**CAAP Quarterly Report**

**September/30/2022**

*Project Name: Pipeline Risk Management Using Artificial Intelligence-Enabled Modeling and Decision Making*

*Contract Number: 693JK32150001CAAP*

*Prime University: Rutgers University*

*Prepared by: Bingyan Cui (PhD student), Emad Farahani (PhD student), Dr. Hao Wang (PI), Dr. Qindan Huang (Co-PI)*

*Reporting Period: 7/1/2022 – 9/30/2022*

**Project Activities for Reporting Period:**

*Data-Driven Probabilistic Modeling of Pipeline Defect Generation and Growth*

The time-dependent performance evaluation of buried pipelines considering corrosion could help develop pipeline integrity corrosion management strategies, where corrosion evolution needs to be understood. The corrosion-induced damage evolution on pipelines is known to be influenced by many factors such as the physical and mechanical properties of the pipeline as well as the pipeline’s surrounding environment. The objective of this task is to develop a probabilistic predictive corrosion growth model of buried steel pipelines using in-line inspection (ILI) obtained from pigging technology and soil survey data. Such growth model thus reflects the impact of soil properties, which can then be used to evaluate the performance of non-piggable buried pipelines. While the growth model will be developed for a specific pipeline, the framework is applicable to other buried pipelines.

# Description of the Studied Pipeline and Inspection Data

The studied on-shore pipeline with a total length of 112 km has been in service since 1969 near the Gulf of Mexico and Table 1 shows basic characteristics of the studied pipeline. The pigging technology has been applied to detect the pipeline wall thickness loss (i.e., corrosion defects) at six different times. Table 2 shows the number of defects found in each pigging. Figure 1 further indicates the pigging location and time.

Although the pipeline has been inspected six times, each time the application areas were not the same. Overall, there are overlapping areas between Jan. 2005 pigging and Mar. & May 2010 pigging, and there are overlapping areas between Oct. & Dec .2005 pigging and Dec. 2011 pigging. It is noted that different pigging technologies were employed at two inspections, i.e., MLF technology in 2005 and straight beam ultrasound technology in 2010. Ideally and theoretically, the pigging should inspect the same defects that have been detected in the previous pigging if these two pigging detections go through the same areas. However, the increase trend of defect size was not always found at each location for the inspected pipeline. The variations can be caused by 1) the changes in measurement accuracy, tolerance, and reporting criteria of ILI tools and the advances in technology in different inspections, and 2) the mismatch of absolute distance from odometer readings during different inspections. This issue will be further checked and investigated with industry partner.

To be conservative, it is assumed that all the defects detected by the pigging inspection are the true defects and the inspection error is ignored in the analysis here due to the lack of relevant information on tool accuracy or tolerance. Furthermore, the measured depth was considered as the maximum depth of the defect.

Table 1. Pipeline characteristics

|  |  |
| --- | --- |
| Total Length | 122 km |
| Outside Diameter | 457.20 mm (18 in.) |
| Nominal Wall thickness | 6.40 mm (0.252 in.) |
| Steel type | API 5L X52 |
| Minimal yielding stress | 52 ksi |

Table 2. Summary of number of defects found in pigging

|  |  |
| --- | --- |
| Pigging date | Total number of defects |
| Jan 2005 | 794 |
| Oct 2005 | 109 |
| Dec 2005 | 43 |
| Mar 2010 | 1613 |
| May 2010 | 331 |
| Dec 2011 | 65 |



Figure 1. Pigging time vs. pigging location

# Damage Evolution Modeling

As the length of the studied pipeline is 112 km, the growth of the corrosion damage is varied with the length, as the environmental conditions of the pipeline changes with the length. To capture this change, there are two potential approaches:

1. One could divide the whole length of the pipeline into *n* separate segments, and for each segment, the corrosion growth can be described by the same damage evolution model. Obviously, this option simplifies the phenomenon and assumes the homogenous corrosion behavior within each segment;
2. Another approach, which is adopted here, is to develop one model for whole length of the studied pipeline, but incorporating influential variables of the pipeline’s surrounding environment in the model.

To describe the damage (corrosion defect) growth over time, the damage state at a time instant and/or damage rate (corrosion rate) are usually used. To evaluate the corrosion rate, two damage states (defect dimensions) for the same defect are needed over time. However, as discussed in the previous section, the matched defects are lacking in the available field inspection data in this study, and this limitation is found to be pretty common. Thus, one can build the model to predict the damage state directly, when either matched or nonmatched defects data are available.

In this study, for both corrosion maximum depth and length, a power-law function of time model formulation, as expressed in Equation (1), is adopted which considers nonconstant damage growth rate over time. To incorporate the environmental impact, the model parameters will be considered through a linear or quadratic function of the field measured physicochemical variables of the soil, including soil moisture, pH, resistivity, redox potential, etc., along the pipeline.

|  |  |
| --- | --- |
|  | (1) |
|  | (1a) |
|  | (1b) |

Where, *m* = types of defect quantity (e.g., *m* = *D* for the maximum defect depth and *m* = *L* for the maximum defect length), *Ym*= defect quantity (e.g., maximum defect depth or defect length) at a time instant *t*,**θ** = unknown model parameters, *g*i and *hj* influencing environmental variables, *t*0m *=* initiation time of each defect, *εm* = Normal random variable with zero mean and unit variance, and *σm* = standard deviation of the model error.

A Poisson process is considered for the occurrence of defects, indicating the initiation time of each individual defect is considered to follow a Gamma distribution. Therefore, the proposed methodology does not assume uniform (equal) corrosion initiation time for all the defects, which can, as a result, predict the number of newly generated defects since last inspection. To evaluate *t*0*m,* we assume the number of defects that occurs follows the same homogeneous Poisson process characterized by a rate parameter *λ*. Additionally, it is assumed that a defect with a larger detected dimensions occurs earlier than other ones; accordingly, the defects are sorted based on their detected dimension values and the initiation time of each defect follows a Gamma distribution, where the scale parameter, *β*, is treated as an unknown model parameter to be estimated and the shape parameter, *α*, is assumed to be the ranking of each defect.

Bayesian statistics (Box and Tiao 1992) is used to assess the joint probability density function (PDF) *p*(**Θ**) of the unknown model parameters, **Θ**, used in the damage evolution model. If **X** denotes the vector of data used to update the model parameters, the posterior joint PDF of **Θ**, *p′*(**Θ**), can be written as

|  |  |
| --- | --- |
|  | (2) |

Where, , and likelihood function. As the model error in Equation (1) follows a Normal distribution after applying appropriate transformation of the defect quantity, the likelihood function can be written as a bi-variant Normal distribution considering the correlation between depth and length model error standard deviations (Miran et al. 2016). Here, for length model logarithm transformation was used to ensure normality of the model error distribution (Miran et al. 2016). However, calculating the normalizing factor can be challenging, especially when the dimension of ***X*** is high. To effectively compute the posterior statistics, one can use a sampling-based technique like the Markov Chain Monte Carlo (MCMC) simulations (Gilks et al. 1996). The MCMC technique generates a sequence of random variables called Markov chain such that the current value or state of the sequence depends only on the previous one. Given certain conditions, the chain will forget its initial state and converge to a stationary distribution. For the convergence criteria, the Geweke method (Geweke 1992) is used in this study, which provides unbiased estimates of the posterior statistics.

# Preliminary Results and Discussion

So for corrosion depth and length growth models are developed for the whole length of the pipeline, where constructed likelihood function considers the correlation between defect depth and length growth, and the Bayesian inference framework is implemented in MATLAB. The results will serve as a benchmark to evaluate the further improvements made by incorporating the environmental factors in the model parameters. Therefore, the Equation (1) is reduced to

|  |  |
| --- | --- |
|  | (3) |

Table 3 summarizes the statistics of the model parameters posterior distribution. As implied by the values of *geweke* in the last column close to unity, the number of simulations in the Bayesian updating process has been large enough and the results have been converged and reach a stationary state. The posterior distributions of the parameters are also shown in Figure 2. In addition, the posterior distribution of the depth and length models error standard deviation, which were also considered as model parameters, are shown in Figure 3. The correlation between depth and length model errors is estimated to be around 5%.

Table 3. Posterior distribution statistics of model parameters in the damage evolution model

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model for | Model parameter | Mean | Std. | Median | geweke |
| Depth | *θ*1*D* | 0.116 | 0.051 | 0.105 | 0.921 |
|  | *θ*2*D* | 0.786 | 0.118 | 0.792 | 0.985 |
|  | *βD* | 0.570 | 0.051 | 0.572 | 0.996 |
|  | *σD* | 0.286 | 0.007 | 0.285 | 0.996 |
| Length | *θ*1*L* | 0.162 | 0.066 | 0.146 | 0.953 |
|  | *θ*2*L* | 0.831 | 0.099 | 0.840 | 0.992 |
|  | *βL* | 0.170 | 0.019 | 0.167 | 0.984 |
|  | *σL* | 2.255 | 0.054 | 2.248 | 0.995 |



Figure 3. Posterior distribution of model parameters

|  |  |
| --- | --- |
|  |  |
| (a) | (b) |

Figure 3. Posterior distribution of model error standard deviation: a) depth, *d*, and b) ln(*l* + 1)

Figure 4 also shows the predicted corrosion damage evolution over time for all inspected defects. As can be seen the initiation time for each defect is unique, and also the corrosion rate tends to decrease over time, as expected. Finally, the predicted corrosion depth and length and the associated actual measured quantities obtained through ILI are compared in Figure 5. For a perfect prediction model, the predicted data should line up along the 1:1 line. One can see from Figure 5 that most of the predicted data are located around the 1:1 line within the ± 1 standard deviation band (i.e., dashed lines). This indicates that the proposed defect growth models provide unbiased prediction with sufficient accuracy despite only one single growth model is used for the whole length of the pipeline.



Figure 4. Predictive corrosion depth evolution for all inspected defects over time



(a)



(b)

Figure 5. Comparison between predicted versus measured defect depth and length: a) depth and b) length

As shown in other studies (including laboratory experiments and filed survey data analysis) that corrosion rate is usually high at the beginning and then decreases over time, because the corroded metal film formed tend to prevent the further corrosion of the pipe. Therefore, the mathematical modeling of the corrosion growth takes into account this behavior. The power-law function of time employed here in and expressed in Equations (1) and (3) can be used to capture such phenomena by setting the power parameter bounded between zero and one.

**Project Activities with Cost Share Partners:**

Cost share is provided by Rutgers University and Marquee University during this quarterly period as budgeted in the proposal.

**Project Activities with External Partners:**

N/A

**Potential Project Risks:**

N/A

**Future Project Work:**

The research team will continue working on Data-Driven Probabilistic Modeling of Defects.

1. The soil parameters will be incorporated into the growth models; in which we will investigate the most appropriate model formulation to incorporate those influencing variables. In addition, we will also search for the appropriate transformation of the damage quantity, both depth and length, to ensure the assumption satisfaction of Normal model error distribution.
2. We will develop Bayesian Neural Network (BNN) based algorithm to predict defect growth considering soil parameters as input features. Compared to Gaussian process regression to establish Bayesian prediction models, BNN is more suitable for dataset with high-complexity and high-dimensionality.

**Potential Impacts to Pipeline Safety:**

The in-line inspection data will be used to develop probabilistic growth models of pipeline defects, which can aid pipeline operators better predict failure risk and make repair decisions.