7TH 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
February 1, 2020 through April 30, 2020

Report Issued
April 30, 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 (see Appendix A) 6

Task 6 – Project Management 6

Risks 6

BUSINESS AND MILESTONE STATUS (Project Quarter #7: 2/1/2020 to 4/30/2020) **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 #7: 2/1/2020 to 4/30/2020) **11**

Task 3 - Develop and Execute Testing Matrix 11

Task 4 - Data Analysis and Model Development and Optimization 11

Task 6 - Project Management 12

PLANS FOR FUTURE ACTIVITY (Project Quarter #8: 5/1/2020 to 7/31/2020) **13**

Task 3 - Develop and Execute Testing Matrix 13

Task 4 - Data Analysis and Model Development and Optimization 13

Task 6 - Project Management 13

APPENDIX A – ASU Task 4 (7TH Quarter Update) **14**

Abstract 15

Introduction 15

Model Selection, Updating and Averaging 16

Bayesian Linear Regression Using the Reference Prior 16
Model Selection, Updating and Averaging 17
Results and Discussions 18
BMA for 53-point dataset 18
Comparison Between Linear Model and SGP model 21
Toy problem 22
Validation for Yield Strength Prediction 25
Preliminary results of using manifold learning 26
Conclusion 27
References 28

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: Methodology flow chart 16
Figure 5: Rank list of variables (linear model) 19
Figure 6: Best 20 linear models 19
Figure 7: Rank list of variables (quadratic model) 20
Figure 8: Best 20 quadratic models 21
Figure 9: Toy problem case #1 23
Figure 10: Toy problem case #2 23
Figure 11: Toy problem case #3 24
Figure 12: Pipeline embedding 26

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) 19
Table 4: 20 Variables for BMA (quadratic model) 20
Table 5: RMSEs for linear model and SGP model 21
Table 6: Posteriors and RMSEs for toy problem 24
Table 7: Posteriors and RMSEs for yield strength prediction 25
Table 8: General case results for yield strength prediction 25
Table 9: RMSEs of testing prediction 26

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

Task 3 – Develop and Execute Testing Matrix

Remaining Nondestructive Testing

a. The NDE testing phase of the project from MMT, Frontics, and SciApps is *completed* with the exception of the 30 small coupons that will be tested later in the project at the MMT labs using their prototype NDTT toughness testing system.

Lab-Based (Destructive) Testing

a. This lab testing continued in Quarter 7 until a mandatory shelter-in-place restriction due to the COVID-19 pandemic was activated in the second half of the quarter.
b. Lab testing completed in Quarter 7 included full wall tensile, hardness, chemistry, and grain size:
   - Chemistry, grains size, and steel grading were *completed* and entered in the master flat table.
   - For the remaining 35 tensile test samples (105 total specimens) we completed deburring, blasting, measuring and painting all the tensile specimens - they are ready for tensile testing next.

All testing is currently on hold until the mandatory shelter-in-place restrictions due to the COVID-19 pandemic are relaxed.

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

- During this quarter, a new rank list of useful variables obtained by using BMA with new 53-point dataset was developed.
- Following that, the best linear model and the best quadratic were achieved. And the best GP model consisted of the top 3 variables in the list.
- A "Toy problem" was used to show basic procedures of doing model selection, updating and averaging. The results indicated that the justification of using a particular model is case-dependent and the weighted model has potential to achieve a better performance.
- Both specific case and general cases were analyzed to evaluate the performance of this new method in terms of doing yield strength prediction.
- It was found that the linear model is the best among the three model types, while it is not the dominant one, i.e. there was not a single dominant model.
- The preliminary results of using manifold learning demonstrated the potential ability to handle a high dimension regression problem..
- A detailed report of this quarter's modeling efforts and results is presented in **Appendix A** of this report.

Task 6 – Project Management

- Submitted Milestone Item **M13**, the seventh 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 remaining lab testing is on hold for now. We can however continue with the modeling tasks using the partial data set we have so far, but this will eventually need to be placed on hold in the future if the remaining testing results are not completed in a timely manner.

We think we can keep the project on schedule as long as the effects of the pandemic do not last too long, and we can start the testing back within approximately the next quarter. If the delays continue longer than that, then there will be a domino effect that could impact the project schedule.

BUSINESS AND MILESTONE STATUS
(Project Quarter #7: 2/1/2020 to 4/30/2020)

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
- 7TH Quarter Project Report	2/1/2020	4/30/2020	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 if lab work is not allowed to resume in a few months; 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.

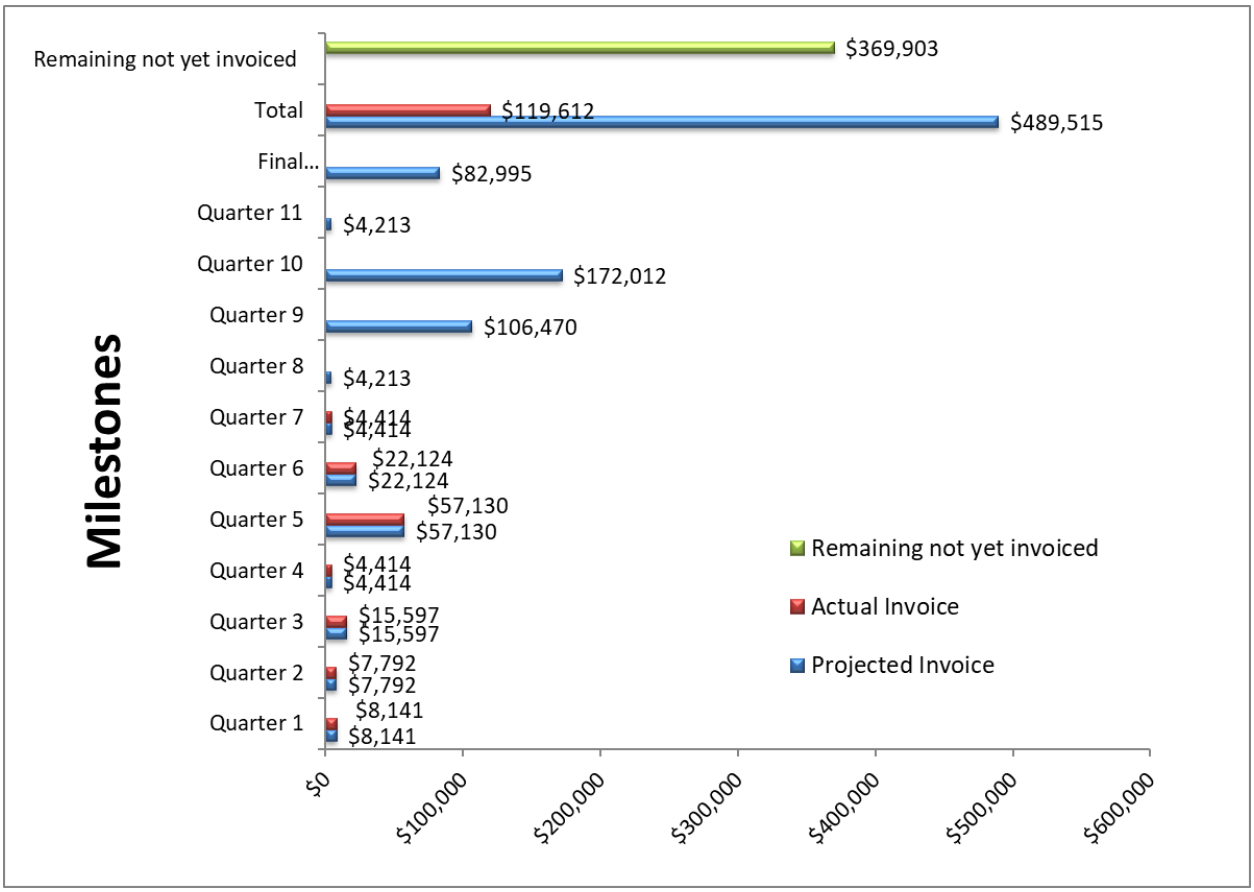
Table 2: Lists of Milestones Linked to Activities/Deliverables/Tasks this Reporting Period

Item No.	Task No.	Activity/Deliverable	Quarter No.	Scheduled Due Date	Completion Date	Payable Milestone/Title
13	6	7th Quarterly Status Report	7	4/30/2020	4/30/2020	Submit 7th quarterly report

Resource Status

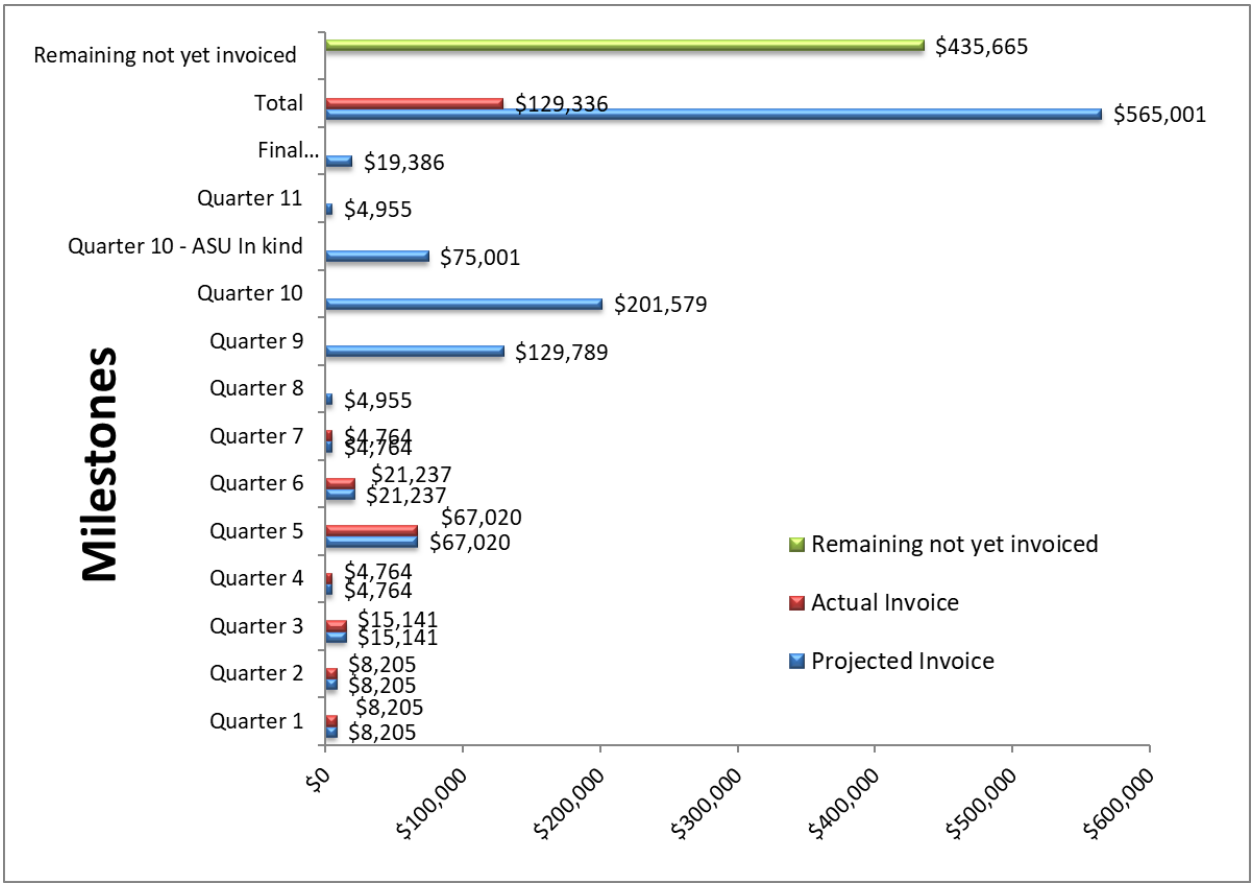
The nature of the contract for this research effort is fixed price, with clearly defined milestone/deliverable payments. Please see Figure 1 for an update on the fund status for DOT/PHMSA, and Figure 2 for the fund status of OTD and ASU.

Figure 1: Quarterly Payable Milestones/Invoices – DOT PHMSA



The seventh quarter invoice will be issued shortly after the end of the 7th project quarter.

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



The seventh quarter invoice will be issued shortly after the end of the 7th project quarter.

Project 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 remaining lab testing is on hold for now. We can however continue with the modeling tasks using the partial data set we have so far, but this will eventually need to be placed on hold in the future if the remaining testing results are not completed in a timely manner.

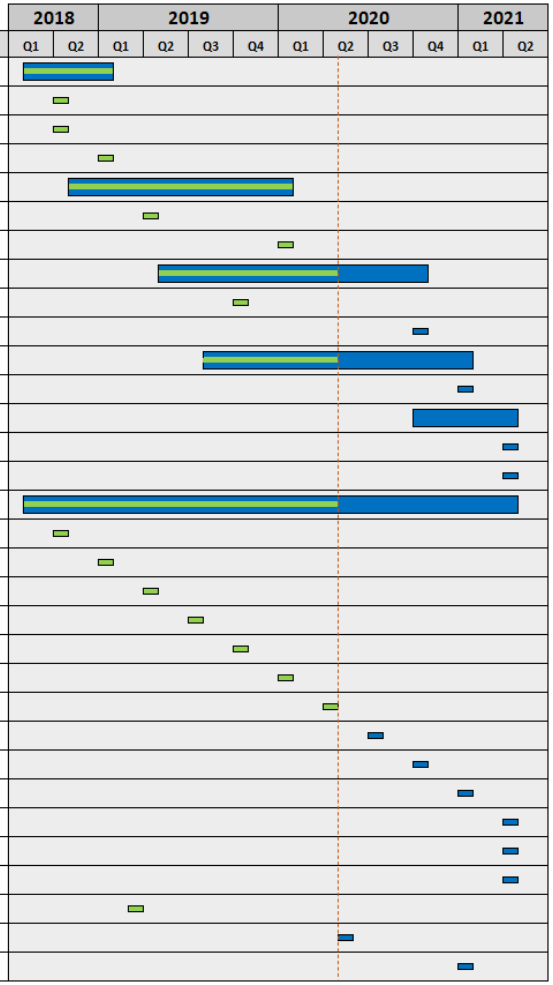
We think we can keep the project on schedule as long as the effects of the pandemic do not last too long, and we can start the testing back within approximately the next quarter. If the delays continue longer than that, then there will be a domino effect that could impact the project schedule.

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 #7: 2/1/2020 to 4/30/2020)

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

1. **Remaining Nondestructive Testing**

b. The NDE testing phase of the project from MMT, Frontics, and SciApps is *completed* with the exception of the 30 small coupons that will be tested later in the project at the MMT labs using their prototype NDTT toughness testing system. This is on hold until the mandatory shelter-in-place restrictions due to the COVID-19 pandemic are relaxed.

c. Once the destructive testing in item (2) directly below is completed, GTI will select the 30 samples to section coupons out and send to MMT for NDTT testing. This is on hold until the mandatory shelter-in-place restrictions due to the COVID-19 pandemic are relaxed.

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

    c. This lab testing continued in Quarter 7 until a mandatory shelter-in-place restriction due to the COVID-19 pandemic was activated in the second half of the quarter.

    d. Lab testing completed in Quarter 7 included full wall tensile, hardness, chemistry, and grain size:
    - Chemistry testing was *completed* and entered in the master flat table.
    - Initial review of chemistry data was *completed*; more detailed quality check is in progress and will continue into the future quarters.
    - Grain size and hardness testing and analysis were *completed*, and data was entered into the master flat table.
    - Also *completed* the assignment of steel grade and type based on lab testing.
    - For the remaining 35 tensile test samples (105 total specimens) we completed deburring, blasting, measuring and painting all the tensile specimens - they are ready for tensile testing next. This is on hold until the mandatory shelter-in-place restrictions due to the COVID-19 pandemic are relaxed.

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:

1. Statistical inference methods were used to explain the difference of bulk yield strength and surface indentation-based yield strength with respect to many measured quantifies, including chemical composition, microstructure, and hardness, etc.
2. Linear model, quadratic model, and Gaussian Process model were evaluated using Bayesian weight updating and averaging.
3. It was observed that the combined model with updated weights had the most robust performance.
4. In order to further enhance the accuracy and efficiency, dimension reduction using manifold fold learning was explored and preliminary results developed.

Task 6 - Project Management

1. Submitted Milestone Item **M13**, the seventh project quarterly report.

PLANS FOR FUTURE ACTIVITY
(Project Quarter #8: 5/1/2020 to 7/31/2020)

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. Once the lab reopens from the COVID-19 pandemic stand down, and staff returns to work, the remaining tensile testing will be scheduled and executed.
b. Once the tensile testing is done, the data will be entered in the flat table and be quality checked.
c. Then a complete review of the YS, UTS, Microstructure, and Chemistry will be conducted to select a subset of pipe samples for Charpy impact toughness testing.
d. After tensile testing is completed, 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.
e. GTI will also complete Charpy toughness testing to generate full S-curves for impact energy, lateral expansion, and %-shear failure.

Task 4 - Data Analysis and Model Development and Optimization

a. Project modeling tasks can continue for now while sheltering in place but will eventually might have to be put on hold if the stand down from the COVID-19 pandemic prevents the other 50% of the test data (see Task 3 above) from being completed in a timely manner.
b. In the next quarter ASU will focus on the fine-tuning of uniform approximation and projection (UMAP) parameters and estimate the yield strength difference between strength measured by tensile testing and strength measured by surface indentation technique.

Task 6 - Project Management

a. Submit eighth project quarterly report to DOT - Milestone Item **M14**, due 7/31/2020.

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 (7TH Quarter Update)

7TH 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
February 1, 2020 through April 30, 2020

Report Issued
April 30, 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

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

Bayesian Model Averaging (BMA) is an application of Bayesian inference to the problems of model selection, combined estimation and prediction. However, in terms of applying the BMA method, different datasets may lead to different results even dealing with the same problem. Thus, in this research, we start with using BMA to do the variable selection with the new 53-points dataset.

A new rank list of variables is obtained which refers to the usefulness for pipeline yield strength prediction. The best linear model and the best quadratic model are selected at the same time. Moreover, many uncertainties exist in the inference process including model uncertainty, parameter uncertainty and measurement uncertainty which are really matter but easy to be ignored. Based on Bayesian methods, model selection, updating and averaging are combined well to address these uncertainties.

The overall procedure is demonstrated by a toy problem and the pipeline yield strength prediction problem.

Furthermore, a novel manifold learning-based Gaussian Process regression model is proposed to conduct non-linear dimensionality reduction and non-parametric prediction. Some preliminary results are shown to validate the performance of the method.

Introduction

Bayesian Model Averaging (BMA) is an extension of the usual Bayesian inference methods in which one does not only models parameter uncertainty through the prior distribution, but also model uncertainty obtaining posterior parameter and model posteriors using Bayes' theorem and therefore allowing for allow for direct model selection, combined estimation and prediction [1]. Sometime, the predictions of pipeline yield strength are not reliable under realistic service conditions due to various uncertainties. Those uncertainties come from material property, manufacturing process, model choice, model parameters, mechanism modeling, measurement data, as well as numerical evaluations [2]. For a specific model, the associated model parameters also have statistical uncertainties introduced by regression analysis with experimental data. The justification of using a particular model depends on the actual problem. One approach to justify using one particular model is to update the initial belief on that model using measurement data from experiments. Bayesian updating is one of the most commonly used methods [3]. The model determination can be made based on the results of Bayes factors in a hypothesis testing context [4]. The pipeline data involves 20 features. A critical challenge is the curse-of-dimensionality to analyze this type of high dimensional data by using most existing methods. Manifold learning-based Gaussian Process (GP) regression model has a great potential for strength prediction which can easily achieve surface information fusion of the aged pipelines. The model is constructed by a new nonlinear dimensionality reduction technique, called uniform approximation and projection (UMAP), and Gaussian Process. 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 based on model selection, updating and averaging. First, we applied BMA with the new 53-point dataset to choose the most useful variables, the best linear model, the best quadratic model and the best GP model. Then conduct model selection, updating and averaging by using these three best models. Methodology flow chart can be seen in Figure 4. Lastly, Manifold learning-based Gaussian Process (GP) regression model is introduced to explore the possibility of combining non-linear dimensionality reduction with non-parametric GP prediction.

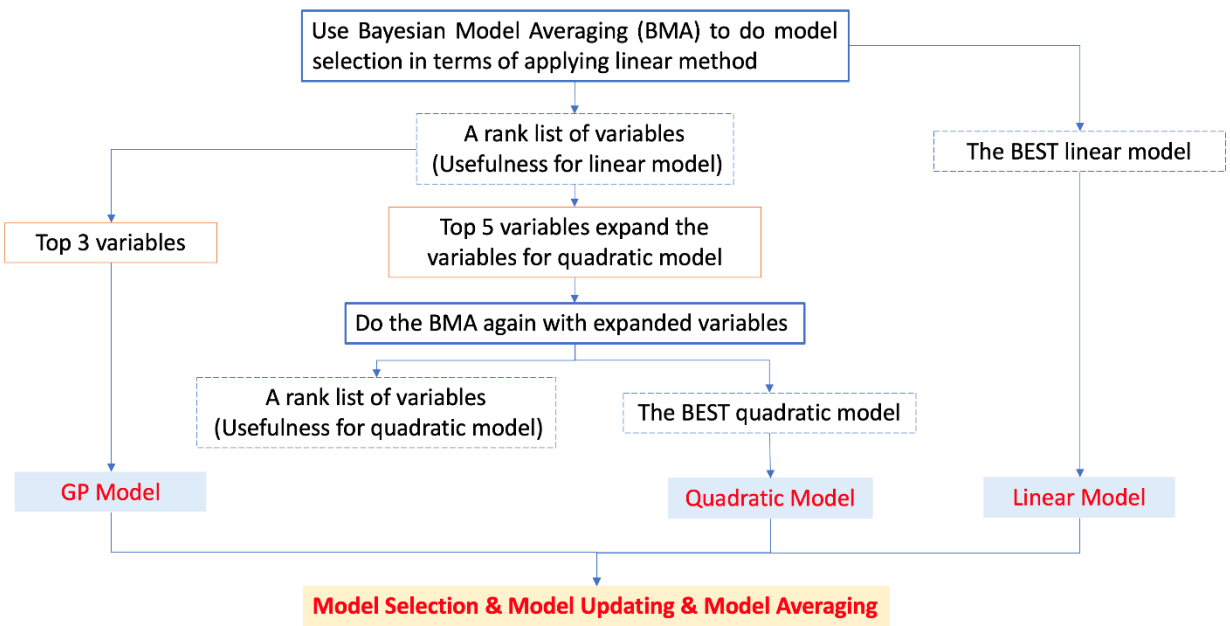


Figure 4: Methodology flow chart

Model Selection, Updating and Averaging

Bayesian Linear Regression Using the Reference Prior

The Bayesian linear model starts with the same model as the classical frequentist approach:

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \cdots, n. \tag{1}$$

With the assumption that the errors, ϵ_i, are independent and identically distributed as normal random variables with mean zero and constant variance σ^2. This assumption is exactly the same as in the classical inference case for testing and constructing confidence intervals for α and β. Our goal is to update the distributions of the unknown parameters α, β, and σ^2, based on the data $x_1, y_1, \cdots, x_n, y_n$, where n is the number of observations.

Under the assumption that the errors ϵ_i are normally distributed with constant variance σ^2, we have for the random variable of each response Y_i, conditioning on the observed data x_i and the parameters α, β, and σ^2, is normally distributed:

$$Y_i|x_i, \alpha, \beta, \sigma^2 \sim Normal(\alpha + \beta x_i, \sigma^2) \tag{2}$$

We first consider the case under the reference prior, which is out standard noninformative prior. Using the reference prior, we will obtain familiar distributions as posterior distributions of α, β, and σ^2, which gives the analogue to the frequentist results. Here we assume the joint prior distribution of α, β, and σ^2 to be proportional to the inverse of σ^2

$$p(\alpha, \beta|\sigma^2) \propto \frac{1}{\sigma^2}. \tag{3}$$

Using the hierarchical model framework, this is equivalent to assuming that the joint prior distribution of α, and β under σ^2 is the uniform prior, while the prior distribution of σ^2 is proportional to $\frac{1}{\sigma^2}$. That is

$$p(\alpha, \beta|\sigma^2) \propto 1. \tag{4}$$

$$p(\sigma^2) \propto \frac{1}{\sigma^2}. \tag{5}$$

Combining the two using conditional probability, we will get the same joint prior distribution.

Model Selection, Updating and Averaging

The probabilistic inference is usually associated with a specific model M_k and it is conditional on the assumption that the model is the correct one which can fully describe the physical phenomenon. However, when the mechanism is not exactly clear, multiple models may be available to simplify the actual complicated mechanism.

The joint distribution of an event X and model M_k with a n_k – dimensional parameter or hyperparameter can be expressed as

$$p(X, M_k) = P(M_k)p(X|M_k), \tag{6}$$

where $k \in K$ is the model index, $P(M_k)$ is the prior probability assigned to model M_k, and $p(X|M_k)$ is the conditional probability distribution of event X given model M_k. Notice that we treat each model as a black box which simplified the uncertainties of model parameters. According to Bayes theorem, the total probability of event X as

$$P(X) = \sum_{k \in K} P(M_k)\, P(X|M_k), \tag{7}$$

where $P(X|M_k)$ is the global likelihood for model M_k. And the posterior probability of model M_k reads

$$P(M_k|X) = \frac{P(M_k)P(X|M_k)}{P(X)}. \tag{8}$$

It is of more interest to consider the model posterior probability ratios of two models. The ratio is referred to as the odds ratio,

$$O_{ij} = \frac{P(M_i|X)}{P(M_j|X)} = \frac{P(M_i)}{P(M_j)} B_{ij}. \quad (9)$$

Where $\frac{P(M_i)}{P(M_j)}$ is the prior odds ratio and $B_{ij} = \frac{P(X|M_i)}{P(X|M_j)}$ is called the Bayes factor in favor of model M_i over model M_j. Given equal prior model probabilities, the odds ratio is reduced to the Bayes factor. Model selection or determination is usually based on the Bayes factors in a Bayesian hypothesis testing context. In this research, we use the equal prior and the likelihood to update the model posterior. By comparing the posterior values to determine which model is more reasonable.

As for model averaging, we added a weight to each model and got a global weighted model as follow,

$$M_{weighted} = \sum_{k \in K} w_k \cdot M_k. \quad (10)$$

where w_k equals to posterior and $\sum_{k \in K} w_k = 1$.

Results and Discussions

BMA for 53-point dataset

In last quarter, we did the BMA with the 20-point dataset and selected three most useful variables (Cu, Hardness, Cr). Then, these three variables used as inputs for linear model and SGP model. Both two-dimensional and three-dimensional analysis are conducted to evaluate the performance of the linear model and the SGP model. From the results, the SGP model has a similar performance of linear model. However, when the dimension goes up with unchanged training set or the training set becomes smaller with unchanged dimensions, the linear regression will be slightly better than SGP regression. Now, we totally have 53 samples including 20 samples we had before and other 33 new samples. Since different datasets may lead to different results even dealing with the same problem when applying the BMA method, so we need to do the BMA again with the new 53-point dataset to see what will happen this time.

The same 15 variables used as inputs for the BMA are shown in Table 3. And the difference between YS_t - YS_s is the output, where YS_t is the yield strength measured by tensile testing which can be seen as the benchmark of predictions and YS_s is the yield strength got from surface indentation technique. The posterior inclusion probabilities for all input variables are show in Fig. PIP is for posterior inclusion probabilities which indicates the rank list of usefulness in terms of using linear method to do the regression. As can be seen in Figure 4, the top three useful variables are C, Si and Mo which are different with the results we had for the 20-point dataset. Likewise, we will use these three variables to compare linear model and SGP model in the next subsection.

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	GSsurface

	PIP	Post Mean	Post SD	Cond.Pos.Sign	Idx
sC	0.9904444	-44.259169112	10.4992594	0.00000000	1
sSi	0.8964444	-22.936739169	10.3834107	0.00000000	11
sMo	0.8792222	155.220027915	78.8463838	1.00000000	8
sS	0.6907778	-97.294327998	79.9428553	0.00000000	4
GSsurface	0.5972222	0.730347479	0.7179824	1.00000000	15
sCu	0.4630000	-18.646832351	23.9235227	0.00143988	7
sMn	0.3754444	1.535866415	2.4676123	1.00000000	2
sNi	0.3232222	17.095338912	30.7441691	0.99484359	10
sCr	0.2926667	15.242916783	30.2537035	1.00000000	6
sTi	0.1597778	-21.821658518	82.4977899	0.02086231	12
sV	0.1384444	-1.833450869	20.5753877	0.29213483	13
HRBWod	0.1354444	0.008676003	0.0614173	0.75717801	14
sAl	0.1314444	-1.118723261	10.3293544	0.19442096	5
sNb	0.1312222	-8.175794812	47.5421636	0.11769687	9
sP	0.1084444	2.125319621	40.8282853	0.64856557	3

Figure 5: Rank list of variables (linear model)

Besides the rank list of variables, we also selected the best 20 models based on posterior model probabilities as shown in Figure 6. The color indicates signs of coefficients for variables where blue refers to positive and red is negative. From the figure, it is very easy to have the best linear model with five variables (C, Mo, Si, S, GSsurface). We will use the best linear model, the best quadratic model and the best GP model to do the model selection, updating and averaging later.

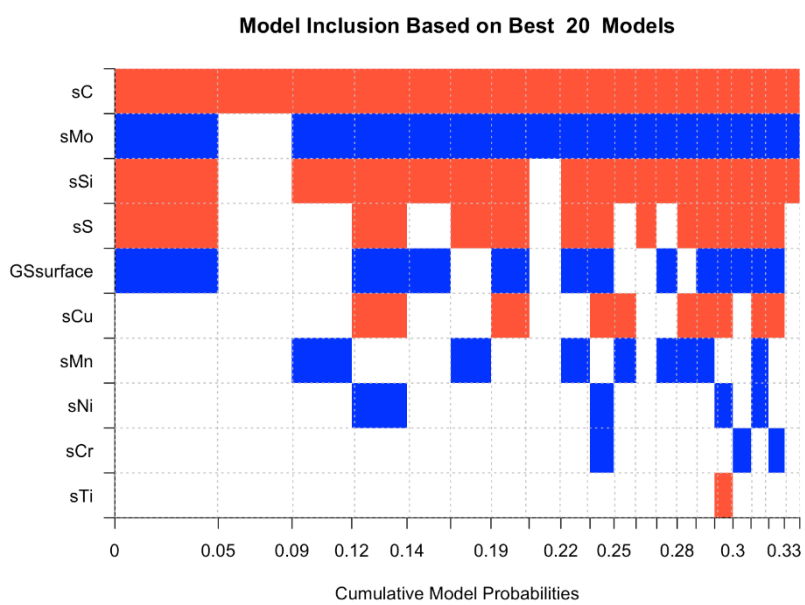


Figure 6: Best 20 linear models

In order to have the best quadratic model, we used the top five variables in Fig to do the expansion. Then we got 20 expanded variables including 5 quadratic terms, 10 interactive terms and 5 linear terms. The variables are shown in Table 4. Similarly, we did another BMA with these 20 inputs and got a new rank list of variables shown in Figure 7 as well as the best 20 quadratic model in Figure 8. The best quadratic model consists with three interactive terms.

In this subsection, two rounds of BMA are applied to analyze the usefulness of variables for both linear model and quadratic model, and to select the best linear model and the best quadratic model.

Table 4: 20 Variables for BMA (quadratic model)

Quadratic terms: sC^2, sS^2, sMo^2, sSi^2, GS^2
Interactive terms: sCsS, sCsMo, sCsSi, sCGS, sSsMo, sSsSi, sSGS, sMosSi, sMo GS, sSiGS
Linear terms: sC, sS, sMo, sSi, GS

	PIP	Post Mean	Post SD	Cond.Pos.Sign	Idx
sC2	0.53644444	-75.27466207	89.7249815	0.00600663	6
sCsS	0.48911111	-501.72428613	628.0273919	0.00022717	11
sMoGs	0.38511111	7.69718826	15.8776914	0.96912868	19
sSi	0.36877778	-15.29748637	32.7783768	0.01265441	4
GS2	0.33966667	0.04601396	0.1356914	0.99640170	10
sMo	0.30700000	60.51204110	167.4965846	0.96308360	3
sSsSi	0.28555556	-225.11562425	463.5128085	0.07392996	16
sCsSi	0.28033333	-32.67426114	69.1572896	0.05192231	13
GS	0.27777778	-0.22208511	2.7045362	0.79520000	5
sS	0.21277778	-39.69358400	180.4538943	0.20522193	2
sSsMo	0.19477778	-110.62110739	3494.0012275	0.59212778	15
sCGS	0.17466667	0.35063784	2.3289591	0.68893130	14
sSGS	0.16422222	4.66747694	17.1666183	0.80108254	17
sMosSi	0.16422222	129.00876804	403.9439221	0.90866035	18
sSiGS	0.16277778	0.25792936	2.5964152	0.43003413	20
sSI2	0.15733333	13.04511822	54.1135416	0.77330508	9
sS2	0.15244444	-123.99618210	1242.2543971	0.29008746	7
sCsMo	0.13911111	-78.95077138	419.6429814	0.28194888	12
sC	0.13855556	-0.29132129	26.0416824	0.48676824	1
sMo2	0.08544444	50.01477876	1137.2061445	0.57477243	8

Figure 7: Rank list of variables (quadratic model)

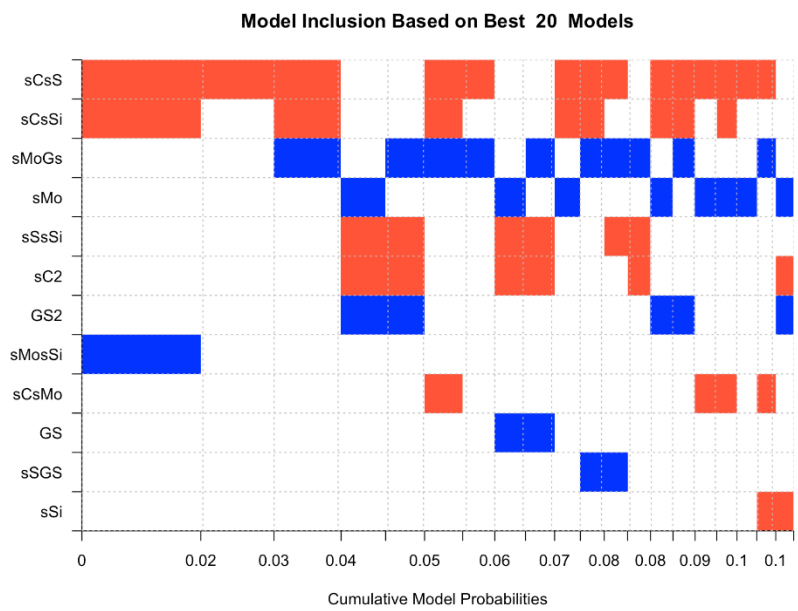


Figure 8: Best 20 quadratic models

Comparison Between Linear Model and SGP model

In this subsection, the procedures of validation are exactly the same as we did before in terms of evaluating the performance of linear model and SGP model but with the different input variables. First, randomly selected 5 or 8 samples from the dataset to be a testing set, and the rest samples would work as the training set correspondingly. Second, train the SGP model and the linear model with the training set. Third, predict the output values with the testing set. Finally, analyze the results of the predictions and evaluate the predictive performance. Numerical comparison results are shown in Table 5. In two-dimensional case, the averaged root mean square error (AVG. RMSE) of SGP model is slightly smaller than linear model's. While in three-dimensional case, the linear model has the smaller AVG. RMSE. This validation results show the correctness of the conclusion we got before that the SGP modeling has a great potential to achieve better predictive performance as long as having more training data. But when the dimension goes up with unchanged training size, the performance of SGP model will get worse. Hence, having larger training set is really critical to SGP prediction. Otherwise, from the table, we can select the best GP model with three variables sC, sSi and sMo.

Table 5: RMSEs for linear model and SGP model

Input Variables	# Training points	AVG. RMSE (500 iterations)	
		Linear	SGP
sC, sSi	48	3.983	3.918
	45	4.0358	4.0054
sC, sSi, sMo	48	3.7325	3.8318
	45	3.8224	3.9455

Toy problem

In this subsection, we made up a toy problem to illustrate how to do the model selection, updating and averaging with several different mechanism models. We use an exponential function to general a bunch of data with some noises. Then randomly selected 5 test points, and the rest data used as training set. Three different mechanism models including linear model, quadratic model and GP model are applied to do the regression and prediction. The predictive mean values will be got for the 5 test points under each model. Suppose every response Yi is distributed to a normal distribution with mean y_i and standard deviation σ. y_i is true value or the benchmark. Then, the likelihood of each Yi given input Xi, model Mk and σ is given by

$$P(Y_i|X_i, M_k, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} exp\left(-\frac{(y_i - Y_i)^2}{2\sigma^2}\right). \quad (11)$$

According to the equation (11), the posterior is proportional to prior times likelihood,

$$P(M_k|X_i, Y_i, \sigma) \propto P(M_k) \cdot P(Y_i|X_i, M_k, \sigma). \quad (12)$$

Without any clear information about the mechanism, we applied uniform priors to the three models which means at the beginning the priors of each model are exactly the same,

$$P(M_L) = P(M_Q) = P(M_{GP}) = 1/3. \quad (13)$$

After using the first test point to calculate, we got posteriors for each model. And then, we use these posteriors as new priors for the second test point. Following the same procedure, the posteriors would be updated for five times and the final posteriors show the preference of each model. There are three cases shown in Figure 9 to Figure 11, and numerical results for posteriors and RMSEs are reported in Table 6. A larger posterior or a smaller RMSE indicates a better model. When we analyze posteriors, linear model is the best in the first case, quadratic model is the best in the second case and the GP model is the best in the third case. Notice that the weighted model is the best in the second case if considering a smaller RMSE. From these results, we would like to conclude that the justification of using a particular model is case-dependent. And the weighted model has potential to achieve a better performance.

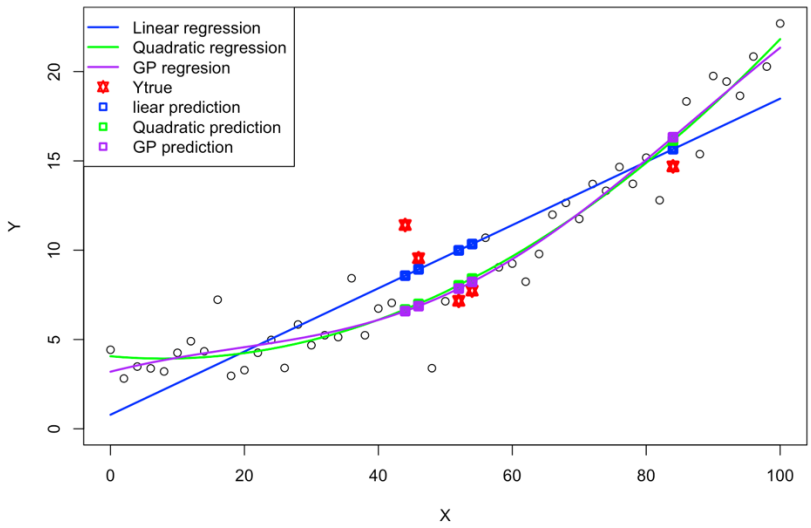


Figure 9: Toy problem case #1

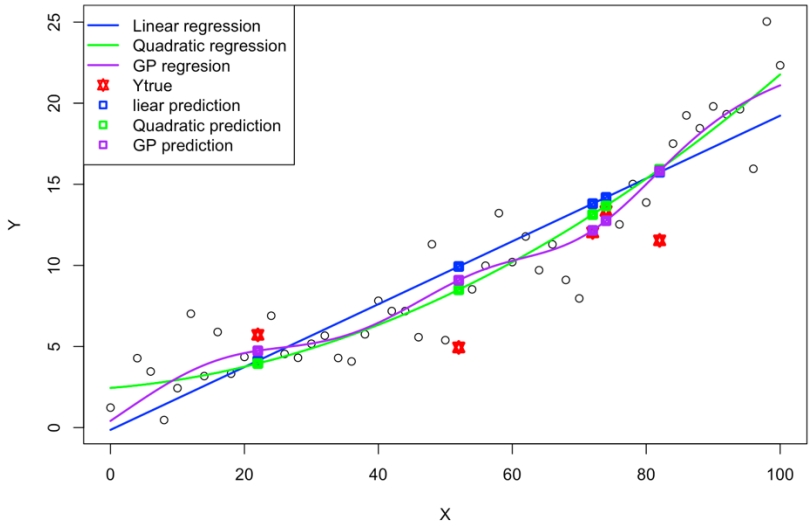


Figure 10: Toy problem case #2

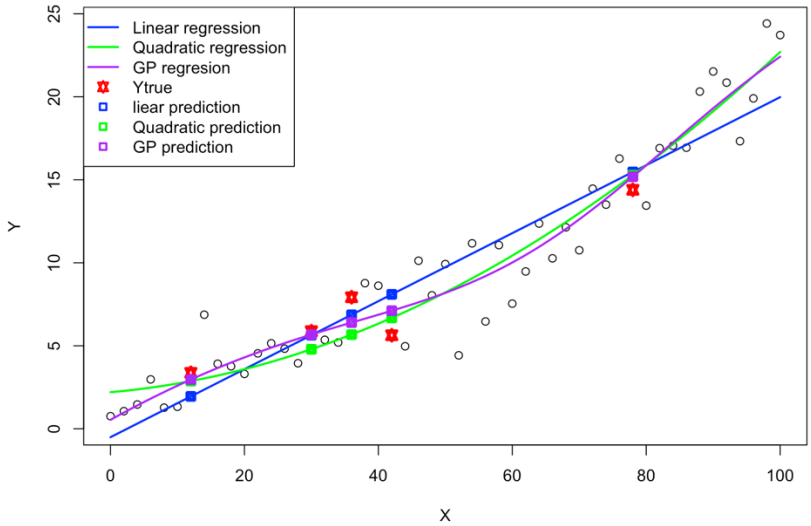


Figure 11: Toy problem case #3

Table 6: Posteriors and RMSEs for toy problem

# Case	Post. L	Post. Q	Post. GP	LRmse	QRmse	GPRmse	WeightedRmse
1	0.9754948	0.0167936	0.007711597	2.196704	2.539742	2.600307	2.188918
2	0.001022391	0.5619517	0.4370259	3.129167	2.695919	2.714507	2.679235
3	0.06089923	0.1694588	0.7696419	1.443569	1.294038	1.034026	1.078101

Validation for Yield Strength Prediction

Specific Case Study

When doing model selection, updating and averaging for 53-point pipeline dataset, the whole procedures are the same as what we did for toy problem. From BMA, we already got the best linear model, the best quadratic model and the best GP model. In this subsection, these three models are used to predict the yield strength for pipelines. The results for specific case study are shown in Table 7. Linear model is the best in the first case, quadratic model is the best in the second case and the GP model is the best in the third case. In both the second case and the third case, the RMSE for weighted model is the smallest.

Table 7: Posteriors and RMSEs for yield strength prediction

# Case	Post. L	Post. Q	Post. GP	LRmse	QRmse	GPRmse	WeightedRmse
1	0.6842218	0.2789746	0.03680361	3.175349	3.648647	4.539219	3.293996
2	0.1740768	0.4263896	0.3995336	2.142408	1.168258	1.264526	1.116928
3	0.3930287	0.1305902	0.4763811	3.731313	4.229568	3.637346	3.339771

General Case Study

In General Case Study, we did 200 iterations for each group of validation and counted the number of the largest posterior for each model. The results are shown in Table 8. The linear model did the best in Group 1, Group 2 and Group 4, but in Group 2 the quadratic model had the same performance with linear model. And the GP model always did the worst. It is safe to say that the linear model is the best among these three models while it is not the dominant one. Now, it seems more valuable to check the performance of the weighted model. We will do this part of work in the next quarter.

Table 8: General case results for yield strength prediction

	Group 1	Group 2	Group 3	Group 4
num_Linear	104	84	96	105
num_Quadratic	69	84	69	66
num_GP	27	32	35	29

Preliminary results of using manifold learning

Manifold learning-based Gaussian Process regression model can be separated into two procedures, UMAP embedding and GP regression. For this stage, all features of pipeline data were used for dimension reduction with respect to yield strength measured by tensile testing. Then GP regression is used to fit the embedded 2D coordinate and the yield strength measure by tensile testing. The data used for this model include all provided 53 points. The low dimensional representation of pipeline data is shown in Figure 12. The color of points indicates different yield strength. It is easy to conclude that low dimensional representation can capture the overall trend of pipeline data. To evaluate the performance of the proposed model, five folds cross-validation were used, where 10 data points were used for testing and rest of 43 data points were used for training each time. Table 9 shows the root mean square error of testing prediction for each run. The average RMSE of 5 folds cross-validation is 4.74.

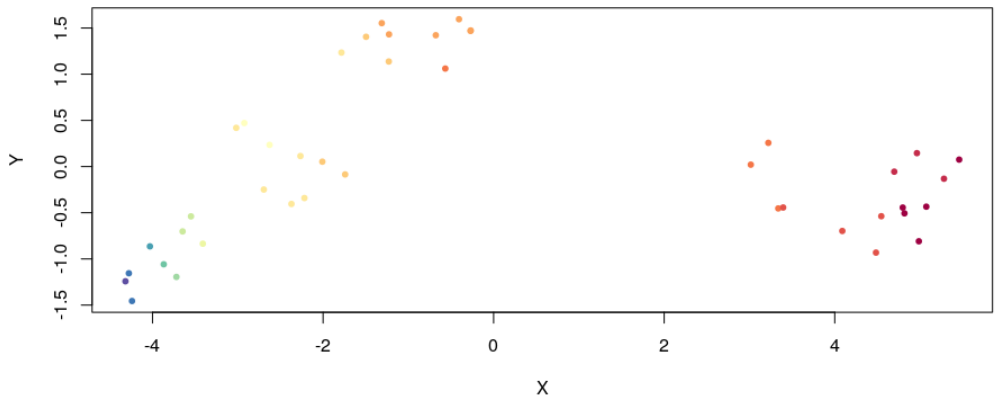


Figure 12: Pipeline embedding

Table 9: RMSEs of testing prediction

	Run 1	Run 2	Run 3	Run 4	Run 5
RMSE	6.89	4.60	3.37	4.32	4.52

The RMSE shows the potential ability to handle to high dimension regression problem by using the proposed method. For this preliminary model, the RMSE is still so high to apply on engineering problems. This may result on the non-optimize parameters setting for UMAP as well as regression model.

The future work will focus on the fine-tuning of UMAP parameters and estimate the yield strength difference between strength measured by tensile testing and strength measured by surface indentation technique.

Conclusion

Different datasets may lead to different results even dealing with the same problem when applying the BMA method. A new rank list of useful variables obtained by using BMA with new 53-point dataset. Following that, the best linear model and the best quadratic have been achieved. And the best GP model consists top 3 variables in the list. Toy problem shown basic procedures of doing model selection, updating and averaging. The results indicated that the justification of using a particular model is case-dependent and the weighted model has potential to achieve a better performance. Both specific case and general case are analyzed to evaluate the performance of this new method in terms of doing yield strength prediction. It is safe to say that the linear model is the best among these three models while it is not the dominant one. Besides, the preliminary results of using manifold learning shown the potential ability to handle to high dimension regression problem.

References

[1] X. Guan, R. Jha and Y. Liu, "Model selection, updating, and averaging for probabilistic fatigue damage prognosis," *Structural Safety,* vol. 33, pp. 242-249, 2011.

[2] P. Gregory, Bayesian logical data analysis for the physical sciences, 2005.

[3] S. Jeffreys, Theory of probability, Oxford University Press, 1998.

[4] M. Clyde, M. Cetinkaya-Rundel and C. Rundel, An Introduction to Bayesian Thinking, 2020.

[5] H. Bijl, "Gaussian Process Regression Techniques," August 2016. [Online]. Available: http://www.hildobijl.com/Research.php.

[6] C. E. Rasmssen and C. K. I. Williams, Gaussian Process for Machine Learning, London: The MIT Press, 2006.

[7] T. Wu and J. Movellan, "Semi-parametric Gaussian Process for Robot System Identification," in *2012 IEEE/RSJ Internetional Conference on Intelligent Robots and Systems*, Vilamoura, 2012.

[8] K. P. Murphy, Machine Learning, London: The MIT Press, 2012.

End of Report