5TH QUARTERLY REPORT - PUBLIC
DOT Project Number: 729 (GTI 22428)
DOT Contract Number: 693JK31810003
OTD Project Number: 4.14.c.2 (GTI 22429)

Validating Non-Destructive Tools for Surface to Bulk Correlations of Yield Strength, Toughness, and Chemistry

Reporting Period
August 1, 2019 through October 31, 2019

Report Issued
October 31, 2019

Prepared For
U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration (DOT/PHMSA)
Joe Pishnery
404.832.1151
joseph.pishnery@dot.gov

Operations Technology Development (OTD)
Michael Adamo
847.544.3428
michael.adamo@otd-co.org

Technical Lead
Daniel Ersoy
Element Resources, LLC
847.343.9755
dersoy@elementresourcesllc.com

GTI Technical Team
Daniel Ersoy, Brian Miller, Marta Guerrero Merino, Ernest Lever

1700 S. Mount Prospect Rd.
Des Plaines, Illinois 60018
www.gti.energy

LEGAL NOTICE

This information was prepared by Gas Technology Institute ("GTI") for U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration.

Neither GTI, the members of GTI, the Sponsor(s), nor any person acting on behalf of any of them:

Makes any warranty or representation, express or implied with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately-owned rights. Inasmuch as this project is experimental in nature, the technical information, results, or conclusions cannot be predicted. Conclusions and analysis of results by GTI represent GTI's opinion based on inferences from measurements and empirical relationships, which inferences and assumptions are not infallible, and with respect to which competent specialists may differ.

Assumes any liability with respect to the use of, or for any and all damages resulting from the use of, any information, apparatus, method, or process disclosed in this report; any other use of, or reliance on, this report by any third party is at the third party's sole risk.

The results within this report relate only to the items tested and/or reviewed.

TABLE OF CONTENTS

Page

LEGAL NOTICE ..... **i**

TABLE OF CONTENTS ..... **ii**

LIST OF FIGURES ..... **iv**

LIST OF TABLES ..... **iv**

PROJECT OBJECTIVES AND REPORT SUMMARY **5**

Project Objectives 5

Report Summary 5

Task 2 - Develop Project Database and Pipeline Sample Library 5

Task 3 – Develop and Execute Testing Matrix 5

Task 4 – Data Analysis and Model Development and Optimization (see Appendix B) 6

Task 6 – Project Management 6

Risks 6

BUSINESS AND MILESTONE STATUS (Project Quarter #5: 8/1/2019 to 10/31/2019) **7**

Tasks Scheduled in this Reporting Period 7

Milestones for Activities/Deliverables Completed 7

Resource Status 8

Project Risks 9

Schedule Update 10

TECHNICAL STATUS (Project Quarter #5: 8/1/2019 to 10/31/2019) **11**

Task 2 - Develop Project Database and Pipeline Sample Library 11

Task 3 - Develop and Execute Testing Matrix 11

Task 4 - Data Analysis and Model Development and Optimization 14

Task 6 - Project Management 14

PLANS FOR FUTURE ACTIVITY (Project Quarter #6: 11/1/2019 to 1/31/2020) **15**

Task 2 - Develop Project Database and Pipeline Sample Library 15

Task 3 - Develop and Execute Testing Matrix 15

Task 4 - Data Analysis and Model Development and Optimization 15

Task 6 - Project Management 16

APPENDIX A – Project Testing Plan/Matrix (M9) **17**

APPENDIX B – ASU TASK 4 (5TH Quarter Update) **26**

Abstract **27**

Introduction **28**

Bayesian Inference with Missing Data **29**

Ignorability 29

Data augmentation 30

Results and Discussion **32**

Dataset with missing data 32

Results 32

General predictive performance 33

Conclusions **35**

References **35**

LIST OF FIGURES

Page

Figure 1: Quarterly Payable Milestones/Invoices – DOT PHMSA 8

Figure 2: Quarterly Payable Milestones/Invoices – OTD and ASU 9

Figure 3: Project Schedule 10

Figure 4: Box Plot of M.E. vs. V.E. Yield Strengths (ksi) 12

Figure 5: Normality Plot of Percent Difference Between M.E. and V.E. 13

Figure 6: Prediction results 33

Figure 7: Root mean square errors of 100 times validations. 34

LIST OF TABLES

Page

Table 1: Lists of Tasks in this Reporting Period 7

Table 2: Lists of Milestones Linked to Activities/Deliverables/Tasks this Reporting Period ... 7

Table 3: Percent Difference Yield Strength (ksi) for Mechanical v. Video Extensometer 12

Table 4: Project Testing Plan 17

Table 5: Dataset with missing data. 32

Table 6: Root mean square errors 33

Table 7: Mean root mean square errors over 100 times validation. 34

PROJECT OBJECTIVES AND REPORT SUMMARY

Project Objectives

The deliverables of this project will facilitate the use of non-destructive surface testing: micro-indentation, micro-machining, in situ chemistry, and replicate microscopy analysis as accurate, efficient, and cost-effective tools for material property confirmation.

This work will provide benefits to pipeline safety, energy continuity, and integrity assessment programs since the developed techniques and models and validated testing technology will not require a line to be taken out of service or destructively cut out samples from the in-service pipeline.

The results of this project will also be applicable to pending DOT/PHMSA regulations that require operators to backfill their material property records for grandfathered pipeline segments and/or those that do not have adequate material records.

Report Summary

During the fifth project quarter the following items were completed (see the **Technical Status** section for greater detail):

Task 2 - Develop Project Database and Pipeline Sample Library

- GTI continued the development of the project sample library. The overall library has been established with a robust set of 70 pipeline samples and has been divided into two groups of 35 pipe samples to allow the Task 4 modeling activities to continue model development with half the project results while the other half of the testing is being completed.
- Reorganized the Master Flat Sheet into a digital analysis "ready" format conducive to importation into statistical analysis software for the upcoming Task 4 activities.

Task 3 – Develop and Execute Testing Matrix

- GTI completed the Project Testing Plan/Matrix (Milestone Number 9), see Appendix A.
- GTI completed a statistical analysis check of the difference of tensile results on five diverse pipeline samples using a mechanical extensometer (M.E.) vs. a video extensometer (V.E.). On average, the %-difference on the yield strength between the V.E. and M.E. is +0.97%. The 90% double-sided confidence interval is 2.98% wide between the 5% and 95% confidence limits. The testing and analysis showed that the methods are statistically and practically equivalent.

- Completed the data entry from Nondestructive Test Technology into the project Master Flat Table, MFT (Database)
- The NDE testing phase of the project from MMT, Frontics, and SciApps has now been completed with the exception of the 30 coupons that will be tested later in the project at the MMT labs using their prototype NDTT toughness testing system (see Task 2 details above).
- GTI received MMT's draft summary report and remaining test data (of bulk tensile predictions and weld classifications). The bulk tensile property predictions were done using the standard MMT linear regression mode, as well as predictions from two new models: a Bayesian regression model and an artificial neural network (ANN) model.
- This lab testing continued in Quarter 5, which included full wall tensile, hardness, chemistry, and grain size.
- During this quarter GTI achieved a milestone by completing 35 of the 70 project samples, both nondestructive and destructive. Final quality checks are being performed prior to the results being turned over to the Task 4 Modeling effort.

Task 4 – Data Analysis and Model Development and Optimization (see Appendix B)

- A detailed report of this quarter's modeling efforts and results is presented in Appendix B of this report.
- This research this quarter demonstrated a general Bayesian methodology to handle missing data (typical in practice) when estimating the strength of pipeline samples. The missing data can be caused by various reasons, e.g., experimental abnormalities, wrong documentation, human errors.
- It was shown that the proposed Bayesian multimodality inference with missing data can successfully handle this issue using the provided data. Further, the proposed method has better/similar performance with surface-only indentation technique and linear regression with complete data.

Task 6 – Project Management

- Submitted Milestone Item M9 (Appendix B of this report) - Project Testing Plan/Matrix.
- Submitted Milestone Item M10, the fifth project quarterly report.

Risks

- There are no current risks to scope, schedule, budget.

BUSINESS AND MILESTONE STATUS
(Project Quarter #5: 8/1/2019 to 10/31/2019)

Tasks Scheduled in this Reporting Period

Table 1 shows the project tasks performed in this reporting period. All tasks are on time or ahead of schedule.

Table 1: Lists of Tasks in this Reporting Period

Task	Scheduled Start	Scheduled Completion	Completion Status
Task 2 - Develop Project Database and Pipeline Sample Library	11/1/2018	1/31/2020	Ongoing (1)
Task 3 - Develop and Execute Testing Matrix	5/1/2019 (2)	10/31/2020	Ongoing
Task 4 - Data Analysis and Model Development and Optimization	8/1/2019 (2)	1/31/2021	Ongoing
Task 6 - Project Management	8/1/2018	4/30/2021	Ongoing
- 5TH Quarter Project Report	8/1/2019	10/31/2019	Completed

(1) GTI is ahead of schedule on Task 2 due to the acceleration of Task 3 (see note (2) below this one).

(2) GTI started Tasks 3 early to allow early coordination with NDE subcontractors and service providers; ASU and GTI started work in Task 4 early to develop a framework for model experimentation and testing of model performance with the calibration data set provided early by GTI. Task 3 is additionally ahead of schedule. Task 4 is on schedule.

Milestones for Activities/Deliverables Completed

Table 2 shows the project milestones for activities/deliverables/tasks completed in this reporting period. All Milestones/Deliverables are on time or ahead of schedule.

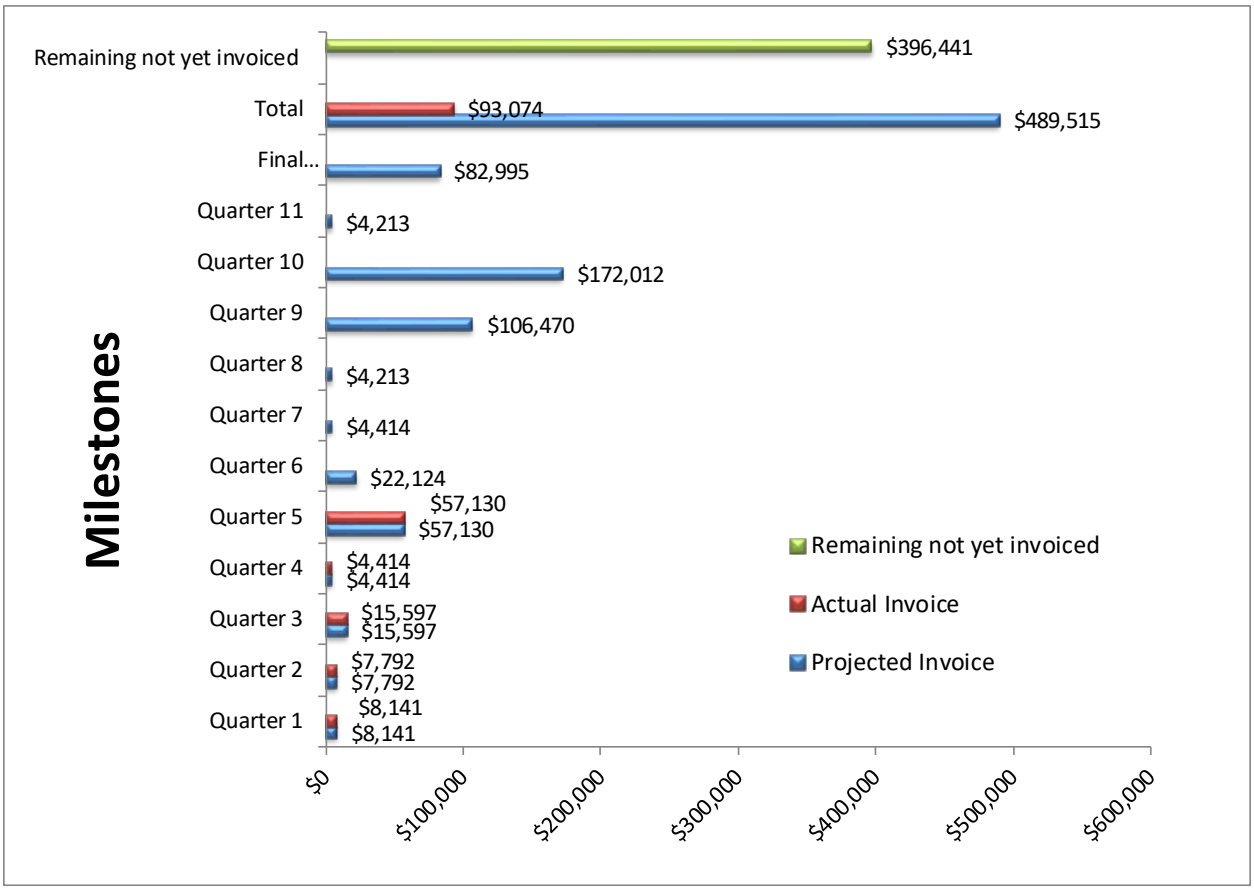
Table 2: Lists of Milestones Linked to Activities/Deliverables/Tasks this Reporting Period

Item No.	Task No.	Activity/Deliverable	Quarter No.	Scheduled Due Date	Completion Date	Payable Milestone/Title
9	3	Develop Testing Matrix and Execute Testing	5	10/31/2019	10/31/2019	Submit project testing plan/matrix
10	6	5th Quarterly Status Report	5	10/31/2019	10/31/2019	Submit 5th quarterly report

Resource Status

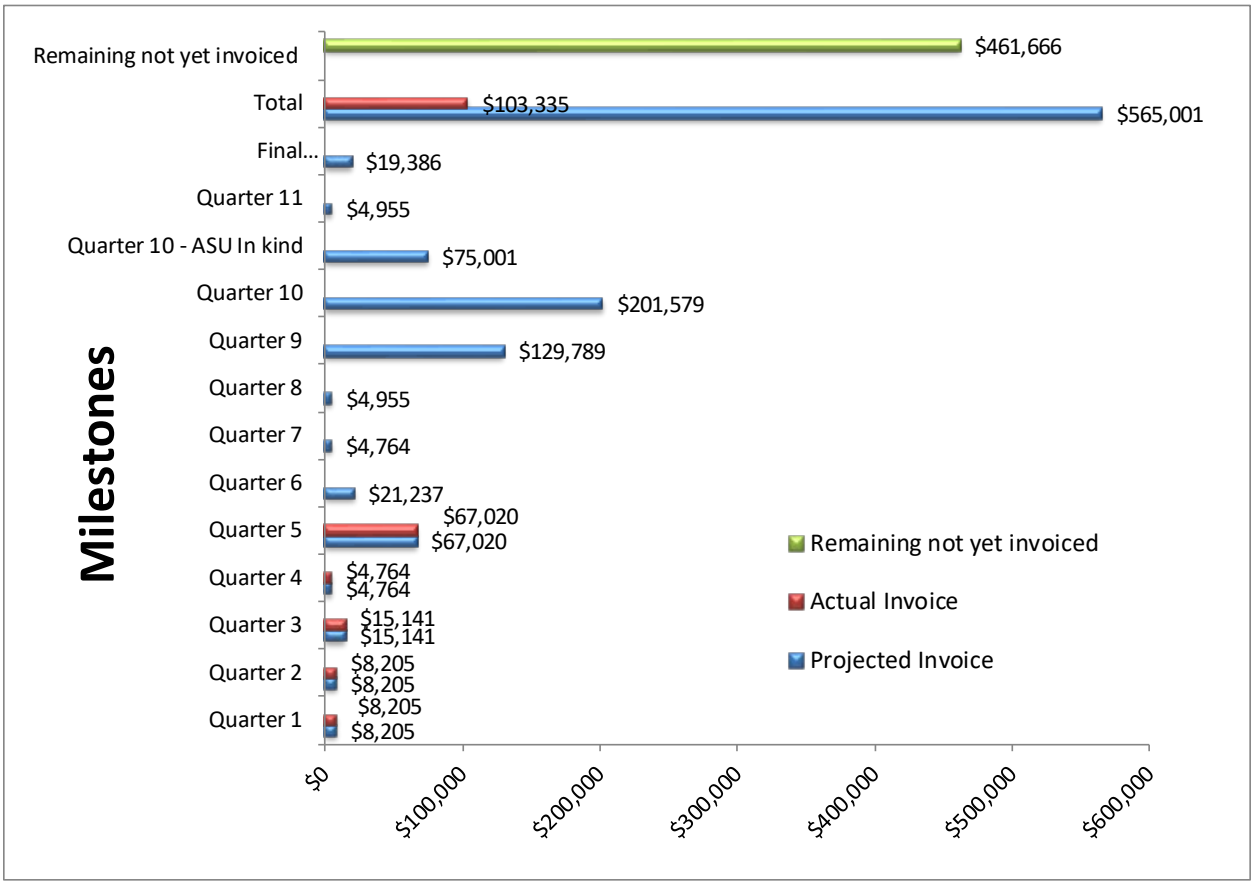
The nature of the contract for this research effort is fixed price, with clearly defined milestone/deliverable payments. Please see Figure 1 for an update on the fund status for DOT/PHMSA, and Figure 2 for the fund status of OTD and ASU.

Figure 1: Quarterly Payable Milestones/Invoices – DOT PHMSA



The third quarter invoice will be issued shortly after the end of the 5th project quarter.

Figure 2: Quarterly Payable Milestones/Invoices – OTD and ASU



The third quarter invoice will be issued shortly after the end of the 5th project quarter.

Project Risks

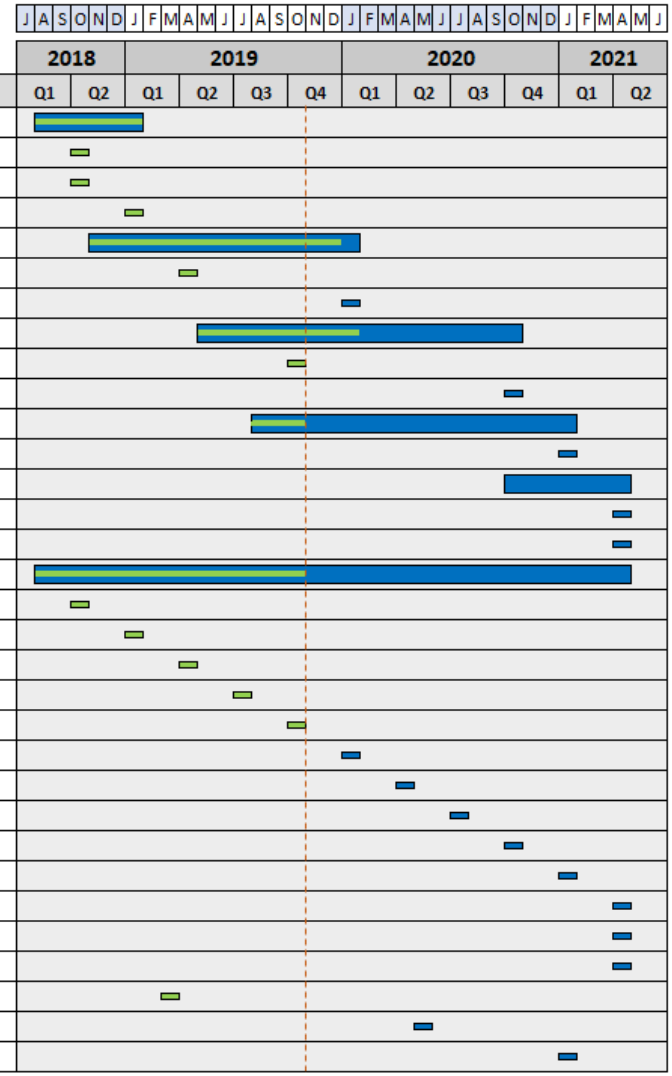
There are currently no risks to the project scope, schedule, or budget.

Schedule Update

The updated project schedule is shown in Figure 3.

Figure 3: Project Schedule

ID	Month	Task / Milestone Name	Start	Finish	% Complete
1	---	Task 1 - Form Technical Advisory Panel and Scope Confirmation	8/1/2018	1/31/2019	100%
2	1	M1 - Conduct kick-team meeting and present scope of work and objectives	10/31/2018	10/31/2018	100%
3	3	M2 - Submit final TAP list and finalize tasks and methods	10/31/2018	10/31/2018	100%
4	6	M4 - Confirmation of subcontract SOW in place	1/31/2019	1/31/2019	100%
5	---	Task 2 - Develop Project Database and Pipeline Sample Library	11/1/2018	1/31/2020	90%
6	9	M6 - Develop a project database on existing data	4/30/2019	4/30/2019	100%
7	18	M11 - Submit summary of pipeline sample library	1/31/2020	1/31/2020	0%
8	---	Task 3 - Develop & Execute Test Matrix	5/1/2019	10/31/2020	50%
9	15	M9 - Submit project testing/plan matrix	10/31/2019	10/31/2019	100%
10	27	M15 - Submit summary of executed testing plan data input into project database	10/31/2020	10/31/2020	0%
11	---	Task 4 - Data Analysis and Model Development and Optimization	8/1/2019	1/31/2021	20%
12	30	M17 - Summary of data analysis and model generation and optimization submitted	1/31/2021	1/31/2021	0%
13	---	Task 5 - Final Report and Implementation Guide	10/31/2020	4/30/2021	0%
14	33	M25 - Prepare and submit final draft report	4/30/2021	4/30/2021	0%
15	NA	M26 - Address comments and submit final report	4/30/2021	4/30/2021	0%
16	---	Task 6 - Project Management	8/1/2018	4/30/2021	40%
17	3	M3 - Submit quarterly report - Q1	10/31/2018	10/31/2018	100%
18	6	M5 - Submit quarterly report - Q2	1/31/2019	1/31/2019	100%
19	9	M7 - Submit quarterly report - Q3	4/30/2019	4/30/2019	100%
20	12	M8 - Submit quarterly report - Q4	7/31/2019	7/31/2019	100%
21	15	M10 - Submit quarterly report - Q5	10/31/2019	10/31/2019	100%
22	18	M12 - Submit quarterly report - Q6	1/31/2020	1/31/2020	0%
23	21	M13 - Submit quarterly report - Q7	4/30/2020	4/30/2020	0%
24	24	M14 - Submit quarterly report - Q8	7/31/2020	7/31/2020	0%
25	27	M16 - Submit quarterly report - Q9	10/31/2020	10/31/2020	0%
26	30	M18 - Submit quarterly report - Q10	1/31/2021	1/31/2021	0%
27	33	M19 - Submit quarterly report - Q11	4/30/2021	4/30/2021	0%
28	NA	M20 - Present at public event or publish a paper	4/30/2021	4/30/2021	0%
29	NA	M21 - Final virtual meeting	4/30/2021	4/30/2021	0%
30	NA	M22 - Peer review #1 (date TBD)	5/1/2019	5/1/2019	100%
31	NA	M23 - Peer review #2 (date TBD)	5/1/2020	5/1/2020	0%
32	NA	M24 - Peer review #3 (date TBD)	1/1/2021	1/1/2021	0%



TECHNICAL STATUS
(Project Quarter #5: 8/1/2019 to 10/31/2019)

This section provides a high-level overview of the technical activities during the reporting quarter and the status.

Additional details, reports, presentations and other support material is placed in this report's Appendix and referenced below.

Task 2 - Develop Project Database and Pipeline Sample Library

1. GTI continued the development of the project sample library. The overall library has been established with a robust set of 70 pipeline samples.
2. The sample library of 70 was divided into two groups of 35 pipe samples. This will allow the Task 4 modeling activities to continue model development with half the project results while the other half of the testing is being completed. Then the remaining 35 sample data set will be turned over to complete the modeling activities.
3. Reorganized the Master Flat Sheet into a digital analysis "ready" format conducive to importation into statistical analysis software for the upcoming Task 4 activities.
4. Once the tensile full-wall thickness testing and bulk and through-wall chemistry testing of the lab samples are complete (see Task 3 below) we will be able to select 30 (of the 70 total pipeline samples) to extract toughness coupons for MMT NDTT testing scheduled for later this year or early next year. This will complete the sample/test assignments for the projects - all other tests have been assigned by sample.

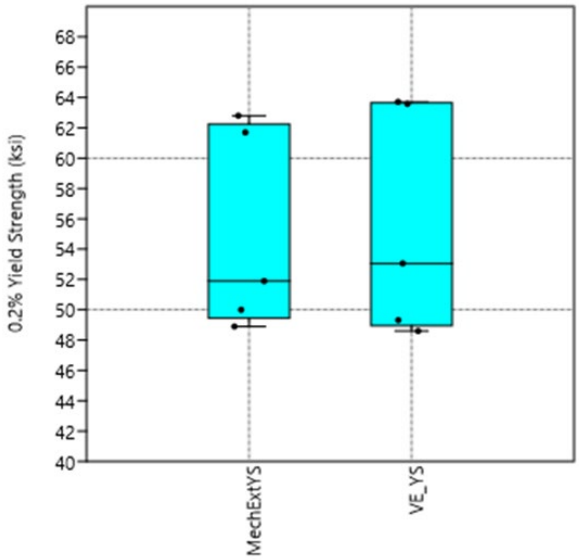
Task 3 - Develop and Execute Testing Matrix

1. GTI completed the Project Testing Plan/Matrix (Milestone Number 9) and this is included in this quarterly report in **Appendix A**.
2. GTI completed a statistical analysis of the difference of tensile results on five diverse pipeline samples using a mechanical extensometer (M.E.) vs. a video extensometer (V.E.).
3. On average, the %-difference on the yield strength between the V.E. and M.E. is +0.97%, i.e., the V.E. is < 1% higher than the M.E. The data overlaps between the methods. The descriptive statistics are summarized in Table 3 and Figure 4 below.

Table 3: Percent Difference Yield Strength (ksi) for Mechanical v. Video Extensometer

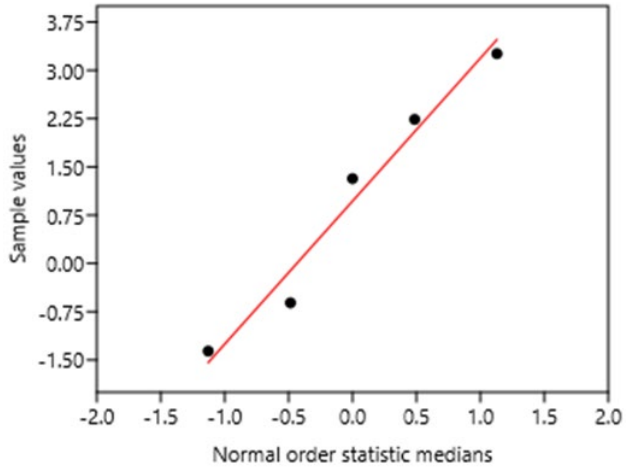
Parameter	All	Bootstrap Ave	
		Lower conf.	Upper conf.
N	5	5%	95%
Min	-1.36		
Max	3.26		
Sum	4.85	-2.62	12.32
Mean	0.97	-0.52	2.46
Std. error	0.86	0.60	1.20

Figure 4: Box Plot of M.E. vs. V.E. Yield Strengths (ksi)



a. The 90% double-sided confidence interval is 2.98% wide between the 5% and 95% confidence limits using the bootstrap average calculation.
b. The normal plot of the %-difference plotted on a normal order statistic, see Figure 5, showed the variation in the percent difference is normally distributed across the range, another sign that we are just seeing random variation between all the process/equipment errors vs. some kind of bias.

Figure 5: Normality Plot of Percent Difference Between M.E. and V.E.



c. In short, the testing and analysis showed that the methods are statistically and practically equivalent. The use of the V.E. allows for very accurate and complete *True* Stress - Strain plots to be generated in addition to the standard *Engineering* Stress - Strain plots.

4. **Data entry from Nondestructive Test Technology into the project Master Flat Table (Database)**

   a. Completed input of all remaining MMT's and Frontics' data into to the MFT.
   b. The NDE testing phase of the project from MMT, Frontics, and SciApps has now been completed with the exception of the 30 coupons that will be tested later in the project at the MMT labs using their prototype NDTT toughness testing system (see Task 2 details above).

5. **MMT Testing Draft Report of Testing**

   a. GTI received MMT's draft summary report and remaining test data (of bulk tensile predictions and weld classifications). The bulk tensile property predictions were done using the standard MMT linear regression mode, as well as predictions from two new models: a Bayesian regression model and an artificial neural network (ANN) model.
   b. GTI is currently working through a quality control review of the report and data with the target date for completion the first half of Nov. 2019.

6. **Lab-Based (Destructive) Testing**

    a. This lab testing continued in Quarter 5, which included full wall tensile, hardness, chemistry, and grain size. Ongoing testing includes:
        - Full wall tensile tests per ASTM A370 with 1"-gauge length longitudinal specimens and an average of 3 specimens
        - Lab GDS chemistry at 4 different depths. (0.005", 0.020", 3/4 thickness and mid thickness. Includes C, S, and P by GDS
        - Bulk ICP + Bulk LECO E1019 C & S Spark Arc
        - Bulk N by Leco fusion technique.
        - Grain size near surface, 1/4 pt, and center for both longitudinal and transverse sections, average of 6 readings
        - Average of near surface grain sizes (~0.005" deep) longitudinal and transverse specimens
        - OD Rockwell B hardness after ~ 0.005" surface grind
        - ID Rockwell B hardness after ~ 0.005" surface grind

    b. During this quarter GTI achieved a milestone by completing 35 of the 70 project samples, both nondestructive and destructive. Final quality checks of this set are being performed prior to the results being turned over to the Task 4 Modeling effort.

Task 4 - Data Analysis and Model Development and Optimization

A detailed report of this quarter's modeling efforts and results is presented in **Appendix B** of this report. A high-level summary of these results is presented below:

1. This research this quarter demonstrated a general Bayesian methodology to handle missing data (typical in practice) when estimating the strength of pipeline samples.
2. The missing data can be caused by various reasons, e.g., experimental abnormalities, wrong documentation, human errors.
3. It was shown that the proposed *Bayesian multimodality inference* with missing data can successfully handle this issue using the provided data.
4. Further, the proposed method has better/similar performance with surface-only indentation technique and linear regression with complete data.

Task 6 - Project Management

1. Submitted Milestone Item **M9** (Appendix B of this report) - Project Testing Plan/Matrix.
2. Submitted Milestone Item **M10**, the fifth project quarterly report.

PLANS FOR FUTURE ACTIVITY
(Project Quarter #6: 11/1/2019 to 1/31/2020)

In the next quarter the project team will continue work on Tasks 2, 3, and 4. The planned activities are listed below.

Task 2 - Develop Project Database and Pipeline Sample Library

a. Complete and submit Milestone Item **M11** - Summary of Pipeline Sample Library, due 01/31/2020.
b. As the full-wall thickness baseline and surface-based chemical testing is completed and organized, GTI will continue the selection process for the 30 toughness coupons that will be sent to MMT for NDTT (a lab prototype surface-based toughness testing unit) calibration testing. GTI will also complete Charpy toughness testing to generate full S-curves for impact energy, lateral expansion, and %-shear failure.

Task 3 - Develop and Execute Testing Matrix

a. Complete quality control reviews of nondestructive testing reports and data.
b. Baseline (referee) lab-based testing will continue and run well into 2019. This testing includes full wall tensile, hardness, chemistry, Charpy, and grain size, specifically, the tests include:
    - Full wall tensile tests per ASTM A370 with 1"-gauge length longitudinal specimens and an average of 3 specimens
    - Lab GDS chemistry at 4 different depths. (0.005", 0.020", 3/4 thickness and mid thickness. Includes C, S, and P by GDS
    - Bulk ICP + Bulk LECO E1019 C & S Spark Arc
    - Bulk N by Leco fusion technique.
    - Charpy impact tests at 4-6 temperatures (Need to complete the selection of 30 samples that is pending the full-wall tensile and bulk chemistry testing)
    - Grain size near surface, 1/4 pt, and center for both longitudinal and transverse sections, average of 6 readings
    - Average of near surface grain sizes (~0.005" deep) longitudinal and transverse specimens
    - OD Rockwell B hardness after ~ 0.005" surface grind
    - ID Rockwell B hardness after ~ 0.005" surface grind

Task 4 - Data Analysis and Model Development and Optimization

a. Based on the results in project quarter four and the results presented in Appendix B of this fourth quarter report, ASU planned work will concentrate on further validation of the developed models with possible additional data and are going to

investigate the general nonlinear regression using Gaussian Process (GP) model in the next quarter.

Task 6 - Project Management

a. Submit sixth project quarterly report to DOT - Milestone Item **M12**, due 01/31/2020.

Respectfully submitted by,

Daniel A. Ersoy

Principal Engineer
Element Resources, LLC
dersoy@elementresourcesllc.com
847.343.9755

End of Quarterly Report Body

APPENDIX A – Project Testing Plan/Matrix (M9)

In Table 4 below each row is a unique item that is used to track the group of pipe samples or a unique nondestructive (field) or destructive (lab) test being done on the pipe sample collection. The item no. column relates to the database (flat table) column where the data from the testing task is entered in the row corresponding to the particular pipe sample.

Table 4: Project Testing Plan

Item No.	Short Title	Description
00A	Pipe_Number	States the pipe number given by GTI for sample identification
00B	Set_1_Pipes	Pipe numbers of 35 samples selected from main set of 70 samples to perform first analysis
00C	Set_2_Pipes	Pipe numbers of 35 samples selected from main set of 70 samples to perform validation analysis
00D	Data_Set_Pipes	Pipe numbers from main set of (70) samples for NDE from MMT, Frontics, and SciAps
0A	MMT_Weld_Type_Pipes	Pipe numbers corresponding to subset of (25) samples from main set of 70 samples for HSD weld characterization along HAZ
0B	Charpy_Pipes	Pipe numbers corresponding to subset of (30) samples from main set of 70 samples for MMT NDTT and Charpy Lab tests
0C	ASU_Training_Set_Pipes	Pipe numbers of samples included in the training set given to Arizona State University
0D	TensileTracker	Pipe numbers of samples that have full-wall tensile tests done on them
0E	GDSTracker	Pipe numbers of samples on which Glow Discharge Spectroscopy chemistry has been performed at 0.005”, 0.020”, ¾”, and mid thickness has been performed
0F	ICPTracker	Pipe numbers of samples on which Inductive Coupled Plasma and LECO chemistries have been performed at the bulk of the pipe wall as specified in E1019 by GTI or by an external laboratory

Item No.	Short Title	Description
0G	Ntracker	Pipe numbers of samples on which bulk nitrogen has been determined by LECO fusion technique
0H	CharpyTracker	Pipe numbers of samples on which Charpy impact tests are 4 to 6 temperatures have been performed
0I	GScompositeTracker	Pipe numbers of samples which have undergone grain size measurements at the surface ¼ pt, and center in longitudinal and transverse section
0J	GSsurfaceTracker	Pipe numbers of samples for which the average of grain size at the surface has been calculated from the longitudinal and transverse specimens
0K	HRBWodTracker	Pipe numbers of samples on which Rockwell B hardness has been performed at the outer diameter surface after 0.005” grinding approximately
0L	HRBWidTracker	Pipe numbers of samples on which Rockwell B hardness has been performed at the inner diameter surface after 0.005” grinding approximately
0M	MMTbulkTracker	Pipe numbers of samples that have chemistry compositions obtained from grindings by MMT
0N	MMTweldTracker	Pipe numbers of samples that have undergone the non-destructive testing at the Heat Affected Zone of the weld
0O	MMTndttTracker	Pipe numbers of samples on which the non-destructive toughness testing has been done
0P	FRONTICSbulkTracker	Pipe numbers of samples that Frontics has tested with their non-destructive technology
0Q	FRONTICSweldTracker	Pipe numbers of samples that Frontics has tested at the Heat Affected Zone
0R	SCIAPSTracker	Pipe numbers of samples on which chemistry composition has been obtained by the SciAps LIBS OES handheld device and XRF
1A	PIPE_NUMBER	Pipe number given by GTI for sample identification

Item No.	Short Title	Description
2A	DIAMETER_NOMINAL	Nominal diameter of the pipe in inches
3A	WALL_THICKNESS_AS_RECIEVED	States the wall thickness of the pipe in inches without any treatment to the sample, i.e.: including corrosion products, adhesions…
3B	WALL_THICKNESS_LAB_SAMPLE	States the wall thickness of the pipe in inches after preparation of the sample for testing
3C	WALL_THICKNESS_MMT	Wall thickness in inches provided by MMT in their data
4A	INSTALLATION_YEAR	Specifies the installation date as provided by the utilities
5A	GRADE_REPORTED	Shows the steel grade as known by the utilities
6A	STEEL_TYPE_ESTIMATE	Specifies the GTI estimate of steel (rimmed, capped, killed Si, killed, Al, semi-killed)
7A	GRADE_CHEMISTRY_GROUP	States the steel grade as per SAE-ASTM
8A	WELD_TYPE_GTI	Determines whether the pipe has a weld or not and what type of weld it is according to GTI inspections
8B	WELD_TYPE_MMT	Determines the type of ERW weld, either High Frequency (HF), High Frequency Normalized (HFN), or Low Frequency (LF), identified by MMT on 25 pipes of the 70 set
9A	YS_LAB_FULL_WALL	Yield Strength mini: Shows the yield strength in ksi as evaluated from a mini 1” gauge length longitudinal sample
9B	UTS_LAB_FULL_WALL	Tensile Strength mini: Shows the tensile strength in ksi as evaluated from a mini 1” gauge length longitudinal sample
10A	YS_FRONTICS_BASE	Bulk Yield Strength as calculated by Frontics from their non-destructive Instrumented Indentation Testing (IIT) using AIS2100
10B	UTS_FRONTICS_BASE	Bulk Ultimate Tensile Strength as calculated by Frontics from their non-destructive IIT using AIS2100

Item No.	Short Title	Description
10C	TOUGHNESS_FRONTICS_BASE	Bulk Toughness as calculated by Frontics from their non-destructive IIT using AIS2100
10D	YS_FRONTICS_WELD	Yield Strength at the weld as calculated by Frontics from their non-destructive IIT using AIS2100
10E	UTS_FRONTICS_WELD	Ultimate tensile Strength at the weld as calculated by Frontics from their non-destructive IIT using AIS2100
10F	TOUGHNESS_FRONTICS_WELD	Toughness at the weld as calculated by Frontics from their non-destructive IIT using AIS2100
10G	YS_FRONTICS_HAZ1	Yield Strength at the Heat Affected Zone on side 1 as calculated by Frontics from their non-destructive IIT using AIS2100
10H	UTS_FRONTICS_HAZ1	Ultimate Tensile Strength at the Heat Affected Zone on side 1 as calculated by Frontics from their non-destructive IIT using AIS2100
10I	TOUGHNESS_FRONTICS_HAZ1	Toughness at the Heat Affected Zone on side 1 as calculated by Frontics from their non-destructive IIT using AIS2100
10J	YS_FRONTICS_HAZ2	Yield Strength at the Heat Affected Zone on side 2 as calculated by Frontics from their non-destructive IIT using AIS2100
10K	UTS_FRONTICS_HAZ2	Ultimate Tensile Strength at the Heat Affected Zone on side 2 as calculated by Frontics from their non-destructive IIT using AIS2100
10L	TOUGHNESS_FRONTICS_HAZ2	Toughness at the Heat Affected Zone on side 2 as calculated by Frontics from their non-destructive IIT using AIS2100
11A	YS_MMT_SURFACE	Surface measurement of yield strength as 0.5% Elongation Under Load (EUL) obtained from MMT HSD Tester
11B	YS_MMT_SURFACE_SDV	Standard deviation for the surface measurement of yield strength as 0.5% EUL obtained from MMT HSD Tester
11C	UTS_MMT_SURFACE	Surface measurement of Ultimate Tensile Strength (UTS) obtained from MMT HDS Tester

Item No.	Short Title	Description
11D	UTS_MMT_SURFACE_SDV	Standard deviation for the surface measurement of UTS obtained from MMT HSD Tester
12A	YS_MMT_BRM	Predicted yield strength as 0.5% EUL using a Bayesian Linear Regression model
12B	YS_MMT_BRM_SDV	Standard deviation for the predicted yield strength as 0.5% EUL using a Bayesian Linear Regression model
12C	UTS_MMT_BRM	Predicted UTS using a Bayesian Linear Regression model
12D	UTS_MMT_BRM_SDV	Standard deviation for the predicted UTS using a Bayesian Linear Regression model
13A	YS_MMT_LRM	Predicted yield strength as 0.5% EUL using a Multiple Linear Regression model
13B	YS_MMT_LRM_SDV	Standard deviation for the predicted yield strength as 0.5% EUL using a Multiple Linear Regression model
13C	UTS_MMT_LRM	Predicted UTS using a Multiple Linear Regression model
13D	UTS_MMT_LRM_SDV	Standard deviation for the predicted UTS using a Multiple Linear Regression model
14A	YS_MMT_ANN	Predicted yield strength as 0.5% EUL using an Artificial Neural Network model
14B	YS_MMT_ANN_SDV	Standard deviation for the predicted yield strength as 0.5% EUL using an Artificial Neural Network model
14C	UTS_MMT_ANN	Predicted UTS using an Artificial Neural Network model
14D	UTS_MMT_ANN_SDV	Standard deviation for the predicted UTS using an Artificial Neural Network model
15A	C_BULK	Bulk chemical composition by weight percentage of carbon as per ASTM E1019
15B, 15C	MN__BULK, P__BULK	Bulk chemical composition using ICP (Inductively Coupled Plasma). Lab weight percentage of elements, Manganese and Phosphorus
15D	S__BULK	Bulk chemical composition by weight percentage of sulfur as per ASTM E1019

Item No.	Short Title	Description
15E - 15N	*Element*_BULK	Bulk chemical composition using ICP (Inductively Coupled Plasma). Lab weight percentage of elements, Al, Cr, Cu, Mo, Nb, Ni, Si, Ti, V, and B
15O	N_BULK	Bulk chemical composition by weight percentage of nitrogen as per ASTM E1019 and E1409 using Fusion analysis (LECO) technique
15P	PEARLITE_PERC_BULK	Percentage pearlite estimated by the lever rule using the bulk carbon content as per column 15A
16A - 16O	*Element*_5MIL	Chemical composition by OES (optical emission spectroscopy) using a laboratory spectrometer, SPECTROMAXX, after grinding 0.005” from surface for C, Mn, P, S, Al, Cr, Cu, Mo, Nb, Ni, Si, Ti, V, B, and N in wt.%
16P	PEARLITE_PERC_5M	Percentage pearlite estimated by the lever rule using the carbon content at 0.005” from surface as per column 16A
17A - 17O	*Element*_20MIL	Chemical composition by OES (optical emission spectroscopy) using a laboratory spectrometer, SPECTROMAXX, at a depth of 0.020” from surface for C, Mn, P, S, Al, Cr, Cu, Mo, Nb, Ni, Si, Ti, V, B, and N in wt.%
18A - 18O	*Element*_QTRTHK	Chemical composition by OES (optical emission spectroscopy) using a laboratory spectrometer, SPECTROMAXX, at a depth of one quarter of the thickness from O.D. surface for C, Mn, P, S, Al, Cr, Cu, Mo, Nb, Ni, Si, Ti, V, B, and N in wt.%
19A - 19O	*Element*_MIDWALL	Chemical composition by OES (optical emission spectroscopy) using a laboratory spectrometer, SPECTROMAXX, at midwall for C, Mn, P, S, Al, Cr, Cu, Mo, Nb, Ni, Si, Ti, V, B, and N in wt.%
20A - 20O	*Element*_AVEGDS	Average of each element at the: 0.005”, 0.020”, and ¼ thickness, and midwall from the surface in wt.%
20P	PEARLITE_PERC_AVE_GDS	Percentage pearlite estimated by the lever rule using the average carbon content as per column 20A

Item No.	Short Title	Description
21A - 21O	*Element*_MMT_GRNDGS	Chemical composition provided by MMT and obtained from pipe grindings for C, Mn, P, S, Al, Cr, Cu, Mo, Nb, Ni, Si, Ti, V, and B in wt.%. C and S measured by combustion analysis and rest of elements measured by ICP-OES
22A - 22O	*Element*_LIBS	Surface chemical compositions obtained from SciAps Z-200 series hand-held device through Laser Induced Breakdown Spectroscopy (LIBS) for C, Mn, Al, Cr, Cu, Mo, Nb, Ni, Si, Ti, V, B, and N in wt.%
23A - 23O	*Element*_LIBS_ERR	Instrument error for the surface chemical compositions obtained from SciAps Z-200 series hand-held device LIBS for C, Mn, Al, Cr, Cu, Mo, Nb, Ni, Si, Ti, V, B, and N in wt.%
24A - 24O	*Element*_XRF	Surface chemical composition obtained from SciAps X-series hand-del device through X-Ray Fluorescence (XRF) spectroscopy for C, Mn, P, S, Al, Cr, Cu, Mo, Nb, Ni, Si, Ti, V, B, and N in wt.%
25A - 25O	*Element*_XRF_ERR	Instrument error for the surface chemical composition obtained from SciAps X-series hand-del device through X-Ray Fluorescence (XRF) spectroscopy for C, Mn, P, S, Al, Cr, Cu, Mo, Nb, Ni, Si, Ti, V, B, and N in wt.%
26A	HRBW_OD	Rockwell hardness at the outer diameter (OD) of the pipe as per ASTM E18
26B	HRBW_ID	Rockwell hardness at the inner diameter (ID) of the pipe as per ASTM E18
27A	CVN_SIZE_PERC	Charpy V-Notch Test Sample Size: States the size of the sample used to perform the Charpy V-Notch test as a percentage of a full-size sample
28A - 28F	CVN_TEMP_1-6	Charpy V-Notch Test Temperature: Shows the temperature in degrees F at which the test was performed as per API
29A - 29F	CVN_FTLBS_1-6	Charpy V-Notch Test Energy Absorbed: Shows the energy absorbed in ft-lbs at the respective temperature in columns 18A-F as per API

Item No.	Short Title	Description
30A- 30F	CVN_LATEXP_1-6	Charpy V-Notch Test Lateral Expansion: Shows the lateral expansion in mils at the respective temperature in columns 18A-F as per API
31A - 31F	CVN_SHEARPERC_1-6	Charpy V-Notch Test Shear Area: Shows the shear area of the sample at the respective temperature in columns 18A-F as per API
32A	CVN_ENERGY_ITT	Energy absorbed value of 15 ft-lb for a full-sized specimen or its equivalent value at a sub sized specimen
32B	CVN_UPSHELF_ENERGY	Upper energy shelf in ft-lbs evaluated from the energy absorbed vs temperature plot of CVN testing data
32C	CVN_LOWSHELF_ENG	Lower energy shelf in ft-lbs evaluated from the energy absorbed vs temperature plot of CVN testing data (note pipe 6 has NA since curve is flat).
33A	GS_COMPOSITE	Grain size represented as an ASTM number measured as an average as per ASTM E112
33B	GD_COMPOSITE	Average grain diameter calculated in mm, from the ASTM grain size in 23A using relation by Gladman T
33C	GMI_COMPOSITE	Mean linear intercept calculated in mm, from the ASTM grain size in 23A as per ASTM E112
33D	GS_SURFACE	Grain size represented as an ASTM number measured at the surface as per ASTM E112; typically, at about 0.005” from the surface.
33E	GD_SURFACE	Average grain diameter calculated in mm, from the ASTM grain size in 23D using relation by Gladman T
33F	GMI_SURFACE	Mean linear intercept calculated in mm, from the ASTM grain size in 23D as per ASTM E112
34A	GS_MMT	Grain Size calculated by MMT using the mean intercept per equation in Table 6 of ASTM E112 at the same location of the HSD base metal testing

Item No.	Short Title	Description
34B	GMI_MMT	Mean intercept calculated by MMT per Section 13 in ASTM E112 (Heyn Lineal Intercept Procedure) at the same location of the HSD base metal testing

APPENDIX B – ASU TASK 4 (5TH Quarter Update)

5TH QUARTER – INPUT FROM ASU
DOT Project Number: 729 (GTI 22428)
DOT Contract Number: 693JK31810003
OTD Project Number: 4.14.c.2 (GTI 22429)

Validating Non-Destructive Tools for Surface to Bulk Correlations of Yield Strength, Toughness, and Chemistry

Reporting Period
August 1, 2019 through October 31, 2019

Report Issued
October 31, 2019

Prepared For
U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration (DOT/PHMSA)
Joe Pishnery
404.832.1151
joseph.pishnery@dot.gov

Operations Technology Development (OTD)

Technical Contact
Daniel Ersoy
Element Resources, LLC
847.343.9755
dersoy@elementresoucesllc.com

GTI Technical Team
Daniel Ersoy, Brian Miller, Marta Guerrero Merino, Ernest Lever

Subcontractor - Arizona State University Team
Dr. Yongming Liu, Jie Chen, Sonam Dahire

1700 S. Mount Prospect Rd.
Des Plaines, Illinois 60018
www.gti.energy

Abstract

This research focuses on the yield strength prediction by fusing the incomplete surface information. The input surface variables are yield strength obtained from surface indentation technique, chemical composition (P, Cr, Cu, Si), grain size and hardness. In this research, it is considered that not every variable is observed for each pipe. The Bayesian inference with missing data provides a tool for strength prediction while imputing the unknown missing values. The predictive performance is compared among direct surface indentation technique, linear regression with complete data and Bayesian inference with missing data. Results show that the Bayesian method can deal with missing data problem properly and makes better predictions than the other two methods. This research provides a method to make bulk strength estimation in the case of incomplete input surface data.

Introduction

Missing data are fairly common in engineering. Unfortunately, many software packages either throw away all subjects with incomplete data or impute missing values with a population mean or some other fixed value, then proceed with the analysis. The first approach is bad because a potentially large amount of useful information is thrown away. The second is statistically incorrect, as it says people are certain about the values of the missing data when they are not actually observed [1].

This quarter's work focuses on the yield strength prediction from the incomplete surface information. The input surface variables are yield strength obtained from surface indentation technique, chemical composition (P, Cr, Cu, Si), grain size and hardness. These variables are selected due to their relatively high predictive usefulness shown from previous work. In this research, it is considered that not every variable is observed for each pipe. The Bayesian inference with missing data provides a tool for strength prediction while imputing the unknown missing values. The research provides a method to make bulk strength estimation in the case of incomplete input surface data.

Bayesian Inference with Missing Data

Let **Y** be the $n \times p$ matrix of all the potential data (complete data), observed and unobserved. Here n and p denote the total number of pipes and variables respectively. Let **I** be the $n \times p$ inclusion matrix which indexes which potential data are observed. $I_{ij} = 1$ if Y_{ij} is observed and $I_{ij} = 0$ if Y_{ij} is missing. The matrix **Y** can then be thought of as consisting two parts: $\mathbf{Y}_{\text{obs}} = \{Y_{ij} \mid I_{ij} = 1\}$, the observed data, and $\mathbf{Y}_{\text{miss}} = \{Y_{ij} \mid I_{ij} = 0\}$, the unobserved data.

When considering data collection, it is useful to break the joint probability model into two parts: (1) the model for the underlying complete data, **Y**, including observed and unobserved components, and (2) the model for the inclusion matrix, **I**. The complete-data likelihood is defined as the product of the likelihoods of these two factors; that is, the distribution of the complete data, **Y**, and the inclusion matrix, **I**, given the parameters in the models [2]:

$$p(\mathbf{Y}, \mathbf{I} \mid \phi, \xi) = p(\mathbf{Y} \mid \phi)\, p(\mathbf{I} \mid \mathbf{Y}, \xi), \tag{1}$$

where ϕ and ξ denote the parameter vectors of the distributions of the complete data and the inclusion matrix, respectively. In the context of pipe strength prediction, the estimates of primary interest are functions of the complete data **Y** and the parameters ϕ. The parameters ξ that index the missingness model are characteristic of data collection but are not generally of interest.

Since **Y** is not completely observed and the actual information available is ($\mathbf{Y}_{\text{obs}}$, **I**), the appropriate likelihood for Bayesian inference is

$$p(\mathbf{Y}_{\text{obs}}, \mathbf{I} \mid \phi, \xi) = \int p(\mathbf{Y}, \mathbf{I} \mid \phi, \xi)\, d\mathbf{Y}_{\text{miss}}, \tag{2}$$

which is marginal probability of the observed variables after integrating out the missing variables.

The joint posterior distribution of model parameters ϕ and ξ given the observed information ($\mathbf{Y}_{\text{obs}}$, **I**) is

$$\begin{aligned} p(\phi, \xi \mid \mathbf{Y}_{\text{obs}}, \mathbf{I}) &\propto p(\phi, \xi)\, p(\mathbf{Y}_{\text{obs}}, \mathbf{I} \mid \phi, \xi) \\ &= p(\phi, \xi) \int p(\mathbf{Y}, \mathbf{I} \mid \phi, \xi)\, d\mathbf{Y}_{\text{miss}} \\ &= p(\phi, \xi) \int p(\mathbf{Y} \mid \phi)\, p(\mathbf{I} \mid \mathbf{Y}, \xi)\, d\mathbf{Y}_{\text{miss}} \end{aligned}, \tag{3}$$

The posterior distribution of ϕ alone is this expression averaged over ξ:

$$p(\phi \mid \mathbf{Y}_{\text{obs}}, \mathbf{I}) = p(\phi) \iint p(\xi \mid \phi)\, p(\mathbf{Y} \mid \phi)\, p(\mathbf{I} \mid \mathbf{Y}, \xi)\, d\mathbf{Y}_{\text{miss}}\, d\xi, \tag{4}$$

Ignorability

When the missing data pattern supplies no information, the data collection mechanism is ignorable. In this case, the posterior distribution of ϕ and the posterior predictive distribution

of $\mathbf{Y}_{\text{miss}}$ are entirely determined by the specification of a data model and the observed values of $\mathbf{Y}_{\text{obs}}$. The Eq. (4) can be expressed as

$$p(\phi \mid \mathbf{Y}_{\text{obs}}, \mathbf{I}) = p(\phi \mid \mathbf{Y}_{\text{obs}}) = p(\phi)\int p(\mathbf{Y} \mid \phi) d\mathbf{Y}_{\text{miss}} \ , \tag{5}$$

Two conditions are sufficient to ensure ignorability of the missing data mechanism for Bayesian analysis:

(1) The condition of missing at random requires that the distribution of the missing-data mechanism does not depend on the missing values. That is, given ξ, the missingness depends only on $\mathbf{Y}_{\text{miss}}$: $p(\mathbf{I} \mid \mathbf{Y}, \xi) = p(\mathbf{I} \mid \mathbf{Y}_{\text{miss}}, \xi)$.

(2) The condition of distinct parameters is satisfied when the parameter of the missing data process is independent of the parameters of the data generating process in the prior distribution: $p(\xi \mid \phi) = p(\xi)$.

Data augmentation

The data augmentation process is conducted through Bayesian inference with missing data [3].

Assume the multivariate normal distribution for pipe data. In Bayesian inference, the probability distributions are used to describe the information about unknown quantities. For multivariate normal model with missing data, the unknowns include missing values as well as model parameters (mean $\boldsymbol{\theta}$ and covariance matrix Σ). Gibbs sampling is used to make inference on these unknowns.

To avoid evaluating the integrals in Eq. (5), Gibbs sampling is used to draw posterior simulations of the joint vector of unknown parameters and unobserved quantities, $p(\boldsymbol{\theta}, \Sigma, \mathbf{Y}_{\text{miss}} \mid \mathbf{Y}_{\text{obs}})$. Given starting values $\{\Sigma^{(0)}, \mathbf{Y}^{(0)}_{\text{miss}}\}$, $\{\boldsymbol{\theta}^{(s+1)}, \Sigma^{(s+1)}, \mathbf{Y}^{(s+1)}_{\text{miss}}\}$ can be generated from $\{\boldsymbol{\theta}^{(s)}, \Sigma^{(s)}, \mathbf{Y}^{(s)}_{\text{miss}}\}$ by

1. sampling $\boldsymbol{\theta}^{(s+1)}$ from $p(\boldsymbol{\theta} \mid \mathbf{Y}_{\text{obs}}, \mathbf{Y}^{(s)}_{\text{miss}}, \Sigma^{(s)})$, the complete-data posterior of $\boldsymbol{\theta}$;

2. sampling $\Sigma^{(s+1)}$ from $p(\Sigma \mid \mathbf{Y}_{\text{obs}}, \mathbf{Y}^{(s)}_{\text{miss}}, \boldsymbol{\theta}^{(s+1)})$, the complete-data posterior of Σ;

3. sampling $\mathbf{Y}^{(s+1)}_{\text{miss}}$ from $p(\mathbf{Y}_{\text{miss}} \mid \mathbf{Y}_{\text{obs}}, \boldsymbol{\theta}^{(s+1)}, \Sigma^{(s+1)})$, the conditional predictive distribution of $\mathbf{Y}_{\text{miss}}$.

Steps 1 and 2 are together called the Posterior or P-step [3, 4]. In steps 1 and 2, the fixed value of $\mathbf{Y}_{\text{obs}}$ combines with the current value of $\mathbf{Y}^{(s)}_{\text{miss}}$ to form a current version of a complete data matrix $\mathbf{Y}^{(s)}$ having no missing data and a value of model parameter ($\boldsymbol{\theta}$ and Σ) are drawn from $\mathbf{Y}^{(s)}$. Step 3 is referred to as the Imputation or I-step and corresponds to imputing a value of the missing data $\mathbf{Y}_{\text{miss}}$ [3, 4]. Step 3 can be expressed in the following

$$\begin{aligned} p(\mathrm{Y}_{\text{miss}} \mid \mathrm{Y}_{\text{obs}}, \theta, \Sigma) &\propto p(\mathrm{Y}_{\text{miss}}, \mathrm{Y}_{\text{obs}} \mid \theta, \Sigma) \\ &= \prod_{i=1}^{n} p(y_{i,miss}, y_{i,obs} \mid \theta, \Sigma) \ , \\ &\propto \prod_{i=1}^{n} p(y_{i,miss} \mid y_{i,obs}, \theta, \Sigma) \end{aligned} \tag{6}$$

For each i, the missing elements of the data vector are sampled conditional on the observed elements. This is achieved by following result about multivariate normal distributions. Let $y \sim$ multivariate normal ($\boldsymbol{\theta}$, Σ), let a be a subset of variable indices $\{1, \ldots, p\}$ and let b be the complement of a. It can be shown from the inverse of partitioned matrices that $\{y_{[b]} \mid y_{[a]}, \boldsymbol{\theta}, \Sigma\} \sim$ multivariate normal ($\boldsymbol{\theta}_{b|a}$, $\Sigma_{b|a}$), where

$$\theta_{b|a} = \theta_{[b]} + \Sigma_{[b,a]}(\Sigma_{[a,a]})^{-1}(\boldsymbol{y}_{[a]} - \boldsymbol{\theta}_{[a]}), \tag{7}$$

$$\Sigma_{b|a} = \Sigma_{[b]} - \Sigma_{[b,a]}(\Sigma_{[a,a]})^{-1}\Sigma_{[a,b]}, \tag{8}$$

For a value of s that is suitably large, $\boldsymbol{\theta}^{(s)}$, $\Sigma^{(s)}$ and $\mathbf{Y}_{\text{miss}}^{(s)}$ can be regarded as an approximate draw from posterior distributions $p(\boldsymbol{\theta} \mid \mathbf{Y}_{\text{obs}})$, $p(\Sigma \mid \mathbf{Y}_{\text{obs}})$ and posterior predictive distribution $p(\mathbf{Y}_{\text{miss}} \mid \mathbf{Y}_{\text{obs}})$, respectively.

Results and Discussion

Dataset with missing data

The variables used are the difference between strengths measured from tensile testing and surface indentation technique (YS_t - YS_s), chemical composition (P, Cr, Cu, Si), grain size and hardness. They are assumed to follow multivariate normal distribution. Table 5 shows the dataset used for yield strength prediction with missing data. The ***NA***'s in Table 5 are indicated as missing data. The missing pattern is as following. For the column of YS_t - YS_s, 8 data are randomly selected for prediction. For the rest variables, 5 data are missing randomly in each column. The goal is to predict the yield strength even though the dataset is not complete.

Table 5: Dataset with missing data.

Pipe no.	YS_t - YS_s	Chemical composition				Grain size	Hard-ness
		P	Cr	Cu	Si		
1	***NA***	***NA***	0.045	***NA***	***NA***	0.008	223.1
2	-1.70	0.011	0.043	***NA***	0.17	0.009	214.7
8	***NA***	0.008	0.011	0.028	0.0049	***NA***	172.65
9	4.39	0.011	***NA***	0.11	0.19	0.005	221.3
11	***NA***	0.01	0.021	0.013	0.03	0.014	155.3
12	-6.62	0.009	0.01	0.01	0.15	0.010	180.5
14	2.06	***NA***	***NA***	0.041	***NA***	0.008	217.2
15	-5.04	0.014	0.02	0.016	0.18	***NA***	***NA***
16	-2.00	0.011	***NA***	0.013	0.023	0.016	***NA***
17	***NA***	***NA***	***NA***	0.009	***NA***	0.017	198.3
20	***NA***	0.013	0.029	***NA***	0.009	0.013	187.5
21	-0.57	0.014	***NA***	0.042	***NA***	0.010	215.2
22	***NA***	0.01	0.02	0.09	0.001	***NA***	165.6
23	***NA***	0.011	0.038	0.029	***NA***	***NA***	***NA***
30	0.66	0.009	0.055	***NA***	0.014	***NA***	***NA***
34	-3.78	0.005	0.03	0.08	0.0049	0.019	152.9
36	-6.67	0.008	0.01	0.013	0.041	0.013	215.8
38	-5.77	***NA***	0.013	***NA***	0.011	0.021	185.9
39	2.37	0.009	0.02	0.038	0.0049	0.024	***NA***
41	***NA***	***NA***	0.029	0.024	0.009	0.007	215.7

Results

Figure 6 show the prediction results. The black square points are yield strength obtained from surface indentation technique, the blue triangles are predictions from linear regression with complete data, the red dots are predictions using the data above, and the error bars represent the 95% credible interval. The 45° line indicates the points closer to it has better prediction. The root means square errors (RMSE) are given in Table 6. The RMSE of prediction in the case of

missing data is lower than that of the surface indentation technique and is close to that of linear regression with complete data.

Figure 6: Prediction results.

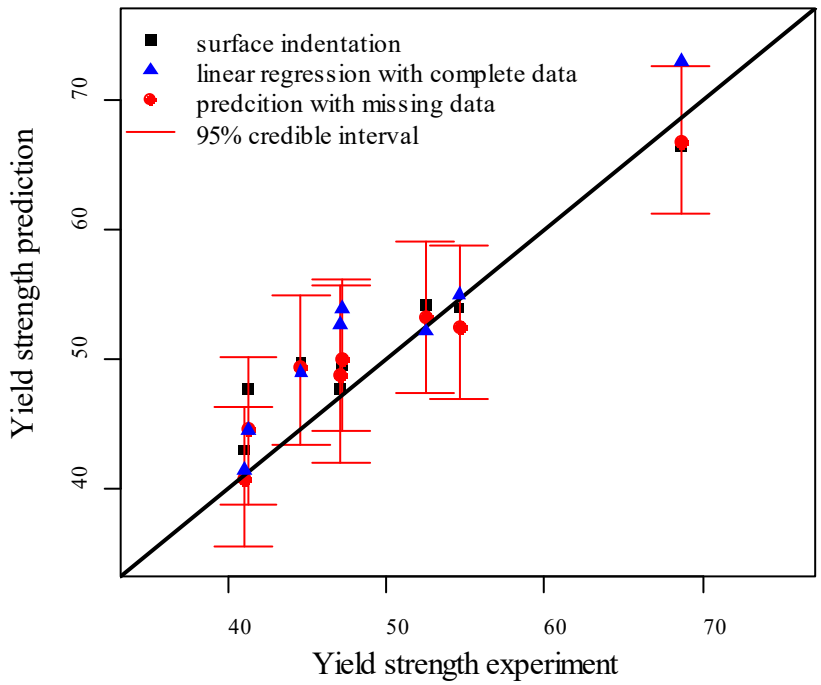


Table 6: Root mean square errors.

	Surface indentation technique	Linear regression with complete data	Prediction with missing data
RMSE	3.64	2.87	3.05

General predictive performance

The results shown above is a specific case of missing data pattern. To show the general predictive performance, a total of 100 times repeating validations is carried out. For the repeating validations, the data are randomly chosen as missing data following the same missing pattern, i.e. 8 data are missing for YS_t - YS_s and 5 data are missing for each of the rest variables.

The 100 RMSE's are shown in Figure 7. The horizontal lines are the mean values of RMSE's. The mean RMSE's for surface indentation technique, linear regression with complete data and prediction with missing data are 3.89, 4.38 and 3.57, respectively, as shown in Table 7.

Figure 7: Root mean square errors of 100 times validations.

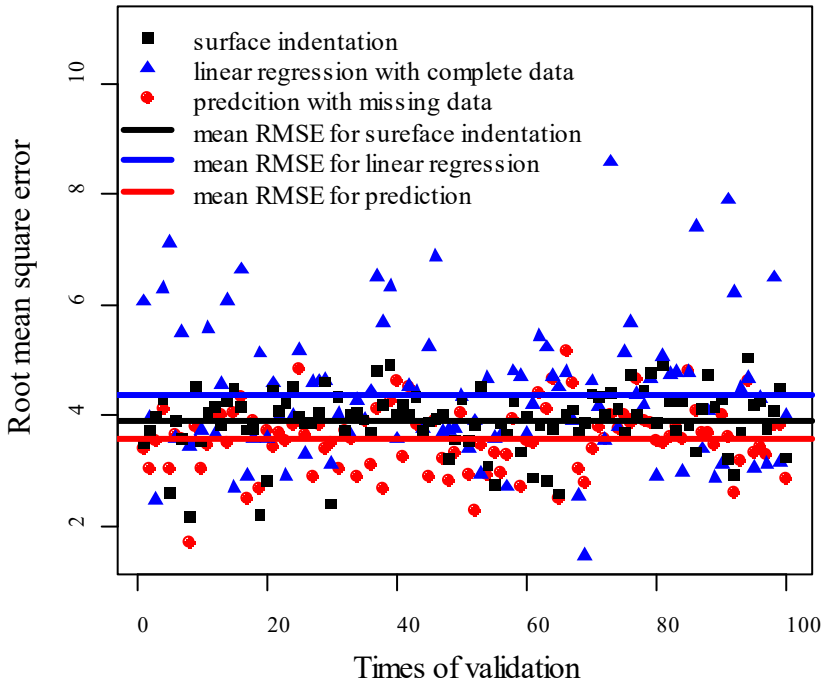


Table 7: Mean root mean square errors over 100 times validation.

	Surface indentation technique	Linear regression with complete data	Prediction with missing data
Mean RMSE	3.89	4.38	3.57

Conclusions

In this research the Bayesian inference with missing data method for multimodality incomplete information fusion is utilized to estimate the probabilistic aging pipe strength. The predictive performance is compared among direct surface indentation technique, linear regression with complete data and Bayesian inference with missing data. First, one specific missing pattern is used for validation. Next, a total of 100 times repeating validations is carried out each of which the missing data are randomly chosen following the same missing data pattern. The following conclusions can be drawn.

1. Results show that the Bayesian method can deal with missing data problem properly.

2. Bayesian inference with missing data makes better predictions than surface indentation technique, linear regression with complete data.

References

[1] P. D. Hoff, *A first course in Bayesian statistical methods*. Springer, 2009.

[2] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*. Chapman and Hall/CRC, 2013.

[3] J. L. Schafer, *Analysis of incomplete multivariate data*. Chapman and Hall/CRC, 1997.

[4] M. A. Tanner and W. H. Wong, "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association,* vol. 82, no. 398, pp. 528-540, 1987.

End of Quarterly Report