Proceedings of the 2016 11th International Pipeline Conference
IPC2016
September 26 – 30, 2016, Calgary, Alberta, Canada

IPC2016-64284

NEW CLASSIFICATION APPROACH FOR DENTS WITH METAL LOSS AND CORROSION ALONG THE SEAM WELD

J. Bruce Nestleroth
Kiefner
Columbus, Ohio, USA

James Simek
TD Williamson
Salt Lake City, Utah, USA

Jed Ludlow
Salt Lake City, Utah, USA

ABSTRACT

The ability to characterize metal loss and gouging associated with dents and the identification of corrosion type near the longitudinal seam are two of the remaining obstacles with in-line inspection (ILI) integrity assessment of metal loss defects. The difficulty with denting is that secondary features of corrosion and gouging present very different safety and serviceability scenarios; corrosion in a dent is often not very severe while metal loss caused by gouging can be quite severe. Selective seam weld corrosion (SSWC) along older low frequency electric resistance welding (ERW) seams also presents two different integrity scenarios; the ILI tool must differentiate the more serious SSWC condition from the less severe conventional corrosion which just happens to be near a low frequency ERW seam. Both of these cases involve identification difficulties that require improved classification of the anomalies by ILI to enhance pipeline safety.

In this paper, two new classifiers are presented for magnetic flux leakage (MFL) tools since this rugged technology is commonly used by pipeline operators for integrity assessments. The new classifier that distinguishes dents with gouges from dents with corrosion or smooth dents uses a high and low magnetization level approach combined with a new method for analyzing the signals. In this classifier, detection of any gouge signal is paramount; the conservatism of the classifier ensures reliable identification of gouges can be achieved. In addition to the high and low field data, the classifier uses the number of distinct metal loss signatures at the dent, the estimated maximum metal loss depth, and the location of metal loss signatures relative to dent profile (e.g. Apex, Shoulder).

The new classifier that distinguishes SSWC from corrosion near the longitudinal weld uses two orientations of the magnetic field, the traditional axial field and a helical magnetic field. In this classifier, detection of any long narrow metal loss is paramount; the conservatism of the classifier ensures that high identification of SSWC can be achieved. The relative amplitude of the corrosion signal for the two magnetization directions is an important characteristic, along with length and width measures of the corrosion features.

These models were developed using ILI data from pipeline anomalies identified during actual inspections. Inspection measurements from excavations as well as pipe removed from service for lab analysis and pressure testing were used to confirm the results.

INTRODUCTION

In the last decade, the detection capability for ILI tools has improved, enabling the reporting of smaller corrosion and shallower dents. Also many tools are better at detecting the seam weld in well-trimmed ERW pipe. However, the reporting of smaller corrosion that is coincident with dents or the long seam has caused an increase in excavations per the regulations in many countries. The goal of the regulations is to ensure that mechanical damage in dents and selective corrosion of the long seam, both potentially injurious anomalies, are always detected.

Keeping with the spirit of the regulations, the goal of the work presented herein is to build classifiers that combine the measurements from multiple sensing systems to detect mechanical damage in dents and selective corrosion of the long seam, while dismissing many of the smaller corrosion features that do not impact pipeline performance. The classifiers are designed to be conservative, meaning some non-injurious corrosion anomalies are designated for excavation so that the likelihood of catching all potentially injurious mechanical damage and selective seam corrosion is greater. This paper discusses the development and verification of these classifiers.

Copyright © 2016 by ASME
The United States Government retains, and by accepting the article for publication, the publisher acknowledges that the United States Government retains, a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for United States Government purposes.

NOMENCLATURE

A – MFL Signal Amplitude
B – Background MFL Amplitude
$\mathbb{A}$ - Normalized Amplitude
w –MFL Signal Width
$\mathbb{w}$ – Normalized MFL Signal Width
X_{ILI} – X is any of the above variables, the subscript ILI indicates data type

- mfl – axial MFL
- lfm – low field axial MFL
- smfl – helical MFL

t – wall thickness
F – general function
x – general variable
β_n – Weighting constants n = 0,1,2,3,4

BACKGROUND

Dent Assessment by ILI

Accepted codes, standards, and governmental regulations address the secondary features of corrosion and gouging in dents. As an example, per 49 CFR 192.933 (d) natural gas pipeline operators in the United States "…must treat the following conditions as immediate repair conditions: (ii) A dent that has any indication of metal loss, cracking or a stress riser." A similar requirement can be found for liquid pipelines in 49 CFR 195.452(h)(4) "(i) immediate repair conditions (C) dents on the top of the pipeline (. . .) with any indicated metal loss." Many international codes have similar repair requirements.

The regulations were shaped by three incidents in the 1990s in which three pipeline failures were caused by dents with gouging. These failures highlighted the need to detect gouging associated with mechanical damage. Gouges, as defined by API 1160, are "the elongated grooves or cavities usually caused by mechanical removal or smearing of metal". The metal loss associated with a gouge can be detected by conventional MFL tools. For two of the three incidents, the pipeline had been inspected by MFL pigs, and post failure analysis of the data showed a metal loss anomaly present in the dent.

Dents, which are primarily detected by caliper tools, often are caused by mechanical damage. Therefore, dents with associated metal loss could have gouging. But, more often than not, the metal loss in a dent is caused by corrosion at damaged or disbonded coating. Unfortunately, traditional applications of MFL tools cannot reliably discern gouging from simple metal loss due to corrosion.

Dual field magnetization approach. An alternative MFL approach that was developed with DOT PHMSA funding, dual field MFL, can distinguish dents with gouging from dents with metal loss. This technology has been commercialized in different forms by ILI vendors. Common MFL-based inspection tools for detecting corrosion use high magnetic fields to saturate the pipeline material [1]. Such high magnetic field-based magnetizers help suppress noise due to local stress variations and changes in the microstructure of the metal [2]. At metal-loss defects, such as those caused by corrosion, an increased amount of magnetic flux attempts to flow through the remaining material, however some flux leaks from the pipeline wall due to saturation of the remaining material. In addition, in magnetically saturated materials, an increase in flux causes the flux-carrying capability (permeability) to decrease [3] resulting in additional leakage. The dual effect of increased magnetic flux and decreased flux-carrying capacity results in significant flux leakage at metal loss defects.

Stress and material variations can also change the flux-carrying capacity of the pipe [4-5]. A local decrease in flux-carrying capacity causes leakage similar to that resulting from metal-loss defects. A local increase in flux-carrying capacity causes a decrease in flux leakage relative to the nominal, magnetic field level. For example, for tensile stresses, the overall flux levels in the pipeline increase. For compressive stresses, such as cold-worked areas, the flux levels decrease. The flux density variations between tensile stresses and compressive stresses are small for magnetic field levels greater than about 6,400 A/m (80 Oersted) and particularly for magnetic field levels greater than about 9,600 A/m (120 Oersted). These general values may vary with pipeline wall thickness, chemical composition, grain structure, and fabrication methods [6]. As discussed previously, most MFL-based apparatuses for corrosion are designed to operate above these levels to reduce signals due to stress, typically considered noise. To detect changes in stress and cold working in the pipe wall, however, the magnetic field must be at lower, unsaturated levels, typically between 4,000 and 5,600 A/m (50 and 70 Oersted). Unfortunately, field levels in this range can produce results that are difficult to interpret because they can be affected by corrosion anomalies, stresses, and changes in material composition. A low-field-strength MFL based pig can be used to detect stresses and material variations using fields in the range of 4,000 and 5,600 A/m (50 to 70 Oersted); however, corrosion anomalies are also detected at this field level and assessment of these anomalies would be inaccurate as demonstrated by early MFL pigs.

Thus, using two magnetic field levels and methods to process both signals can improve the detection and assessment of pipeline anomalies [7-11]. The high magnetic field employed in most inspection tools detects and sizes metal loss such as corrosion. A second low magnetic field must also be applied to detect the metallurgical changes caused by mechanical damage (e.g., from excavation equipment). To apply a two magnetization approach, it is possible (but rarely practical) to send more than one pig separately through the pipeline. A single pig having two separate sets of magnetizers, while also technically feasible, results in a tool length that can be prohibitive for some pipeline systems.

Copyright © 2016 by ASME

Seam Corrosion Assessment by ILI

Similar to dents, accepted codes, standards, and governmental regulations address corrosion near the longitudinal seam weld. For example, per 49 CFR 195.452(h)(4)(iii)(H) liquid pipeline operators in the United States must address corrosion of or along a longitudinal seam weld in 180-days. Many international codes have similar repair requirements. The primary concern of these regulations is selective seam weld corrosion (SSWC), which accounts for about 1.0 percent of the failure incidents reported in the United States [12]. SSWC is a localized corrosion attack along the bond line of low-frequency ERW and electric flash welding (EFW) piping that leads to the development of a narrow groove. Since the weld bond line chemistry may be more susceptible to corrosion processes, the presence of corrosion at the weld is considered an indicator of potential SSWC. This is particularly true if the pipeline has the following conditions present:

- Exposure to corrosive conditions due to poor or absent coating;
- Ineffective cathodic protection; and
- The presence of non-metallic inclusions in the weld bond line region (e.g., contaminants present during the manufacturing process).

SSWC is generally not considered to be a concern for pipe manufactured subsequent to 1970 due to the use of cleaner steels with greatly reduced sulfur content and the replacement of low frequency welding equipment with high frequency equipment in the manufacturing process.

Pipeline companies use in-line inspection (ILI) technology to detect and assess the potential impact of various corrosion threats. SSWC is detectable with both liquid coupled angle beam ultrasonic crack detection (UTCD) and electromagnetic acoustic transducer (EMAT) ultrasonic methods when anomalies are isolated. However, when surface corrosion surrounds the SSWC, detection and sizing can be impacted. For UTCD methods, the adjacent corrosion may change the reflection angle, so energy that reflects from an anomaly may not return to the sending sensor for detection, identification and sizing. EMAT methods detect both the general corrosion and selective corrosion; however, the long wave length nature of these tools make it difficult to always distinguish between the two corrosion types.

Circumferential MFL methods are most applicable for SSWC since this feature has width. The signal amplitude is a function of crack width combined with length and depth, so sizing of SSWC is not as accurate as other corrosion anomalies

In summary, the regulatory requirement of investigating any corrosion of or along a longitudinal seam weld is the result of the difficulty that ILI tools have with detecting and identifying SSWC. The industry and public would benefit if ILI tools could reliably differentiate the more serious SSWC condition from the less severe conventional corrosion near a low frequency ERW seam. These ILI tool limitations require improved anomaly classification algorithms to better identify dents with gouges and SSWC.

APPROACH

Five different ILI measurements made on a single tool were used to develop the discrimination algorithms:

- Conventional axial MFL technology (high field)
- ID versus OD discrimination sensors
- Deformation measurements
- Helical MFL technology
- Reduced field axial MFL measurements (low field)

The advantage of using data from a single tool run is to avoid the additional step of aligning the data between runs. The goal of this work was to develop algorithms that combine all five methods to differentiate severe mechanical damage from less severe metal loss in dents and separate SSWC from general corrosion near the long seam.

Previous algorithms for dents with metal loss and gouging attempted to quantify mechanical damage severity. The algorithms established five levels of importance from benign to severe; however this approach was not implemented by industry nor accepted by regulators. The more conservative approach presented herein considers all mechanical damage anomalies to be severe. As the process evolves and improved mechanical damage assessment methods become robust, the assessment of mechanical damage from ILI data may become part of the process. For the SSWC challenge, combining helical and axial MFL should allow metal loss on an ERW longitudinal seam to be identified and better characterized by determining the exact location with respect to the seam and improving sizing algorithms for corrosion associated with the seam.

Use of data from actual pipeline anomalies collected at excavation sites was a key part in this development. Pipeline companies provided results from ILI tool runs and in-the-ditch field assessment data to correlate with ILI data. Pipes were removed from service and burst tests were conducted to determine the remaining strength for selected anomalies. For the dent discrimination classifier, field samples were augmented with some carefully manufactured mechanical damage features.

ALGORITHM DEVELOPMENT

Algorithm for Dent Discrimination

Previous work on dents with gouges and the use of two MFL field levels for prioritization demonstrated a process for identifying signal characteristics that were indicative of cold work and metal removal during the gouge creation [13]. The gouges used in that study were severe, both in depth and length, and obvious “gouge” and “plow” signatures were discernable in the low-field MFL (LFM) signatures after a suitable signal

Copyright © 2016 by ASME

decoupling algorithm was applied. To meet the goal of detecting all mechanical damage, the gouges in the current study were generally shallow and much less severe. Most ILI tools in use today are quite capable of detecting very shallow metal loss. That said, the resulting MFL and LFM signatures were very small and subtle. Signal decoupling was attempted early on in this study but was deemed unreliable because of the strong sensitivity to alignment between the two MFL and LFM data sets. Shifts in sample alignment in both axial and circumferential direction were often as significant as the decoupled signal. Consequently a new approach was required.

Metal Loss Types. The field measurements describe five distinct metal loss types (and associated number of occurrences):

1. Corrosion (24)
2. Gouge (50)
3. Mill Grinding (1)
4. Puddle Weld (3)
5. No Metal Loss (14)

It was determined that the puddle weld ILI signatures would be sufficiently recognizable as something other than true metal loss in future inspections, and these samples were eliminated from the list, removing one type entirely. A single mill grind sample is insufficient to adequately train a classifier, so this sample was also removed from consideration, leaving 88 out of the original 92 samples, and three of the original five types. The remaining three types are labeled as follows in the discussion below:

- Corrosion
- Gouge
- None

Feature Engineering. The selection of type labels addressed the problem of choosing meaningful input features. LFM can identify locations where magnetic permeability has changed as a result of cold work. At a gouge, enough through-the-thickness cold work will have to take place in order for a LFM signature to be affected significantly. The LFM also has a response associated with bulk metal loss, so a method for separating the LFM response due to permeability versus the LFM response due to removed metal when both LFM and MFL responses are already quite small was needed.

As discussed earlier, full signal decoupling was deemed unreliable for this study because of the small signal sizes. However, it was decided that the ratio of MFL response to LFM response amplitudes might be a suitable stand-in as an indicator of how much the observed LFM signal is due to bulk metal loss versus magnetic permeability change. First, the signal amplitude in the LFM and MFL signals are normalized by their local background level to account for local magnetization strength.

$$\mathbb{A}_{lfm} = \frac{A_{lfm}}{B_{lfm}} \qquad \mathbb{A}_{mfl} = \frac{A_{mfl}}{B_{mfl}} \tag{1}$$

Then an overall amplitude ratio is computed using these normalized amplitudes as

$$\mathbb{A}_{ratio} = \frac{\mathbb{A}_{mfl}}{\mathbb{A}_{lfm}} \tag{2}$$

This amplitude ratio is used as an input to the classifier. In practice, whenever $\mathbb{A}_{lfm}$ is zero, $\mathbb{A}_{mfl}$ is also zero. Under these circumstances, $\mathbb{A}_{ratio}$ is set to zero by definition. Other feature inputs from the ILI record are:

- Number of distinct metal loss signatures at the dent (ML Sig Count)
- Estimated maximum metal loss depth (ML Depth %)
- Location of metal loss signatures relative to dent profile (Apex, Shoulder, Both, or None)

The location data are first converted to binary indicator variables before classifier training.

Just as important in this study are the features which were eliminated from consideration. Orientation with respect to top-of-pipe was not included in the feature set since the manufactured dents do not have a meaningful orientation, and those dents constitute a significant fraction of the samples in this study. Dent depth was also excluded since dent depth was strongly correlated with real versus manufactured anomalies. The dents from operating pipelines were all very shallow. The dents in the manufactured set were generally deeper. Dent depth was removed to avoid the potential of the classifier exploiting this somewhat artificial correlation that happens to be present in this data set but would not be representative of the general mechanical damage population in actual use.

The classifier input features are plotted as pairs against each other in Figure 1. The separation between the classes is quite subtle, and simple linear decision boundaries would be insufficient to provide good classifier performance. A more advanced classification model is appropriate.

Model Selection. A random forest classifier was selected because of its ability to model a complex decision boundary while still avoiding the tendency to over fit the training data.

Copyright © 2016 by ASME

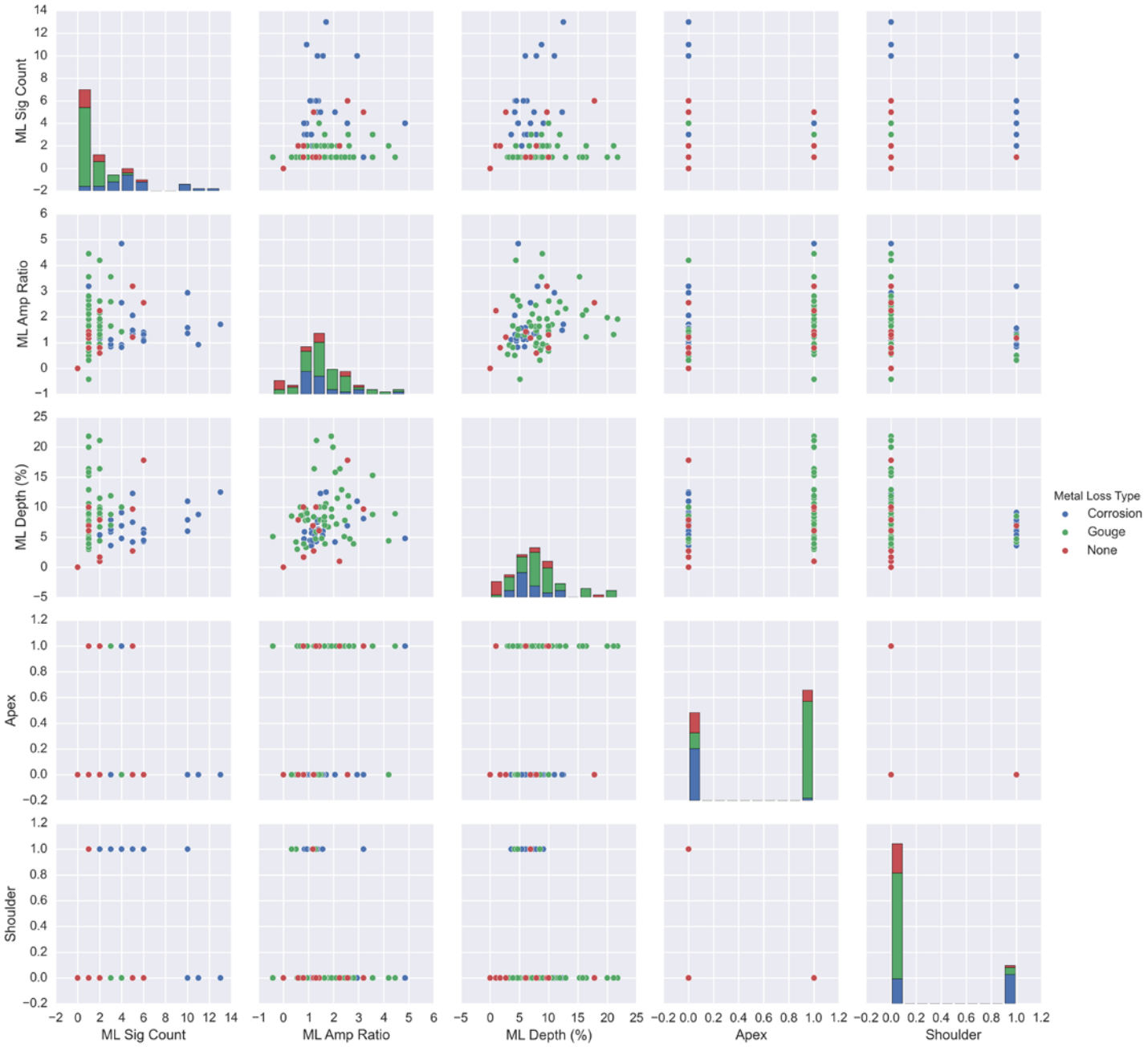

FIGURE 1. PAIR PLOT OF SOME OF THE CLASSIFIER INPUTS FOR THE THREE METAL LOSS TYPES.

Copyright © 2016 by ASME

Random forests are created by training many binary decision trees in a special fashion and then averaging the predictions of these trees. A full discussion can be found in the literature and is beyond the scope of this paper [14]. When trained with a technique called "bootstrap aggregation," random forest classifiers generate their own estimates of generalization accuracy based on samples that are excluded from the training of individual trees in the ensemble. These samples are sometimes referred to as "out-of-bag" (OOB) samples. We take advantage of this feature in the performance metrics that follow.

Model Training. A random forest classifier with 100 individual binary tree estimators was trained on the 88 observations with the goal of predicting the three metal loss types: Corrosion, Gouge, and None. Increasing the number of trees beyond 100 produced no obvious improvement in classifier performance.

The classifier itself estimates the relative importance of the various input features as part of the training process. Table 1 shows relative input feature importance for the metal loss classifier. Metal loss signal count, amplitude ratio, and estimated metal loss depth appear to be of nearly equal importance while metal loss location indicator variables are somewhat less important unless the metal loss is located at the apex of the dent.

TABLE 1. CLASSIFIER INPUT FEATURE IMPORTANCE

Input Feature	Relative Importance
ML Sig Count	0.25
ML Amp Ratio	0.22
ML Depth (%)	0.21
Apex	0.17
Both	0.05
None	0.02
Shoulder	0.08

Custom Decision Function. In a multi-class problem such as this one, the classifier produces a probability that each observation belongs to each of the types. With type probabilities in hand, the most commonly used decision function is to assign each observation to the type that has the highest probability. However, in this study we wish to influence the decision function to err on the side of assigning observations to the "Gouge" type to be conservative. As a result, we adopt a decision function that assigns each observation to whichever type has the highest probability except where the probability of "Gouge" is above a certain threshold in which case "Gouge" is assigned.

As a brief illustrative example, consider that the classifier might return the following type probabilities after evaluating the feature inputs for a given sample:

- Corrosion: 0.45
- Gouge: 0.35
- None: 0.20

The conventional decision function would assign this sample to the "Corrosion" type since its probability was the highest. However, to reflect a desire for conservatism, suppose the decision function is modified to assign "Gouge" to any sample that has a gouge probability above 0.3 regardless of the other type probabilities. For this sample the more conservative decision function would assign this sample to the "Gouge" type.

More precisely, there are quantitative measures of classifier performance that are useful to examine in this study (see [14]).

- Precision: The fraction of samples that were correctly called "type" out of all samples that were called "type."
- Recall: The fraction of samples that were correctly called "type" out of all samples that were truly "type."

Under these definitions, high precision implies a low false positive rate, and high recall implies a low false negative rate. The simultaneous maximization of both is generally not possible. We would seek to maximize precision when calling false positives would be most detrimental. Conversely, we seek to maximize recall when false negatives would be most detrimental. A classifier with high precision might be labeled as efficient while a classifier with high recall might be labeled as conservative.

Out-of-Bag Performance Metrics. As discussed previously, the OOB samples in a random forest model can be used to provide unbiased estimates of classifier performance similar to what would be obtained through the use of a hold-out test set. The OOB samples are used first in the selection of a suitable gouge probability threshold followed by a more general examination of overall classifier performance.

To select a reasonable gouge probability threshold, the precision and recall of the classifier and decision function as a function of gouge probability are examined. Figure 2 shows the results of computing both precision and recall for various gouge probability thresholds in the decision function. A threshold of 0.3 provides reasonable classifier precision around 80% for all types while maintaining a recall of over 95% for gouges.

Copyright © 2016 by ASME

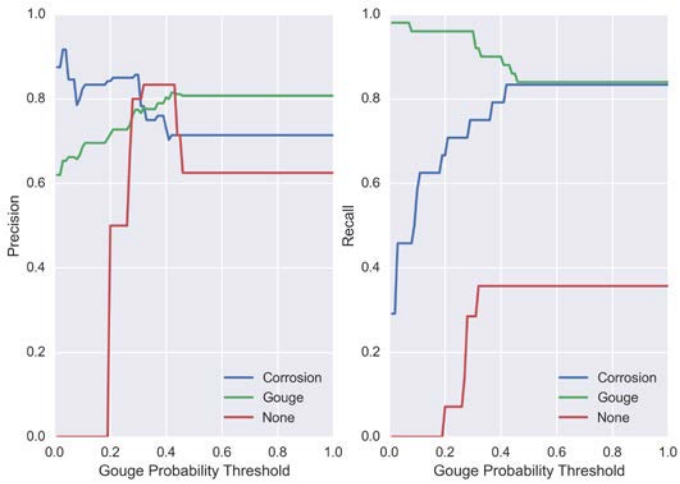

FIGURE 2. . PRECISION AND RECALL FOR THE THREE METAL LOSS TYPES AS A FUNCTION OF GOUGE PROBABILITY THRESHOLD.

Summary details on the number of correctly classified features can be obtained from examining the classifier confusion matrix as presented in Table 2. Table 3 shows a summary of precision and recall for the classifier and decision function combination.

TABLE 2. CLASSIFIER CONFUSION MATRIX

	Called "Corrosion"	Called "Gouge"	Called "None"
Is "Corrosion"	18	6	0
Is "Gouge"	1	48	1
Is "None"	2	8	4

TABLE 3. CLASSIFIER PERFORMANCE SUMMARY

	Precision	Recall	Samples
"Corrosion"	0.86	0.75	24
"Gouge"	0.77	0.96	50
"None"	0.80	0.29	14
average / total	0.80	0.80	88

In summary, a random forest classifier was developed that can distinguish dent with gouge from dent with corrosion. A custom decision function was developed that provides a means to vary the conservatism of the classifier so that high recall on gouges can be achieved. A gouge probability threshold of 0.3 achieves better than 95% recall for OOB samples.

Algorithm for SSWC Discrimination

Many ILI vendors have historically employed a process for identifying potential SSWC based largely on the scrutiny of a subject matter expert (SME). The objective of the current classifier model development is to arrive at a quantitative model that resembles the process that a SME would follow. While there are many forms that such a classifier can take, one of the most commonly used models is logistic regression. It is especially appropriate where the training data set is relatively small and can be separated by a linear decision boundary.

In logistic regression, a linear combination of continuous input variables, often called "features," is mapped to a continuous output value between 0 and 1 using the logistic function. It is commonly employed when the goal is to place candidates into one of two classes, the first corresponding to a model output of 0, and the second corresponding to 1. As such, the output of such a model can be treated as a probability. The continuous probability can be interpreted as the likelihood that the candidate anomaly belongs to the second class. The decision to label each observation as one class or the other is set by selecting a probability level as the decision boundary, often called the discrimination threshold. This is commonly set at 0.5, but it can be set differently to emphasize either false positive or false negative rates [15].

The features selected for inputs in this study mirror those used by subject matter experts in their qualitative classification. Qualitatively, a large SMFL amplitude response that is sharp in the width (circumferential) direction is an indicator of narrow, axial character to the anomaly. Furthermore, if the SMFL signal response is uncharacteristically large relative to the corresponding MFL response at the same location, then the anomaly is more likely to be SSWC rather than corrosion along the long seam. Accordingly, the input features selected for development of the classifier were signal widths and signal amplitudes from both the SMFL and MFL signatures. Commercial ILI analysis software is used to extract the best estimates of signal widths and signal amplitudes for each anomaly.

To make the model more general, the raw signal measurements were first nondimensionalized. Signal width and signal amplitude are nondimensionalized by wall thickness and local background flux density, respectively.

$$\mathbb{w}_{smfl} = \frac{w_{smfl}}{t} \quad \mathbb{A}_{smfl} = \frac{A_{smfl}}{B_{smfl}}$$
$$\mathbb{w}_{mfl} = \frac{w_{mfl}}{t} \quad \mathbb{A}_{mfl} = \frac{A_{mfl}}{B_{mfl}} \qquad 3)$$

These four nondimensionalized features are plotted as pairs against each other in Figure 3. Opportunities for developing a boundary between the two classes are apparent when examining these parameters, especially the signal widths.

Copyright © 2016 by ASME

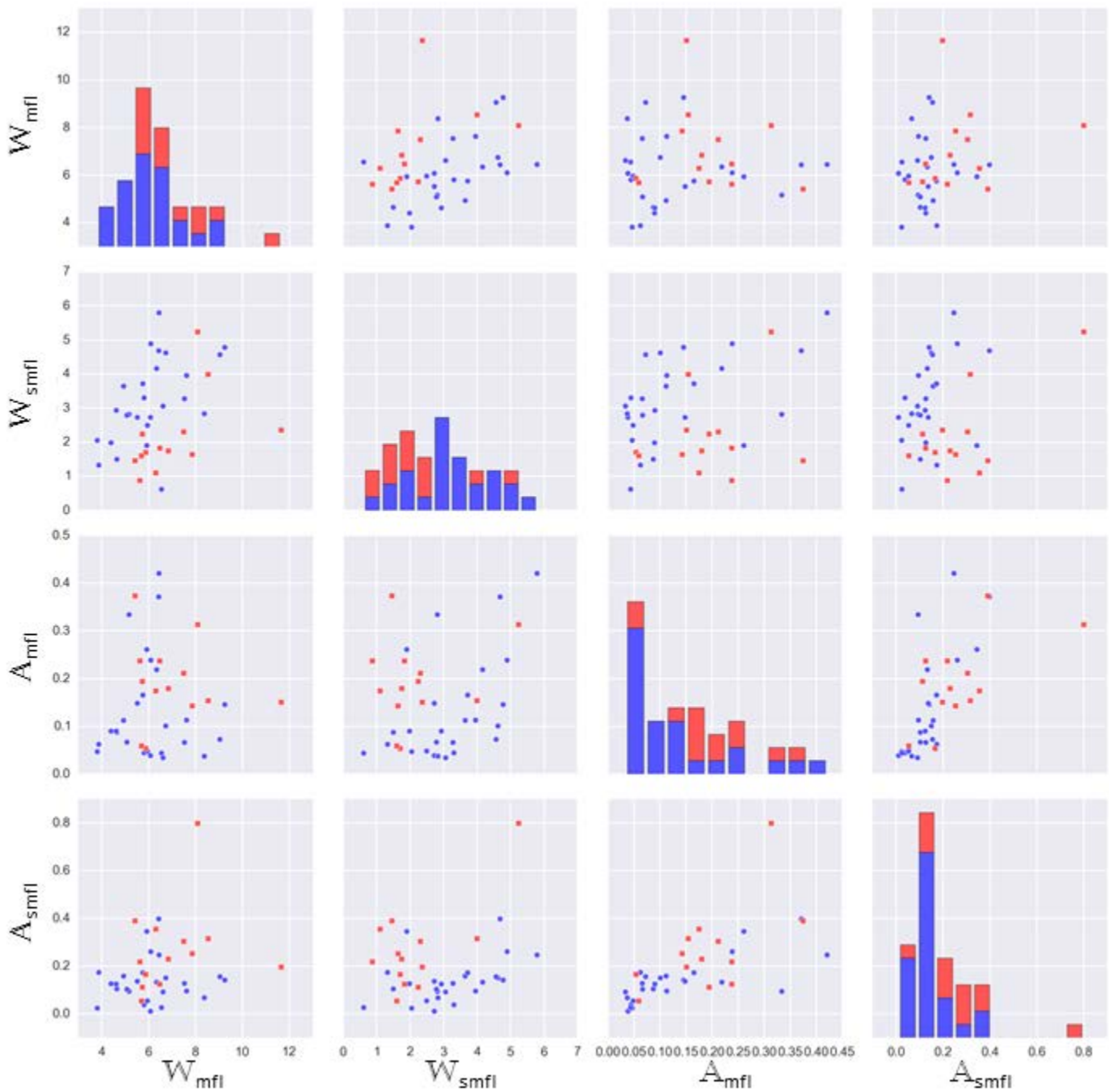

FIGURE 3. PAIR PLOT OF THE NONDIMENSIONALIZED CLASSIFIER INPUTS FOR THE TWO CLASSES. RED SQUARES REPRESENT SSWC, AND BLUE CIRCLES REPRESENT CCLS.

Copyright © 2016 by ASME

With input features defined, a logistic regression model of the following form was fit to the full set of 39 training anomalies:

$$F(z) = \frac{1}{1+e^{-z}} \qquad 4)$$

$$z = \beta_0 + \beta_1 \mathrm{w}_{mfl} + \beta_2 \mathrm{w}_{smfl} + \beta_3 \mathrm{A}_{mfl} + \beta_4 \mathrm{A}_{smfl}$$

where F(z) represents the likelihood that a particular anomaly is SSWC. Best fit model coefficients are

$\beta_0 = -5.21$
$\beta_1 = 1.08$
$\beta_2 = -1.90$
$\beta_3 = 5.42$
$\beta_4 = 9.10$

Raw training accuracy is 87% against the original training data using a discrimination threshold of F(z)=0.5. The classifier commits five errors, three of which are false negatives and two of which are false positives. For the false negatives, meaning the algorithm missed a SSWC anomaly, the missed anomalies were quite small and SSWC anomalies, true positives, were detected by the algorithm in the same joint of pipe.

With a small set of data it is usually necessary to use all of the available data to train the classifier rather than hold out some of the data in a separate model test set. Some form of model cross validation using sets resampled from the training set is pursued instead. For this study, a stratified k-fold cross validation was employed using two folds. A stratified approach preserves the fraction of observations from each class in each fold and also shuffles the observations randomly before dividing the full set into training and test sets. With two folds, half the observations are used to train the model, and then the other half are used to test the model. The process is then repeated with the training and test sets swapping places. Since the process involves random sampling, it is only possible to report representative accuracies on the test sets since rerunning the cross validation exercise produces a different random sample in the two sets. For one representative random sampling, the test accuracies on the two folds were 70% and 79%, somewhat lower than the training accuracy which is to be expected.

Another valuable indicator of model performance is the receiver operating characteristic (ROC). The ROC curve shows the relationship between a classifier's true positive and false positive rates as the discrimination threshold varies from 0.0 to 1.0. The area under the ROC curve is an indicator of overall model performance. An ideal model would have a false positive rate of 0.0 and a true positive rate of 1.0, which would produce an area under the ROC curve of 1.0. At the other extreme, random assignment of each observation to one of the two classes would have roughly equal false positive and true positive rates, yielding an area under the "random guess" ROC curve of 0.5. A ROC curve was developed for the SSWC classifier model using the same stratified k-fold cross validation split described above. The results are presented in Figure 4.

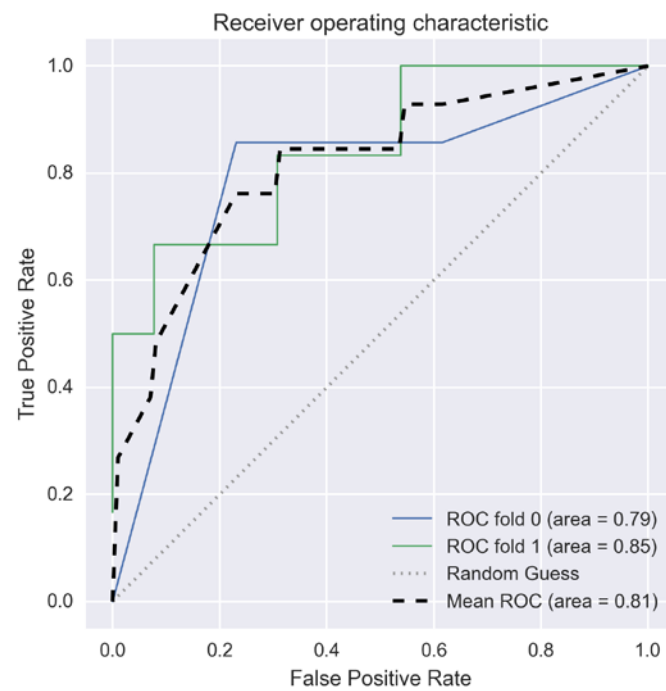

FIGURE 4. RECEIVER OPERATING CHARACTERISTIC FOR THE SSWC CLASSIFIER

VERIFICATION

A conservative approach to verifying both classifiers used pressure testing of the smallest detected anomaly. If the smallest detected anomaly did not fail, then smaller anomalies that have a lower failure pressure would be safe; larger anomalies would be selected for excavation, assessed and repaired as necessary. This approach is inherently conservative as larger anomalies would be identified for repair even though they may pass a pressure test. However, the number of tests is still too small to be statistically significant.

For the mechanical damage sample, one dent with a gouge that was identified as a gouge by the classifier was removed from service. The pipe properties were Grade X52, 508mm (20 inch) diameter and 6.35mm (0.250 inch). The dent measured 1.2 percent of the diameter and contained three gouge marks. The pipe was subjected to a pressure test to 100 percent of specified minimum yield stress (SMYS). The pipe did not fail after completing five pressurizations to simulate five hydrostatic tests.

For the SSWC classifier, three pipe samples containing SSWC were removed from service. The pipe with the small SSWC was pressure tested and failed at a pressure of 23,028 kpa (3,340 psig) and did not fail in the longitudinal seam. This pressure exceeded the calculated yield strength pressure of

Copyright © 2016 by ASME

14,940 kpa (2,167 psi). It also exceeded the pressure required to reach the pipe flow stress, the average of the yield stress (YS) and ultimate tensile stress (UTS), (YS+UTS)/2, of 22,470 kpa (3,259 psig). Therefore, it was concluded that the ILI tool and classifier are capable of identifying very small non-threatening SSWC anomalies. The SSWC algorithm is currently being testing by one pipeline company. In this test, all corrosion near the long seam is being investigated per US DOT regulatory requirements. Detailed measurements are being made at each excavation site to determine the presence of SSWC and the susceptibility of the pipe to this pipeline threat. The assessment is being performed with DOT PHMSA oversight.

SUMMARY

Classifiers that combine the measurements from multiple sensing systems to detect mechanical damage in dents and selective corrosion of the long seam were developed. These classifiers dismiss many of the smaller corrosion features that do not impact the performance of the pipeline while still conservatively identifying dent with gouge and SSWC features. The classifiers were designed to be conservative, meaning some non-injurious corrosion anomalies are designated for excavation in order to catch all potentially injurious mechanical damage and SSWC. These anomaly classification improvements will continue to maintain pipeline safety while eliminating unnecessary excavations of non-threatening anomalies.

ACKNOWLEDGMENTS

The authors would like to thank the U.S. Department of Transportation, Pipeline and Hazardous Material Safety Administration for their financial support of this program under contract DTPH56-13-T-000009. The authors thank Koch Pipeline Company, Shell Pipeline Company, Phillips 66 and ConocoPhillips for supplying ILI data and permitting access to excavation sites to attain the data used in these developments. Data from actual pipeline anomalies was crucial.

REFERENCES

1. Nestleroth, J.B. and Bubenik, T. A., Magnetic Flux Leakage (MFL) Technology for Natural Gas Pipeline Inspection, Battelle, Report Number GRI-00/0180 to the Gas Research Institute, February 1999.
2. Nestleroth, J.B. and Davis, R.J., “Stress, Cold Work, and Metal-Loss Induced Magnetic Flux Leakage Signals,” Proceeding of the 11th PRCI-EPRG Joint Technical Meeting on Pipeline Research, PRC International, April 1997.
3. Dobmann, G. and Höller, P., “Physical Analysis Methods of Magnetic Flux Leakage,” in Research Techniques in Nondestructive Testing, Volume IV, edited by R. S. Sharpe, Academic Press, 1980.
4. Atherton, David L., et al, “Stress Induced Magnetization Changes of Steel Pipes-Laboratory Tests,” IEEE Transactions on Magnetics, Volume MAG-19, Number 4, July 1983.
5. Atherton, D. L., “Effect of Line Pressure on the Performance of Magnetic Inspection Tools,” Oil & Gas Journal, October 27, 1986.
6. Nestleroth, J. B. and Crouch, A. E., Variation of Magnetic Properties in Pipeline Steels, Report No. DTRS56-96-C-0010, Subtask 1.1 Report, to the U. S. Department of Transportation, March 1998
7. Davis, R. J., et al., “The Feasibility Of Magnetic Flux Leakage In-Line Inspection as a Method To Detect and Characterize Mechanical Damage,” GRI Report GRI-95/0369, 1996.
8. Davis, R. J. and Nestleroth, J. B., “The Feasibility of Using the MFL Technique to Detect and Characterize Mechanical Damage In Pipelines,” Review of Progress in Quantitative Nondestructive Evaluation, Volume 16, Plenum New York, 1997.
9. Nestleroth, J.B., and Davis, R.J., “Pipeline Mechanical Damage Characterization By Multiple Magnetization Level Decoupling” Review of Progress in Quantitative NDE, Volume 18, Plenum New York, July 1998.
10. Bubenik, T.A., et al, “In-Line Inspection Technologies for Mechanical Damage and SCC in Pipelines – Final Report” Department of Transportation Report DTRS56-96-C-0010, June 2000.
11. Bubenik, T.A., J. B. Nestleroth, R. J. Davis, Lloyd Ulrich, and Harvey Haines, “Multiple Magnetization Level MFL For Pipeline Mechanical Damage Characterization” IPC00-0218 Proceedings of the International Pipeline Conference 2000, October 1-5, 2000, Calgary, Alberta, Canada.
12. Fact Sheet: Selective Seam Corrosion (SSC) Retrieved from //primis.phmsa.dot.gov/comm/FactSheets/FSSelectiveSeamCorrosion.htm, December 2011
13. Peter Massopust, Carl Torres, and Alan Dean, "Improving In-line Inspection For Mechanical Damage in Natural Gas Pipelines," PRCI, L52084, 2003.
14. Leo Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
15. Gareth James, An Introduction to Statistical Learning with Applications in R. New York, NY: Springer, 2013.

The United States Government retains, and by accepting the article for publication, the publisher acknowledges that the United States Government retains, a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for United States Government purposes.

 Copyright © 2016 by ASME