

Prepared for: U.S. DOT Pipeline and Hazardous Materials Safety Administration

Contract Number: 693JK31950004CAAP

Project Title: Multi-modal NDE Assisted Probabilistic Pipeline Performance Evaluation under Interactive Anomalies

Prepared by: Kiswendsida Jules Kere, Karthik Gopalakrishnan, Dr. Vivek Rathod, Qindan Huang, Qixin Zhou, and Yiming Deng

Contact Information: Qindan Huang, qhuang@uakron.edu, 330-972-6972
Yiming Deng, dengyimi@egr.msu.edu, 517-884-0926
Qixin Zhou, qzhou@uakron.edu, 330-972-7159

For quarterly period ending: 06/30/2020

Business and Activity Section

(a) Contract Activity

No contract modification was made or proposed in this quarterly period.

(b) Status Update of Past Quarter Activities

In the past quarter, we continued making progress in Task 1, Task 2, and Task 3. In particular, due to COVIN-19 the exposure testing in Task 1 had to be paused in the middle of March, and we resumed the testing in early June. For Task 2, we exploited different signal processing algorithms to correlate different defect features with the non-destructive detection responses. For Task 3, we developed probabilistic burst pressure prediction models for pipeline with isolated corrosion defects, and compared the prediction performance of the proposed models with the existing prediction models based on the experimental and numerical data and reliability analysis.

(c) Cost share activity

Partial support for one graduate student tuition was provided by The University of Akron as per the cost share agreement.

(d) Technical approach

The goal of this proposed study is to develop probabilistic pipeline performance evaluation framework based on multi-modal NDE assisted by physical and mechanical modeling under interactive anomalies. This study will utilize the experimental testing and numerical analysis to generate more realistic defect shapes and colony profiles, which will be used for characterization and validation of interactive defects NDE. Meanwhile, the identified defect profile will be used for the probabilistic defect time-evolution model development, which is crucial for reliability evaluation of pipeline performance under interactive defects. In addition, probabilistic models of failure pressure of a pipeline containing corrosion and cracking-like defects will be developed, achieving predictions that are unbiased with reduced variability and considering defect interaction. Specific technical objectives are proposed as follows:

- Generate realistic corrosion and cracking defect profiles through laboratory testing and electrochemical simulation;
- Establish an expanded NDE framework for interactive anomalies by probabilistic characterization of defect profiles;

- Establish a comprehensive failure pressure database including corrosion and cracking defects, and isolated and colony defects;
- Develop probabilistic time-evolution models for defect profile quantities based on NDE defect characterization;
- Develop probabilistic failure pressure prediction models incorporating defect interaction;
- Investigate the impact of various physical quantities and uncertainty sources on pipeline reliability.

To achieve the proposed research objectives described in Section 2, four tasks are developed.

- Task 1. Realistic defect profiles generation using experiments and COMSOL (Drs. Zhou & Deng)
 - Task 1a. Defect configuration in a corrosive environment
 - Task 1b. Defect configuration with cracks in a corrosive environment
- Task 2. NDE framework development and validation for interactive defect detection and state characterization in both lab and field environments (Dr. Deng)
 - Task 2a. Multi-scale and multi-physics modeling
 - Task 2b. NDE system development
 - Task 2c. Multi-modal data processing for interactive anomalies
- Task 3. Probabilistic capacity model development considering interactive anomalies (Dr. Huang)
 - Task 3a. Establishment of a failure pressure database
 - Task 3b. Probabilistic failure pressure model development
- Task 4. Probabilistic model development of anomaly time-evolution and reliability evaluation (Dr. Huang)

1. Task 1: Realistic defect profiles generation using experiments and COMSOL

1.1 Background and Objective in the 3rd Quarter

The objective of Task 1 is to generate realistic corrosion profiles through environmental exposure testing. The defect shapes and colony profiles will be used for non-destructive evaluation (NDE) and for the probabilistic defect time-evolution model development.

1.2 Research Progress in the 3rd Quarter

Due to the pandemic of COVID-19, The University of Akron has stopped all the non-essential research work. The exposure testing had to be stopped in the middle of March. We got back to our research lab in early June, keeping social distance and following the back-to-lab protocol. We are only allowed to have one person per 100 ft² in the lab. The exposure testing resumed on June 15 and is scheduled to run at least a month. During the exposure testing, an infinite microscopy characterization and electrochemical measurements are conducted. The completed data sets are expected to generate in the next Quarter.

In this Quarter, we have purchased COMSOL Multiphysics with an academic-use license. The whole package includes AC/DC module, Corrosion module, Fatigue module, Nonlinear Structural Materials module, and Structural Mechanics module.

1.3 Future Work (Next Quarter)

In the next Quarter, the research team will continue working on Task 1. The realistic corrosion defect profile as a function of time is expected to be generated. It will include a defect shape and a corrosion rate. Besides, the research team will start on the COMSOL simulation that includes the COMSOL model setup and simulation parameter identification.

2. Task 2: NDE framework development and validation for interactive defect detection and state characterization in both lab and field environments

2.1 Background and Objectives in the 3rd Quarter

Background

In the previous quarters, we have shown two different FEM models in which guided waves were utilized to detect corrosion pits and cracks of varying lengths and depths. Figure 1 shows the two different models that were used for this study. Specifically, the axial guided wave model used axial guided waves to detect corrosion defects along the length of the surface of the pipe, while the circumferential guided waves were used to detect circumferential defects in the circumferential guided wave model. In order to classify and characterize the corrosion defects, a dataset was created using the axial model alone. The axial model was considered as the first step, as it involves lesser complexity compared to the circumferential guided wave model. Once a successful study is established for the axial model, it can be extended to the other model with further modifications. The dataset created contained the guided wave responses of healthy samples, as well as samples with defects of varying lengths, depths and quantities. The dataset created is discussed in further detail in this report. Finally, a simple classifier using Multi-Layer Perceptron Networks (MLP) was developed to accurately classify a response from a healthy pipeline or a response from a pipe with defects. The accuracy of the developed network was found to be about 95%.

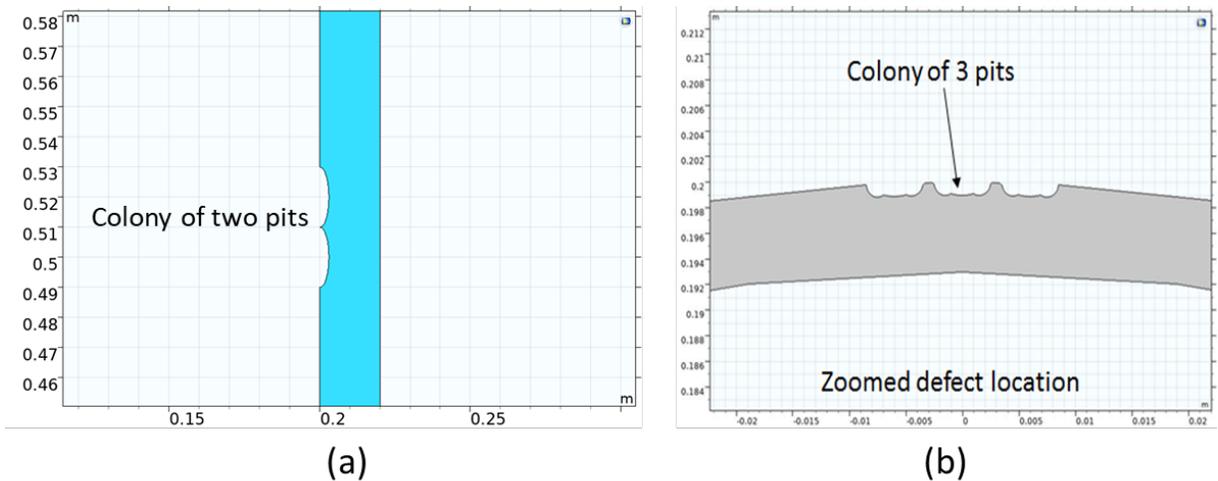


Figure 1: (a) Axial guided wave model and (b) circumferential guided wave model for corrosion pits

Objectives in the 3rd Quarter

There are two objectives in this quarter. One is to determine different types of damage Index (DI) for quantifying the corrosion defects using the dataset created. We exploited different signal processing algorithms and some of the methods that are used are the Hilbert Transform Analysis (HT), Wavelet Transform (WT) and Multi-scale Cross Entropy Analysis (MCSE). In addition, different features of corrosion defects that can be potentially used in a feature engineering-based characterization algorithm.

The other objective is to develop an efficient automated defect classification and characterization algorithm. As mentioned, in the previous quarter, we used multi-layer perceptron (MLP) to accurately classify defect responses. In this quarter, we make the model more robust by training the network with different levels of noise to mimic real world experiments as much as possible. We also propose a 1D Convolutional Network (1D-CNN) to predict defect parameters that can effectively characterize corrosion pitting. This is in view of developing an overall learning paradigm to develop featured based Machine Learning Algorithms and Deep Learning Algorithms, where we can feed in raw sensor data to characterize corrosion defects.

2.2 Research Progress in the 3rd Quarter

Dataset

The dataset consists of 57 healthy responses and 93 defect responses. All the responses are collected from the axial guided wave model for corrosion pits. For simplicity, three different configurations have been considered. The first configuration is the healthy pipe, which is defect-free. The second configuration has one corrosion pit in the pipe, while the third configuration has a colony of two pits in the pipe. The length of the corrosion pits varies from 1 mm to 5 mm, while the depth of the pits varies from 5 mm to 20 mm. The operating frequency of the model is 25 KHz as mentioned in the previous reports. Figure 2 gives the dataset split.

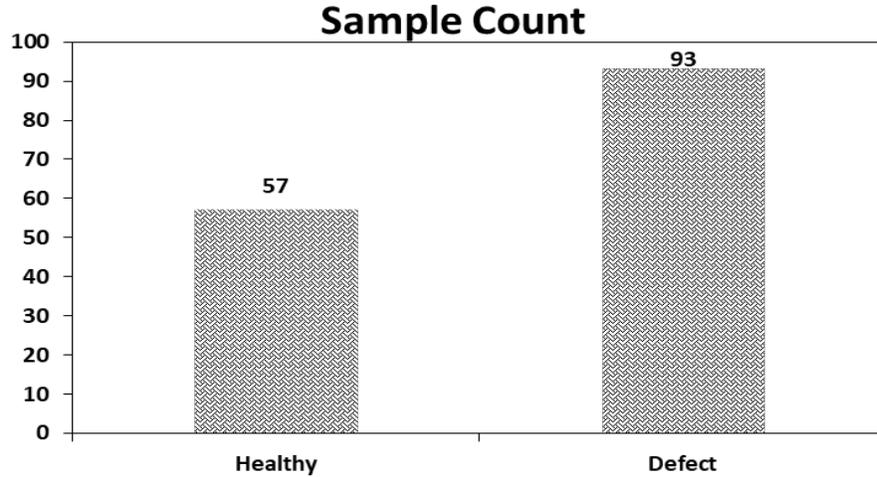


Figure 2: Split of the axial model dataset

In the last quarter, some basic features of the dataset were examined, which is shown in Figure 3. The four features shown in Figure 3 could be potentially used as feature vectors in various characterization algorithms. In the 3rd quarter, more complex features are added using some signal processing tools.

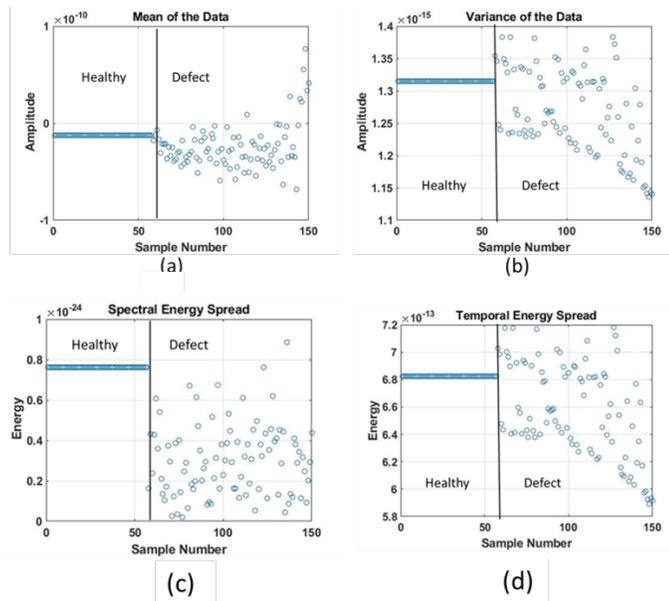


Figure 3: (a) Sample mean (b) sample variance (c) spectral energy and (d) temporal energy spread of the dataset

Cross Entropy Analysis

The Cross Sampling Entropy Method (CSamp-En) is mainly used to evaluate the degrees of asynchrony and dissimilarity of two time series in the same system [1]. Please note that, the responses collected in the dataset are velocity time histories. The CSamp-En method is based on the sampling Entropy method (SampEn) [2] with a concept called Approximate Entropy that is basically a measure of the degree of irregularity or disorder in a measurement time series. When SampEn is lower, the sequence is more regular; the larger SampEn, the more irregular and complex the sequence becomes. SampEn is independent of the length of the data record and the algorithm uses the following parameters: threshold (r), sample length (m), and signal length (N). The length of the time series has no effect on the analysis results, and the results remain relatively consistent.

The analysis step of a CSamp-En algorithm is similar to that in the SampEn analysis method. The difference is that the object of SampEn analysis is mainly a single time series signal system, whereas the CSamp-En method analyzes two different time series signals to establish a template space for each of the two signals. This method basically gives a measure of how similar two time series signals are in terms of a similarity number. In the scope of our work, the similarity of different defect responses with respect to the healthy signal can be computed, which can be used to define a related Damage Index (DI). The procedure of CSamp-En is similar to that of SampEn and can be summarized as follows [3].

Let us define two time series signals: $\{X_i\} = \{x_1, x_2, \dots, x_i, \dots, x_N\}$ and $\{Y_j\} = \{y_1, y_2, \dots, y_j, \dots, y_N\}$. Both the time series signals are of the same length N . The two signals are then divided into templates of size m :

$$u_m(i) = \{x_i, x_{i+1}, \dots, x_{i+m-1}\} \quad 1 \leq i \leq N - m + 1 \quad (1)$$

$$v_m(j) = \{y_j, y_{j+1}, \dots, y_{j+m-1}\} \quad 1 \leq j \leq N - m + 1 \quad (2)$$

A similarity number between $u_m(i)$ and $v_m(j)$ is defined as $n_i^m(r)$, and can be expressed as:

$$n_i^m(r) = \sum_{j=1}^{N-m} d[u_m(i), v_m(j)] \quad (3)$$

where, the maximum distance $d[u_m(i), v_m(j)]$ between the two template spaces $u_m(i)$ and $v_m(j)$ is defined as:

$$d[u_m(i), v_m(j)] = \max\{|x(i+k) - y(j+k)|\} \quad 0 \leq k \leq m-1 \quad (4)$$

$$d[u_m(i), v_m(j)] \leq r \quad 1 \leq j \leq N-m \quad (5)$$

When the distance between the two samples is smaller than the threshold, r , the two samples are considered similar; conversely, when the distance between the two samples exceeds r , the two samples are considered dissimilar. The threshold r can be chosen manually by the user. Through the use of different templates for similarity comparison and the calculation of the number of templates that exhibit the conditions of similarity, the number of similar samples in the i^{th} template to those in the entire template space can be obtained. The similarity probability of the i^{th} template can be calculated as:

$$U_i^m(r)(v||u) = \frac{n_i^m(r)}{(N-m)} \quad (6)$$

The average probability of similarity for template m can then be obtained as:

$$U^m(r)(v||u) = \frac{1}{N-m} \sum_{i=1}^{N-m} U_i^m(r)(v||u) \quad (7)$$

The degree of dissimilarity resulting from the division of the two time-series by m points represents the degree of synchronization between the two template spaces. Finally, the sample space is composed of the sample of length $m+1$, and the average similarity probability is calculated. The formula for calculating CSamp-En is expressed as:

$$CS_E(m, r, N) = -\ln \left\{ \frac{U^{m+1}(r)(v||u)}{U^m(r)(v||u)} \right\} \quad (8)$$

In this investigation, the changes of both final CSamp-En and the similarity number $n_i^m(r)$ are studied for

different corrosion configurations and different corrosion pit depth. The parameters used in this study are listed in Table 1.

Table 1: Parameters for CSamp-En

Signal length, N	1038
Template length, m	4
Threshold (based on healthy response), r	0

A DI based on the mean of the similarity numbers for different configurations is defined. The variation in the DI for different configurations is plotted in Figure 4, while the DI for different defect depths is plotted in Figure 5. From Figure 4, it is observed that there is a direct relationship between number of pits and DI. As the number of pits increase in the colony, the absolute DI value increases, as shown in Figure 4. For defect depth, the DI value does increase for a defect depth of 5mm and 10 mm, but the DI values for defect depths of 15mm and 20mm are similar to that of 10mm. The increasing DI values basically indicate that the responses of a particular group are more dissimilar to the healthy response. This can be directly related to a conclusion that has the samples corresponding to the particular group are relatively more damaged. For example, the samples with 5 mm defect depths have a smaller DI value compared to samples with 10 mm defect depths. This basically suggests that samples with 10 mm defect depths are more damaged, which is truth. A similar study is also conducted for changing the defect length (i.e. corrosion pit length).

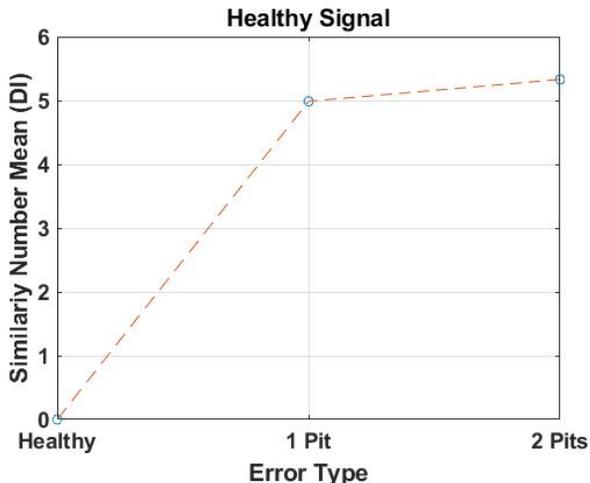


Figure 4: Normalized similarity number $n_i^m(r)$ vs number of pits

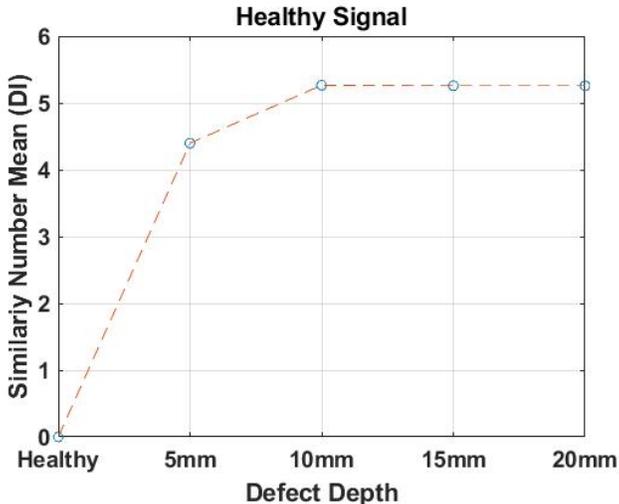


Figure 5: Normalized similarity number $n_i^m(r)$ vs defect depth

When using the CSamp-En analysis, we were able to extract a DI that shows a direct relationship to the number of pits. However, the results are not as promising for different defect depths. Nevertheless, overall the algorithm presented above provides a valuable feature engineering tool.

Hilbert Transform Analysis

Hilbert Transform (HT) is one of the most commonly used signal processing tools to study time signals. Implementing a HT enables us to create an analytic signal based on some original real-valued signal. HT gives the instantaneous amplitude of a signal, and it can be used to find the envelope of harmonic signals, which exactly is the nature of the response in our dataset. HT is mathematically described below for a signal $u(t)$:

$$H(u(t)) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{u(\tau)}{t-\tau} d\tau \quad (9)$$

Similar to how MSCE is studied as a potential DI, HTs of three different configurations (i.e., healthy, single pit and colony of pits) are computed. The results are shown in Figure 6. The two peaks (i.e., Peak I and Peak II) shown in Figure 6(a) are basically due to the two reflections seen in the raw signal (see previous quarter reports). The two reflections are the S0 wavemode, and a combination of A0 and reflected S0 wavemodes. As shown in Figure 6, one can observe that though the S0 peak (i.e., Peak I) has little difference for different configurations, the second peak (i.e., Peak II) shifts slightly to the right with increasing number of pits in the colony. The increase is in the order of 0.1-0.2 ms (as shown in Figure 6(b)), which is significant considering the scale which we work with.

HTs of the signals from different defect depths are also computed and the results are shown in Figure 7. Similar to Figure 6, there is no distinct different among the HT responses at Peak I, but the second peak certainly shows a difference between the responses from a damaged pipeline and a healthy response. Therefore, Hilbert Transform is another valuable tool to compute a feature that distinguishes different configurations and defect depths. A DI based on the HT can be easily defined to quantify the corrosion defects, which is part of the work planned for the future.

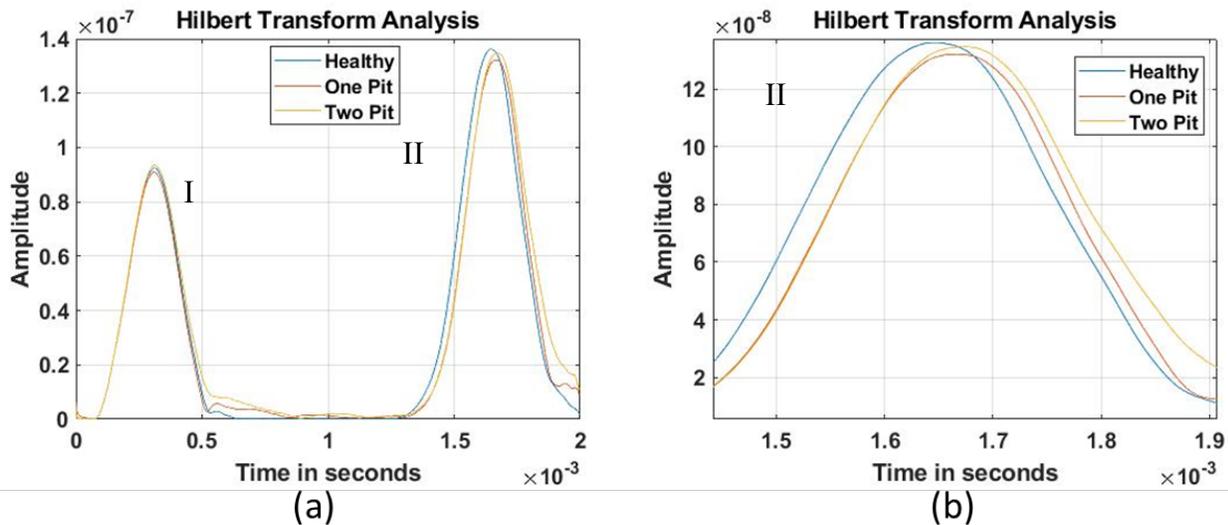


Figure 6: (a) HT responses for increasing number of pits and (b) a zoomed-in version of Peak II

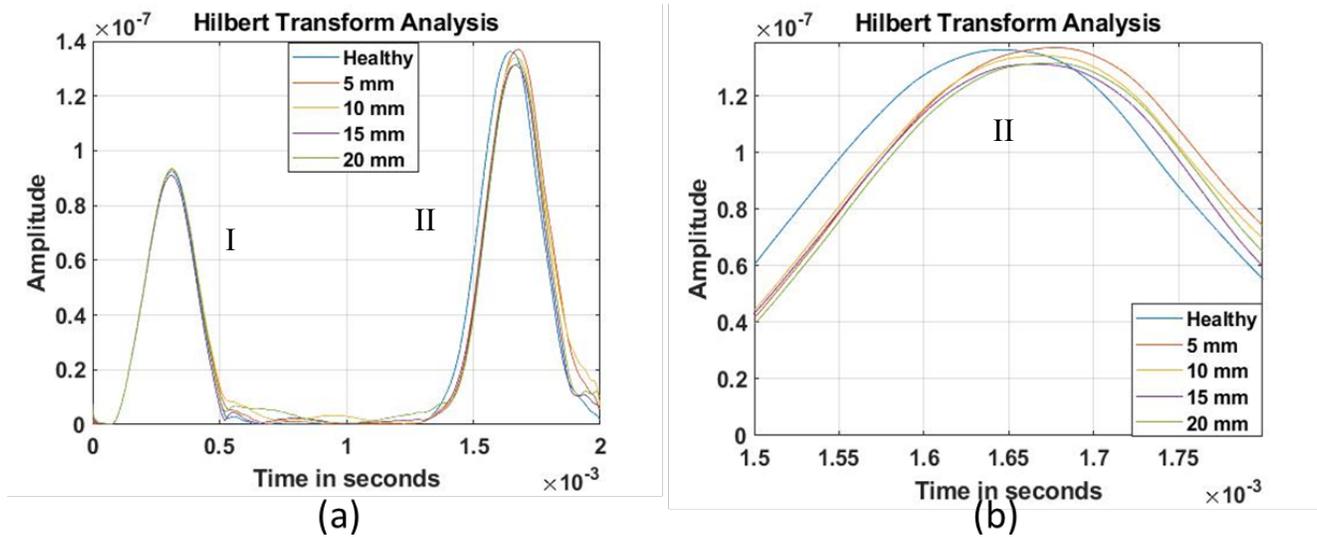


Figure 7: (a) HT responses for increasing defect depth and (b) a zoomed-in version of Peak II

Multi-Layer Perception Classifier Network

In this Quarter, we also tried to improve the robustness of the MLP based classifier network presented last quarter. For a quick recap, Figure 8 shows the architecture of the network that was used. Figure 9 shows the results when the trained model was used to predict on the unseen responses. For the dataset we defined above with well-labeled responses, the classification accuracy is defined as the fraction of correct predictions over the total number of predictions. The accuracy of this particular network was about 95%, which means that the MLP network performs really well.

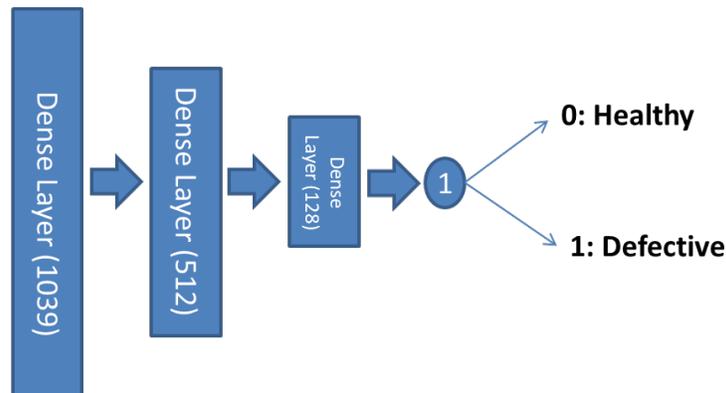


Figure 8: Multi-Layer Perceptron network for defect response classification

In order to simulate a real-world experiment environment, additive white Gaussian noise (AWGN) of different Signal to Noise Ratio (SNR) levels are added to the training dataset. The model is then trained, and predicted on the unseen examples. For this analysis, three different SNR levels of 5, 10 and 20 are considered. The signal with a SNR level of 5 has the highest noise content, while the signal with a SNR level of 20 has the least noise content. Figure 10 shows a signal without noise, and a signal with an artificially added noise of SNR 5. As shown in Figure 10, a signal with SNR 5 is highly distorted with noise, and training on such data is tricky and challenging. However, training with such high-level noises can make a model more robust and more generic in nature. As expected, the performance of the MLP drops as noise is added. Figure 11 shows the performance of the model when it is trained on a dataset with a SNR level of 20. Though the accuracy drops to 93%, the model still correctly classifies healthy and defective responses according to its true ground class as is visible from the confusion chart.

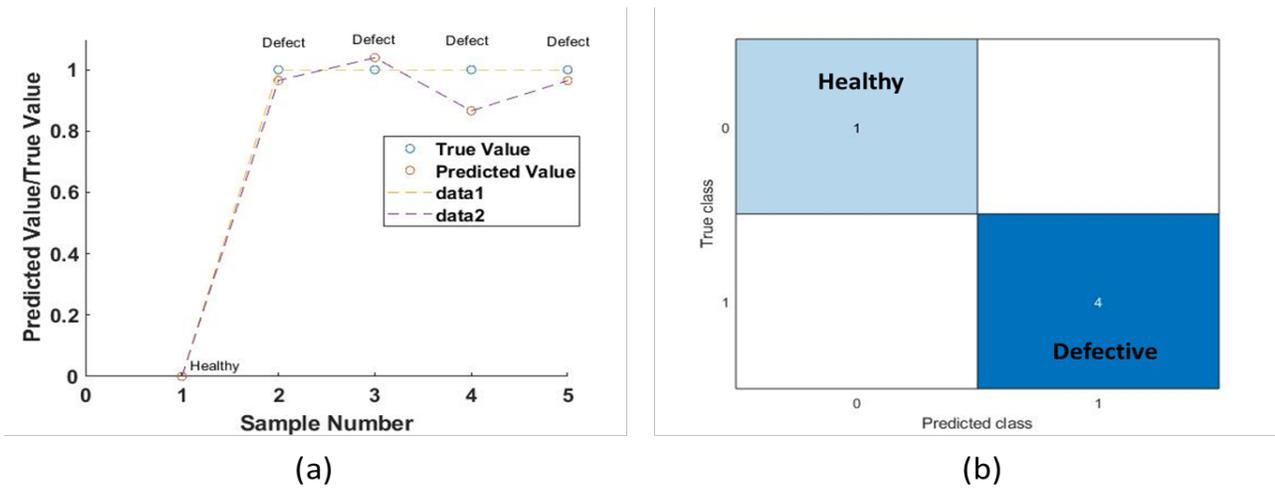


Figure 9: (a) Prediction results for the MLP network and (b) confusion plot for the predicted results

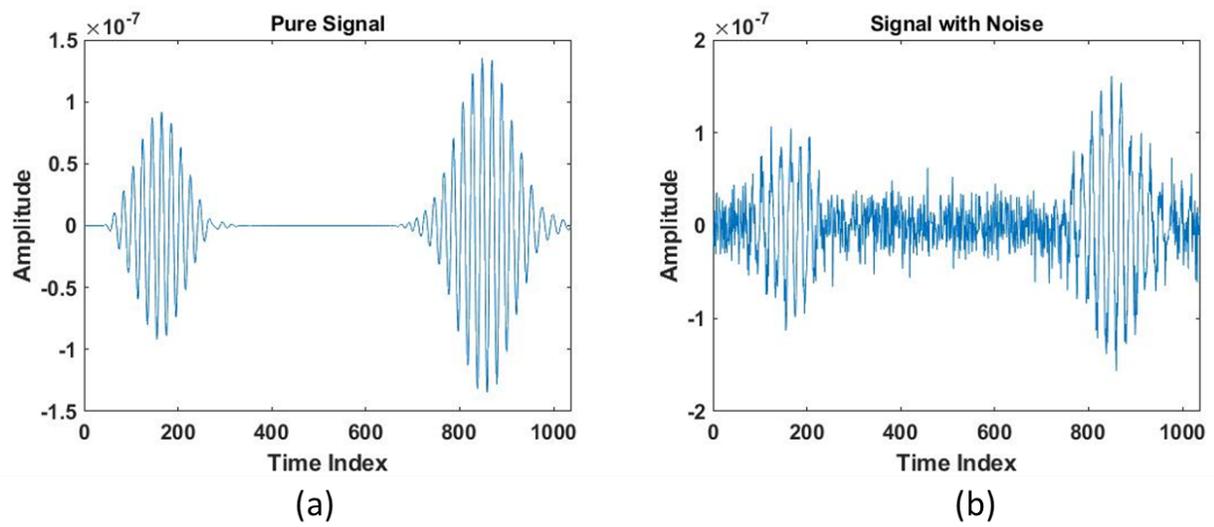


Figure 10: (a) Pure raw signal and (b) signal with additive white Gaussian noise (AWGN) added at SNR 5

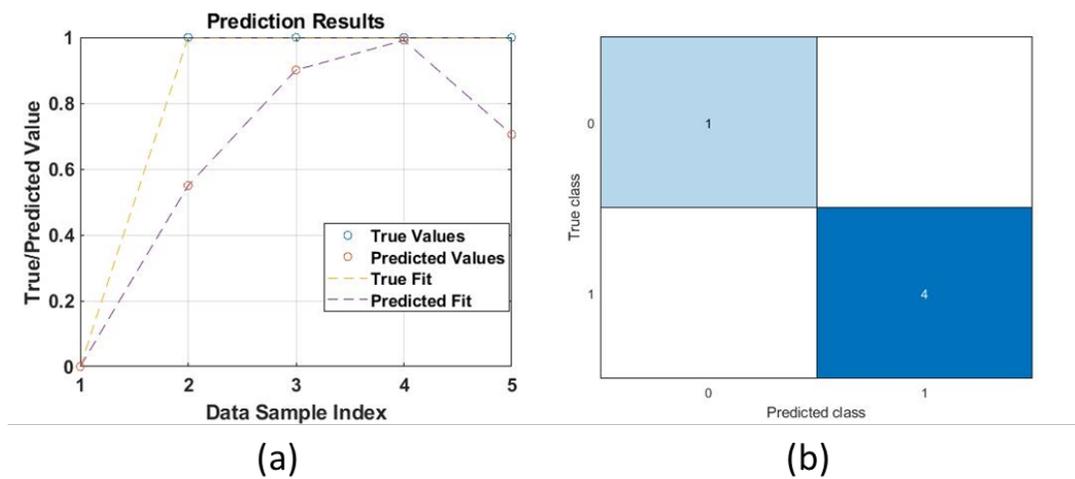


Figure 11: (a) Prediction results for the MLP network trained with a dataset with SNR 20 (b) confusion plot for the predicted results

Similarly, when the noise level is increased by maintaining a SNR level of 10, the performance further drops to 91% accuracy. Though it classifies the responses accurately, the absolute values the MLP is predicting for a defect are further away from 1, as is visible from Figure 12. For a SNR level of 5, the accuracy further drops to 89%, and the network now falsely classifies a healthy response as a defective response as seen in Figure 13. But despite this error, the MLP network correctly classifies all defective responses correctly. Though this is not the best scenario, it is acceptable as when it comes to nondestructive evaluation, it is sometime acceptable to falsely classify healthy signals as long as defective signals are classified correctly. Thus, we are able to attain very good performance at SNR levels of 10 and 20, and acceptable performance at a SNR level of 5.

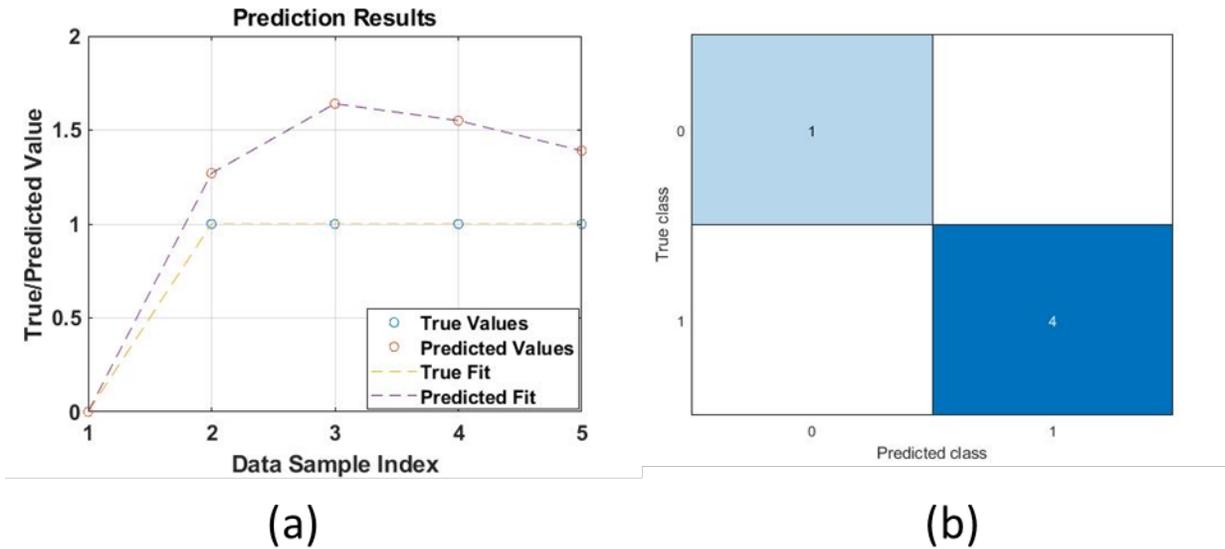


Figure 12: (a) Prediction results for the MLP network trained with a dataset with SNR 10 (b) confusion plot for the predicted results

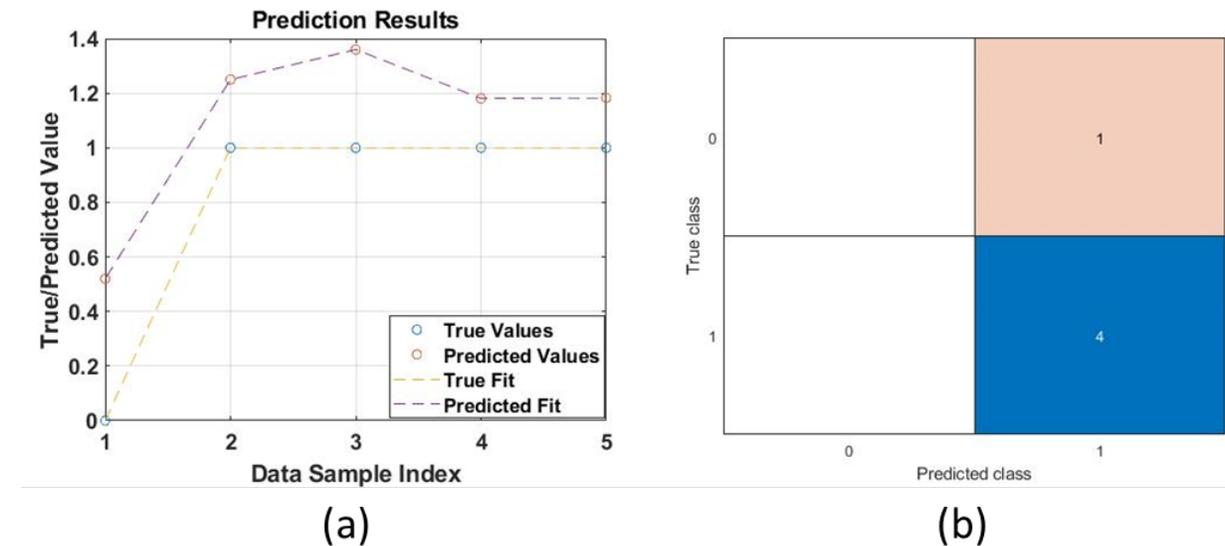


Figure 13: (a) Prediction results for the MLP network trained with a dataset with SNR 5 (b) confusion plot for the predicted results

1D-Convolutional Neural Network

The classifier network has been optimized by training it with signals added with different noise levels,

thereby making the network robust and reliable. The next step is to develop 1D Convolutional Neural Network (1D-CNN) to help characterize the corrosion defects. 1D-CNN is known to extract inherent features from long time series data. Extensive research has gone into using CNNs for defect localization and characterization in composites in the last few years. In the 1D-CNN network proposed in this study, the input will be the velocity time histories obtained, and the output will be the defect parameters such as defect depth, defect length, and number of corrosion defects in the pipe. One of the reasons for adopting 1D-CNN is that it trains faster than recurrent neural networks.

Sparse connections and parameter sharing are two important ideas in CNN, whereas, in a fully connected network (FCN), every neuron interacts with every other neuron [3]. CNN helps to reduce the number of learnable parameters, which eventually saves memory and decreases the training time. CNNs are also very robust to external influences, and generally have been shown to perform well even when there is low level noise in the data. 1D-CNN works similar to a traditional CNN/2D-CNN, the only difference is that the inputs, kernels and feature maps are all in one dimension. Figure 14 shows the framework of the proposed 1D-CNN model. Please note the CNNs are data hungry, and with a current dataset of size 150 samples, it is incredibly hard to attain acceptable performance. Therefore, the current model shown is being fine-tuned using the small dataset, while simultaneously more data is being created using data augmentation techniques along with FEM simulations to populate the dataset. The results of this network are planned to be presented in the next quarter.

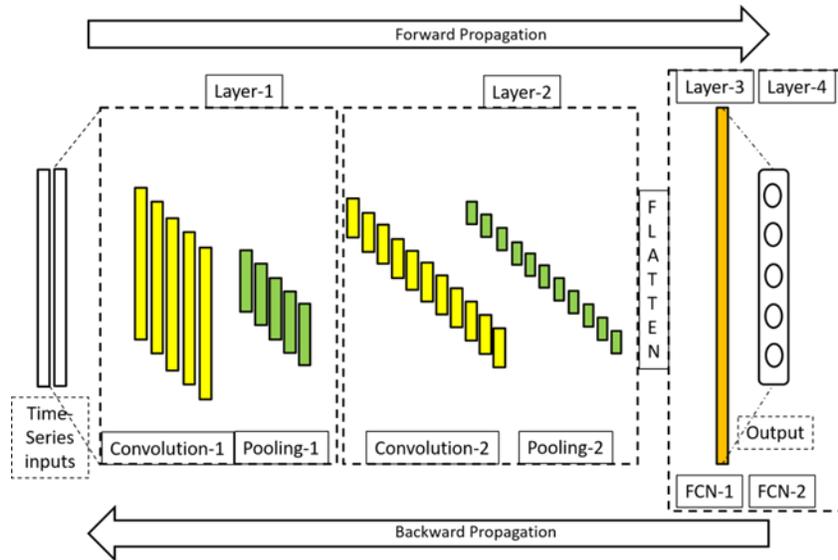


Figure 14: Architecture of the proposed 1D-CNN

2.3 Conclusions

In this quarter, different defect features were established to be used in an overall learning paradigm that leverages the best of Machine Learning based feature engineering algorithms, and Deep Learning based Convolutional Neural Networks (CNNs). A Multi Cross Entropy Analysis (MCEA) was conducted, and a Damage Index (DI) based on the similarity number was established to quantify the defect depths, and the number of corrosion defects. Also, Hilbert Transform (HT) analysis was conducted to establish another feature of interest. The multi-layer perception (MLP) classifier developed last quarter was further improved by training based on the data with different noise levels. It was seen that even at very high noise levels (such as SNR of 5), the performance of the networks was acceptable. Furthermore, a 1D-CNN network was proposed for characterization of corrosion defects, which currently is being fine-tuned.

2.4 Future Work (Next Quarter)

In the next Quarter, we plan to complete the characterization of corrosion pits by training a 1D-CNN network proposed above. Along with this, a feature-based machine learning algorithm will be developed

to characterize corrosion pits. The features studied till now will be used as inputs to these networks. A comparative study between Machine Learning and Deep Learning methods will be undertaken, with a view of creating an overall learning paradigm for the characterization and classification of corrosion pits in pipelines. On the experimental side, experiments will be undertaken with setups similar to the ones that have been simulated. Experimental data can validate the learning models and also be potentially used to further train the models.

3. Task 3. Probabilistic capacity model development considering interactive anomalies

3.1 Background and Objectives in the 3rd Quarter

Background

The inaccurate prediction of failure pressure capacity is one of the critical issues in risk management of pipeline systems, as it can impede the ability to achieve a target margin of safety. The burst failure mechanisms for corrosion and cracking defects are fundamentally different, and even more complex for interactive anomalies. With corrosion, the burst failure is a ductile failure due to plastic collapse; with cracking defect, the failure includes ductile failure (similar to corrosion) and brittle failure due to fracture. For a colony of closely spaced defects, the residual strength of a pipeline becomes much lower than an isolated defect due to the interaction among the adjacent defects.

The limitations of existing work regarding the failure pressure predictions include the following: (1) numerous models are available, but no model is universally accepted; (2) the majority of the models were developed based on the concept of a factor of safety, thus, these models are deterministic and cannot be directly used in reliability analysis; and (3) numerous studies have shown that these models provide over-conservative predictions for both corrosion and cracking-like defects, and the bias needs to be quantified and corrected. This Task 3 is aimed to address the limitations mentioned above, and it includes two subtasks:

- Task 3a. Establishment of a failure pressure database
- Task 3b. Probabilistic failure pressure model development

Objectives in the 3rd Quarter

The overall objective for Task 3a is to establish a database for three groups: isolated and colony of corrosion defects, isolated and colony of crack-like defects, and colony of corrosion and crack-like defects. The overall objective for Task 3b is to develop probabilistic failure pressure models for a pipeline with corrosion anomalies, crack-type anomalies, and interactive anomalies with different types.

The objectives for Task 3 in the 3rd quarter are (1) to develop probabilistic burst pressure prediction models for pipeline with isolated corrosion defects and (2) to evaluate the prediction performance of the proposed models.

3.2 Research Progress in the 3rd Quarter

Proposed prediction models

Based on the established database for pipelines with isolated corrosion defects, three probabilistic prediction models of failure pressure are developed corresponding to three levels of yield strength, σ_y , in order to reduce model errors. The three levels of yield strength are Level 1 – low strength ($\sigma_y = [262\ 433]$ MPa), Level 2 – moderate strength ($\sigma_y = (433\ 508]$ MPa), and Level 3 – high strength ($\sigma_y = (508\ 802]$ MPa). For all three levels, the probabilistic failure pressure models follow the same formulation as:

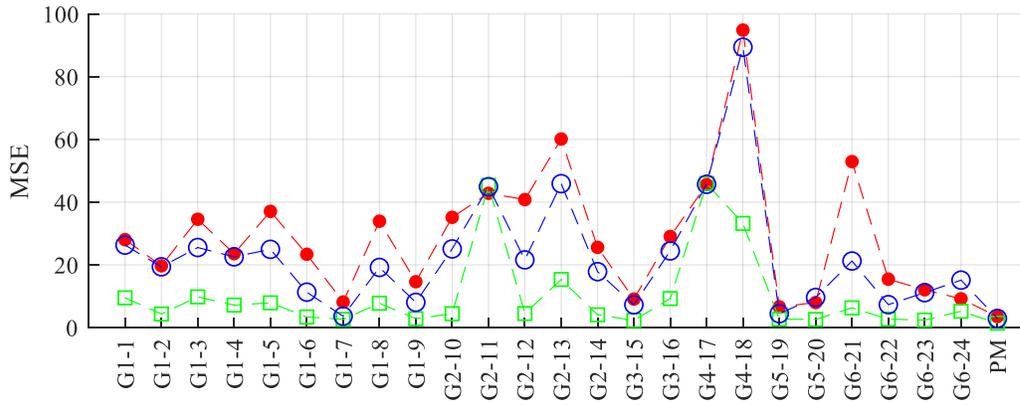
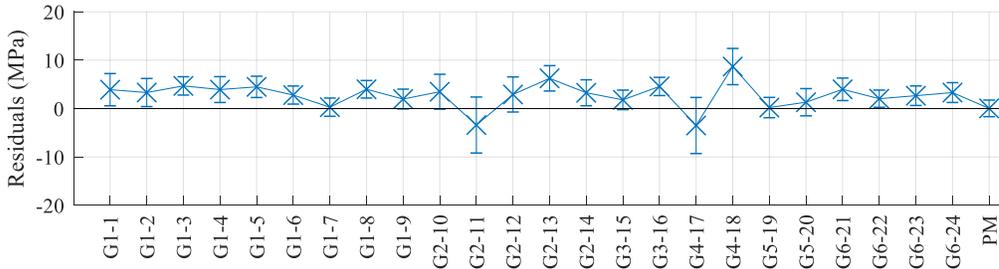
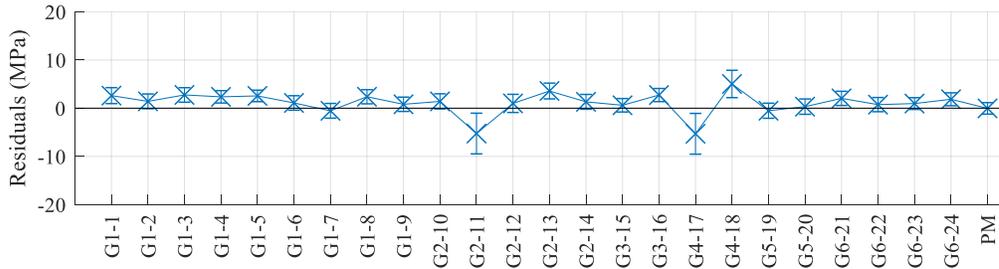
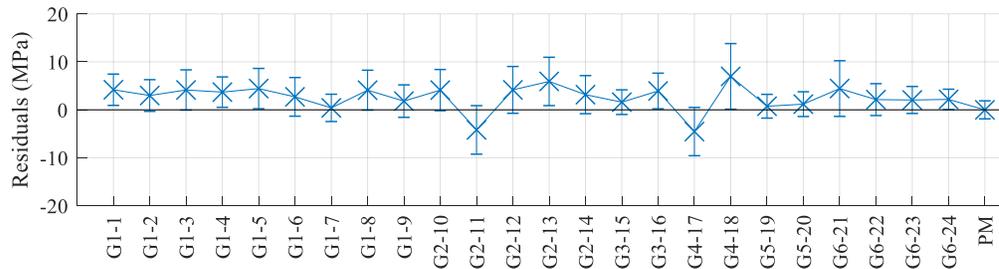
$$Y = \theta_0 + \sum_{i=1}^m \theta_i \hat{y}_i + \sigma \varepsilon \quad (10)$$

where Y = predicted failure pressure or a suitable transformation; θ_i = model parameters; \hat{y}_i = deterministic prediction from the existing prediction model P_i ; and $\sigma \varepsilon$ = model error in which σ is the standard deviation, assumed to be constant and ε is the standard normal random variable (normality assumption). When considering all the existing prediction models in Eq. (10) (i.e., $m = 24$), the model is a full model. Since

not all the terms contribute the model prediction, a model selection procedure is adopted to eliminate the ones that do not contribute statistically significantly. Note that the formulation of all the 24 existing models, P_i ($i = 1 \dots 24$) can be found in the 2nd Quarterly Report, and i refers to the second index of the model number in the 2nd Quarterly Report.

Model Development

Figure 15 shows the performance comparisons of the 24 existing prediction models at the three levels of σ_y , where the crosses refer to μ_{res} (mean of residuals), the horizontal lines refer to $\mu_{res} \pm \sigma_{res}$ (standard deviation of residuals), and solid dots are the MSE (mean squared error) values. As shown in Figure 15, the performance of the existing models changes with different levels of yield strength. Therefore, to reduce the model error, the proposed models are developed for the three levels of yield strength leading to three models adopting the full model as shown in Eq. (10). Each model is assessed based on the data in the corresponding level. In particular, randomly selected 80% of the data (or called training data) from the established database reported in the 2nd Quarterly Report is used for the model development, while rest of the data (i.e., 20% of the data) is used later for validation purposes.



• Level 1 □ Level 2 ○ Level 3

Figure 15: Comparison of residual and MSE of each model for three levels of σ_y

With the full model, an all possible subset model selection procedure is used to reduce the model size to determine the final formulation [4]. For a full model with a model size of 24, the size of the reduced model varies from 1 to 23. In all possible subset model selection, all possible combinations of predictors are evaluated for each model size (or subset) and the best model from that subset is identified. Then the best models from all the subsets are compared to determine the final model.

To compare the model performance, this study uses three statistics measures: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and standard deviation of model error (σ). Both AIC and BIC measures how well the model fits the data through log likelihood, $\log(L)$, with the consideration of the number of predictors used in the model, trading off the complexity of the model formulation with its accuracy. These two measures are calculated as below:

$$ACI = 2k - 2 \log(L) \quad (11)$$

$$BCI = 2 \log(n)k - 2 \log(L) \quad (12)$$

where k = number of estimated parameters in the model, and n = number of data points. The less value of AIC or BIC is, the better performance the model has. The standard deviation of model errors, σ , only measures the model accuracy; and the lower σ is, the better the model is.

For each subset (i.e., the possible models with the same model size), all the statistics measures advocate the same model as the best model. However, when comparing all the best models from all the subsets, these three statistics measures may suggest different models to be the most desirable one. Then engineering judgement has to be applied. Note that since in Eq. (10) each predictor is actually an existing deterministic model that may involve complex formulation already; thus, using number of predictors in the model as the measure of the model complexity in AIC and BIC oversimplified in this case. Consequently, the final model is determined by using engineering judgement for the complexity of the formula and using σ for the accuracy in this study.

Probabilistic models

Regardless the yielding strength level, it is found that when model size goes up to 3 or more, σ decreases insignificantly from model size of 2. This indicates that it is not beneficial to choose a model with model size larger than 2.

Table 2 compares the model accuracy of model sizes 1 and 2 in terms of model error, σ for three levels of yield strength of pipeline. As expected, σ value of model size 2 is smaller the one for model size 1; however, when such decrease becomes marginally, increasing model size is not beneficial. Then, the final model is decided if increasing the model size does not result in substantial decrease in σ . For instance, for Level 2, σ value decreases from model size 1 to model size 2 by 7%; however, the improvement is not significant enough to make it worth the complexity of adding more parameters from the additional existing model formula.

Table 2. Model selection

Yield strength level	Model size	Existing models selected in Eq. (1)	σ
Level 1	1	P_{24}	2.0139
	2	P_9, P_{15}	1.8781
Level 2	1	P_5	1.1995
	2	P_1, P_5	1.1149
Level 3	1	P_{16}	1.6998
	2	P_7, P_{18}	1.5132

Table 3 shows the final model formula of the three proposed models for the three yield strength levels, and the model parameter statistics. The existing models selected in the proposal formula are summarized below:

Phan et al. Modified NG-18 (referring to P_9 in this report and $G1-9$ in the 2nd Quarterly Report)

$$\hat{y}_9 = \frac{2t\sigma_u}{D} \left[\frac{1 - 0.92126 \left(\frac{d}{t}\right)}{1 - 0.92126 \left(\frac{d}{t}\right) M^{-1}} \right] \quad (13)$$

$$M = \left(1 + 0.06361 \left(\frac{l^2}{Dt}\right)^{2.75485} \right) \quad (13a)$$

PCORRC (referring to P_{15} in this report and $G3-15$ in the 2nd Quarterly Report)

$$\hat{y}_{15} = \frac{2t\sigma_u}{D} \left[1 - \frac{d}{t} \left(1 - \exp \left\{ -0.157 \frac{l}{\sqrt{r(t-d)}} \right\} \right) \right] \quad (14)$$

RSTRENG Effective Area (referring to P_5 in this report and $G1-5$ in the 2nd Quarterly Report)

$$\hat{y}_5 = \frac{2t}{D} (\sigma_y + 69[MPa]) \left[\frac{1 - A/A_0}{1 - (A/A_0)M^{-1}} \right] \quad (15)$$

$$M = \begin{cases} \sqrt{1 + 0.6275 \left(\frac{l^2}{Dt}\right) - 0.003375 \left(\frac{l^2}{Dt}\right)^2}, & l^2/Dt \leq 50 \\ 3.3 + 0.032 \left(\frac{l^2}{Dt}\right), & l^2/Dt > 50 \end{cases} \quad (15a)$$

Modified PCORRC (referring to P_{16} in this report and $G3-16$ in the 2nd Quarterly Report)

$$\hat{y}_{16} = \frac{2t(0.9\sigma_u)}{D} \left[1 - \frac{d}{t} \left(1 - \exp \left\{ -0.224 \frac{l}{\sqrt{r(t-d)}} \right\} \right) \right] \quad (16)$$

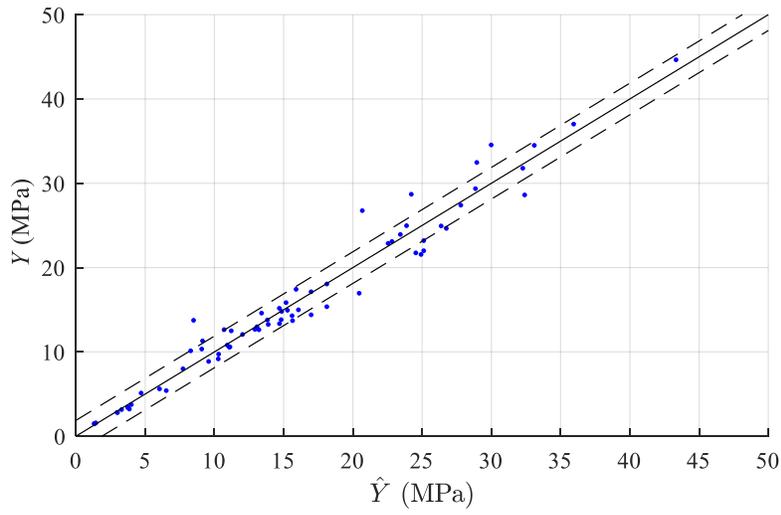
where D : outside diameter of the pipe, r : outside radius of the pipe, t : wall thickness of the pipe, d : maximum depth of the corrosion defect, l : length of the corrosion defect, A : longitudinal area of metal loss, A_0 : original uncorroded area of length l and thickness t , σ_u : ultimate tensile strength of the pipe material, and M : folias or bulging factor.

Table 3. Proposed models with parameter statistics

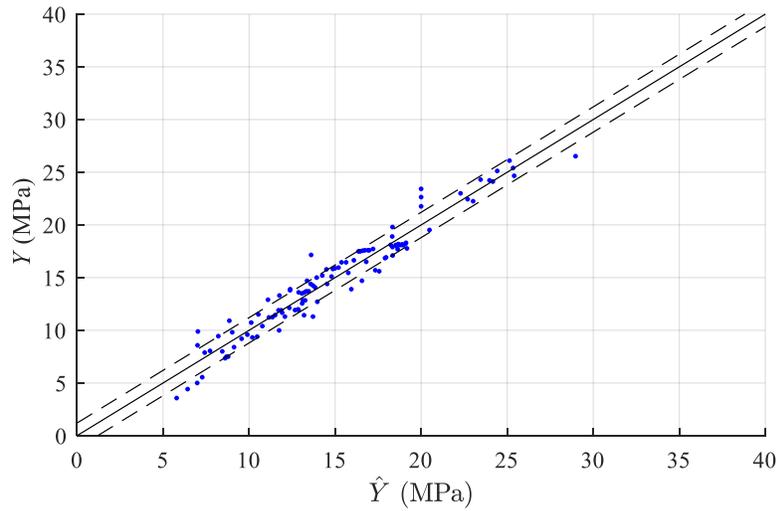
Yield strength level	Formula	Model Parameters						σ
		θ_0		θ_1		θ_2		
		Mean	Std	Mean	Std	Mean	Std	
Level 1	$\theta_0 + \theta_1 \hat{y}_9 + \theta_2 \hat{y}_{15}$	0.2768	0.4508	-1.3288	0.2266	2.4006	0.2207	1.8781
Level 2	$\theta_0 + \theta_1 \hat{y}_5$	2.5728	0.3133	1.0009	0.0237	-	-	1.1995
Level 3	$\theta_0 + \theta_1 \hat{y}_{16}$	1.8711	0.4859	1.1597	0.0275	-	-	1.6998

Model performance evaluation

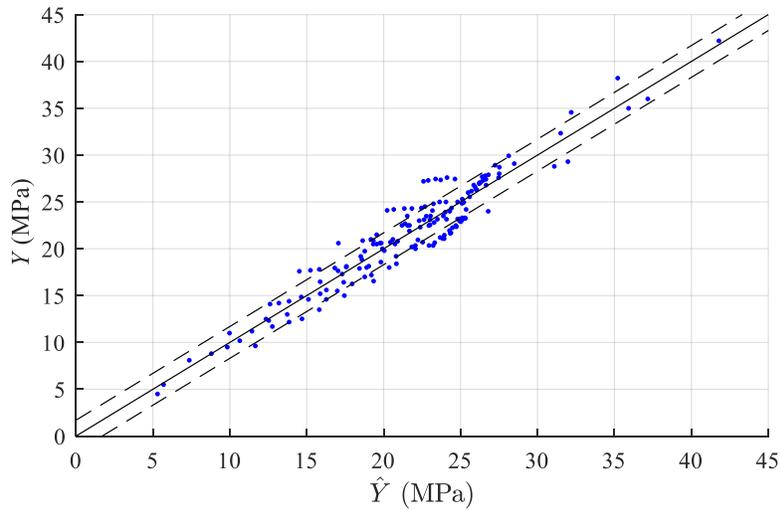
Figure 16 provides the scatter plots of the prediction of the proposed model, \hat{Y} vs. the observed data, Y that is the 80% data used for the model development. If the prediction is perfect, the dots should line up on the 1:1 line, shown as the solid line. The dashed lines are the mean ± 1 standard deviation of the model error. Figure 16 shows that the dots are evenly scatter around the 1:1 line, indicating that the developed model provides the unbiased prediction. The scatter of the dots indicates the accuracy of the model. For example, the plot for Level 3 shown in Figure 16(c) shows a larger scatter compared to the other levels shown in Figures 16(a) and 16(b).



(a) Level 1



(b) Level 2



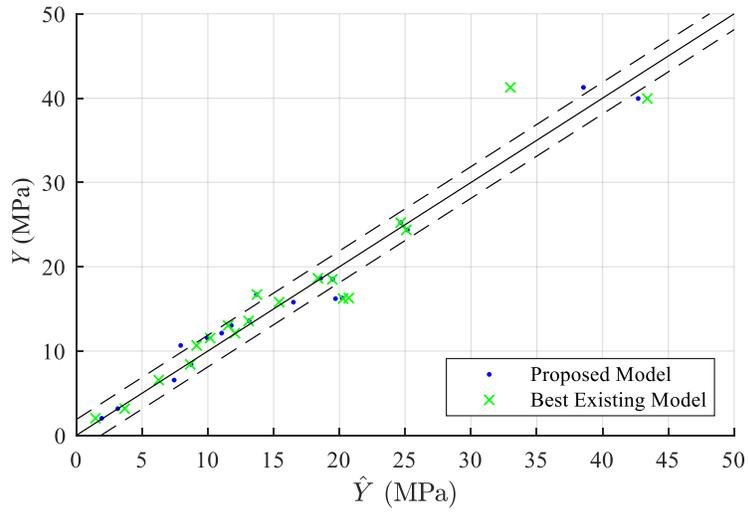
(c) Level 3

Figure 16: Scatter plots of the prediction of proposed model vs. the observed data (80% of data)

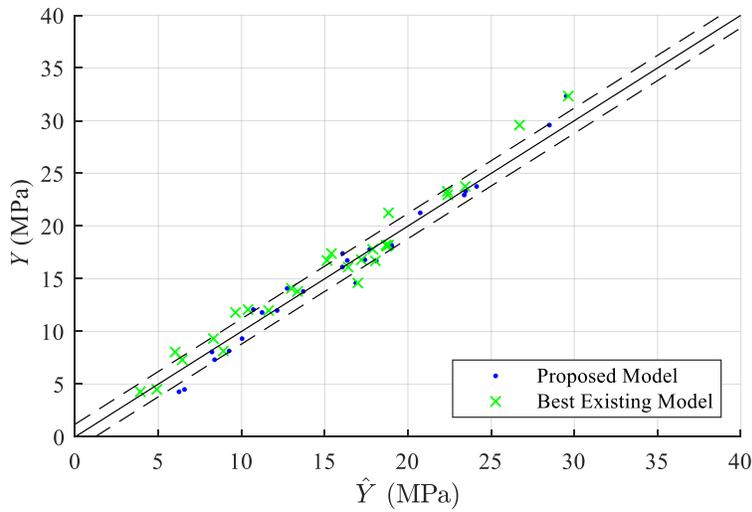
To evaluate the proposed model, Figure 17 shows the scatter plots of the prediction of the proposed model

\hat{Y} vs. the observed data, Y using the validation data (i.e., 20% of the data). The figure displays that the majority of the proposed model predictions using the validation data are within the mean ± 1 standard deviation of the model error, which validate the proposed model predication. In addition, the scatter plots of the prediction of the best existing models are compared with the proposed models in Figure 17. Note that the best existing model for each level of yield strength is the one with the lowest MSE value using the training data. It is found that the best existing models are P_{19} , P_{15} , and P_7 for Level 1, Level 2 and Level 3, respectively. As shown in Figure 3, the scatter plots of the best existing models also show that the dots are evenly scatter around the 1:1 line indicating that the best existing models can provide a certain degree of unbiased failure prediction. However, the scatter of the points based on the best existing models are slightly larger than the proposed models. This indicates that the proposed model improves the accuracy from the existing models.

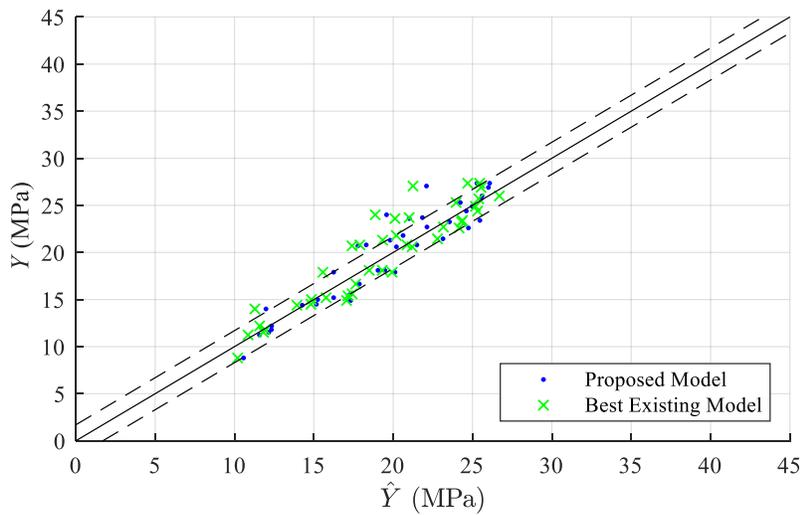
In addition, Figure 15 also shows the performance of the proposed models at the three levels of σ_y , compared with the existing models. A shown in Figure 15, regardless the levels, the proposed models are unbiased and have the lowest MSE. Note that the best existing models (i.e., P_{19} , P_{15} , and P_7 corresponding to the Levels 1, 2, and 3) are not necessarily selected in the proposed models.



(a) Level 1



(b) Level 2



(c) Level 3

Figure 17: Scatter plots of the prediction of proposed model and best existing model vs. the validation data (20% of data)

Reliability analysis

Failure of a pipe generally designates the event that the pipe integrity does not satisfy a specific set of functional requirements. Failure can consist of a complicated sequence of unfortunate events such as an external impact or loss of pipe integrity. There are different types of failure modes. In engineering design, the failure mode distinction is typically made based on different categories of design criteria, which are frequently referred to as limit states. The common categories for these limit state functions are Serviceability Limit State (SLS), the Ultimate Limit State (ULS) and the Fatigue Limit State (FLS) [5]. The probability failure, P_f , can be assessed through conducting the reliability analysis such as Monte Carlo simulations and First/Seconds Order Reliability Methods (FORM/SORM).

The probability of failure, P_f , can be written as:

$$P_f = \int_{\text{failure domain}} f(\mathbf{X}) d\mathbf{X} \quad (17)$$

where $f(\mathbf{X})$ is the joint probability density function of a vector of random variables, \mathbf{X} . The failure domain is formulated using the limit-state function, $g(\mathbf{X})$; thus, failure domain is defined as $g(\mathbf{X}) \leq 0$. In this study, failure is defined as occurring when the operating pressure of the pipe exceeds the burst pressure of the pipe; accordingly, the limit-state function is written as follow:

$$g(\mathbf{X}) = P_b - D_p \quad (18)$$

where P_b is the estimated failure pressure or burst pressure of the corroded pipe, and D_p is the operating pressure of the pipe. In practice, engineering adopts reliability index as performance evaluation, and a generalized reliability index is defined as [6]:

$$\beta = \Phi^{-1}(1 - P_f) \quad (19)$$

where Φ refers to cumulative distribution function of standard normal distribution.

The random variables used in the reliability analysis and their distribution information are listed in Table 4. Note that to have a fair comparison of the proposed models and the best existing models, the model errors of the best existing models are evaluated using the available database and are included in the reliability analysis. Table 4 indicates that the best existing models are slightly biased and have larger model errors compared to the proposed models.

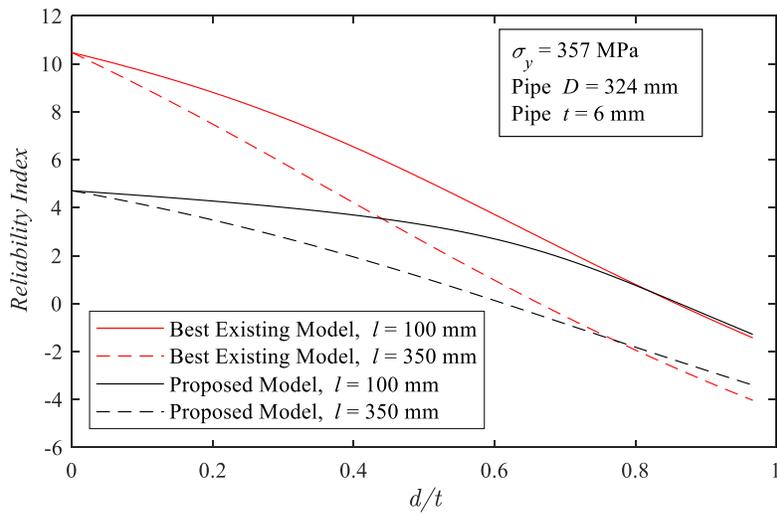
Table 4. Distribution parameters of random variables used in the reliability analysis

			Level 1		Level 2		Level 3	
Random variable	Distribution	COV (%)	Mean	Std	Mean	Std	Mean	Std
Outside diameter of pipe, D (mm)	Normal	5	324	16.2	324	16.2	324	16.2
Nominal wall thickness, t (mm)	Normal	5	6	0.3	6	0.3	6	0.3
Defect depth, d (mm)	Normal	5	-	-	-	-	-	-
Defect length, l (mm)	Normal	5	-	-	-	-	-	-
Yield strength, σ_y (MPa)	Normal	3	357	10.71	452	13.56	589	17.67
Ultimate strength, σ_u (MPa)	Normal	3	458	13.74	542	16.26	731	21.93
Operating Pressure, D_p (MPa)	Normal	5	7.61	0.38	9.64	0.48	12.57	0.63
Model error in the proposed model, $(\sigma\varepsilon)_p$	Normal	-	0	1.88	0	1.20	0	1.70
Model error in the best existing model, $(\sigma\varepsilon)_b$	Normal	-	0.90	2.41	0.61	1.42	0.24	1.87

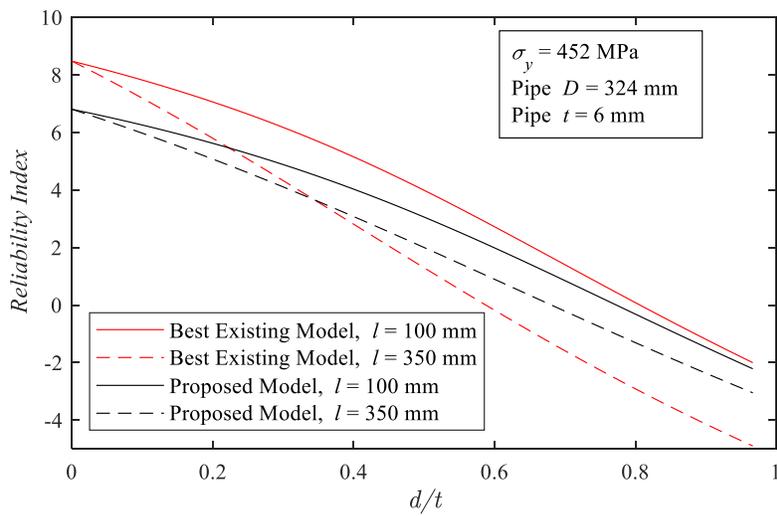
To evaluate the impact of the proposed models on the assessment of corroded pipelines burst pressure, Figure 18 compares the reliability index of the proposed models and the best existing models for the three

levels of yield strength for various levels of corrosion depths and lengths. As expected, the reliability index decreases with the increase of the defect depth on the pipe. The figure also indicates that for long defect length $l = 350$ mm, the reliability index is lower than the one for short defect length $l = 100$ mm. This result shows that both the length and depth of defects have a critical impact on the integrity of pipeline.

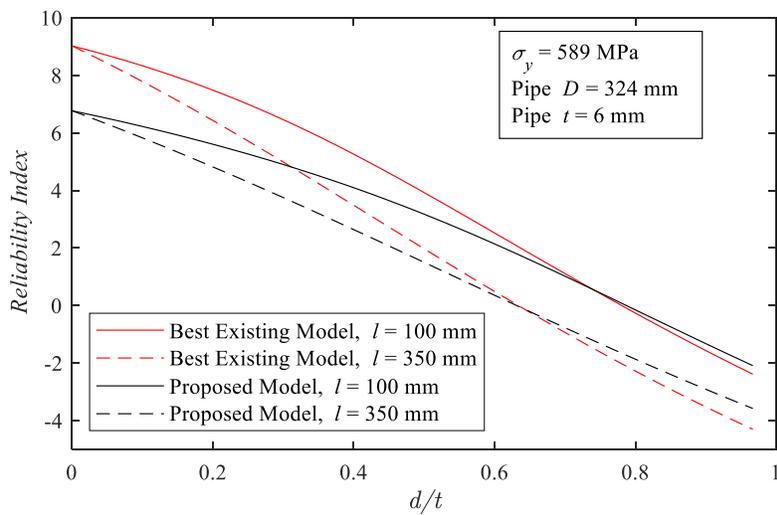
In addition, Figure 18 shows that overall the proposed models have a lower reliability index compared to the best existing models, indicating that the proposed models are more conservative. For example, for level 1 Figure 18(a) shows that for a defect depth of 50% of wall thickness and a defect length of 100 mm the reliability index of the proposed model and best existing model are about 3.3 and 5.2 respectively. In this case, using the best existing model may cause delay of the pipeline maintenance and repair, which may lead to unexpected pipeline failure with tremendous consequences (both economically and environmentally). However, for Level 2 Figure 18(b) shows that the proposed model can be more conservative than the best existing model for a long defect length ($l = 350$ mm) after the defect depth reaches about 35 % of the wall thickness. Furthermore, the rate of reliability decrease based on the proposed model is smaller than the best existing models for all three levels. Overall, one can conclude the prediction model plays a critical role in determining the reliability performance of corroded pipelines.



(a) Level 1



(b) Level 2



(c) Level 3

Figure 18: Defect depth-dependent reliability index based on the proposed models and best existing models

3.3 Future Work (Next Quarter)

In the next Quarter, the research team will continue working on Task 3. In Task 3a, currently we have started collecting the colony defect database from literature review. The FE model will be extended from isolated corrosion defect to colony corrosion defect, and additional data from the numerical analysis will be generated to be added to the database. In Task 3b, the research team will evaluate the prediction performance of existing prediction models based on the collected data.

References

- [1] “Three-Dimensional Structural Health Monitoring Based on Multiscale Cross-Sample Entropy | Request PDF.” n.d. ResearchGate. Accessed July 1, 2020.
https://www.researchgate.net/publication/319095129_Threedimensional_structural_health_monitoring_based_on_multiscale_cross-sample_entropy.
- [2] J. S. Richman and J. R. Moorman, “Physiological time-series analysis using approximate entropy and sample entropy,” *Am J Physiol Heart CircPhysiol, American Journal of Physiology-Heart and Circulatory Physiology*, 278, pp. H2039-H2049, 2000.
- [3] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press; 2016 Nov 10.
- [4] S. Sheather. *A modern approach to regression with R*. Springer Science & Business Media, 2009.
- [5] B. J. Leira, *Optimal stochastic control schemes within a structural reliability framework*, Springer Briefs in Statistics, 2013, doi: 10.1007/978-3-319-01405-0_2
- [6] O. Ditlevsen, and H. O. Madsen. *Structural reliability methods*, Chichester, UK, Wiley, 1996.