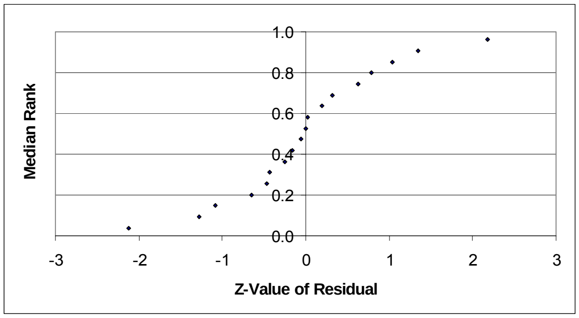


(c) Technology (H and I)'

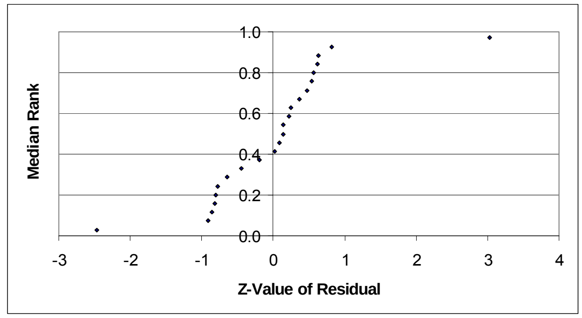
Figure 43: Residuals from regression analyses for metal loss

Figure 44 shows Normal Probability and Tukey Lambda Probability Plot Correlation Coefficient Plots for the residues. Only Technology (H and I)' has a Normal Probability plot that is approximately linear. In addition, it is the only Technology with the maximum Correlation Coefficient occurring at approximately $\lambda = 0.14$. Consequently, it is

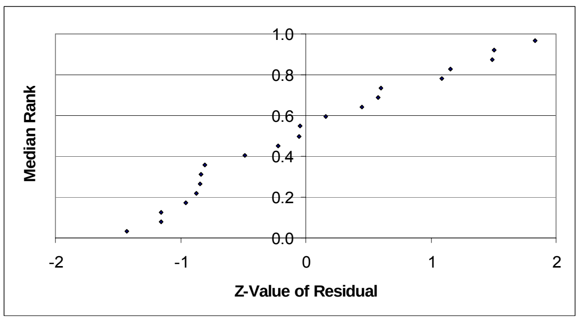
- Reasonable to assume residuals for Technology (H and I)' are normally distributed.
- Not reasonable to assume residuals for the other two technologies are normally distributed.



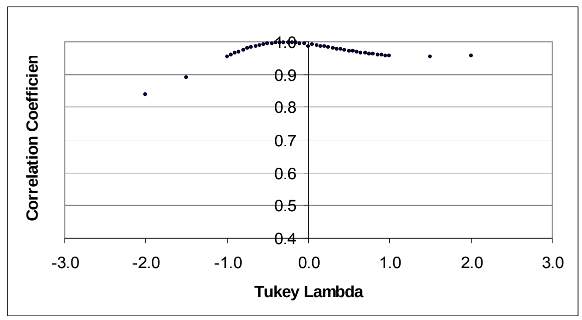
(a) Technology C



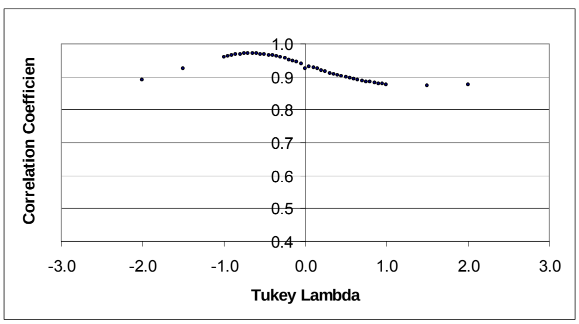
(b) Technology H and I



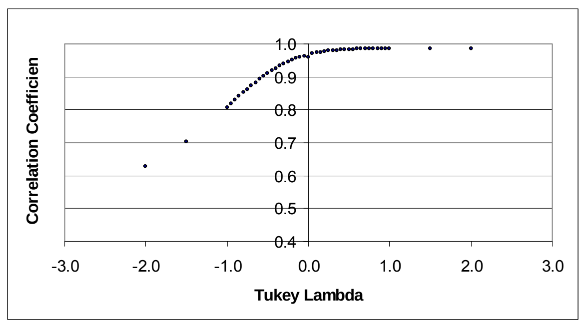
(c) Technology (H and I)'



(d) Technology C



(e) Technology H and I



(f) Technology (H and I)'

Figure 44: Tukey Lambda Probability Plot Correlation Coefficient plots for residuals of dent lengths and widths

Table 25 gives 95% confidence intervals for coefficients of regression lines shown in Figure 41 and Figure 42. None of the Technologies have both 0 in the confidence interval for the intercept and 1 in the confidence interval for the slope. Thus, it is possible to exclude the ideal relationship y = x as being the true relationship between ILI and Dig measurements for all three technologies with 95% confidence level.

			95% Confidence Interval	
		Coefficient	Left End	Right End
Technology C	Intercept	7.907	1.505	14.308
	Slope	0.377	-0.061	0.816
Technology H and I	Intercept	5.680	0.885	10.475
	Slope	0.657	0.341	0.973
Technology (H and I)'	Intercept	2.456	0.214	4.698
	Slope	0.905	0.751	1.060

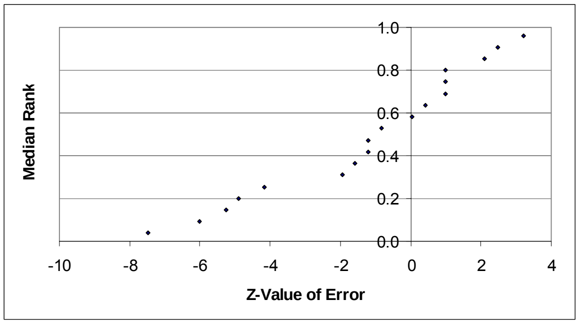
Table 25: Confidence intervals for coefficients of regression lines in Figure 41 and Figure 42

Table 26 gives statistics for errors in the ILI measurements. A Mean less than zero implies that on average the Technology under-predicts metal loss. According to the data received, Technology C under-predicts metal loss with a larger variability in measurements (Standard Deviation) than the other two Technologies. However, this should not be emphasized because of the small data sets and scatter in the data.

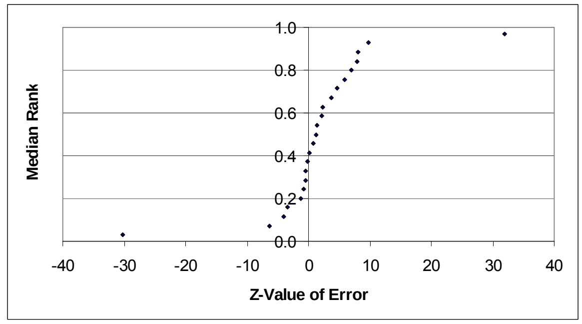
	Mean	Standard Deviation	Skewness
Technology C	-2.233	8.443	-0.537
Technology H and I	1.232	6.540	-0.226
Technology (H and I)’	1.286	2.714	0.186

Table 26: Error statistics (%WT) for metal loss

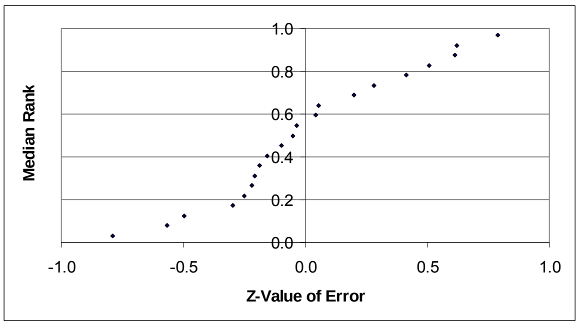
Figure 45 shows the Normal Probability and Tukey Lambda Probability Plot Correlation Coefficient plots for metal loss errors. None of the Technologies appear to have errors that are normally distributed. However, additional data may fill in “bends” in the graphs for Technologies C and (H and I)’, making it reasonable to assume their errors are normally distributed.



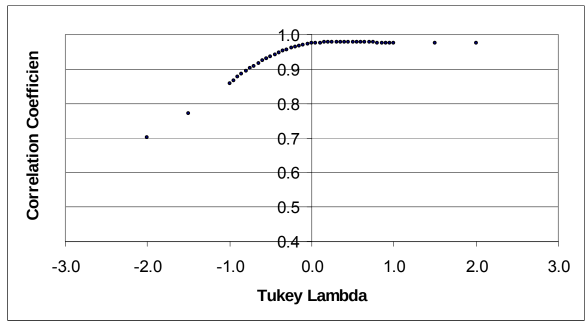
(a) Technology C



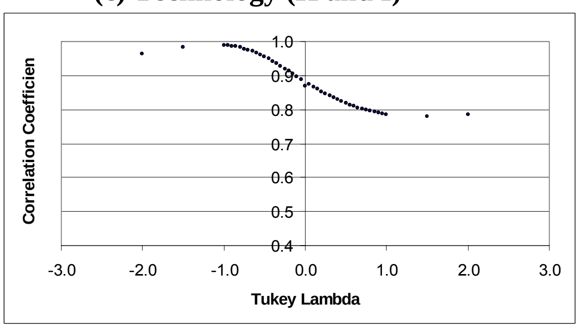
(b) Technology H and I



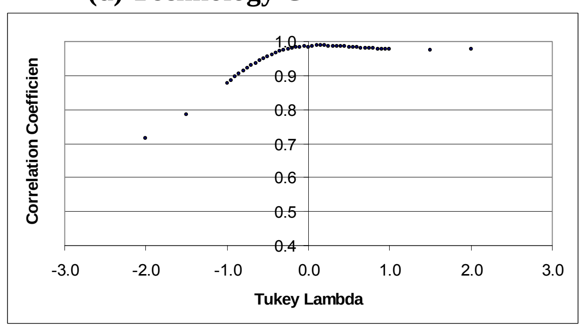
(c) Technology (H and I)’



(d) Technology C



(e) Technology H and I



(f) Technology (H and I)’

Figure 45: Tukey Lambda Probability Plot Correlation Coefficient plots for residuals of dent lengths and widths

Table 27 gives tolerances for the technologies to be “Not Inconsistent” and “Consistent” with respect to metal loss measurements for a certainty of 0.8 at a confidence level of 95%. Technologies (H and I) and (H and I)’ require a much smaller tolerance to be Consistent with a Certainty of 0.8 with 95% Confidence Level than is required by Technology C. This reflects the smaller amount of scatter in Technology (H and I)’, as seen in Figure 41.

	Tolerance Levels for 0.8 ≤ p with 95% Confidence Level			
	Successes		Tool Performance	Tolerance (±% WT)
Technology C	x1	11	Not Inconsistent	7.0
	x2	17	Consistent	15.0
Technology H and I	x1	15	Not Inconsistent	3.1
	x2	21	Consistent	6.4
Technology (H and I)'	x1	14	Not Inconsistent	2.5
	x2	20	Consistent	5.3

Table 27: Tolerance levels to be Not Inconsistent and Consistent with 0.8 ≤ p at 95% confidence level

The vendors reported that for all current technologies, ILI measurement of metal loss, corrosion or gouges feature sizing within deformations is complicated by the tendency for sensors to lift off the pipe wall when traversing deformations. Most of the vendors have employed proprietary mechanical design features to minimize lift off. Using the confidence interval technique the sample for Technology C is consistent with a performance of 80% certainty of ±13% WT depth tolerance at 95% confidence. The performance analysis of the combined data of Technologies H and I indicate a tolerance of about 6.5% WT, or 5.3% WT if outliers are ignored, for the same performance specification. These differences should not be emphasized due to the small data sets used in the analysis.

The metal loss data from Technologies C, H and I together with individual case study examples from Technologies E and J indicate the MFL based technologies have success in detecting metal loss less than 10% wall thickness coincident with plain dents in the range of 2% to 6%. However, the data is insufficient to quantify a detection performance.

C.0 APPENDIX C- MECHANICAL DAMAGE ASSESSMENT CRITERIA AND MODELS

This appendix provides an overview of mechanic damage guidelines, including guidelines that utilize on strain based methods.

C.1 Regulatory & Industry Standard Guidance

Recently, comprehensive reviews of the regulations and industry guidance for managing various forms of mechanical damage have been published by a special Task Group[9,10] and DOT OPS[11]. The first review provided by the special Task Group with the assistance of GTI[9,10] served as a basis for the new criteria in the latest version of ASME B31.8S-2003 for prioritization and repair of mechanical damage in gas pipelines. The second review issued by DOT OPS further provided an overview of the potential effects of dents on the integrity of both gas and liquid pipelines[12], as well as guidance for prioritization and repair as provided by the regulations, industry standards and recommended practices. The regulations and industry standards pertinent to assessment of mechanical damages, including dents, are:

For Gas Pipelines

- 49CFR 192 Transportation of Natural and Other Gas by Pipeline: Minimum Federal Safety Standards
- ASME B31.8 -1995 Gas Transmission and Distribution Systems (1995)
- ASME B31.8 -2003 Gas Transmission and Distribution Systems (2003)

For Liquid Pipelines

- 49CFR 195 Transportation of Hazardous Liquids by Pipeline
- ASME B31.4-1998 Pipeline Transportation Systems for Liquids Hydrocarbons and Other Liquids (1998)
- ASME B31.4-2002 Pipeline Transportation Systems for Liquids Hydrocarbons and Other Liquids (2002)
- API Publication 1156 Effect of Smooth and Rock Dents on Liquid Petroleum Pipelines (1997)
- API Publication 1160 Managing System Integrity for Hazardous Liquids Pipelines (2001)

---

[9] Rosenfeld, M. J.: "Proposed New Guidelines for ASME B13.8 on Assessment of Dents and mechanical damage", GRI Topical Report No. GRI-01/0084, May, 2001

[10] Rosenfeld, M., Pepper, J. and Leewis, K.: "Basis of the New Criteria in ASME B31.8 for Prioritization and Repair of Mechanical Damage", Proceedings of IPC'02, 4th International Pipeline Conference, Paper No. 27122, Sept. 29 to Oct. 3, 2003. Calgary, Canada.

[11] Anon. *"Dent Study - Final Report"*, DOT OPS Integrity Management Program Delivery Order DTRS56-02-D-70036. Report prepared by Michael Baker Jr., Inc., November 2004.

[12] Anon: "Gas Transmission and Distribution Piping System", ASME B31.8 2003, Paragraph 851.4, 2003.

Both reviews indicate that the main parameters used to determine severity of mechanical damage are the

- Nature of mechanical damage (such as plain dent, dent with gauges, cracks, etc)
- Depth of dent, which is expressed in terms of a percentage of the pipe diameter.

In fact, depth is the only dent geometry parameter currently mentioned in 49 CFR 192 and 195 for evaluation of disposition of dents. Table 28 is a summary and comparison of 49 CFR 192 and 195 on the disposition of plan dents and dents associated with other defects for gas and liquid pipelines[12]. The table shows the timescale requirements for operators to take prompt remediation action based on the dent conditions discovered through integrity assessment.

Anomaly	49 CFR 192 Condition	49 CFR 195 Condition
A dent that has any indication of metal loss, cracking or a stress riser.	Immediate	Upper 2/3 of the pipe — Immediate Lower 1/3 of the pipe — 60-day
A dent with a depth greater than 6% of the nominal pipe diameter.	Upper 2/3 of the pipe — One-year[1]. Lower 1/3 of the pipe — Monitored	Upper 2/3 of the pipe — Immediate Lower 1/3 of the pipe — 180-day
A dent with a depth greater than 3% of the nominal pipe diameter on the upper 2/3 of the pipe.	Not defined	60-day
A dent with a depth greater than 2% of the nominal pipe diameter on the upper 2/3 of the pipe.	Not defined	180-day
A dent with a depth greater than 2% of the pipeline's diameter that affects pipe curvature at a girth weld or at a longitudinal seam weld	One-year[1]	180-day
[1] Can be downgraded to a monitored condition providing engineering analyses of the dent demonstrate that critical strain levels are not exceeded. In the case of a dent affecting a weld, the weld properties must also be considered.		

Table 28: Summary of 49 CFR 192 and 49 CFR 195 Regarding Dents]

For the gas pipelines, 49 CFR 192 places dent conditions into three categories: immediate conditions, one-year conditions, and monitored conditions:

- Immediate: A dent with any indication of metal loss, cracking, or a stress riser falls into the immediate repair conditions category. To maintain safety, an operator must, as soon as possible after receiving the ILI report without excavation verification, temporarily reduce operating pressure or shut down the pipeline until all immediate conditions are repaired.
- One year: (i) A smooth dent with depth at least 6% of the nominal diameter and located at upper 2/3 of the pipe or (ii) a dent with depth 2% of the nominal diameter that affects pipe curvature at a girth weld or longitudinal seam weld. The operator must take action to remediate within one year of the discovery of the condition.
- A dent with depth greater than 6% of the nominal diameter located at bottom 1/3 of the pipe falls into the monitored conditions category. An operator does not have to schedule remediation, but must record and monitor the conditions during subsequent risk assessments and integrity assessments for any change that may require remediation.

For liquid pipelines (45 CFR 195.452)*,* the acceptance/rejection criteria are also based on the nature of dents and depth. However, differences can be found in timescale requirements for an operator to take actions to address integrity issues. These are summarized in Table 28 and may be described as follows:

- 49 CFR 195 also places dent conditions into three repair categories in timescale: immediate repair, 60-day condition, and 180-day condition. There is no “monitored condition” for liquid pipelines.
- A dent located at the upper 2/3 of the pipe and associated with a stress riser, such as a crack, or with a depth greater than 6% of the nominal pipe diameter, falls into the immediate repair condition category. However, it becomes a 60 days or 180 days condition if it is located in the lower 1/3 of the pipe. The respective conditions for gas pipelines are immediate, one-year and monitored, respectively.
- A dent located at the upper 2/3 of the pipe with depth greater than 3% or 2% are 60-day or 180 day conditions, respectively, for a liquid pipeline. There are no corresponding requirements for a gas pipeline.
- A dent with depth greater than 2% of the nominal diameter that affects pipe curvature at a girth weld or at a longitudinal seam weld is a 180-day condition, but is a one-year condition for a gas pipeline.

From the above review, it is seen that 49 CFR 192 and 49 CFR 195 are both “depth” based criteria, but differ on the disposition of anomalies in terms of dent assessment. 49 CFR places anomalies into one of three categories: immediate repairs, one-year conditions, and monitored conditions; while 49 CFR defines immediate conditions, 60-days and 180-days conditions.

Both 49 CFR 192 and 49 CFR 195 incorporate ASME codes B31.8 (1995) and B31.4 (1998) by referencing their Repair Procedures (Paragraph 851, B31.8) and “Deposition of Defects” (paragraph 451.6.2, B31.4), respectively, to them. However, the 1995 edition of ASME B31.8 and the 1998 edition of B31.4 have been replaced by new editions, B31.8 (2003) and B31.4 (2002). Even though the new editions of the ASME codes are not currently referenced by 49 CFR 192, the requirements recommended by the new editions are considered to be aligned and compatible with 49 CFR 192.

It is noted, however, that the repair procedures have been largely revised in the new edition of ASME B31.8 2003. The threshold of pipeline operating pressure (expressed as hoop stress level) has been changed to 30% of SMYS from 40% SMYS in the previous B31.8 (1995). B31.8 (2003) (Paragraph 851.4) provides options to use strain criteria and assess corrosion features in dents using remaining strength criteria for corroded dents, see Table 29. For example, a 6% strain in pipe bodies and 4% strain in welds are acceptable for plain and rock dents. This provides operators with safe alternatives, which are particularly important for features located in areas that are difficult to access, such as river-crossings. For strain calculation, B31.8 (2003) provides a non-mandatory formula. Other formulas in the open literature, or derived by a qualified engineer, are also allowed[9,10,11].

There are no similar options in the new addition of B31.4 for liquid pipelines. However, from the point of view of the static behavior of dents, the respective strain based criteria in B31.8 (2003) may be applicable to liquid pipelines.

Table 29 is a summary of the acceptance/rejection criteria of B31.8s (2003) for mechanical damages in gas pipelines.

Feature	Failure Mode	Safe Limit	Mitigation
Plain, unrestrained dents (anywhere on OD)	None	6% OD; deeper up to 6% strain in pipe body, 4% strain in welds	Repair coating if excavated; monitor for corrosion if not excavated
Rock dents	Corrosion, SCC		
External mechanical damage, gouges, scrapes, cracks, or SCC in dents	Rupture Low cycle fatigue	None	Cut out Sleeve Grind out
Internal (pig passage) mechanical damage	None	None	None
Dents affecting ductile girth welds or seam welds	Fatigue	2% OD; deeper per analysis, subject to 4% max strain limit	Cut out Sleeve
Dents on acetylene welds and brittle seams	Brittle fracture	None	
Dents with metal-loss corrosion	Rupture	6% OD and metal loss per corrosion criterion	
Dents with grind repair	Rupture	4% OD and metal loss per grind criterion	

Table 29: ASME 31.8 (2003): Criteria of Mechanical Damage for Gas Pipelines Operating at Hoop Stress Levels At or Above 30% of the Specified Minimum Yield Strengt.

More recently, Dawson et. al.[13] summarized the international code guidance and recommended practices relevant to the assessment of dents in pipelines, see Table 30. It is seen that they are all, except ASME B31.8 (2003), simple depth and dent nature based criteria. For plain dents, most codes adopted a 6% OD criterion with some exceptions, except that PDAM allows 7% OD for unconstrained and 10% OD for constrained plain dents.

[13] Dawson, S. J., Russell A. and Patterson, A.: "Emerging Techniques For Enhanced Assessment and Analysis of Dents", Proceedings of IPC 2006, Paper No. 10264, 6th International Pipeline Conference, September 25-29, 2006, Calgary, Canada.

Published Guidance	Top of Line Dents (8 to 4 o'clock)				Bottom of line Dents (4 to 8 o'clock)			
	Plain Dent	**Dents with Cracks/ Gouges**	**Dents at Welds**	**Dents with Corrosion**	**Plain Dent**	**Dents with Cracks/ Gouges**	**Dents at Welds**	**Dents with Corrosion**
CSA Z662-03 (2005)	Upto 6% OD	Not allowed	Up to 2% for >NPS 12" or upto 6mm for < NPS 12"	As per ASME B31.G up to max depth of 40%wt	As for top of line dents			
AS2885.3 (2001)	Upto 6% OD	Not allowed	Not allowed	Detailed assessment allowed	As for top of line dents			
ASME B31.8* (2003)	Upto 6% OD or unlimited if strain <6%	Not allowed	Upto 2% OD or unlimited if strain <4% for dictule wields (No safe limit for brittle welds)	As per ASME B31.G limits	As for top of line dents			
ASME B31.4 (2004)**	Upto 6% OD	Not allowed	Not allowed	External corrosion<87.5% RWT required for design i.e. <12.5 wt. Internal corrosion as per ASME B31.G	As for top of line dents			
API 1160 (2001)**	Upto 2% OD for 12* NPS (6.35MM <12" NPS)	Not allowed	Not allowed	Not Allowed	Upto 6% OD	Not Allowed	Investigate / mitigate within 6 months	Investigate / mitigate within 6 months
PDAM (2003)	Upto 7% OD (unconstrained) Upto 10% OD(constrained)	Method Provided	Not allowed	Not Allowed	Upto 6% OD (unconstrained)	Method Provided	Not Allowed	Not Allowed
DOT Liquid Rule (Part 195) (2000)**	1. Upto 6% OD (immediate condition) 2. Upto 3% OD for NPS? 12* or>6.35MM for NPS <12" (60 day condition) 3. Upto 2% OD (180 day condition)	Not allowed (immediate condition)	Upto 2% OD (180 day condition)	Not Allowed	Upto 6% OD	Not Allowed (60 day condition)	Upto 2% OD (if > 180 day condition)	Not Allowed (60 day condition)
DOT Gas Rule (Part 192)* (2001)	1. Upto 6% OD for 12" NPS or 12.7mm for <12" NPS (1 year condition) 2.Monitor dents >6% OD with acceptable strain levels	Not allowed (immediate condition)	1. Upto 2% for NPS 12* or up to 6.35mm for <NPS 12" (one year condition) 2. Monitor dents >2% OD with acceptable strain levels	Not Allowed	Monitor Dents >6% OD	Not Allowed	1. Upto 2% for NPS 12* or Upto 6.35mm for <NPS 12" (one year condition) 2. Monitor dents >2% OD with acceptable strain levels	Not Allowed

* Only relevant to gas pipelines

** Only relevant to liquid pipelines

Table 30: Summary of Published Guidance on the Assessment of Dents in Pipelines[13]

It should be noted that the methods and associated criteria described in Table 28, Table 29 and Table 30 are for assessing static dent behavior only[13]. This is adequate for constrained dents because they cannot re-round under internal pressure. API 1156 and its addendum[14] conclude, based on full size fatigue tests, that if dents are constrained, then the time required for the number of cycles necessary for failure is likely to be greater than the normally expected life of the pipeline. Therefore, there is usually little concern for fatigue of constrained plain dents, except in aggressive pressure cycles over a long period of time. These results have some operators of liquids pipelines considering not digging up rock dents at all in order to avoid fatigue in services[9,10]. However, they will have to address long-term corrosion control and monitoring issues at the dents[9,10,14].

An important exception to the above findings is associated with double-centered dents with a flattened or saddle-shaped area between them. The flattened area between the two dent centers is susceptible to pressure cycle fatigue because it is effectively unconstrained and flexes readily in response to pressure cycles. Hence, API 1153 recommends that dents having spacing between centers of one pipe diameter or less should be candidates for excavation and repaired. This rule is primarily applicable to liquid lines because no failure in a gas pipeline due to fatigue has been observed or reported.

For unconstrained plain dents, the major concern is pressure-cycle induced fatigue. This occurs due to high local bending stresses associated with operating pressure fluctuation[13]. An empirical method was proposed by the European Pipeline Research Group (EPRG) for predicting the fatigue life for an unconstrained plain. This method may be the "best" semi-empirical approach in terms of the quality to fit the full scale test data, but could be very conservative[15,16].

Fatigue is a main concern for liquid pipelines, but not for gas pipelines because their operating pressure cycling is not as aggressive as that of liquid lines. PDAM provides a review of plain dent fatigue assessment models[15].

C.2 Strain based Assessment methods

The use of depth-based criteria has served the industry well by demanding that operators at least investigate indications of potentially severe deformations[9,10]. However, the use of depth alone could result in both unnecessary excavations due to many deep dents and ovalities that are not necessarily harmful[9,10] and miss potentially severe dents due to their overall size and sharpness. Since metal damage in principle is related to strain, strain levels derived from the dent profile offer another measure of the severity of a plain dent. This strain-based assessment makes use of the detailed dent profile information obtained from high-resolution geometry tools to calculate

[14] Anon: " Effect of Smooth and Rock Dents on Liquid Petroleum Pipelines", API Publication 1156 (1997).
[15] A. Cosham and P. Hopkins. The Pipeline Defect Assessment Manual (PDAM), a Report to the PDAM Joint Industry Project. May 2003.
[16] Roovers,P., Bood,R., Galli,M., Marewski,U., Steiner,M., and Zaréa,M.; EPRG Methods for Assessing the Tolerance and Resistance of Pipelines to External Damage, Pipeline Technology, Volume II, Proceedings of the Third International Pipeline Technology Conference, Brugge, Belgium, 21-24 May 2000, R. Denys, Ed., Elsevier Science, 2000, pp. 405-425.

local strain and compares the result with engineering judgment based on acceptance criteria proposed by the new edition of ASME 31.8 (2003).

C.2.1 Strain Calculation Equations Recommended by ASME B31.8

So far there are no standard methods for the calculation of dent strains. Appendix R of ASME B31.8 (2003) provides equations to estimate the circumferential bending strain, the longitudinal bending strain and the extensional strain using dent geometry data. However, an error was found in its strain equations (1) and (2)[17], which overestimates the respective bending strains by a factor of 2 (Appendix R of ASME B31.8 2003 omits the "2" in the denominators in equations (1) and (2)). The following are the B31.8 (2003) formulae with the corrected strain expressions of the circumferential and longitudinal strain components proposed by Noronha et. al.[17], and Figure 46 defines the parameters used in the equations:

Circumferential bending strain (ε_1),

$$\varepsilon_1 = \frac{t}{2}\left(\frac{1}{R_o} - \frac{1}{R_1}\right) \tag{1}$$

Longitudinal bending strain (ε_2),

$$\varepsilon_2 = \frac{-t}{2R_2} \tag{2}$$

Extensional strain (ε_3),

$$\varepsilon_3 = \frac{1}{2}\left(\frac{d}{L}\right)^2 \tag{3}$$

Total strain on inside pipe surface (ε_i)

$$\varepsilon_i = \left[\varepsilon_1^2 - \varepsilon_1(\varepsilon_2 + \varepsilon_3) + (\varepsilon_2 + \varepsilon_3)^2\right]^{\frac{1}{2}} \tag{4}$$

Total strain on outside pipe surface (ε_o)

$$\varepsilon_o = \left[\varepsilon_1^2 + \varepsilon_1(-\varepsilon_2 + \varepsilon_3) + (-\varepsilon_2 + \varepsilon_3)^2\right]^{\frac{1}{2}} \tag{5}$$

The overall strain, $\varepsilon_{\max}$, is then defined by equation (6) and used to be compared with an acceptance and rejection criteria[14]:

$$\varepsilon_{\max} = Max\left[\varepsilon_i, \varepsilon_o\right] \tag{6}$$

where

---

[17] Noronha, D. B., Martins, R., Jacob, B. and Souza E.: "The Use of B-Splines in the Assessment of Strain Levels Associated with Plain Dents", Rio Pipeline Conference & Exposition 2005, Paper No. IBP 1245_05, October 2005, Rio de Janeiro, Brazil.

t = Wall thickness
R_o = Initial pipe surface radius
R_1 = Radius of curvature in transverse plane, negative for reentrant dents
R_2 = Radius of curvature in longitudinal plane, negative for reentrant dents
L = Dent length
d = Dent depth

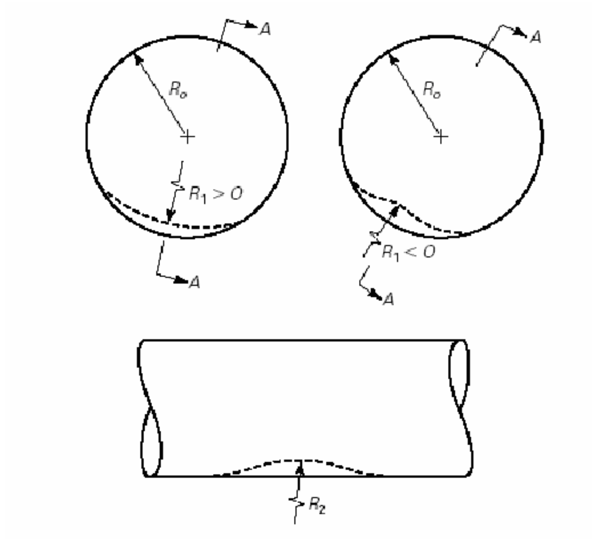


Figure 46: Dent Geometry: Non-Reentrant versus Reentrant Dents

Equations 1-6 can be used to calculate each of the three strain components in circumferential, longitudinal and extensional directions as well as the total strains. The strain calculation method provided in ASME B31.8 (2003) is not mandatory. Other methods published in the open literature can also be used, which is fully up to the operator. Even equations derived by a qualified engineer can be used.

C.2.2 Strain Calculation Methods in the open literature

At present, implementation of strain-based assessment is a challenge due to a lack of efficient and accurate analytical methods for calculating dent strains[17]. Nonetheless some methods have been published in the open literature[18,19,20]. This may be why the method in the ASME B31.8 is not mandatory.

[18] Lukasiewicz, S. A., Czyz, J. A., Sun, C., Adeeb, S. "Calculation od=f Strains in Dents Based on High Resolution In-Line Caliper Survey", IPC2006, Paper 10101, 6th International Pipeline Conference, September 25-29, 2006, Calgary, Canada.
[19] Rosenfeld, M. J.., Porter, P. C., Cox, J. A., "Strain Estimation Using Vetco Deformation Tool Data", ASME 2nd International Pipeline Conference, Calgary, 1998.
[20] Cosham, A. "Assessment Methods for Dents and Other Defects in Pipelines" - A Report to the Pipeline Defect Assessment Manual Joint Industry Project. January 2002.

It is known that a strain in a pipe wall consists of two main components: longitudinal and circumferential. Each of which can be further separated into bending and membrane strains. The membrane strain is constant through the wall, while the bending component changes linearly from the inner to outer surface, see Figure 47. Lukasiewicz et. al.[18] showed that the main difficulty in strain based methods is in determining membrane strains in the dented region. The calculation of the bending component is fairly straightforward as demonstrated in reference 17. The existing techniques are limited to the longitudinal strain, but they are either very inaccurate or inefficient. Lukasiewicz et. al. referred to a method[19] of calculating the longitudinal component of membrane strain in a dented region that is based on the assumption that the circumferential strains are negligible. This assumption is not supported by FEM analysis of actual dents. ASME B31.8 code offers a simplistic approach for estimation of longitudinal membrane strain, i.e., extensional strain given by Equation (3) in the previous section. The accuracy of Equation (3) is, however, extremely poor[18].

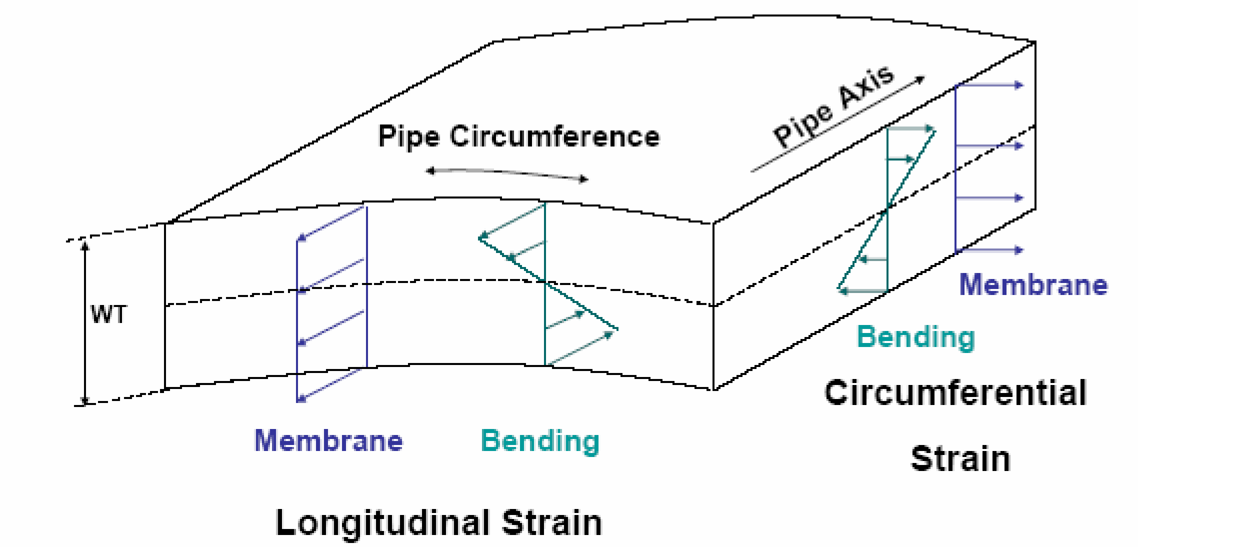


Figure 47: Strain components in a pipe wall.

At present, no exact (analytical) solutions are available for calculation of membrane strains of the dented region in the pipe. Lukasiewicz et. al. proposed a method that uses membrane strain and displacement relationships (differential equations) for large deformations of a cylindrical shell. These equations are solved numerically for the unknown displacements u and v (See Figure 65) using a two dimensional FEA method. The normal displacement w is the only measured parameter, preferably a the high-resolution geometry tool, on the dent geometry. The membrane strain displacement relationships for large deformation of a cylindrical shell are:

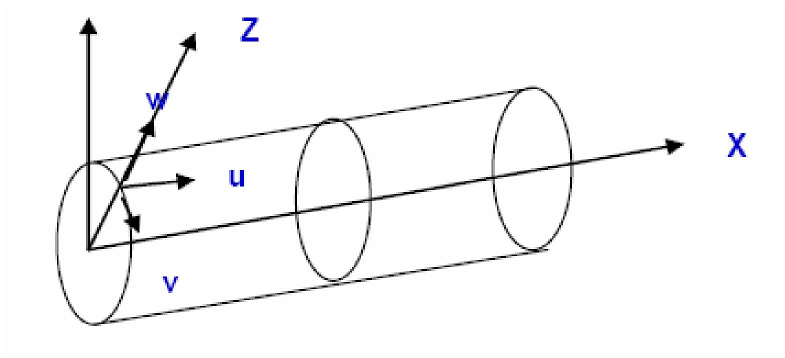


Figure 48: Coordinate system and displacement[18]

$$\varepsilon_x^m = \frac{\partial u}{\partial x} + \frac{1}{2}\left(\frac{\partial w}{\partial x}\right)^2 + \varepsilon_x^o$$

$$\varepsilon_y^m = \frac{\partial u}{\partial y} + \frac{w}{R} + \frac{1}{2}\left(\frac{\partial w}{\partial y}\right)^2 + \varepsilon_y^o \qquad (7)$$

$$\gamma_{yx} = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} + \left[\frac{\partial w}{\partial x}\right]\left[\frac{\partial w}{\partial y}\right]$$

where

> ε_x^m and ε_{yx}^m are the strains in the axial (x) direction, and the circumferential (y) direction
> γ_{xy} is the shear strain in the plane *x, y*
> R is the mean radius of the pipe
> ε_x^o and ε_y^o are the initial strains due to the pressure in the pipe, thermal expansion, etc.

The two dimensional FEA model leads to a set of algebraic equations with the two unknowns u and v at each node point of the mesh on the shell surface. The fundamental equation of FEM is

$$[K]\begin{Bmatrix} u \\ v \\ w \end{Bmatrix} = \{F\} \qquad (8)$$

where $[K]$ is the stiffness matrix of the system and $\{F\}$ is the vector nodal forces.

If the displacement w is known the above equation can be transformed into

$$\left[K_m\right]\begin{Bmatrix} u \\ v \end{Bmatrix} = \{F_M\} \qquad (9)$$

where $\left[K_m\right]$ is the stiffness matrix for a membrane shell problem and $\{F_M\}$is the modified vector of equivalent nodal force. Once Equation (9) is solved for u and v, the membrane strains are then calculated from Equation (3). These strains can be superimposed with the bending components, thereby producing maximum values of strain in the axial and circumferential directions:

$$\varepsilon_x = \varepsilon_x^m \pm \bar{\varepsilon}_x^b$$

$$\varepsilon_y = \varepsilon_y^m \pm \bar{\varepsilon}_y^b$$

where the positive and negative signs refer to the outer and inner wall surface, respectively.

The maximum equivalent strain (both on the inner and outer surface) in the dented area of the pipe is then given by

$$\varepsilon_{eq} = \frac{2}{\sqrt{3}}\sqrt{\varepsilon_x^2 + \varepsilon_x\varepsilon_y + \varepsilon_y^2} \qquad (10)$$

and can be compared against the strain acceptance criteria proposed by ASME B3.8.

This combined mathematical algorithm and two-dimension FEA model allows the calculation of all the strain components based on the measurement of the radial deformation of a pipe with a high-resolution geometry in-line tool. Case studies have been conducted and compared with three-dimensional shell model results. The proposed method is likely to provide a practical tool for assessment of strain in dents. For the best accuracy, the caliper data collected by high-resolution in-line inspection tools with narrow sensors not exceeding 1” is required.

C.2.3 Simulation of Dent Profile – Filtering and Interpolation

Implementation of a strain based assessment, either using ASME B31.8 Appendix R, or other methods, requires accurate knowledge of the dent profile, as well as curvatures in both longitudinal and circumferential directions. The radial displacements obtained from in-line caliper tools are usually affected by the measurement errors, as well as pipe wall surface irregularities, that must be minimized before using them in strain calculations[18]. Small errors in radial (w) measurements introduce large curvature errors. Therefore, appropriately filtering and interpolating of the caliper data are necessary to obtain an accurate geometric description of a dent on which an accurate calculation of the bending strains is based. The adjusted caliper data also produces a more realistic dent shape that allows for the calculation of more accurate membrane strains, although they are not as susceptible to the noise as the bending strains.

There are many methods, all based on piece-wise interpolation techniques, that can be used for filtering. Rosenfeld et al. employed a piece-wise Bessel cubic interpolation method to characterize the geometry of the dent contour and the osculating circle technique to estimate radii of curvature[19]. Noronha et al. made use of fourth-order B-spline curves[17] to approximate the dent profile in both longitudinal and circumferential directions. Since fourth-order B-splines have second-order continuity, radii of curvature can be calculated at any location directly from a classical equation of curvature[21]. Dawson et al.[13] used piece-wise quadratic curves fitted by least squares to simulate the dent profile. More than three points are used for each quadratic equation fit in order to reduce the effect of measurement error. The radii of curvature were also calculated using a classical equation. The filtering algorithm used by Lukasiewicz et al. uses splines that are available in the commercial software MATLAB[22]. There are no radii of curvatures involved in the method of Lukasiewicz et al. for strain calculation.

The accuracy of a piece-wise interpolation techniques for simulation of an actual dent profile and strain calculation depends on three factors:

1) the quality of raw data reported by the geometry tool
2) the algorithm for interpolation
3) the method for curvature calculation

The first factor is intrinsic. The filtering can reduce or minimize, but cannot eliminate, measurement errors. The degree of accuracy achieved by piece-wise smoothing cannot be beyond the quality and resolution of the caliper data. Figure 49 and Figure 50 present geometry interpolations and their associated distributions of circumferential bending strain along the transverse section of the pipe near the dent as a function of the number of sensors, respectively, for a 12.75" OD, 0.188" wall and 12% constrained dent.

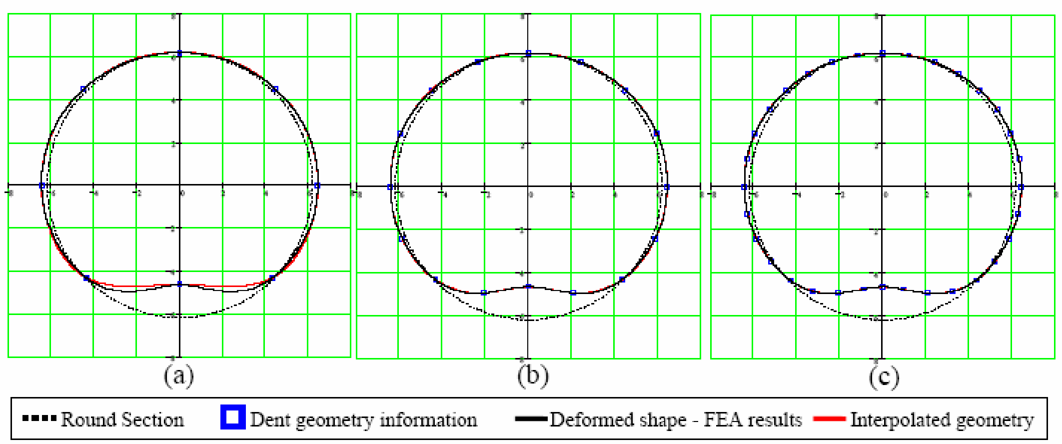


Figure 49: Dent geometry interpolation for an increasing number of sensors, showing the influence of sensor number on the accuracy of the interpolated geometry (a) 6 sensors (b) 16 sensors and (3) 32 sensors[50]

[21] Weisstein, E. W.: "Curvature." from MathWorld – A Wolfram Web Resource. http://mathworld.wolfram.com/curvature.html

[22] MATLAB Spline Toolbox. The MethWorks Inc.

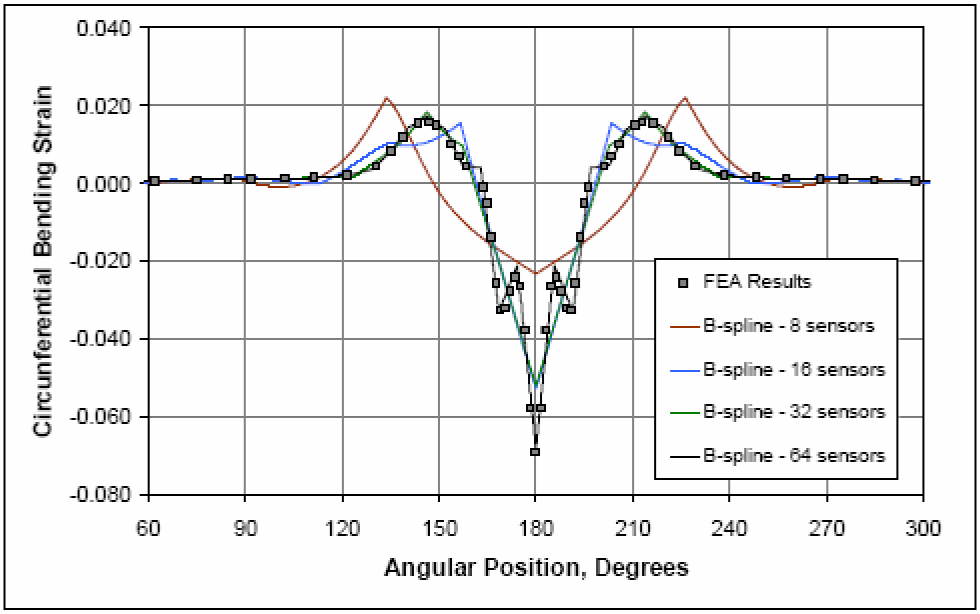


Figure 50: Circumferential bending strain calculated from the interpolated dent geometry and FEA[50].

Figure 67 shows that accuracy of the calculated strain increases with increasing sensor number. The most accurate result relative to FEA analysis is achieved when the number of sensors increases to 64.

It is seen from these figures that, in the case of geometry interpolation, the dent geometry is well represented by 16 sensors, while 64 sensors are needed to achieve an accurate estimation for circumferential strain.

In addition to the influence of the sensor number of the geometry tools (resolution), the algorithm for interpolation and the method for curvature calculation (factors of 2 and 3) can also have an impact on the accuracy of the calculated strains. Figure 51 shows a comparison of the estimated circumferential bending strain between the fourth order B-spline[17] and Bessel Cubic-Osculating Circle[19] methods. The overall agreement between these two methods is good, however, a 15% difference in peak bending strain prediction is apparent.

The influence of the algorithm and simulation model on the estimated accuracy in dent profile and strain estimation is extrinsic. With today's mathematical tools and commercially available software, this influence can be readily minimized.

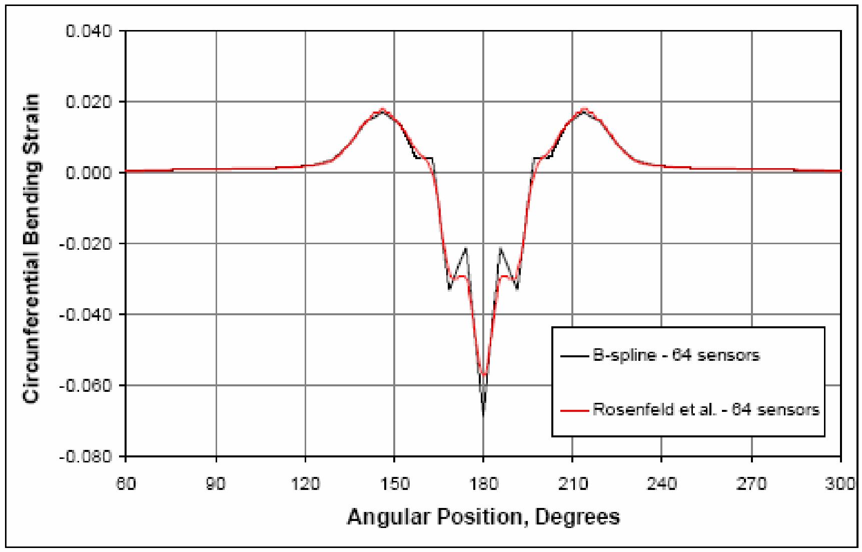


Figure 51: A comparison of the estimated circumferential bending strain between the fourth order B-spline and Bessel Cubic interpolation and Osculating Circle methods.

C.3 Implications: Integrity Assessment and ILI Inspection

As discussed above, the only geometric parameter for mechanical damage in the regulations and industry codes is depth, except for geometrically determined "strain" as an alternative option in the new edition of ASME B31.8 (2003). The effectiveness of an integrity assessment in terms of safety and cost is highly dependent on the accuracy of the ILI reported depth. Therefore, knowledge and understanding of current ILI technologies for depth sizing are critical. In addition to sizing, the probability of detection and discrimination of mechanical damage features by the ILI technology are equally important because conditions for mitigation also depend on the nature of features, in particular, for immediate conditions. The purpose of this project is to evaluate current ILI technologies in terms of sizing accuracy, probability of detection and discrimination, thereby increasing the quality of integrity assessments.

There are two levels of knowledge: general and specific. A general knowledge of in-line inspection technologies for mechanical damage is based on vendor specifications and other vendor supplied information. A general knowledge of ILI technologies is used by operators primarily to select appropriate ILI inspection tools and identify "immediate" conditions, which require mitigative action without excavation verification. Therefore, general specification-based knowledge should be reliable in order for operators to select appropriate in-line inspection technology and minimize risk. This should include knowledge of what the threats are, what assessments are required by regulation and what information is required to make these assessments. This information may include the following

- Probability of detection for
  - Plain dent
  - Dent with corrosion
  - Dent with gouge
  - Dent with crack

    - Dent with other defects
- Probability of discrimination
    - Discrimination between corrosion and gouge
    - Discrimination between metal loss and cracks
- Tolerance of defect sizing, tool certainty and confidence level
    - Dent depth – the most important parameter for depth-based assessment
    - Dent length and width – the necessary parameters for strain based assessment
    - Depth and length of corrosion features in the dents

A specific knowledge of in-line inspection technologies for mechanical damage is established using excavation verification data and tool performance analysis for a specific pipeline after an ILI run. This analysis quantifies the actual performance of the ILI tool for a particular inspection. The results serve as a basis for making engineering decisions on how to use the validated data for integrity assessment. The assessment based on general knowledge should be refined using the specific knowledge of the established tool performance. A re-inspection interval will be based on knowledge of the tool's validated performance.

In addition, ASME B31.8 (2003) provides an option to assess corrosion features in dents using remaining strength criteria for corroded dents. Validation of sizing accuracy of the employed ILI technology should also be performed as part of the entire validation process.

A high-resolution caliper technology is required for strain-based assessment,. This is because the accuracy and reliability of the calculated strains are very sensitive to the tool reported dent geometry. The minimum requirement for sensor density for a reliable strain assessment has not been established yet. The influence of the measurement errors, particularly random noise, on the displacement could be minimized by piece-wise smoothing technology, but this has not been studied in detail. It may be required to develop a specific validation protocol for high-resolution ILI technologies.