Grade Prediction Report: Gosferd Meter & Regulator Station

3/3/23

Table of contents

[Executive Summary 1](#_Toc128987054)

[1. Introduction 2](#_Toc128987055)

[Number of Cases 3](#_Toc128987056)

[2. Model Inputs 4](#_Toc128987057)

[3. Model Results 5](#_Toc128987058)

[3.1 Mean Probabilities 5](#_Toc128987059)

[Tabulated Results 6](#_Toc128987060)

[3.2 Percentage of Cases 7](#_Toc128987061)

[3.3 Boxplots of Probabilities 8](#_Toc128987062)

[4. Conclusion 9](#_Toc128987063)

[5. References 9](#_Toc128987064)

# Executive Summary

The 5 pipe joints from Gosferd Meter & Regulator Station had chemical composition testing completed using non-destructive examination (NDE) and the information was supplied to RSI Pipeline Solutions 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.

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. The predicted grade for all the joints is Grade X42 based on mean probability or percentage of cases.

**Table** **1:** Grade Prediction Results Summary

| **Joint** | **Highest Predicted Mean Probability** | **Highest Percentage of Cases** |
| --- | --- | --- |
| JT-01 | X42 | X42 |
| JT-02 | X42 | X42 |
| JT-05 | X42 | X42 |
| JT-07 | X42 | X42 |
| JT-08 | 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.

# 1. 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. 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 a 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 in **?@tbl-casecount**.

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)

|  |
| --- |
| Figure 1: Confusion Matrix |

## Number of Cases

**Table** **2:** Number of Cases in Confusion Matrix Validation Set

| **Grade** | **Total Number of Cases** |
| --- | --- |
| LTB | 45 |
| B | 116 |
| X42 | 107 |
| X46 | 45 |
| X52 | 338 |
| X60 | 321 |
| X65 | 12 |
| X70 | 7 |
| Total | 991 |

# 2. 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-1) 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-2). 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 in **?@tbl-data**.

**Table** **3:** Joint Characteristics [1]

| **Joint** | **OD (in.)** | **NWT (in.)** |
| --- | --- | --- |
| JT-01 | 4.5 | 0.250 |
| JT-02 | 4.5 | 0.237 |
| JT-05 | 4.5 | 0.237 |
| JT-07 | 4.5 | 0.237 |
| JT-08 | 4.5 | 0.237 |

The values in **?@tbl-samples** are a summary of the information used to generate the Monte Carlo input cases.

**Table** **4:** Joint Material Properties [2]

| **Joint** | **NDE Method** | **Input** | **Mean (wt%)** | **Std. Deviation (wt%)** |
| --- | --- | --- | --- | --- |
| JT-01 | Filings | C | 0.207 | 0.024 |
| Mn | 1.087 | 0.050 |
| S | 0.013 | 0.004 |
| JT-02 | Filings | C | 0.205 | 0.024 |
| Mn | 1.087 | 0.050 |
| S | 0.013 | 0.004 |
| JT-05 | Filings | C | 0.207 | 0.024 |
| Mn | 1.106 | 0.050 |
| S | 0.014 | 0.004 |
| JT-07 | Filings | C | 0.205 | 0.024 |
| Mn | 1.087 | 0.050 |
| S | 0.014 | 0.004 |
| JT-08 | Filings | C | 0.211 | 0.024 |
| Mn | 1.087 | 0.050 |
| S | 0.013 | 0.004 |

# 3. Model Results

## 3.1 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-3) Grade X42 has the highest mean probability, for all joints.

|  |
| --- |
| Figure 2: Mean Probabilities |

## Tabulated Results

Using the compositional data, Grade X42 was the highest predicted probability for the joint as shown in **?@tbl-predstable**.

**Table** **5:** Mean Probabilities

| **Joint** | **LTB** | **B** | **X42** | **X46** | **X52** | **X60** | **X65** | **X70** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| JT-01 | 0.001 | 0.099 | 0.884 | 0.006 | 0.007 | 0.003 | 0.000 | 0.000 |
| JT-02 | 0.001 | 0.068 | 0.910 | 0.010 | 0.007 | 0.004 | 0.000 | 0.000 |
| JT-05 | 0.001 | 0.056 | 0.923 | 0.010 | 0.006 | 0.004 | 0.000 | 0.000 |
| JT-07 | 0.001 | 0.051 | 0.927 | 0.011 | 0.007 | 0.003 | 0.000 | 0.000 |
| JT-08 | 0.001 | 0.067 | 0.909 | 0.010 | 0.009 | 0.005 | 0.000 | 0.000 |

## 3.2 Percentage of Cases

For each case analyzed by the model, it producers a probability is produced for each grade and the one with the highest probability is the chosen grade for that case. **?@tbl-pctclass** and [Figure 3](#fig-maxgrade) display the percentage of cases that each grade was selected as the most probable. Grade X42 was the highest percentage of cases for for all features.

|  |
| --- |
| Figure 3: Percentage of Cases Grade was Selected |

**Table** **6:** Percentage of Cases

| **Joint** | **LTB (%)** | **B (%)** | **X42 (%)** | **X46 (%)** | **X52 (%)** | **X60 (%)** | **X65 (%)** | **X70 (%)** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| JT-01 | 0.00 | 4.64 | 95.36 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| JT-02 | 0.00 | 1.88 | 98.12 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| JT-05 | 0.00 | 1.68 | 98.32 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| JT-07 | 0.00 | 1.22 | 98.78 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| JT-08 | 0.00 | 1.74 | 98.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

## 3.3 Boxplots of Probabilities

Boxplots are a visual aid that demonstrates not just the range of values but where the majority 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 the 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 of the grade predictions were overlapping between Grade B and X42 no other grades had significant probabilities relative to those two.

|  |
| --- |
| Figure 4: Boxplot of Probabilities |

# 4. Conclusion

For the 5 joints analyzed at the Gosferd Station, the predicted grade for all the joints is Grade X42 based on mean probability or percentage of cases.

**Table** **7:** Grade Prediction Results Summary

| **Joint** | **Highest Predicted Mean Probability** | **Highest Percentage of Cases** |
| --- | --- | --- |
| JT-01 | X42 | X42 |
| JT-02 | X42 | X42 |
| JT-05 | X42 | X42 |
| JT-07 | X42 | X42 |
| JT-08 | 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.

# 5. References

1. Pacific Gas & Electric Co. (PG&E) Facility Integrity Management Program (FIMP) Risk, “Pipe Nominal Wall Thickness Report: Gosferd Rd Meter and Regulator Station,” November 7, 2022.
2. Exponent, Inc., “Gosferd Chemical Composition Results and Analysis,” XXXXXX XX, 2022 (Report No. 2202872.000-XXXX).

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-1)
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-2)
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-3)