022-2227

Draft



Pacific Gas and Electric Company

Grade Prediction Report: Syar Industries Meter & Regulator Station

Joel Anderson

Principal Engineer

October 24, 2023

Page Intentionally Left Blank

Draft Report

for

Pacific Gas and Electric Company

**Grade Prediction Report: Syar Industries Meter & Regulator Station**

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Issue date | Lead author | Reviewed by |
| Draft 01 | 10/24/23 | Joel Anderson | Adam Steiner |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**RSI Pipeline Solutions, LLC**

5195 Hampsted Village Center Way

Unit 172

New Albany, OH 43054

Disclaimer

The report provides the results of consultancy services performed by employees of RSI Pipeline Solutions, LLC. The findings of this report were performed using information provided by Client and standards of practice that is commonly accepted within the industry.

This report (including any attachments) has been prepared for the exclusive use and benefit of the Client. Unless RSI Pipeline Solutions provide express prior written consent, no part of this report should be used for the benefit of any party other than the contracting party to RSI Pipeline Solutions. We do not accept any liability if this report is used for an alternative purpose from which it is intended, nor to any third party in respect of this report. Additionally, any assumptions, facts, or information not described in this report may affect the analysis and conclusions presented in this report. No warranties, either expressed or implied, are intended or made.

Table of Contents

[Executive Summary 1](#_Toc149039251)

[1 Introduction 2](#_Toc149039252)

[1.1 Number of Cases 3](#_Toc149039253)

[2 Model Inputs 4](#_Toc149039254)

[3 Model Results 6](#_Toc149039255)

[3.1 Mean Probabilities 6](#_Toc149039256)

[3.2 Percentage of Cases 8](#_Toc149039257)

[3.3 Boxplots of Probabilities 9](#_Toc149039258)

[4 Conclusion 10](#_Toc149039259)

[5 References 11](#_Toc149039260)

List of Tables

[Table 1: Grade Prediction Results Summary 1](#_Toc149039268)

[Table 2: Number of Cases in Confusion Matrix Validation Set 3](#_Toc149039269)

[Table 3: Joint Characteristics [3] 4](#_Toc149039270)

[Table 4: Joint Material Properties [4] 5](#_Toc149039271)

[Table 5: Mean Probabilities 7](#_Toc149039272)

[Table 6: Percentage of Cases 8](#_Toc149039273)

[Table 7: Grade Prediction Results Summary 11](#_Toc149039274)

List of Figures

[Figure 1: Confusion Matrix 3](#_Toc149039284)

[Figure 2: Mean Probabilities 7](#_Toc149039285)

[Figure 3: Percentage of Cases 8](#_Toc149039286)

[Figure 4: Boxplot of Probabilities 10](#_Toc149039287)

Page Intentionally Left Blank

**Grade Prediction Report: Syar Industries Meter & Regulator Station**

# Executive Summary

The 7 pipe joints from Syar Industries Meter & Regulator Station (Syar Ind Lake Herm Station) had chemical composition testing completed using non-destructive examination (NDE) and the information was supplied to RSI Pipeline Solutions (RSI) to predict the most probable grade. The chemical composition data was used in a machine learning model that was trained on an aggregated composition dataset from various third parties as well as the Pacific Gas and Electric Company (PG&E) to predict grade.

All the information and code that was used to generate this report are backed up in a repository using Git which is an open-source version control system. Any changes are “pushed” to the repository which tracks what changes were made to each file and when. This allows the tracking of changes and if at any time if there is a need to go back to a previous version the whole repository can be reverted to any saved version. This eliminates ad hoc methods such as saving different versions of a file under different file names or folders each time there is a change that the developer/writer wants to save as a version. The name of the repository for this report is documented in the references [1].

Several methods of NDE chemical composition testing were used including filings collection followed by laboratory analysis (Filings) and laser-induced breakdown spectroscopy (LIBS) combined with X-ray fluorescence (XRF). For joints that had filings data available, the filings data were used for grade predictions, otherwise the LIBS/XRF data were used.

Table : Grade Prediction Results Summary

| **Joint** | **Highest Predicted Mean Probability** | **Highest Percentage of Cases** |
| --- | --- | --- |
| **JT-02** | X42 | X42 |
| **JT-04** | X42 | X42 |
| **JT-06** | X42 | X42 |
| **JT-07** | X42 | X42 |
| **JT-10** | B | B |
| **JT-100** | B | B |
| **JT-11** | X42 | X42 |

PG&E will make the final determination of the pipe grade to use in downstream analyses. Any disagreement between the highest predicted mean probability and highest percentage of cases will be resolved during PG&E’s grade determination process.

# Introduction

Although yield strength (YS) can be measured through in-situ methods like Instrumented Indentation Testing (IIT) with a reasonable degree of confidence, YS by itself does not address the question of pipe grade (and hence, SMYS) since a single YS can span multiple grades based on the allowable ranges of YS per API 5L. To overcome this challenge, a supervised classification machine learning (ML) model was developed to predict pipe grade using NDE chemical composition measurements as inputs.[2] This ML model is referred to in this report as a pipe grade calculator. While using the ML-based model provides substantial improvement in predicting pipe grade, measurement uncertainty from both composition NDE tools must be considered per 192.607(d)(2). This model uses the wt% of manganese, carbon, and sulfur as well as diameter and wall thickness to predict grade using a Random Forest ML model using a dataset of approximately 10,000 joints with known grade and composition. The data sources came from PG&E’s ECA and TIMP data as well as third party mill inspection reports. The data was split for training and validation at an 85%/15% ratio respectively. The overall accuracy is 88%. The confusion matrix is shown in [Figure 1](#fig-confuse). The total number of cases used to validate the pipe grade calculator are summarized inTable 2.

The following inputs are used in the ML model.

1. Carbon content (wt%)
2. Manganese content (wt%)
3. Sulfur content (wt%)
4. Outside diameter (OD)
5. Nominal wall thickness (NWT)
6. OD to NWT ratio (OD/NWT)

|  |
| --- |
| A table with numbers and a number in it  Description automatically generated  Figure : Confusion Matrix |

## Number of Cases

Table : Number of Cases in Confusion Matrix Validation Set

| **Grade** | **Total Number of Cases** |
| --- | --- |
| **LTB** | 42 |
| **B** | 369 |
| **X42** | 130 |
| **X46** | 75 |
| **X52** | 413 |
| **X60** | 378 |
| **X65** | 45 |
| **X70** | 7 |
| **Total** | 1,459 |

# Model Inputs

A measurement is the process of assigning a value to a physical quantity. Regardless of the precision, measurements will have some variation between the true (but unknown) value and the reported value. To account for this uncertainty of the chemical composition measurements acquired through NDE, a large set of input cases were generated using the uncertainty previously quantified by a third-party vendor and analyzed with the pipe grade calculator using a Monte Carlo analysis.[[1]](#footnote-2) Specifically, these input cases were generated by treating each measurement quantity as a stochastic variable and randomly sampling each quantity’s uncertainty distribution, which was assumed to be a t-distribution with corresponding mean, standard deviation, and number of samples[[2]](#footnote-3). The OD, NWT, and OD/NWT were treated as fixed values and stochastic samples (hereto after called cases) were only generated for chemical composition elements. The chemical composition dataset consisted of Mn, C, and S elements. A total of 5,000 input cases were generated for each joint and had corresponding grade probabilities generated. The wall thickness and diameter information is shown inTable 3.

Table : Joint Characteristics [3]

| **Joint** | **OD (in.)** | **NWT (in.)** |
| --- | --- | --- |
| **JT-02** | 4.500 | 0.237 |
| **JT-04** | 4.500 | 0.237 |
| **JT-06** | 8.625 | 0.312 |
| **JT-07** | 8.625 | 0.312 |
| **JT-10** | 6.625 | 0.280 |
| **JT-100** | 4.500 | 0.237 |
| **JT-11** | 6.625 | 0.156 |

The values in Table 4 are a summary of the information used to generate the Monte Carlo input cases.

Table : Joint Material Properties [4]

| **Joint** | **Chemistry** | **Input** | **Mean** | **Stdev** |
| --- | --- | --- | --- | --- |
| **JT-02** | Filings | C | 0.244 | 0.024 |
| Mn | 0.894 | 0.050 |
| S | 0.009 | 0.004 |
| **JT-04** | C | 0.205 | 0.024 |
| Mn | 1.029 | 0.050 |
| S | 0.009 | 0.004 |
| **JT-06** | C | 0.182 | 0.024 |
| Mn | 0.642 | 0.050 |
| S | 0.014 | 0.004 |
| **JT-07** | C | 0.188 | 0.024 |
| Mn | 0.642 | 0.050 |
| S | 0.018 | 0.004 |
| **JT-10** | C | 0.083 | 0.024 |
| Mn | 0.323 | 0.051 |
| S | 0.015 | 0.004 |
| **JT-100** | C | 0.256 | 0.024 |
| Mn | 0.845 | 0.050 |
| S | 0.011 | 0.004 |
| **JT-11** | C | 0.247 | 0.024 |
| Mn | 0.7 | 0.050 |
| S | 0.022 | 0.004 |

# Model Results

## Mean Probabilities

The machine learning model produces a probability for each grade based on each of the 5,000 cases supplied to it. After running all the cases, the probability for each grade was averaged for each joint. The most probable grade is the one with the highest predicted probability. [Figure 2](#fig-meanpreds) shows the mean probability for each grade. The 95% confidence intervals of the mean were not shown on the plots due to their small size.[[3]](#footnote-4)

|  |
| --- |
| A graph of different colored bars  Description automatically generated  Figure 2: Mean Probabilities |

Using the composition data, the highest probable grade for all joints as shown in Table 5.

Table : Mean Probabilities

| **Joint** | **LTB** | **B** | **X42** | **X46** | **X52** | **X60** | **X65** | **X70** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **JT-02** | 0.005 | 0.351 | 0.478 | 0.094 | 0.042 | 0.029 | 0.000 | 0.000 |
| **JT-04** | 0.009 | 0.183 | 0.605 | 0.027 | 0.128 | 0.048 | 0.000 | 0.000 |
| **JT-06** | 0.145 | 0.314 | 0.493 | 0.017 | 0.024 | 0.006 | 0.000 | 0.000 |
| **JT-07** | 0.163 | 0.291 | 0.515 | 0.011 | 0.017 | 0.003 | 0.000 | 0.000 |
| **JT-10** | 0.273 | 0.652 | 0.069 | 0.001 | 0.005 | 0.000 | 0.000 | 0.000 |
| **JT-100** | 0.006 | 0.412 | 0.399 | 0.145 | 0.030 | 0.009 | 0.000 | 0.000 |
| **JT-11** | 0.001 | 0.261 | 0.485 | 0.189 | 0.056 | 0.008 | 0.000 | 0.000 |

## Percentage of Cases

For each case analyzed by the model, a probability is produced for each grade and the one with the highest probability is the chosen grade for that case.Table 6and [Figure 3](#fig-maxgrade) display the percentage of cases that each grade was selected as the most probable.

|  |
| --- |
| A graph of different colored bars  Description automatically generated with medium confidence  Figure 3: Percentage of Cases |

Table : Percentage of Cases

| **Joint** | **LTB (%)** | **B (%)** | **X42 (%)** | **X46 (%)** | **X52 (%)** | **X60 (%)** | **X65 (%)** | **X70 (%)** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **JT-02** | 0.00 | 26.72 | 73.16 | 0.1 | 0.00 | 0.00 | 0.00 | 0.00 |
| **JT-04** | 0.00 | 0.36 | 99.64 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **JT-06** | 2.2 | 15.84 | 81.96 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **JT-07** | 3.1 | 11.98 | 84.92 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **JT-10** | 1.1 | 98.90 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **JT-100** | 0.00 | 56.76 | 42.72 | 0.48 | 0.00 | 0.00 | 0.00 | 0.00 |
| **JT-11** | 0.00 | 32.00 | 62.72 | 5.28 | 0.00 | 0.00 | 0.00 | 0.00 |

## 

## Boxplots of Probabilities

Boxplots are a visual aid that demonstrates not just the range of values but where most of the data is concentrated. Boxplots consist of four parts. The “box” portion is called the inner quartile range (IQR), which starts at the end of the first through the third quartile of the data where most of the data resides. The line through the middle of the box is the median value. The lines extending out from the box continue to the last data point that is less than or equal to 1.5 times the IQR from the box. Anything that is further than 1.5 times the IQR is considered an outlier and each of the data points that meet that criteria are plotted as a single point. As shown in [Figure 4](#fig-boxplots), all the grade predictions were strongly Grade B or X42. No other grades had significant probabilities relative to those two.

|  |
| --- |
| A graph of different colored boxes  Description automatically generated with medium confidence  Figure 4: Boxplot of Probabilities |

# Conclusion

For the 7 joints analyzed for Syar Ind Lake Herm Station, the predicted grade for all joints is Grade X42 except for JT-10 and JT-100 which were Grade B based on mean probability and percentage of cases. There were no differences in predictions between the mean probability and percentage of cases. JT-06, 07, and 10 had significant probabilities of LTB, especially JT-10. Joints JT-06 and JT-07 are consistent with each other in composition while JT-10 is dissimilar, noticeably in C and Mn which were quite low. The composition for JT-06 and JT-07 were consistent with historical trends for X42, it is possible the higher-than-expected LTB probability was because of the low diameter to wall thickness ratio. The higher-than-expected LTB for JT-10 is probably due to the very low C and Mn values in addition to the low diameter to wall thickness ratio.

Table : Grade Prediction Results Summary

| **Joint** | **Highest Predicted Mean Probability** | **Highest Percentage of Cases** |
| --- | --- | --- |
| **JT-02** | X42 | X42 |
| **JT-04** | X42 | X42 |
| **JT-06** | X42 | X42 |
| **JT-07** | X42 | X42 |
| **JT-10** | B | B |
| **JT-100** | B | B |
| **JT-11** | X42 | X42 |

PG&E will make the final determination of the pipe grade to use in downstream analyses. Any disagreement between the highest predicted mean probability and highest percentage of cases will be resolved during PG&E’s grade determination process.

# References

1. Version Control: syar\_ind.git
2. RSI Pipeline Solutions, “Pipe Grade Calculator Development Report”, June 26, 2023
3. Pacific Gas & Electric Co. (PG&E) Facility Integrity Management Program (FIMP) Risk “Pipe Nominal Wall Thickness Report: Syar Industries Meter & Regulator Station”, XXXXXXX XX, 2023
4. Exponent, Inc., “Syar Industries Meter & Regulator Station Chemical Composition Results and Analysis Exponent Project No. 2303657.000”

1. A Monte Carlo analysis consists of running numerous simulations by repeatedly sampling a model’s inputs based on each input variable’s probability distribution. In this way, the resulting uncertainty of a deterministic model’s output can be explored based on the likely distribution of the model’s input. [↑](#footnote-ref-2)
2. The resulting sampled input variables were constrained to ensure that no spuriously generated values were used. Specifically, the following constraints were utilized when sampling the input quantities: 0.02 C 0.4, 0.23 Mn 1.8, 0 S 0.1, and 1 Mn/C 70. Reported values of S less than the NDE instrument’s limit of detection (LOD) were sampled assuming a uniform distribution between zero and the LOD. [↑](#footnote-ref-3)
3. Since the confidence interval shrinks in proportion to the square root of the number samples, the length of the confidence interval for 5,000 samples is imperceptible at this scale. [↑](#footnote-ref-4)