Quarterly Report – Public Page

Date of Report: 1st Quarterly Report-December 27, 2019
Contract Number: 693JK31910018POTA
Prepared for: DOT PHMSA
Project Title: Mapping Indication Severity Using Bayesian Machine Learning from Indirect Inspection Data into Corrosion Severity for Decision-Making in Pipeline Maintenance
Prepared by: TEES (Texas A&M Engineering Experiment Station), University of Dayton
Contact Information: Homero Castaneda, hcastaneda@tamu.edu, 979 458 9844.
For quarterly period ending: December 29th, 2019

1: Items Completed During this Quarterly Period:

Per the contract, Task 1 is associated with the first quarterly report. The format of the data frame has been designed. Inspection data (including both direct and indirect) and soil information have been collected to start the filtering and check process for consistency, but it is not complete. The data processing is still on-going due to the time taken for the data acquisition.

Item #	***Task #***	***Activity/Deliverable/Title***	***Federal Cost***	***Cost Share***
1	1	Download remote sensing, topography, and geological data of the selected regions (i.e., right-of-way of the selected pipeline interval);	$14,000.00	$5,000.00

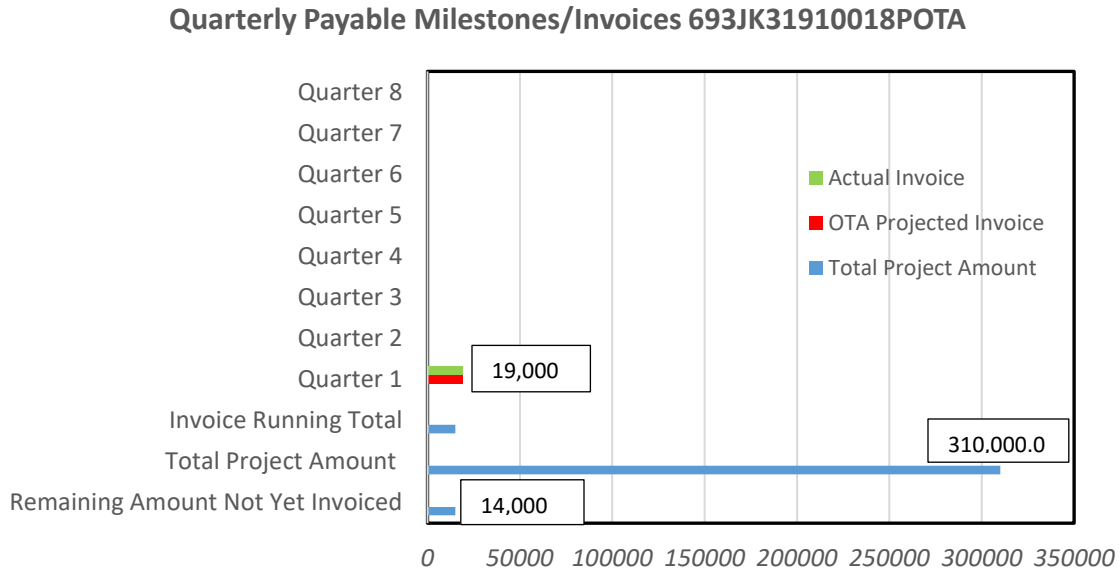
2: Items Not-Completed During this Quarterly Period:

Task number 2, collection and preprocess of the indirect and direct surveys. Potential selection of more pipeline sections to cover more characteristics for the model to be develop. Part of Task 2 will be cover in the following partial report.

Item #	***Task #***	***Activity/Deliverable/Title***	***Federal Cost***	***Cost Share***
2	1	Preprocess the indirect and/or direct measurements from the industry partners	$10,000.00	$3,000.00

3: Project Financial Tracking During this Quarterly Period:

The figure below provides details on the overall financial status of the project.



4: Project Technical Status –

The following tasks are included in the project:

- *Task 1:* **Establishing a database**
- *Task 2:* **Experiments and analyses to bridge gaps in prior knowledge**
- *Task 3:* **Bayesian machine learning to bridge gaps in uncertainty quantification.**
- *Task 4* **Finalize and evaluate/validate the model.**

During the kick off meeting we started to cover task 1
We performed a project discussion with internal team (5 people) and later with industrial partners to seek required information input and select the most representative database and systems that could be relevant to this project. We have discussed the protocol for data inquiry, data content, data transferring, and data management with our industrial partners. During the internal meeting, we discussed 1) the potential tools (spreadsheet and open-source programming language *Python* equipped with its data processing package *Pandas*) for data processing, and 2) the data format with key information needed for the following analysis. Some details are listed below:

Task 1 – Establishing a database. The selected database was aligned, screened, and formatted. The digital database was generated from direct, indirect, and survey technologies with sub-meter location, in-line inspection, and direct characterization. The initial data frame has been created.

We collected and align different sets of direct and indirect inspection databases for RoW conditions with macro measures of soil type, and topography from database sources. This step included the ground check for completeness data quality/consistency. This task include the raw data of two different underground steel pipelines sections. The following characteristics were included in the database (Table 1):

Table 1: Basic information of the selected pipeline systems

Pipeline Section	37.5 miles X65 API steel	68 miles X52 steel
Thickness	0.343 in	0.374 in
Diameter	30 in	18 in
Welded	SMAW	SMAW
Indirect Inspections	CIS, DCVG, resistivity, pH, potential	CIS, DCVG, resistivity, pH, potential
Survey profiles	Soil composition	Soil composition
CIS Inspections	Year 1 , Year 2	Year 1 , Year 2
DCVG Inspections	Year 1 , Year 2	Year 1 , Year 2
Direct Inspections	ILI (NDT, MFL)	ILI (MFL)
Years in operation	40 years	50 years
Coating type	Coal tar	Coal tar
Protection system	CP impressed current	CP impressed current
Atmosphere conditions	Heavy rain area and high temperatures	Heavy to moderate rain and medium high temperatures
Topography	Different slope profiles	TBD

The metrics of corrosivity (resistivity, soil chemistry, pH, redox potential, and water content) have been queried, integrated and aligned with field measurements. The two sets of database included such magnitudes. The raw datasets are originally stored in spreadsheets and have different data formats. We screened and extracted the relevant columns, jointed and matched corrosion defect records, and then performed data cleaning and filtering to eliminate erroneous and outlier records. The screen shots of the created initial data frame is shown below.

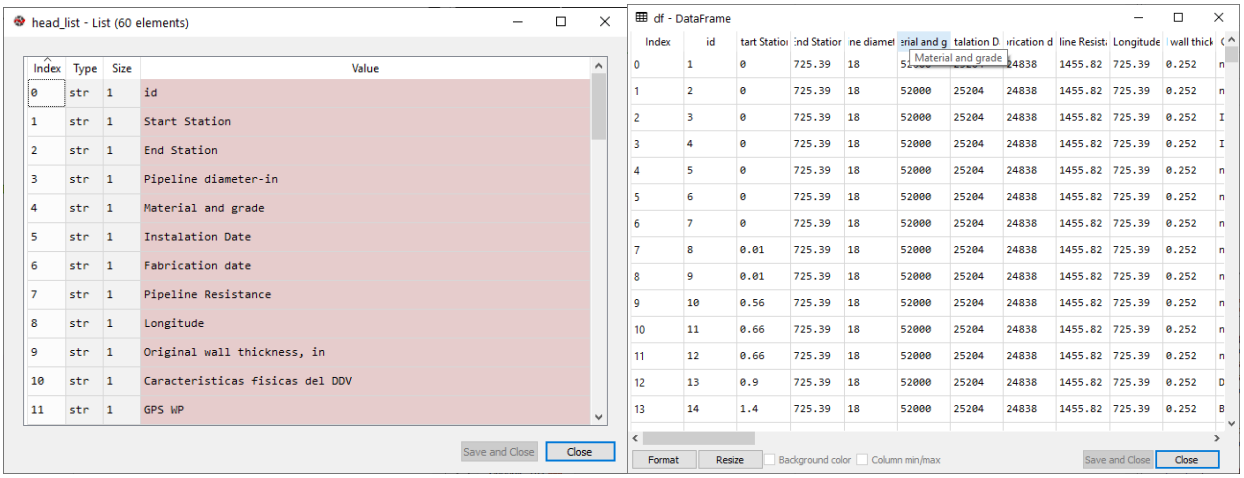


Figure 1. Sample screen shots of the initial data frame

The data frame is currently not completed subject to overlaying temporal rainfall, temperature data from the National Oceanic and Atmospheric Administration, and other operational parameters affecting the coating conditions (raining average, temperature, and other indicators) along with CP data. So far, the elevation profile (Figure 2) of one selected system is known, and base on which

we can start to explore and move on to the next step, i.e., converting the geo-referenced database to an object-based database for performing the following Bayesian machine learning analysis.

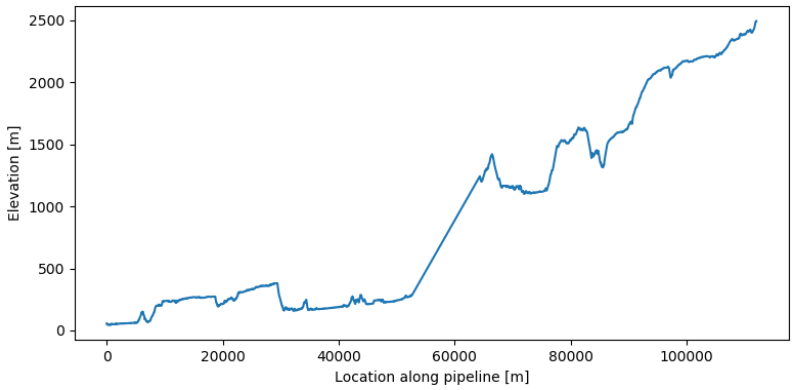


Figure 2. The elevation profile of a selected system

5: Project Schedule –

The project is on-schedule as originally-proposed.