B.Tech Project (MTN - 400B)

"Uniform Corrosion Rate Prediction in Mild Steel in Citric and Acetic acid medium Using Machine Learning."

Under the aegis of

Dr. G.P. Chaudhari

Metallurgical & Materials Engineering Department

Indian Institute of Technology Roorkee

Submitted By

Abhinav Gupta 17118003

Deepansh Goyal 17118020

Contents

Abstract

Corrosion is one of the leading causes of the loss of billions of dollars for the oil and gas industry. It has caused widespread component failures resulting in significant environmental and financial downfalls. Online corrosion monitoring aims to improve sustainability and ensures the safety of the assets. Probes can only access conventional techniques like linear polarization resistance and weight loss measurements invasive as the interior of closed vessels. Several assumptions like an estimation of wall thickness loss rates are considered in these techniques, leading to inaccurate results. Machine Learning can be of immense help trying to model and predict the corrosion rate. Machine Learning(Alpha GO by Google) [1] beat the best Go(which is the most complex game) player in the world by identifying an extraordinarily complex move exceedingly early, which seemed unusual and losing as it was played but proved pivotal in finally defeating the player. Machine learning can help to model complex problems and provide efficient solution to it.

Introduction

Motivation

The durability issues of materials in natural and artificial environments are critical in designing and using structures and devices [2]. The leading cause of material loss is corrosion. It also hampers human health and contributes to pollution in the environment. Corrosion alone is the cause of billions of dollars in the oil and gas industry (In India and the US). The sustainability of assets is improved by corrosion resistance. However, corrosion is a highly nonlinear statement highly influenced by complex equations and models for predicting the corrosion rate of steel currently lacking a practical basis. Researchers still are struggling to find the best model to predict corrosion because of a lack of enough understanding of sources that affect corrosion rate. There are too many factors that make the corrosion challenging to model through a specific equation or method. Machine learning and artificial intelligence can provide some respite and could provide a solution for corrosion prediction. So, into the modeling problem, we have tried leveraging machine learning and let the machine learning do the modeling and find whether it is possible to use this approach for the existing situation. Machine learning (ML) and artificial intelligence (AI) based approaches have attracted a great deal of scientific attention and have been successfully used in nonlinear and optimization problems.

Objective

Given the data of decomposition of material through uniform corrosion, try leveraging a machine learning script which

- Models the corrosion rate of the material in the medium.
- Can predict the time it would take a certain depth to get depleted and vice-versa.
- Compare the various machine learning models for the same medium in which corrosion is studied.

Background

Ultrasonic testing approaches the corrosion monitoring technique in a more precise and direct manner. Only manual UT could be carried out in the past, and this technique, showed poor measurement repeatability because of coupling and transducer positionings. This uncertainty in measurement was corrected using permanently installed transducers which could produce micron-level precision values and could replicate the results of experiments.

The Ultrasonic testing method is the best way to determine the Wall Thickness Loss Rates. Whenever such small wall thickness loss rates are measured, differentiating actual WTL.[1]'s from the noise data becomes the utmost priority. Machine Learning offers a convenient way of evaluating the performances of ultrasonic corrosion monitoring systems.

Structure

The report consists of six chapters, each segregated for a purpose. The *Introduction* outlines the introduction by providing a gist of the corrosion problem faced in the industries, which also mentions the motivation and the objective of the work. The *Literature Review* is based on the theory work which we have used to our advantage. *Dataset preparation* presents a piece of brief information about the dataset and how the authors collected it. The *Methodology* and *Results* gives a quick code and explains the basis of our results. The *Conclusion and Future work* conclude our purpose and results of the experiment. Finally, References ends the report by giving due credit to the authors whose works proved helpful in our research in this area.

[1] wall Thickness loss

Literature Review

Machine Learning

We have used machine learning algorithms i.e., Simple Linear and Polynomial Regression (Quadratic and Cubic).

Linear Regression

Linear Regression is a statistical model with a linear approach to model the relationship between a scalar response and one or more explanatory variables. If there is only one explanatory variable, it is called simple linear regression, which the equation gives.

$$\hat{y}(w, x) = w_0 + w_1 x$$

The general terminology involves calling w_0 as the intercept and w_1 as the coefficient.

Linear regression fits a linear model with coefficients to minimize the residual sum of squares between the observed responses in the dataset and the responses predicted by the linear approximation.

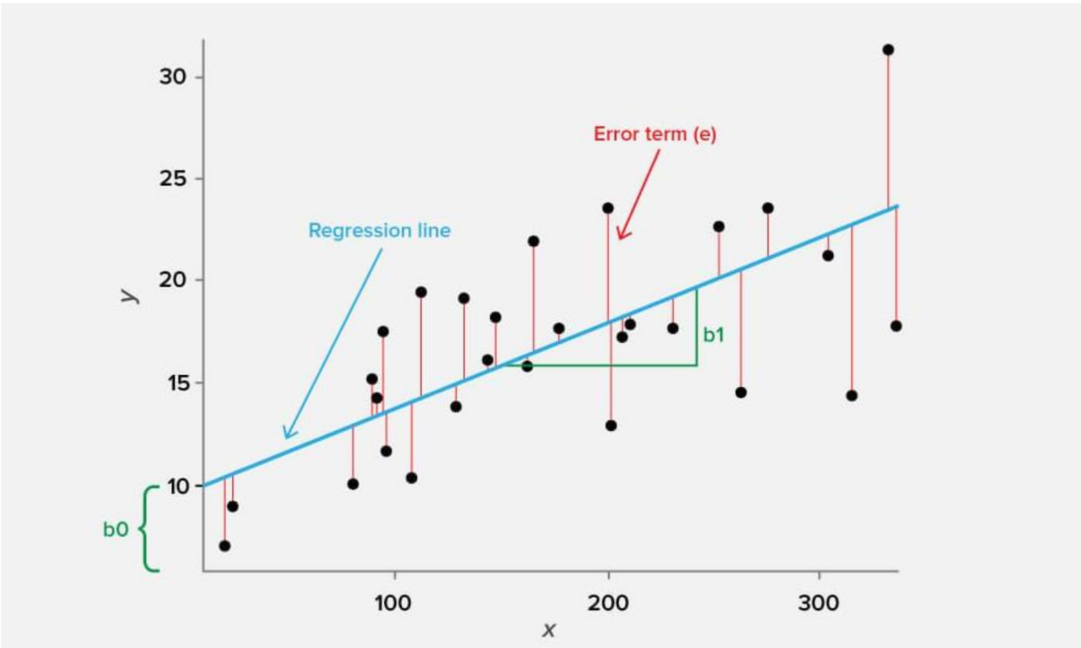


Figure 1: Visualisation of Linear Regression [3]

The aim involves fitting the line(blue) [3] such that the mean of squared distances from each point is minimum.

Quadratic Regression

The quadratic model is a second-order polynomial where we have a squared term. It is the best fit model for data points scattered across in the shape of a parabola. It is an extended version of the linear model. This model's disadvantage is that it requires several data points to make sure that the plot is a ‘U’ shape.

The equation has the form.

$$\mathrm{ax}^2 + \mathrm{bx} + \mathrm{c} = 0$$

The least squared method is used to find this equation manually. The values of a, b and c are generated so that the squared distances between each data (x, y) and the curve of the equation are minimum to get the best fit curve.

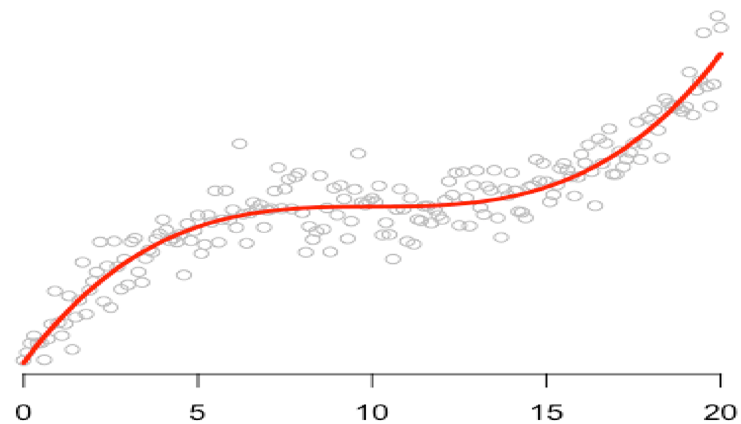


Figure 2 Visualisation of quadratic regression [4]

Cubic Regression

The cubic regression model is used for those data points whose scatter plot curves one way and then the other. The equation used for this model is of the order of three.

$$\hat{y} = ax^3 + bx^2 + cx + d$$

To summarise all three regressions, which are linear regressions only, the following image can be visualized.

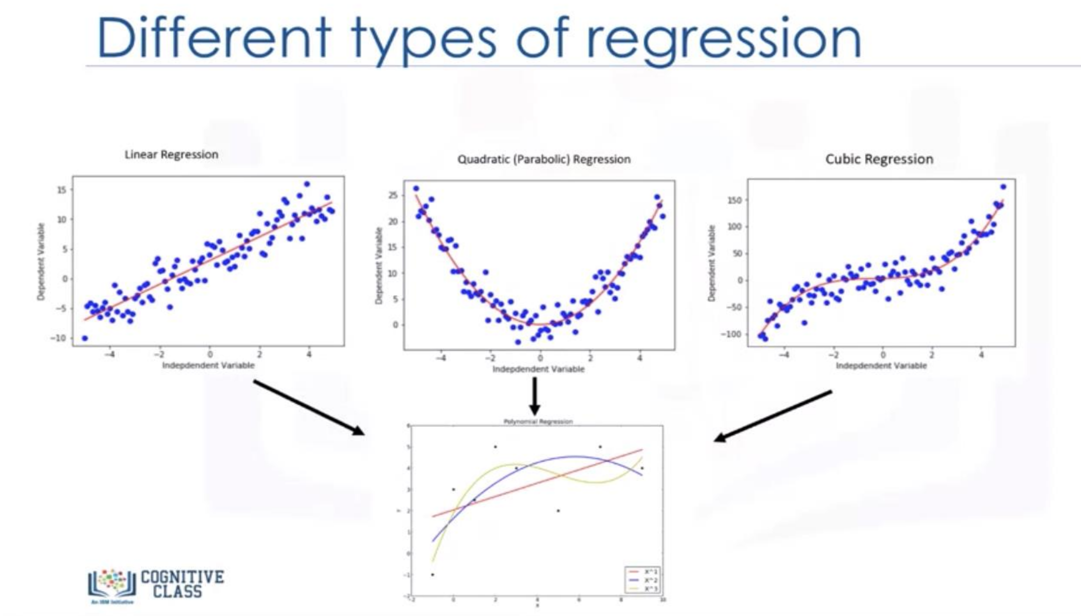


Figure 3- All three regressions are visualized in the same figure. [5]

MSE (Mean Squared Error)

Mathematically,

$$MSE = \frac{1}{n}\sum_{1}^{n}(\hat{y} - y_i)^2$$

$\hat{y}$ is the predicted output and y_i is the ith observed value in the datasets, and n is the total number of data points.

R-squared in quadratic regression

Also known as the coefficient of determination. It tells you how much variation in y is explained by the x variables. Range of values varies from 0(0% variation) to 1(1% variation).

It explains how much one variable depends on the other.

R-Squared Metric

It tells the goodness of fit of any curve to the given data points. More is the value of this metric, better the curve fits the dataset, but keeping the complexity of the model in mind, we must make sure it does not overfit.

Corrosion

Uniform and Pitting Corrosion

Uniform corrosion is the type of corrosion that happens uniformly over the full dimensions of the exposed surface and proceeds at approximately the same rate over the whole surface. In contrast, pitting corrosion is when the protective layer of the steel/metal gets damaged. Metal can act as an electron donor to accelerate the corrosion, and this region becomes susceptible to the initiation of electrochemical reaction and results in the formation of small cavities or otherwise known as *'Pits*'.

Ultrasonic WTL

Ultrasonic Wall Thickness Loss measurement uses ultrasonic waves to measure the Wall Thickness Loss or Electrochemical deposition up to a micron level of precision. It has measurement repeatability of up to 20nm using permanent transducers. This process does not involve any direct contact with the material, the electrolyte, or the vessel's interiors, and it is a non-destructive approach. It can measure the depositions or deteriorations without assuming any of the chemical reactions happening inside the vessel. UT does not consider the change in measurement happening due to loosely adhered particles on the surface, which means it can prove helpful in testing the structural toughness of the material.

Dataset Preparation

Dataset was taken from the [2] paper, which used the apparatus mentioned in the [6] in which ultrasonic testing (UT) was used to get the decomposition of the material. Ultrasonic testing does not disrupt the material and more accurate method of corrosion monitoring. The material under examination is a 10 mm **Mild Steel** in a *Uniform Corrosion* environment. The mild steel used is (**BS 970:1983:080A15, UNS G10160**). All the measurements are acquired in a time interval of 1 minute. The electrolyte used is distilled water, **.1M Acetic Acid, and .1M Citric Acid**. The machine learning analysis was performed on the data points, which are part of the analysis where they measured the decompositions of metal using ultrasounds, which were catalyzed using various kinds of acids, namely acetic and citric. Based on the observations, multiple changes in metal thickness and corresponding time were plotted on the graphs from which the points used were read and used in our Machine Learning analysis.

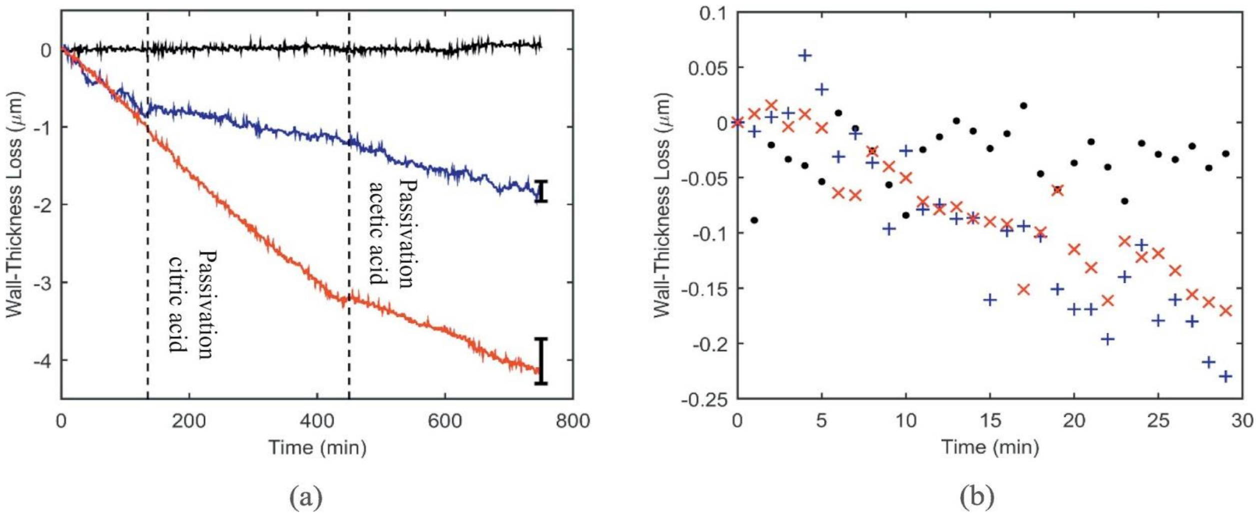


Figure 4(a) shows the loss of thickness under the influence of citric acid and distilled water as reference data. [2]

Figure 4(b) Thickness Decomposition in the acetic acid medium. [2]

We can see that no significant WTL happened in the first experiment, which occurred in distilled water. We can also see the reduction in the corrosion after a certain point in both acetic and citric acid medium, which happens due to **surface passivation**, which caused WTL to change significantly, which occurs due to the continuous deposition of corrosion products on the corroding surface. It forms a layer that inhibits corrosion.

The data points, when imported (using the help of [7]) and plotted on the graph, looked like below:

Acetic Acid

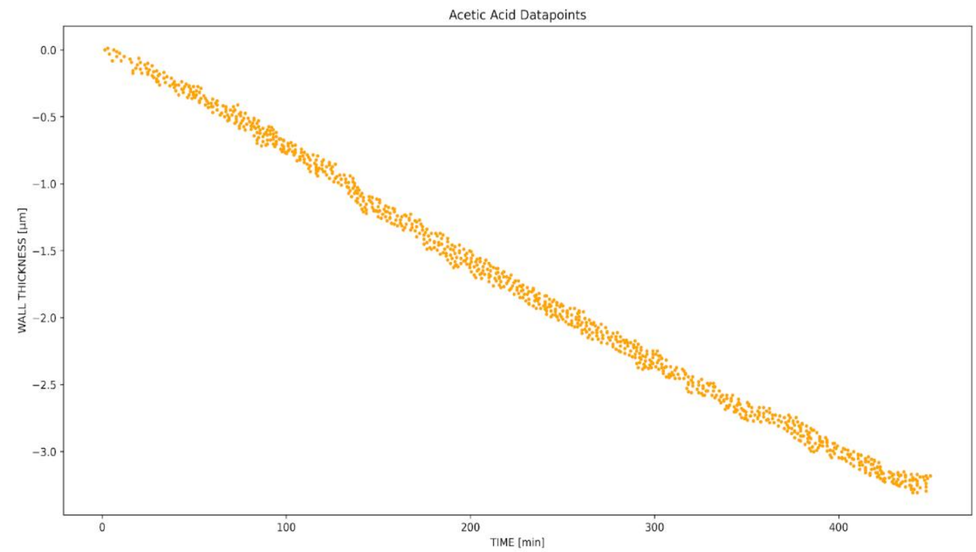


Figure 5 Data points of Acetic acid plotted using matplotlib

Citric Acid

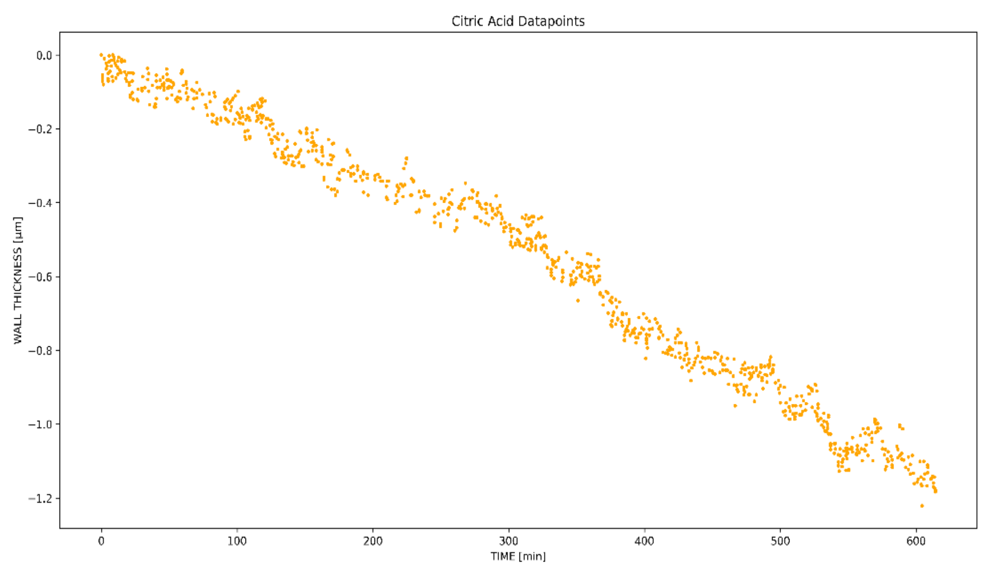


Figure 6 Datapoints of citric acid plotted using matplotlib

Methodology Followed

```
# Reading the datasets using pandas
df = pd.read_csv("Datasets_citric_acid.csv")
# Set x and y and reshape it
x_input = df["Time (min)"].values.reshape(-1,)
y_input = df["Wall – Thickness Loss (µm)"].values.reshape(-1,1)
```

The first step involved in the import the dataset, which is done using the above commands getting help from the pandas documentation. [8]

```
x_modified=modify_input(x_input,x_input.size,n)
# Create model
Y_pred=tf.add(tf.matmul(X,W),b)

#algortihm
# Calculate the loss mse
loss = tf.losses.mean_squared_error(Y, Y_pred)
#training algorithm
optimizer = tf.train.GradientDescentOptimizer(0.01).minimize(loss)
#initializing the variables
init = tf.global_variables_initializer()

#starting the session
sess = tf.Session()
sess.run(init)
# Reset the graph
tf.reset_default_graph()
# stes number of epochs
epoch= 100000
# Run the model
for step in range(epoch):
        _, c=sess.run([optimizer, loss], feed_dict={X: x_modified, Y: y_input})
        if step%1000==0 :
            print(c)
# print model paramters
print("Model paramters:" )
print(sess.run(W))
print("Bias:%f" %sess.run(b))
# Calculate the prediction value (linear regression points)
y_test=sess.run(Y_pred, feed_dict={X:x_modified})
# Calculate the R^2
r2 = r2_score(y_input, y_test)
print(r2)
# Append data to empty list
Y_test.append(y_test)
```

Figure 7 Training Loop

The above code was written using the help of TensorFlow Documentation [9] [10]. We can follow the small comments written after each line and understand what each line does.

```
plt.figure(figsize=(16,8))
plt.scatter(x_input, y_input, label='training points', color='orange',s=5)
plt.plot(x_input, y_test[0],
         label='linear (n=1), $R^2=%.5f$ $mse=%.5f$' % (R2_score[0], mse[0]),
         color='black',
         lw=2,
         linestyle=':')
plt.title('Tensorflow model-Linear Regression-citric Acid')
# set x and y labels
plt.xlabel('TIME [min]')
plt.ylabel('WALL THICKNESS [µm]')
# Add legend
plt.legend(loc='upper right')
plt.savefig('Linear_citric.png',bbox_inches='tight',dpi=600)
plt.close()
```

Figure 8 Code for plotting the desired graphs

Using the Matplotlib library [11], the above code lets us export the results and plot them on a graph. First, we initialize the plot of a particular dimension and plot all the training data on the plot. We then plan the corresponding result of the algorithm with desired color and style. We then name the x and y-axis of the plot and choose where to keep the legend, and finally save the plot giving the desired name, followed by closing the plot. To get all three combined, we can stack all three curves on the top of training points and export the plot giving all three curves assorted color and style and mentioning all of them in the legend box.

Results

The experiment was performed separately for each dataset and the corresponding findings and graphs obtained are shown below.

Acetic Acid

Linear Regression

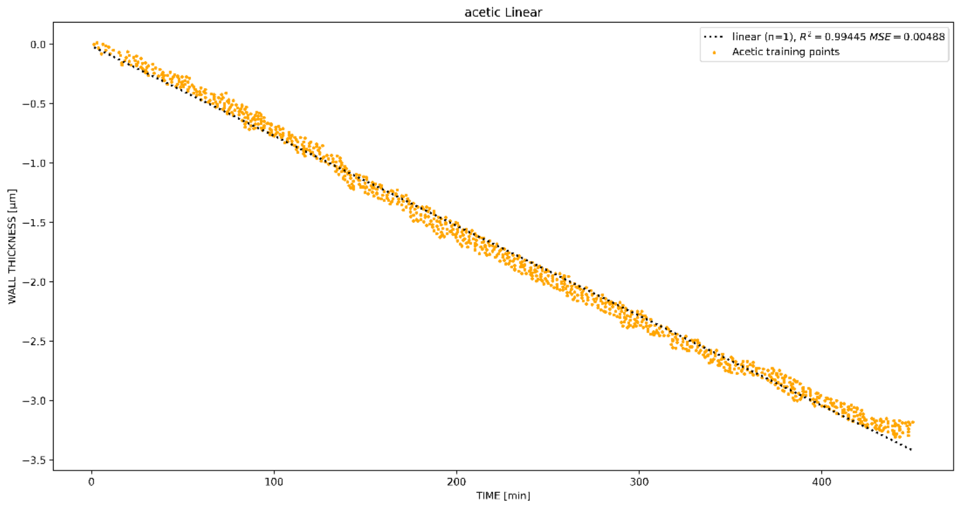


Figure 9 Linear Regression Plot for Acetic Acid

Coefficient	-0.00757141
Intercept	-0.013791589773339163

Table 1 Coefficient and Intercept of Linear regression for Acetic Acid

We fitted the simple linear regression model on the given data points, and as we can see on the graph, the R-squared matrix, which tells the goodness of fit, is .99445, and the mean squared loss came out to be .00488.

Quadratic Regression

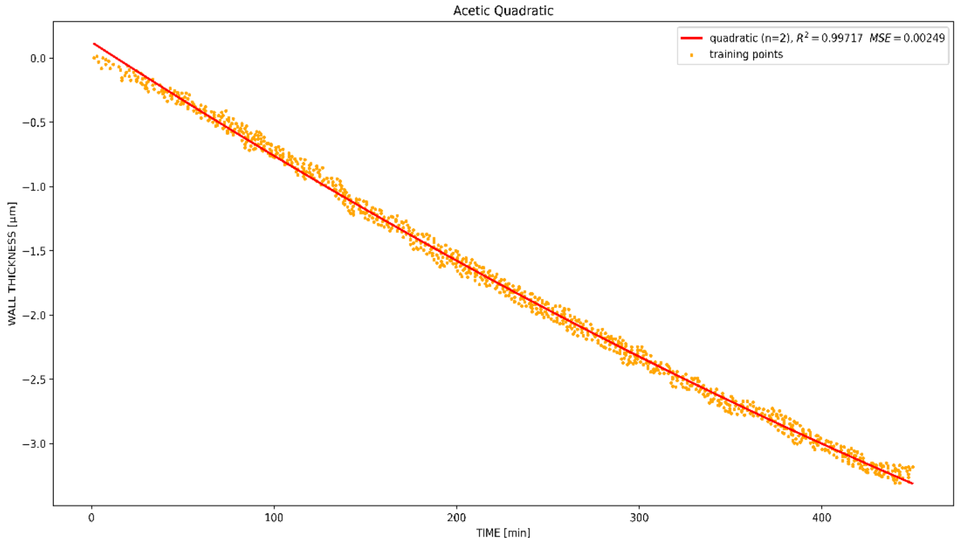


Figure 10 Quadratic Regression Plot for Acetic Acid

Coefficient	$-9.20111467e^{-3}$	$3.46923662e^{-6}$
Intercept	0.12468431591615636	

Table 2 3 Coefficients and Intercept of Quadratic Regression for Acetic Acid

Cubic Regression

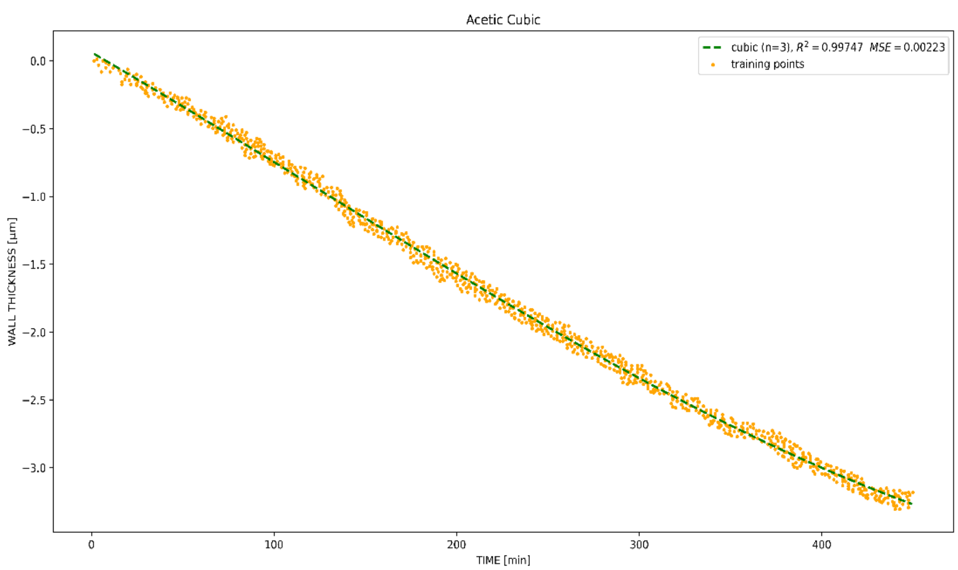


Figure 11 Cubic Regression Plot for Acetic Acid

Coefficient	$-7.82871702e^{-3}$	$-3.69292573e^{-3}$	$1.02790185e^{-8}$
Intercept	0.0631601328978888		

Table 4 Coefficients and Intercept of Cubic regression for Acetic Acid

Comparison of all three algorithms on the same plot

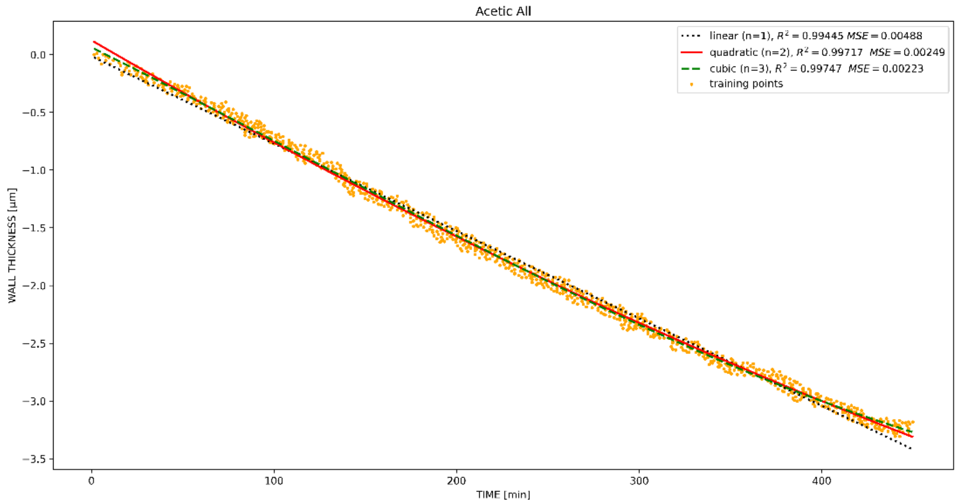
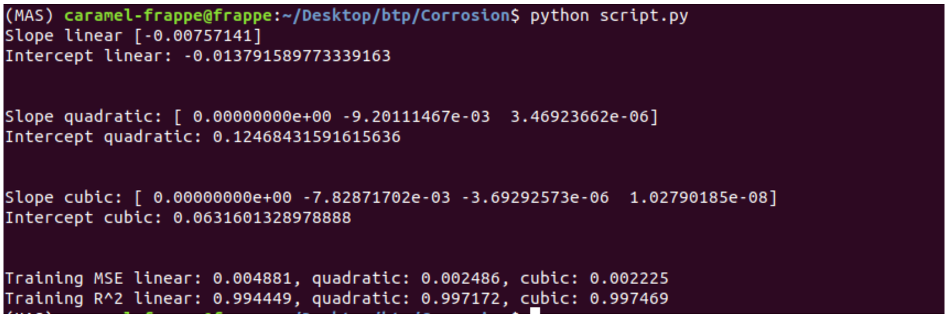


Figure 12 Comparison of all three plots

	MSE	R-squared Metric
Linear	.004881	.994449
Quadratic	.002486	.997172
Cubic	.002225	.997469

Table 5 All the evaluation metrics of all three algorithms

Following the script ran for generating results for acetic acid, we can see the values of all the intercepts and MSE and R-squared Metrics for each regression.



```
(MAS) caramel-frappe@frappe:~/Desktop/btp/Corrosion$ python script.py
Slope linear [-0.00757141]
Intercept linear: -0.013791589773339163


Slope quadratic: [ 0.00000000e+00 -9.20111467e-03  3.46923662e-06]
Intercept quadratic: 0.12468431591615636


Slope cubic: [ 0.00000000e+00 -7.82871702e-03 -3.69292573e-06  1.02790185e-08]
Intercept cubic: 0.0631601328978888


Training MSE linear: 0.004881, quadratic: 0.002486, cubic: 0.002225
Training R^2 linear: 0.994449, quadratic: 0.997172, cubic: 0.997469
```

Figure 13 A snapshot of the script which produced the outputs, and you can see corresponding coefficients and metrics.

Citric Acid

Linear Regression

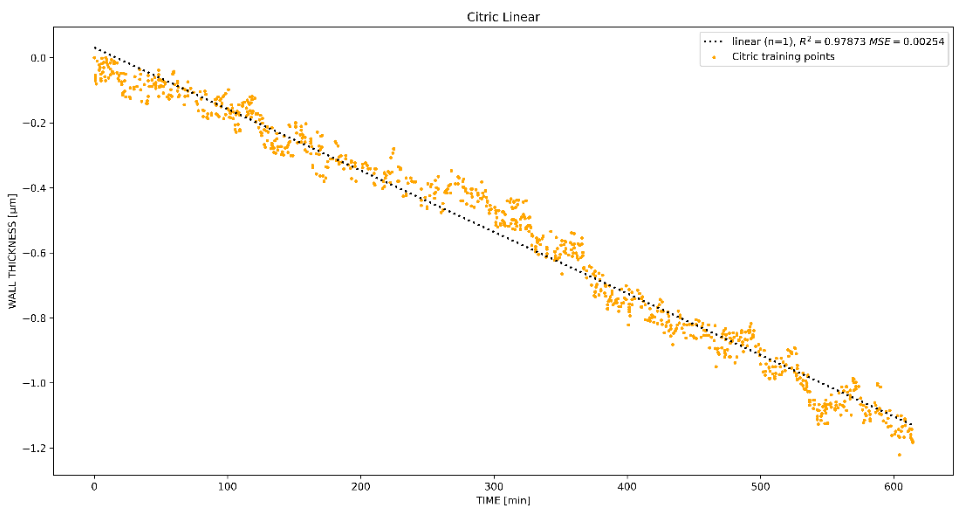


Figure 14 Linear Regression Plot for Citric Acid

Coefficient	-0.00189021
Intercept	0.0312122164095322

Table 6 Coefficient and Intercept of Linear regression for Citric Acid

Quadratic Regression

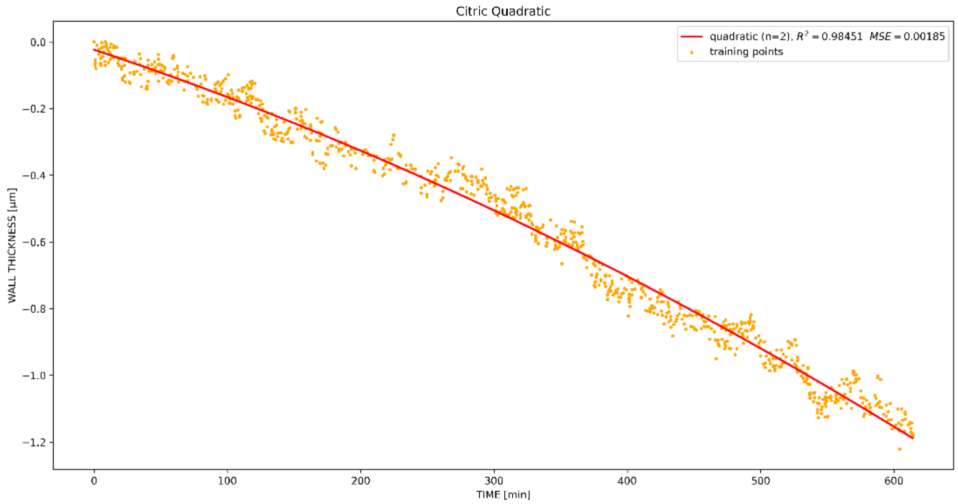


Figure 15 Quadratic Regression Plot for Citric Acid

Coefficients	$-1.32822681e^{-3}$	$-9.25867882e^{-7}$
Intercept	-0.0237842699997609	

Table 7 Coefficients and Intercept of Quadratic Regression for Citric Acid

Cubic Regression

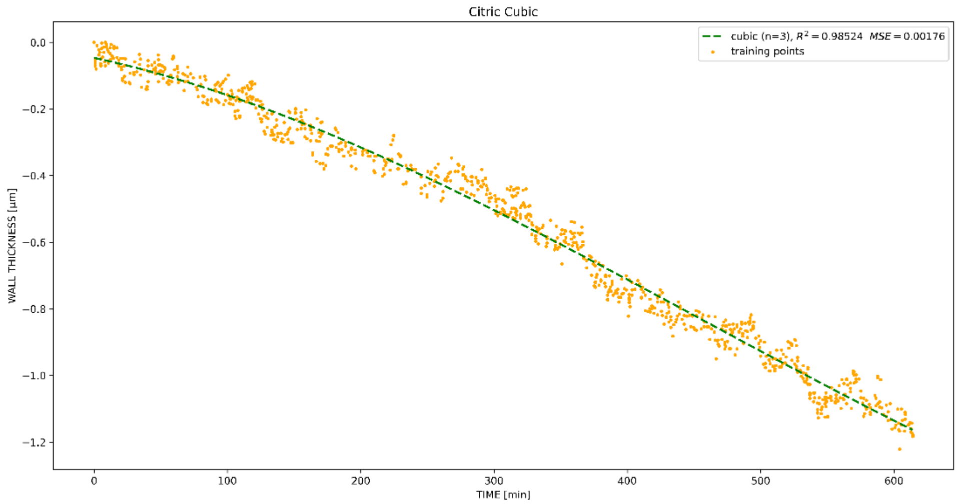


Figure 16 Cubic Regression Plot for Citric Acid

Coefficients	$-8.54837379e^{-4}$	$-2.86525255e^{-6}$	$2.10923304e^{-9}$
Intercept	-0.04679988967427973		

Table 8 Coefficients and Intercept of Cubic Regression for Citric Acid

Comparison of all three algorithms on the same plot

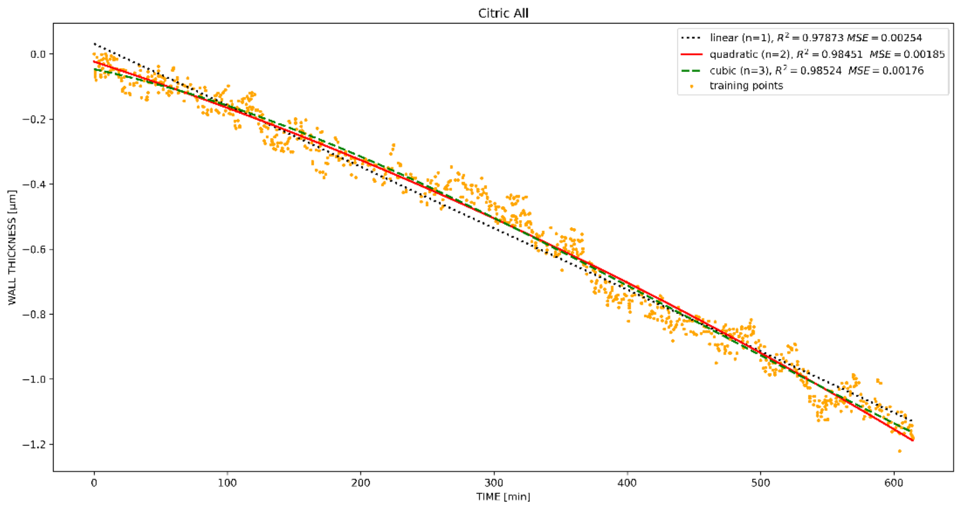


Figure 17 Comparison of three plots

	MSE	R-squared Metric
Linear	.002536	.978727
Quadratic	.001846	.984512
Cubic	.001759	.985244

Table 9: MSE and R-squared metric of all three algorithms

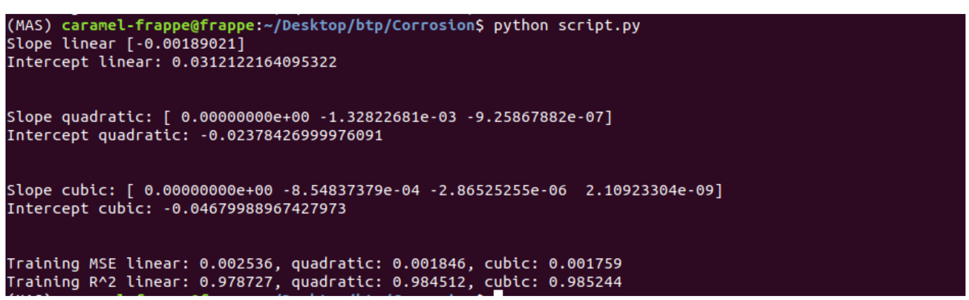


Figure 18: A snapshot of the script which produced the outputs, and you can see corresponding coefficients and metrics.

Conclusion and Future Work

Conclusion

As we started with the objective to map and find the corrosion rate, in *Citric acid* as the medium, the corrosion rate comes out to be **-.979776 mm/year,** with the negative sign suggesting the material gets eroded during corrosion. For the other medium that is *Acetic Acid*, the same metric, i.e., Corrosion rate, comes out to be **-3.92501894 mm/year,** which is like the results mentioned in the paper, which were {~1mm/year, ~4mm/year} [2]. We can say that machine learning was successful in modelling the corrosion for this specific material. However, linear regression was best at predicting the corrosion rate as other models tended to overfit on the dataset that was available, Decrease in losses as well as increase in modelling complexity led to overfitting. For the given data simple regression proved to be sufficiently accurate and increasing complexity did not affect the results.

Future work

- We can extend the work to model the corrosion in the presence of corrosion inhibitors such as Methionine, or other green inhibitors, etc.
- However, as we have used simple machine learning algorithms, it remains unseen how Neural Networks perform if given this task.

References

[1] Google, “AlphaGo is the first computer program to defeat a professional human Go player, the first to defeat a Go world champion, and is arguably the strongest Go player in history.,” [Online]. Available: https://deepmind.com/research/case-studies/alphago-the-story-so-far.

[2] F. B. C. Fangxin Zou, “On quantitative corrosion rate monitoring with ultrasound,” *Journal of Electroanalytical Chemistry,* no. ISSN 1572-6657, pp. Pages 115-121, 2018.

[3] Y. Gavrilova, “Regression Analysis Overview: The Hows and The Whys,” [Online]. Available: https://serokell.io/blog/regression-analysis-overview.

[4] B. Hartmann, “How you can use linear regression models to predict quadratic, root, and polynomial functions,” [Online]. Available: https://www.freecodecamp.org/news/learn-how-to-improve-your-linear-models-8294bfa8a731/.

[5] M. Toprak, “Non Linear Regression,” 24 September 2020. [Online]. Available: https://medium.com/@toprak.mhmt/non-linear-regression-4de80afca347.

[6] F. B. C. Fangxin Zou, “High accuracy ultrasonic monitoring of electrochemical processes,,” *Electrochemistry Communications,* no. 1388-2481, pp. 134-138, 2017.

[7] G. SINGH, “Introduction to Regression Splines (with Python codes),” [Online]. Available: https://www.analyticsvidhya.com/blog/2018/03/introduction-regression-splines-python-codes/.

[8] “Documentation of Pandas,” [Online]. Available: https://pandas.pydata.org/docs/.

[9] aymericdamien, “A linear regression learning algorithm example using TensorFlow library.,” [Online]. Available: https://github.com/aymericdamien/TensorFlow-Examples/blob/master/tensorflow_v1/examples/2_BasicModels/linear_regression.py.

[10] Tensorflow team, “API Documentation,” [Online]. Available: https://www.tensorflow.org/api_docs.

[11] Matplotlib Team, “Matplotlib Documentation,” [Online]. Available: https://matplotlib.org/stable/contents.html.