8TH 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
May 1, 2020 through July 31, 2020

Report Issued
July 31, 2020

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 3 - Develop and Execute Testing Matrix 5

Task 4 - Data Analysis and Model Development and Optimization 5

Task 6 - Project Management 5

Risks 6

BUSINESS AND MILESTONE STATUS (Project Quarter #8: 5/1/2020 to 7/31/2020) **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 #8: 5/1/2020 to 7/31/2020) **11**

Task 3 - Develop and Execute Testing Matrix 11

Task 4 - Data Analysis and Model Development and Optimization 11

Task 6 - Project Management 11

PLANS FOR FUTURE ACTIVITY (Project Quarter #9: 8/1/2020 to 10/31/2020) **12**

Task 3 - Develop and Execute Testing Matrix 12

Task 4 - Data Analysis and Model Development and Optimization 12

Task 6 - Project Management 12

APPENDIX A – ASU Task 4 (8TH Quarter Update) **13**

Abstract 14

Introduction 14

Manifold-based GP Regression Model 15

Uniform Approximation and Projection 15
Gaussian Process Regression 15
Results and Discussions 16
Dataset 16
General Cases Study 16
Specific Case Study 18
Final Results for the Averaged Model (Q8) 19
Conclusion 19
References 20

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: Low Dimensional Representation Obtained by True Yield Strength 17

Figure 5: Low Dimensional Representation Obtained By Surface Yield Strength 18

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: 15 Variables for BMA (linear model) 16

Table 4: RMSEs for Two Embedding Ways 17

Table 5: RMSEs for linear model 18

Table 6: Minimum RMSE for 3 Models 18

Table 7: RMSEs of the Averaged Model General Case Study 19

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

Task 3 - Develop and Execute Testing Matrix

1. Completed tensile testing for the project.
2. Used the new tensile data to "down-select" the Charpy and NDTT test specimens.
3. Rough cut: (a) Charpy toughness specimens for 20 pipes (10 pipes have Charpy tests already completed); and (b) MMT NDTT specimens for the same 30 specimens. These have both been scheduled for testing.

Task 4 - Data Analysis and Model Development and Optimization

1. Continued with finalizing the Bayesian averaging model.
2. A journal article for our previous Bayesian modeling averaging work has been accepted in the top journal in this field – Structural Safety.
3. During the quarter, the team also compared the performance of the best linear model, the best quadratic model and the best GP model in general case study.
4. Manifold learning work: Continued the preliminary work related to manifold learning modeling and started investigating the dimension reduction capability for regression analysis of the employed manifold learning techniques.

Task 6 - Project Management

1. Submitted Milestone Item **M14**, the eighth project quarterly report.

Risks

Scope and Budget. There are no current risks to the scope and budget.

Schedule. Unfortunately, due to the COVID-19 shelter-in-place orders the lab testing was slowed down. We think we can keep the project on schedule as long as the effects of the pandemic do not last too long. We have finished the bulk of all testing with the remaining lab tests being Charpy toughness testing (in Chicago, IL) and NDTT NDE toughness testing at MMT (Boston, MA). The critical path will be the MMT NDTT testing and if that testing is delayed at MMT then there will be a domino effect that could impact the project schedule. We will keep DOT and OTD informed of any changes in status related to these items.

BUSINESS AND MILESTONE STATUS
(Project Quarter #8: 5/1/2020 to 7/31/2020)

Tasks Scheduled in this Reporting Period

Table 1 shows the project tasks performed in this reporting period. Task 2 was completed last quarter. All tasks are on schedule.

Table 1: Lists of Tasks in this Reporting Period

Task	Scheduled Start	Scheduled Completion	Completion Status
Task 3 - Develop and Execute Testing Matrix	5/1/2019 (1)	10/31/2020	Ongoing
Task 4 - Data Analysis and Model Development and Optimization	8/1/2019 (1)	1/31/2021	Ongoing
Task 6 - Project Management	8/1/2018	4/30/2021	Ongoing
- 8TH Quarter Project Report	5/1/2020	7/31/2020	Completed

(1) 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 on schedule but could be affected by the **COVID-19** pandemic shelter-in-place restrictions; 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 schedule.

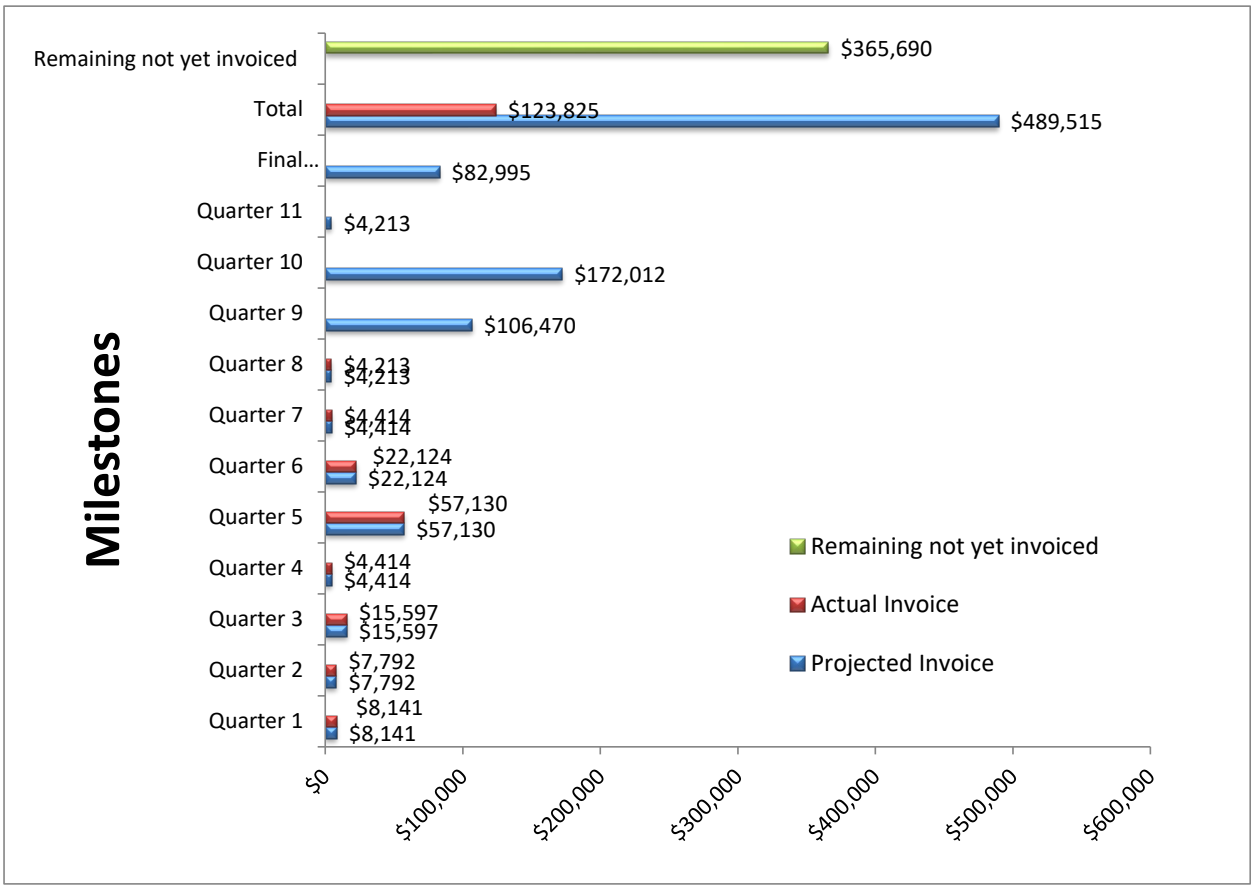
Table 2: Lists of Milestones Linked to Activities/Deliverables/Tasks this Reporting Period

Item No.	Task No.	Activity/Deliverable	Quarter No.	Scheduled Due Date	Completion Date	Payable Milestone/Title
14	6	8th Quarterly Status Report	8	7/31/2020	7/31/2020	Submit 8th quarterly report

Resource Status

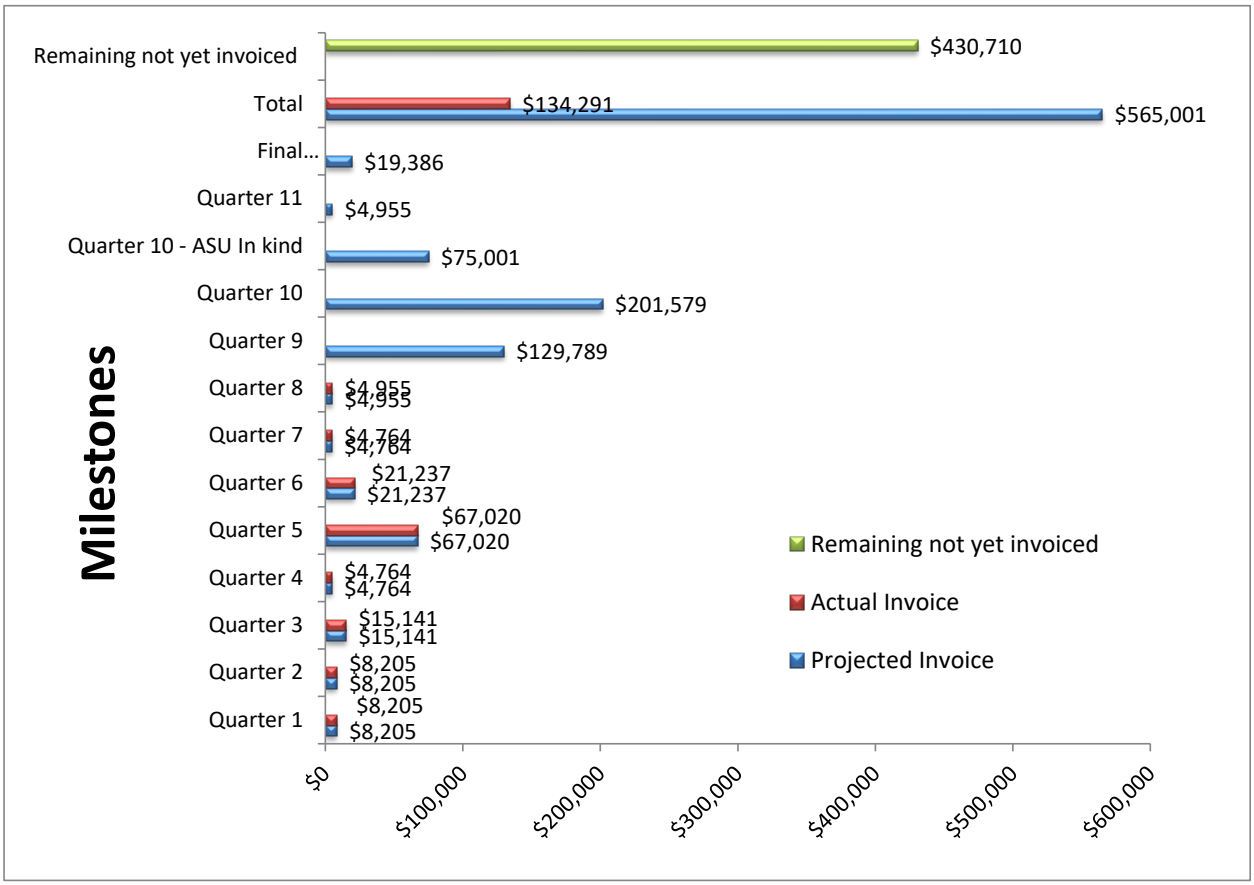
The nature of the contract for this research effort is fixed price, with clearly defined milestone/deliverable payments. Please see Figure 1 for an update on the fund status for DOT/PHMSA, and Figure 2 for the fund status of OTD and ASU.

Figure 1: Quarterly Payable Milestones/Invoices – DOT PHMSA



The eighth quarter invoice will be issued shortly after the end of the 8th project quarter.

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



The eighth quarter invoice will be issued shortly after the end of the 8th project quarter.

Project Risks

Scope and Budget. There are no current risks to the scope and budget.

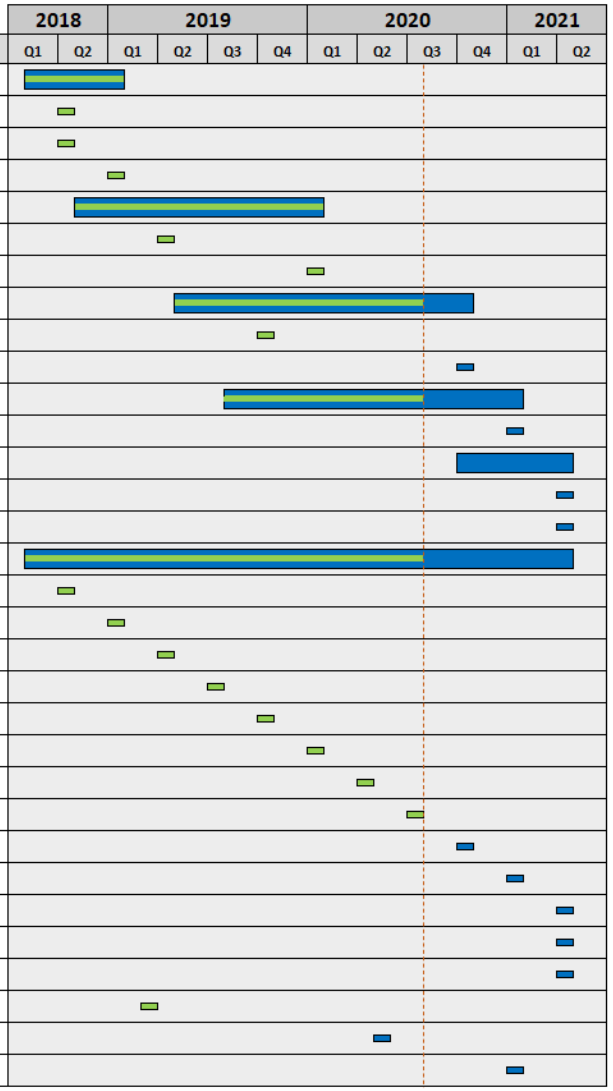
Schedule. Due to the COVID-19 shelter-in-place orders the lab testing was slowed down. The team currently believes it can keep the project on schedule as long as the effects of the pandemic do not last too long. We have finished the bulk of all testing with the remaining lab tests being Charpy toughness testing (in Chicago, IL) and NDTT NDE toughness testing at MMT (Boston, MA). The critical path will be the MMT NDTT testin,g and if that testing is delayed at MMT then there will be a domino effect that could impact the project schedule - but would be limited to the toughness modeling portion of the scope of work and not the yield strength and other properties. We will keep DOT and OTD informed of any changes in status related to these items.

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	100%
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	100%
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	30%
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	50%
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	100%
23	21	M13 - Submit quarterly report - Q7	4/30/2020	4/30/2020	100%
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 #8: 5/1/2020 to 7/31/2020)

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 3 - Develop and Execute Testing Matrix

4. Completed tensile testing for the project. This was a significant accomplishment, since COVID-19 protocols had to be followed to ensure personnel safety.
5. Used the new tensile data to "down-select" the Charpy and NDTT test specimens.
6. Rough cut: (a) Charpy toughness specimens for 20 pipes (10 pipes have Charpy tests already completed); and (b) MMT NDTT specimens for the same 30 specimens. These have both been submitted to the respective labs and are awaiting final sample preparations and testing expected in the next quarter.

Task 4 - Data Analysis and Model Development and Optimization

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

5. Continued with finalizing the Bayesian averaging model: The root mean square error (RMSE) of the averaged model with Bayesian updating was analyzed for the general case and compared to linear model, quadratic model, and GP model. The averaged model has the smallest RMSE which shows the best predictive performance in terms of predicting yield strength for pipes.
6. A journal article for our previous Bayesian modeling averaging work has been accepted in the top journal in this field – Structural Safety.
7. During the quarter, the team also compared the performance of the best linear model, the best quadratic model and the best GP model in general case study. From frequency results, it is fair to say that the linear model is the best among these three models while it is not the dominant one. This inspired the team to check the performance of the weighted model. The RMSE of the weighted model was the smallest, which means it can achieve the best predictive performance.
8. Manifold learning work: Continued the preliminary work related to manifold learning modeling and started investigating the dimension reduction capability for regression analysis of the employed manifold learning techniques.

Task 6 - Project Management

2. Submitted Milestone Item **M14**, the eighth project quarterly report.

PLANS FOR FUTURE ACTIVITY
(Project Quarter #9: 8/1/2020 to 10/31/2020)

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

Task 3 - Develop and Execute Testing Matrix

a. Complete final quality checks and enter final set of tensile data into the flat table.
b. Send ASU the new 35 pipe specimen data sets to add to the prior 35 data sets. This will allow them to conduct the modeling with all 70 data sets moving forward (and the additional calibration set if desired).
c. Complete Charpy baseline (destructive) toughness testing.
d. Complete (tentatively planned) the MMT NDTT NDE toughness testing of the same 30 specimens as the tests being done in (a.) above.
e. Submit summary of executed testing plan data input into project database - Milestone Item **M15**, due 10/31/2020.

Task 4 - Data Analysis and Model Development and Optimization

a. ASU will run the optimized models with the full data set of 70 pipe samples that will now include the second half of 35 new pipeline samples.
b. ASU will continue programming the Bayesian Network model in R with both continuous and discrete variables.
c. GTI will commence with Structural Equation Modeling of the project data set to compliment the ASU modeling work.

Task 6 - Project Management

a. Submit summary of executed testing plan data input into project database - Milestone Item **M15**, due 10/31/2020.
b. Submit ninth project quarterly report to DOT - Milestone Item **M16**, due 10/31/2020.

Respectfully submitted by,

Daniel A. Ersoy

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

End of Quarterly Report Body

APPENDIX A – ASU Task 4 (8TH Quarter Update)

8TH 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
May 1, 2020 through July 31, 2020

Report Issued
July 31, 2020

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

Subcontractor - Arizona State University Team
Dr. Yongming Liu, Jie Chen, Qiongfang Zhang, Nan Xu

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

Abstract

Gaussian Process (GP) is a non-parametric statistical learning model which introduces uncertainly on prediction. However, GP suffers from the curse of dimensionality when the features of data increase. In this research, manifold learning-based GP regression model is proposed for yield strength prediction in order to handle all features of gas pipes. There are two procedures of proposed method, dimension reduction by manifold learning and GP regression for obtained low dimensional representation. The predictive performance of proposed method is compared with linear regression based-on validation results. The results show that the proposed method can handle high dimensional data without any coefficients selection techniques required.

Introduction

Aging pipe data involve 22 features. A critical challenge is the curse-of-dimensionality to analyze this type of high dimensional data by using most existing methods. In other words, the amount of data needed for capturing tendency will increase with dimensions increasing, while the performance will exponentially decrease as well. The most efficiency way to avoid curse of dimensionality is dimension reduction for this stage. There are several dimension reduction techniques available to obtain the low dimensional embedding. Principal Components Analysis (PCA) is one of the most popular linear dimension reduction method, which requires all objects are statistically independent. t-distributed stochastic neighbor embedding (t-SNE) is the commonly used nonlinear dimension reduction method which majorly focus on local structure. Recently, a new nonlinear dimensionality reduction technique, called uniform manifold approximation and projection (UMAP) was introduced, which claimed to preserve as much of local and more of global structure than t-SNE [1, 2]. Therefore, the proposed model for this research is constructed by UMAP and Gaussian Process (GP) regression, where UMAP is used to find the low dimensional representation of high dimensional pipes' data and Gaussian Process (GP) regression model is used to study the low dimensional data points. The major benefit of using manifold learning-based GP regression is to analyze high dimensional/super-high dimensional data from the recognized low-dimensional pattern.

This quarter's work focuses on the yield strength prediction by GP regression based on low dimensional representation obtained by manifold learning. And root mean square errors (RMSE) from both proposed method and linear regression were used to compare the performance of prediction. The result shows relative high precision according to demonstrated examples of aging pipes yield strength prediction.

Another reported activity is on the finalizing of the task for Bayesian network modeling. We have summarized the previously developed model and am working on developing a demo code in R for future deliverables. Example for the development of Bayesian Network using correlational analysis and expert's judgement will be illustrated using the available data. Final prediction accuracy with Bayesian Model Averaging, linear/quadratic regression is compared. All results will be included in the final report.

Manifold-based GP Regression Model

Uniform Approximation and Projection

The theoretical foundations of UMAP include topological data analysis and manifold theory. For brief mathematical preliminaries, Topology is a branch of mathematics focusing on whether certain properties of geometric objects are preserved by homeomorphisms or not, where a homeomorphism is a map between two topological spaces that is bijective (i.e., one-to-one and onto) and bicontinuous (i.e., with the map itself and its inverse both being continuous). A manifold is a topological space that "locally" resembles Euclidean space near each point; that is, any point in the space has a neighborhood in the space which is homeomorphic to an open ball of a Euclidean space.

UMAP has three major assumptions about the data: (a) the data is uniformly distributed on Riemannian manifold; (b) the Riemannian metric is locally constant; (c) the manifold is locally connected [1]. Based on those assumptions, a low dimensional projection can be obtained from constructed fuzzy simplicial complex, the obtained low dimensional projection should be topological equivalent to the constructed fuzzy simplicial complex based on Nerve Theorem [3] A detailed description can be found in in [1]. Here, UMAP is firstly used to find the low dimensional representation of high dimensional data in a L2-space. L2 spaces is known as the Lebasque space, in which a basic regression can be presented. We assume that the low dimensional patterns of high-dimensional data are distributed in this L2-space. For this stage, UMAP can work as both supervised learning and unsupervised learning. In the proposed method, supervised features are used for dimension reduction. As introduced above, it was claimed that UMAP can preserve as much of local and more of global data structure than t-SNE. In other words, the low dimensional embedding calculated by UMAP have more topological information than other dimensionality reduction techniques.

Gaussian Process Regression

GP regression is a stochastic process which specified by the mean function $m(t)$ and the covariance function as reference [4].

$$f(t) = GPR(m(t), k(t, t')) \tag{1}$$

In this project, GP regression is defined with a constant estimated trend as [5 [1] [1]].

$$Y = \epsilon + f(t) \tag{2}$$

Where ϵ is the noise with standard normal distribution and Y is the observed response. The key idea of GP regression is to evaluate the covariance matrix for testing sample x_m itself. Generally, the evaluation is depending on a covariance function or a kernel. A commonly used covariance function is following

$$k(x_i, x_j) = \sigma_f^2 exp\left(-\frac{1}{2l^2}\left\|x_i - x_j\right\|\right) + \sigma_n^2 \delta_{pq} \tag{3}$$

Where $[l,\ \sigma_f, \sigma_n]$ are optimal hyper-parameters obtained by the maximum likelihood method [6]. A GPR model can be constructed based on given covariance function. The posterior distribution can be computed according Bayes' rule by equation below

$$Y \sim N(\bar{Y}(x_t, \sigma(x_t)) \quad (4)$$

Where mean value $\bar{Y}(x_t) = \Sigma_{j=1}^{m}(K + \sigma_n^2 I)^{-1} y$ is also the predicted value for x_t. Here, GP regression is applied on points embedded in L2-space by UMAP.

General procedures can be formulated as below.

$$(x \in R^2) \sim UMAP(X \in R^n) \quad (5)$$

$$e \sim GPR(x, y_i) \quad (6)$$

$$y_t = e + y_s \quad (7)$$

Where R^2 indicates a 2-dimensional L2-space, R^n is a n-dimensional space $(n \gg 3)$, y_t is the true yield strength from tensile test, y_s is yield strength from surface indentation, y_i is either true yield strength or surface yield strength and e is the difference between y_s and y_t.

Results and Discussions

Dataset

In this quarter, 53 samples were used as previous quarter. Manifold learning can handle the high dimensional features, therefore, all features of the given dataset were used in this study as shown in Table 3.

Table 3: 15 Variables for BMA (linear model)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
sC	sMn	sP	sS	sAl	sCr	sCu	sMo	sNb	sNi	sSi	sTi	sV	HRBWod	HRBWid
16	17	18	19	20	21	22								
GSs	ND	WT	FWTND	ST	GC	Y/r								

General Cases Study

The low dimensional representations of general case study are shown in Figure 4 – 2, which obtained by two different variables. For Figure 4, the low dimensional embedding was computed with respect to true yield strength (notated by EBT), while the low dimensional embedding was calculated with respected to surface yield strength (notated by EBS) in Figure 5. In both figures, the color of each point indicates the corresponding yield strength.

To evaluate the performance of the proposed method, 5 test points were randomly selected and the rest data was used as training points to present the 10-fold cross-validation, where the testing embedding of one trial was also presented in Figure 4 – 2. And the corresponding numerical results are given in Table 4, in which the prediction performance of EBS is much better than another one. The reason probably is that errors between two yield strengths are more relatively to the surface yield strength. Therefore, an equation can use to show the relationship among true yield strength, surface yield strength and errors between two strengths as following

$$y_t \sim e(y_s) + y_s \tag{8}$$

Table 4: RMSEs for Two Embedding Ways

Embedded By	Training set	AVG. RMSE (10 Runs)
True Yield Strength (EBT)	48	4.05
Surface Yield Strength (EBS)	48	3.41

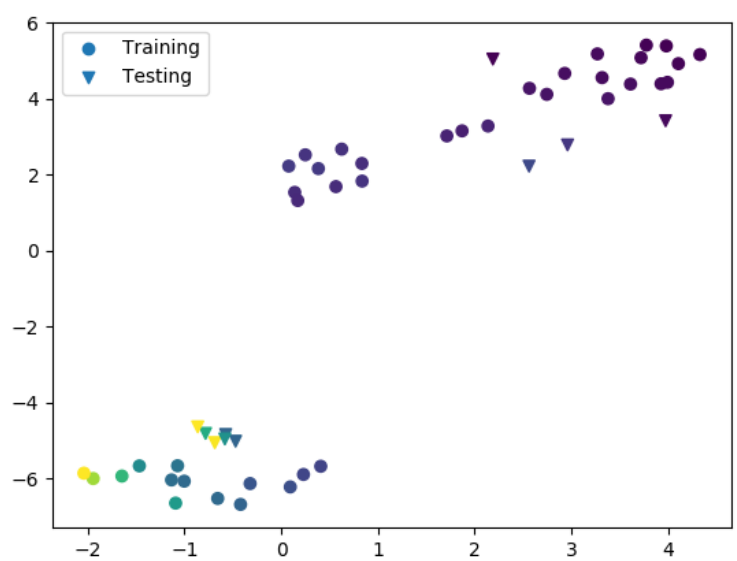


Figure 4: Low Dimensional Representation Obtained by True Yield Strength

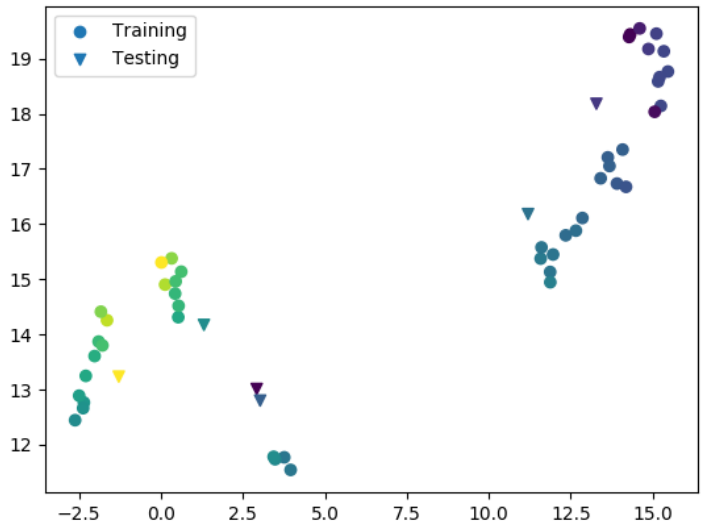


Figure 5: Low Dimensional Representation Obtained By Surface Yield Strength

Similarly, linear regression was used for comparison without preselection. The corresponding result is shown in Table 5. It is very clearly that the proposed method has the better performance.

Table 5: RMSEs for linear model

Training set	AVG. RMSE (10 Runs)
48	5.52

Specific Case Study

The best predictions the minimum root mean square errors of using linear regression, using EBT and using EBS were pick up as shown in Table 6. In the scenario as general case by using 48 training points and 5 test sample, the minimum RMSEs of EBT is 0.3102 while the minimum RMSE of EBT is 1.6641 and the minimum RMSE of linear regression is 1.1740. A conclusion can be made that for total samples of 53, manifold-based GP regression with low dimensional representation embedded by surface yield strength is better than EBT and linear regression with respect to yield strength prediction.

Table 6: Minimum RMSE for 3 Models

Models	Min RMSE
LR	1.1740
EBT	1.6641
EBS	0.3102

Final Results for the Averaged Model (Q8)

In the last quarter work, we compared the performance of the best linear model, the best quadratic model and the best GP model in general case study. We did 200 iterations for each group of validation and counted the number of the largest posterior for each model. From frequency results, we knew that it is fair to say that the linear model is the best among these three models while it is not the dominant one. This inspired us to check the performance of the weighted model. Then 100 runs hold-out cross validation was conduct and the comparison results are shown in Table 7. The RMSE of weighted model is the smallest 2.7632 which means it can achieve the best predictive performance.

Table 7: RMSEs of the Averaged Model General Case Study

	LRmse	QRmse	GPRmse	WeightedRmse
RMSE	3.0895	3.2039	3.7141	2.7632

Conclusion

In this research, for small dataset with 53 samples in total, manifold-based Gaussian Process regression was proposed to predict the yield strength of pipes, and compared with the linear regression and different inputs for manifold learning. Based on results from both general case study and specific case study, it is clearly to conclude that the manifold-based GP regression model by using surface yield strength to obtain the low dimensional representation has a better performance than EBT and linear regression. The benefit of the proposed study for engineering application shows the high computational efficiency for large parameters' system. In addition, UMAP, as a non-deterministic method, when integrated with GP regression, can partially bring the uncertainty into the prediction.

References

[2] McInnes, Leland, John Healy, and James Melville. "Umap: Uniform manifold approximation and projection for dimension reduction." arXiv preprint arXiv:1802.03426 (2018).

[3] Becht, Etienne, et al. "Dimensionality reduction for visualizing single-cell data using UMAP." Nature biotechnology 37.1 (2019): 38-44.

[4] Ghrist, Robert W. Elementary applied topology. Vol. 1. Seattle: Createspace, 2014.

[5] Su, G., Peng, L., and Hu, L., “A Gaussian process-based dynamic surrogate model for complex engineering structural reliabilityanalysis,” Structural Safety, Vol. 68, 2017, pp. 97–109

[5] Rasmussen, C., and C. Williams. "Gaussian Processes for Machine Learning. Massachusetts." (2006).

End of Report