Computers and Chemical Engineering 177 (2023) 108358

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Development of machine learning algorithms for predicting internal corrosion of crude oil and natural gas pipelines

Jian Fang [a], Xiao Cheng [b], Huilong Gai [a], Sidney Lin [a], Helen Lou [a,*]

[a] Dan F. Smith Department of Chemical and Biomolecular Engineering, Lamar University, 4400 MLK Blvd., PO Box 10053, Beaumont, TX 77710, USA
[b] The Village School, 2005 Gentryside Dr, Houston, TX 77494, USA

ARTICLE INFO

Keywords:
Internal corrosion rate
Machine learning models
Crude oil pipeline
Natural gas pipeline

ABSTRACT

Energy industry is losing billions of dollars each year due to corrosion. However, publicly accessible datasets on pipeline corrosion are limited, making it difficult to accurately predict and prevent corrosion. In this research, thousands of corrosion scenarios of crude oil pipelines and natural gas pipelines were simulated using OLI Studio Corrosion Analyzer, a software based on the rigorous first principles. Thousands of datasets were generated with simulation. Then two ML algorithms, Random Forest (RF) and CatBoost were developed and used to predict the internal corrosion rate (CR) based on pipeline operating conditions. The models for crude oil pipeline CR predictions reached high prediction accuracy using five important features: temperature (T), molar concentrations ($CO2$, $CCl-1$), velocity (v), and pH value (pH). Similarly, the models for natural gas pipeline CR predictions achieved high prediction accuracy using six important features: stream compositions ($nO2$, $nH2S$, $nCO2$), pressure (P), velocity (v), and temperature (T). Both the RF- and CatBoost-Corrosion Rate models demonstrated good performance in predicting the internal corrosion rates of crude oil and natural gas pipeline.

1. Introduction

Pipelines transport crude oil and natural gas from oil fields to various distribution points, and then to refineries where they are processed into fuels, chemicals, and other products. The leakage of pipelines may have many negative consequences such as environmental pollution, production loss, and even safety incidents. Historically, approximately 60% of all pipeline incidents were caused by corrosion (Baker Jr. 2008).

Corrosion has been causing the energy industry billions of dollars per year. Monitoring and controlling corrosion in pipelines are essential for pipeline integrity. Various technologies have been developed to control and prevent corrosion, such as weight loss corrosion coupons and Electrical resistance.

Weight loss corrosion coupons (Singh, 2017) are a well-established method of monitoring corrosion and have been in use for a long time. They involve exposing a metal specimen to a particular environment, then weighing it before and after exposure to determine the weight loss corrosion that has occurred. The national association of corrosion engineers (NACE) has a standard method for determining weight loss in miles per year (MPY), which is the most widely used reference. One of the advantages of weight loss corrosion coupons is their relatively low cost, making them accessible to many industries. Additionally, they provide a positive result, and samples of corrosion products or bacteria can be obtained from the surface for further analysis. Moreover, unlike some electronic methods of corrosion monitoring, coupons are not subject to instrument failure. However, one of the disadvantages of weight loss corrosion coupons is that their size is not critical, but more surface area exposed leads to more accurate readings. Additionally, while the coupons may indicate the effectiveness of an inhibitor, they may not necessarily reflect the corrosion rates of the pipe itself.

Electrical resistance (ER) (Royer and Unz, 2002; Legat, 2007; Xu et al., 2016) probes are commonly used in various industrial fields, particularly in the petrochemical industry, for corrosion monitoring. This technique measures the thickness reduction caused by corrosion and is highly sensitive to general corrosion. While ER probes may not be as effective in detecting localized corrosion types and transient events, they have proven to be an accurate and reliable method for evaluating the state of corrosion in steel and cementitious materials. One advantage of using ER sensors is that they can provide a direct measurement of metal loss and corrosion rate. Additionally, these sensors can operate in almost any environment, making them highly versatile and suitable for use in high-resistance media such as gas and crude oil. However, ER

* Corresponding author.
E-mail address: Helen.lou@lamar.edu (H. Lou).

https://doi.org/10.1016/j.compchemeng.2023.108358
Received 18 January 2023; Received in revised form 3 July 2023; Accepted 15 July 2023
Available online 21 July 2023

sensors have a major drawback in that they have a lengthy response time. The sensitivity of an ER sensor is determined by the thickness of the element, with thicker sensors having slower response times. As a result, it can be challenging to detect corrosion in the initial stages or in mild corrosive environments using ER sensors.

In the field of corrosion rate prediction, various statistical methods are employed to develop prediction models based on actual corrosion rate data. Examples of such methods include linear regression (LR) (Al-Fakih et al., 2016), multivariate regression (Velázquez et al., 2009), and time series models including autoregressive moving average (ARMA) (Du et al., 2005) and autoregressive integrated moving average (ARIMA) (Liu and Mu, 2011). However, statistical models are limited in their ability to capture complex non-linear relationships between variables. These models often assume linear relationships or require specific functional forms, which can cause them to miss important patterns and dynamics in the data. Furthermore, statistical models typically struggle to make accurate predictions on unseen or future data unless the underlying assumptions are strictly met. Due to these limitations, statistical models have a restricted range of applicability and may not generalize well across different domains and scenarios.

To overcome the drawbacks associated with statistical methods, alternative approaches, such as machine learning algorithms or computational modeling techniques, can be considered. These methods have the potential to offer better extrapolation capabilities, broader prediction ranges, and increased accuracy, even in the presence of nonlinear data patterns.

More specifically, the advancement of Big Data Analytics and Artificial Intelligence can facilitate the prediction of internal corrosion in crude oil and natural gas pipelines. Machine Learning (ML), a subset of artificial intelligence, is a data mining technique to create data-derived models. ML enables extraction of complex and often hidden correlations, insights, patterns, rules, and guidance from given datasets and encapsulate them in mathematical form (Winkler, 2020). While numerous ML algorithms have found applications in chemical industry, supervised learning has been most commonly used (Haghighatlari and Hachmann, 2019). One limitation of ML model development is the involvement of simplistic and unrealistic assumptions, which often lead to major uncertainties and prediction errors.

In the field of oil pipeline corrosion prediction, several studies have been conducted using machine learning techniques (Liu et al., 2012; Liao et al., 2012; Chern-Tong and Aziz, 2016; Hatami et al., 2016; Ossai, 2019; Peng et al., 2021; Ben Seghier et al., 2022). For example, Liu et al. (2012) proposed a support vector machine (SVM) regression model with a radial basis function (RBF) kernel to predict the inner corrosion rate of oil pipelines. They used particle swarm optimization (PSO) algorithm to optimize needed parameters in the SVM regression, reducing the dimensionality of data space and preserving important features. Their PSO-SVM model achieved higher accuracy and speed compared to the BP neural network model, using 25 individual experiments that were normalized before training and testing with 23 samples for training and three samples for testing. Chern-Tong and Aziz (2016) proposed the CMARGA model that utilized a genetic algorithm to build an optimal decision tree model. The model achieved an average accuracy of 80.20% on 15 datasets and a 96.67% accuracy on a simulated dataset of three features: partial pressure of carbon dioxide, velocity, and temperature. Ossai (2019) developed a machine learning approach using principal component analysis (PCA) and boosting machine (GBM) to estimate the corrosion defect depth growth of aging pipelines. The PCA-GBM model was shown to achieve an accuracy 3.52 to 5.32 times greater than models without PCA transformation. Peng et al. (2021) developed a new hybrid intelligent algorithm model called PCA-CPSO-SVR for predicting the internal corrosion rate of multiphase pipelines. The model combined support vector regression (SVR), principal component analysis (PCA), and chaos particle swarm optimization (CPSO) algorithms and achieved low error values of 0.083 (RMSE), 2.7% (MAE), and 0.053 (MAPE) with a running time of 17.49 s. These studies demonstrate the effectiveness of machine learning techniques in predicting pipeline corrosion and provide insight into the development of more accurate and efficient prediction models.

The selection of input features is a crucial step in constructing the ML model. In general, the data for integrity assessment of the corrosion rate in crude oil and natural gas pipelines has been acquired from three major sources: (1) Field data are obtained from inline inspection and other operational and environmental parameters, (2) Experimental data are obtained after conducting the inhouse experimental setups, for example, the full scale burst pressure test and corrosion rate prediction, and (3) Simulation data generated to predict the pipe failure or reliability assessment of pipe using simulation software, such as OLI Studio Corrosion Analyzer and Monte Carlo Simulation (Soomro et al., 2022).

All the ML models rely on the data available. However, publicly accessible datasets are limited, and some models (Liu et al., 2012; Chern-Tong and Aziz, 2016; Ossai, 2019) were lack of sufficient data. Additionally, obtaining the integrity-related parameters from the dataset can be challenging, which can lead to inaccurate modeling predictions.

OLI Studio Corrosion Analyzer allows the users to address the cause of electrolyte-based corrosion by understanding and predicting the corrosion environment. OLI Studio Corrosion Analyzer is based on rigorous first principles (electrochemistry, reaction kinetics, thermodynamics, etc.). It predicts the stability of metals, metal ions, oxides, etc. as a function of T, P, and solution composition. It draws conclusions about the ranges of immunity to corrosion, possible passivation and dissolution of metals in the presence of species that promote or inhibit corrosion (OLI Studio, 2011). Since it is difficult for the academic researchers to get field data, the corrosion rate data used in this research were generated by running OLI Studio Corrosion Analyzer.

Ensemble learning is a branch of machine learning, and it is mainly divided into Bagging and Boosting algorithms (Gomez et al., 2010; Prokhorenkova et al., 2019). Bagging algorithms, such as Random Forest (RF), aim to reduce variance and improve the generalization of the model. RF is a popular Bagging-based ensemble learning algorithm for regression or classification (Breiman, 2001; Louppe, 2014; Scornet, 2015; Chrysafis et al., 2017). It is a meta estimator that fits a number of decision trees on various sub-samples of the dataset and uses majority voting (for classification problems) or averaging (for regression problems) to improve the predictive accuracy and control over-fitting. The sub-sample size is always the same as the original input sample size, but the sub-samples are drawn randomly. The predictions from the forests are averaged using bootstrap aggregation and random feature selection. RF models have been demonstrated to be robust predictors for both small sample sizes and high dimensional data (Biau and Scornet, 2016).

Catboost, an adaptive boosting ensemble learning algorithm based on decision trees, is powerful and very efficient (Hancock and Khoshgoftaar, 2020). It is widely applied to multiple data types to solve a wide range of problems including fraud detection, recommendation items, and forecasting. Catboost can also return highly accurate results with relatively less data compared to classical algorithms, which need to learn from a massive amount of data (Luo et al., 2021).

The hybrid models combining the strength of ML model with the first principles-based model can provide powerful solutions efficiently. It is natural to develop such hybrid approaches for Asset Integrity Management. This research aims to bridge the gap between “post-corrosion” measurement and proactive corrosion management, enhance the reliability of midstream assets, and reduce potential risks. In this research, RF and CatBoost algorithms were used to train ML models to predict the internal corrosion rate using the OLI simulated data. The data-driven ML algorithm can fully utilize the available data, while the first principles-based model can verify the reasonableness and explore unknown operation range.

2. The general methodological framework

ML prediction models are developed utilizing simulated data from the OLI corrosion analyzer. The prediction systems in crude oil and natural gas pipelines are characterized by specific parameters, which are detailed in Tables 4-7, respectively. Each component within the prediction systems serves a distinct function. To address overfitting concerns, the early stopping method is employed during the prediction process. Fig. 1 presents a comprehensive methodological flowchart depicting the development of the OLI simulation for accurate corrosion rate prediction in both crude oil and natural gas pipelines.

3. Generation of crude oil and natural gas properties

3.1. Crude oil properties

Typical pipeline crude oil specifications were obtained from open literature (Trench, 2001; Xinru et al., 2006; Xu et al., 2006; Zhang et al., 2006; WANG, 2008; Knowino, 2010; Javadli and Klerk, 2012; El-Abbasy et al., 2014; Mosavat et al., 2014; PetroWiKi, 2015; Zhang and Lan, 2017; López and Lo Mónaco, 2017; Mohammed, 2017; Jechura, 2018; Pasban and Nonahal, 2018; Wang et al., 2019; Grabner Instruments, 2019; bp, 2022). Tables 1 lists typical specifications of pipeline crude oil and Table 2 lists the specifications of Alaska North Slope crude oil. The simulation of the crude oil compositions was generated based on the variation of compositions within the range of each specification.

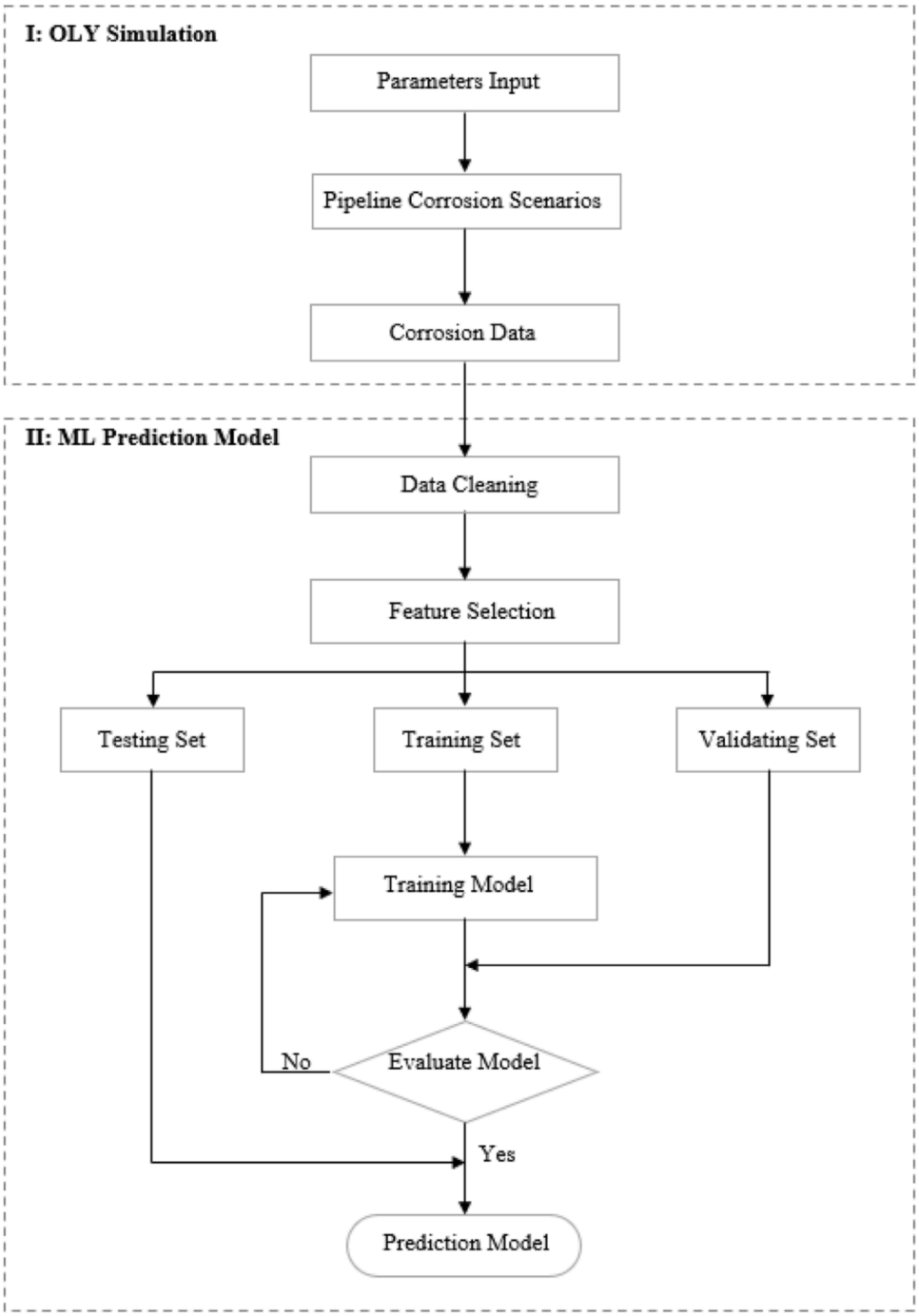


Fig. 1. The methodology of OLI simulation and ML prediction model.

Table 1
Typical pipeline crude oil specifications.

Characteristic	Specification
Water content	1 wt%
NaCl, $MgCl_2$, and $CaCl_2$	5000 - 250,000 ppm, 75 wt% NaCl, 15 wt% $MgCl_2$, and 10 wt% $CaCl_2$
Total H_2S and HS^{-1}	$\leq$ 100 ppm
Total HCO_3^{-1} and CO_2	50 - 500 ppm
SO_4^{-2}	4730 mg/L Na_2SO_4 solubility in water
o_2	dissolved oxygen 8 mg/l
sulfur	0.909 wt%
Ni^{+2}	30 - 250 ppm
V^{+2}	40 - 1500 ppm
Assay	Alaska North Slope sample
pH	7.0 - 9.0
Typical delivery temperature	Ambient
Typical delivery pressure	600 - 1000 psia
Typical delivery velocity	1.3 - 3.6 m/s

Table 2
Specifications of Alaska north slope crude oil assay summary (bp, 2022).

#	Yield on Crude	%wt	%vol
1	Gas to C4	3	4.45
2	Light Distillate to 149 °C	19	22.45
3	Kerosene 149 - 232 °C	12.85	13.7
4	Gas Oil 232 - 369 °C	22.4	21.95
5	Vacuum Gas Oil 369 °C - 550 °C	24.75	22.6
6	Residue above 550 °C	18	14.85
Total		100	100

* Gravity (°API) = 33.

3.2. Natural gas properties

Typical pipeline natural gas specifications were obtained from open literature as shown in Table 3 (Poe and Mokhatab, 2017). Different natural gas compositions were simulated based on the variation of compositions within the range of each specification.

4. OLI corrosion chemistry simulation

4.1. OLI simulation of crude oil pipeline corrosions

In the base case of the simulation, the amount of the crude oil stream resulted from the mix of the brine stream and the ASSAY stream is 100 kg. The pipeline material is Carbon steel 1018. The pipe diameter is set to be 48 inches. The base case settings are given in Tables 4 (a) and (b).

The operating parameters of crude oil pipelines and the compositions of species are listed in Table 5. According to engineering experience, the most important contributors to corrosion are compositions of oxygen (O_2), hydrogen sulfide (gaseous H_2S and aqueous HS^{-1}), carbon dioxide (gaseous CO_2 and aqueous HCO_3^{-1}), sulfate (SO_4^{-2}), and salts (NaCl, $CaCl_2$, and $MgCl_2$). In addition, temperature (T), pressure (P), velocity (v), and pH value (pH) also make noticeable impacts. Thousands of datasets were generated by varying the compositions of these sensitive species: T, P, v, and pH. Then the first principles-based software package, OLI Studio Corrosion Analyzer, was used to simulate the corrosion rate under specific conditions.

Table 3
Typical pipeline gas specifications (Poe and Mokhatab, 2017).

Characteristic	Specification
Water content	4 – 7 lbm H_2O/MMscf of gas
Hydrogen sulfide content	0.25 – 1.0 grain/100 scf
Gross heating value	950 – 1200 Btu/scf
Hydrocarbon dewpoint	14 – 40°F at specified pressure
Mercaptans content	0.25 – 1.0 grain/100 scf
Total sulfur content	0.5 – 20 grain/100 scf
Carbon dioxide content	2 – 4 mol%
Oxygen content	0.01 mol% (max)
Nitrogen content	4 – 5 mol%
Total inert content (N_2 + CO_2)	4 – 5 mol%
Sand, dust, gums, and free liquid	None
Typical delivery temperature	Ambient
Typical delivery pressure	400 – 1200 psig

Table 4
Specifications for base case pipeline crude oil.

(a) Characteristics of Brine Stream

Brine Stream	Unit	Amount
Stream Weight	*g*	1059.6
Stream Volume	*L*	1
T emperature (*T*)	*°C*	50
Pressure (*P*)	*psia*	800
Measured pH (*pH*)		8.0
CO_2 Concentration	*mg/L*	100
H_2S Concentration	*mg/L*	5
O_2 Concentration	*mg/L*	4
NaCl Concentration	*mg/L*	100,000
$CaCl_2$ Concentration	*mg/L*	
$MgCl_2$ Concentration	*mg/L*	
Ni^{+2} Concentration	*mg/L*	200
V^{+2} Concentration	*mg/L*	300
SO_4^{-2} Concentration	*mg/L*	1000
HCO_3^{-1} Concentration	*mg/L*	100
HS^{-1} Concentration	*mg/L*	5

(b) Characteristics of Crude Oil Assay Stream

ASSAY Stream	Unit	Amount
Stream Weight	*g*	98,940.4
T emperature (*T*)	*°C*	50
Pressure (*P*)	*psia*	800
H_2O Weight	*g*	0
CH_4S* Weight	*g*	19.0984
C_4H_4S** Weight	*g*	41.4524
$C_{12}H_8S$*** Weight	*g*	126.7469
ASSAY Weight	*g*	98,753.1023
Specific Gravity		0.8597

Note:.
* CH_4S is Methanethiol.
** C_4H_4S is Thiophene.
*** $C_{12}H_8S$ is Dibenzothiophene.

Table 5
The operating parameters and concentrations of sensitive components.

Parameter	Unit	Range
T emperature (*T*)	*°C*	20 – 80
Pressure (*P*)	*psia*	600 – 1000
Velocity (*v*)	*m/s*	1.5 – 3.5
pH value (*pH*)		7.00 – 9.00
Concentrations of Sensitive Components		
H_2S	*mg/L*	0 – 10.0
HS^{-1}	*mg/L*	0 – 10.0
CO_2	*mg/L*	25 – 200
HCO_3^{-1}	*mg/L*	25 – 200
O_2	*mg/L*	0 – 8.0
SO_4^{-2}	*mg/L*	100 – 2000
NaCl	*mg/L*	5000 – 200,000*
$CaCl_2$	*mg/L*	
$MgCl_2$	*mg/L*	

Note:.
* 75 wt% NaCl, 15 wt% $MgCl_2$, and 10 wt% $CaCl_2$.

4.2. Investigation of the impacts of parameters on corrosion rates of crude oil pipelines

4.2.1. Effect of temperature, pressure, and velocity on corrosion rate

Fig. 2 presents the heat map of *T*, *P*, and corrosion rate. It showed that corrosion rate increases as temperature increases, while the effect of pressure on the corrosion rate was negligible. High temperature crude oil can increase pipeline corrosion rate because it can accelerate the chemical reactions between the oil and the pipeline material, leading to increased corrosion. At higher temperatures, the oil can become more acidic and reactive, and can also contain higher levels of sulfur and other corrosive compounds. On the other hand, increasing the pressure in the pipeline may not necessarily have a direct effect on the corrosion rate. This is because pressure does not directly affect the chemical reactions. Moreover, the corrosion rate increased with higher temperature and higher flow velocity, as shown in Fig. 3. High crude oil velocity can cause erosion on the surface of the pipeline, leading to the removal of the protective oxide layer and exposing the metal surface to the corrosive environment. Compared with the color changes of temperature and velocity, the temperature has a greater influence on the corrosion rate over the velocity.

4.2.2. Effect of O_2, H_2S, and CO_2 concentrations on corrosion rate and pH value

Simulation results showed that the corrosion rate increased linearly as O_2 concentration increased, but pH value stayed constant within certain O_2 ranges. O_2 in crude oil can react with metal surfaces of the internal pipeline, causing corrosion. Corrosion can be accelerated by the presence of water or acidic compounds in the crude oil. O_2 can also react with other compounds in the crude oil, forming acidic species such as organic acids and peroxides. These acidic species can then attack the internal pipeline surface, leading to increased corrosion rates. This indicated that the change in O_2 concentration in these ranges didn't affect the pH value, as shown in Fig. 4. In the prior research, Ismail and Adan (2014) found out that the corrosion rates were 0.150 and 0.008 mm/year in 3.5% NaCl solution with O_2 and without O_2. The corrosion rates were 0.364 and 0.0034 mm/year in 3.5% H_2SO_4 solution with O_2 and without O_2. The corrosion rates were 0.1301 and 0.0011 mm/year in 3.5% HCl solution with O_2 and without O_2 respectively. The OLI Studio Corrosion Analyzer simulation results were consistent with these reported data.

As shown in Fig. 5, Cox and Roetheli (1931) showed that as the O_2 concentration increased at any definite velocity in the range studied, the corrosion rate increased as well. This literature report showed coherency with the OLI Studio Corrosion Analyzer simulation.

The OLI Studio Corrosion Analyzer simulation data showed that same corrosion rate resulted from the same concentration of gaseous H_2S and aqueous HS^{-1}, respectively. As shown in Fig. 6, the x-axis represented the total concentration of the gaseous H_2S and the aqueous HS^{-1}. While the corrosion rate increased with the increase of H_2S concentration, pH value decreased. This was because the mild H_2S acid gasses may be absorbed in the crude oil aqueous solution and then dissociated to H^+ which causes the decrease of pH value. These cations may catalyze the anodic dissolution of carbon steel and the cathodic reduction process, both of which generated hydrogen.

The OLI Studio Corrosion Analyzer simulation data showed that the same corrosion rate resulted from the same concentration of gaseous CO_2 and aqueous HCO_3^{-1} respectively. As shown in Fig. 7, the x-axis represented the total concentration of gaseous CO_2 and aqueous HCO_3^{-1}. As the CO_2 concentration increases, the corrosion rate increases, but pH

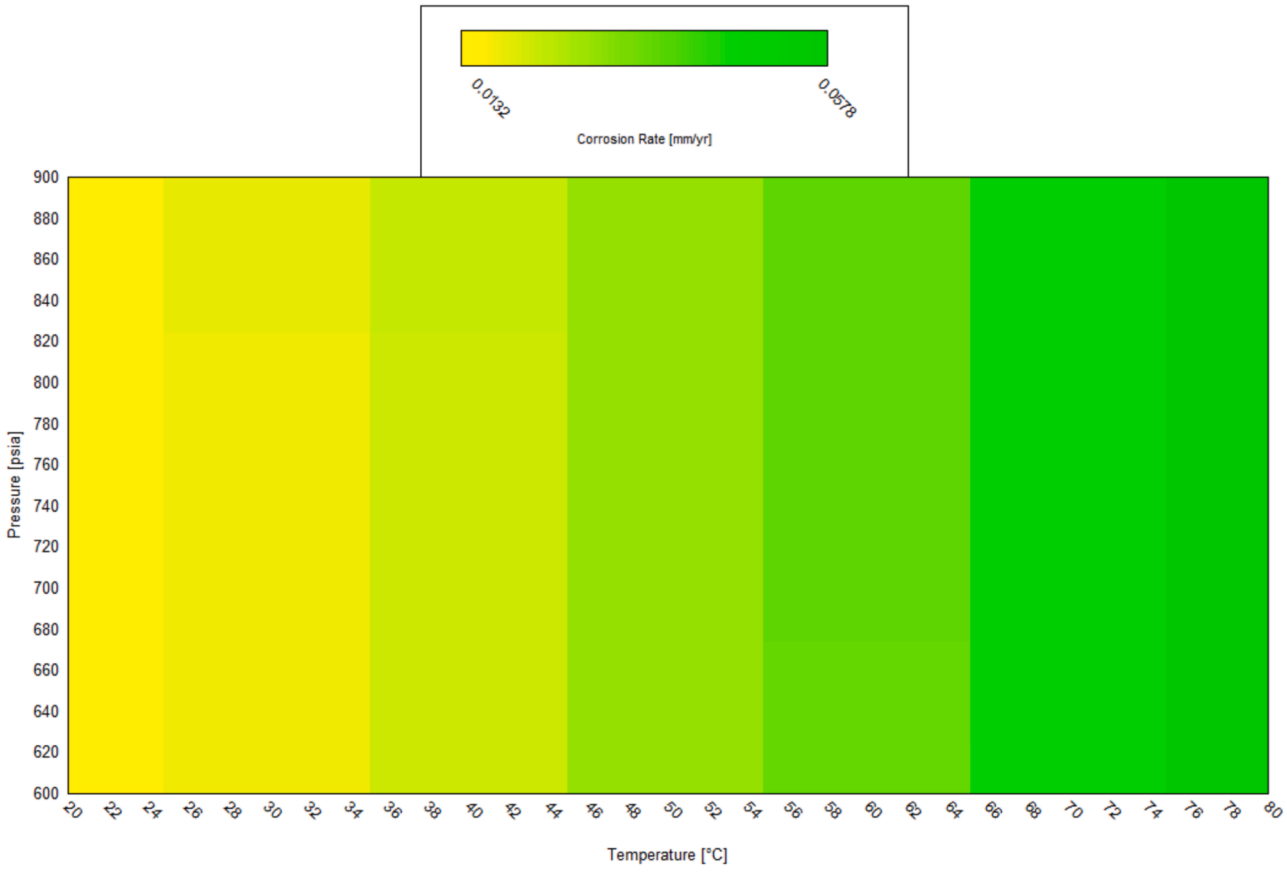


Fig. 2. The heat map of temperature, pressure, and corrosion rate.

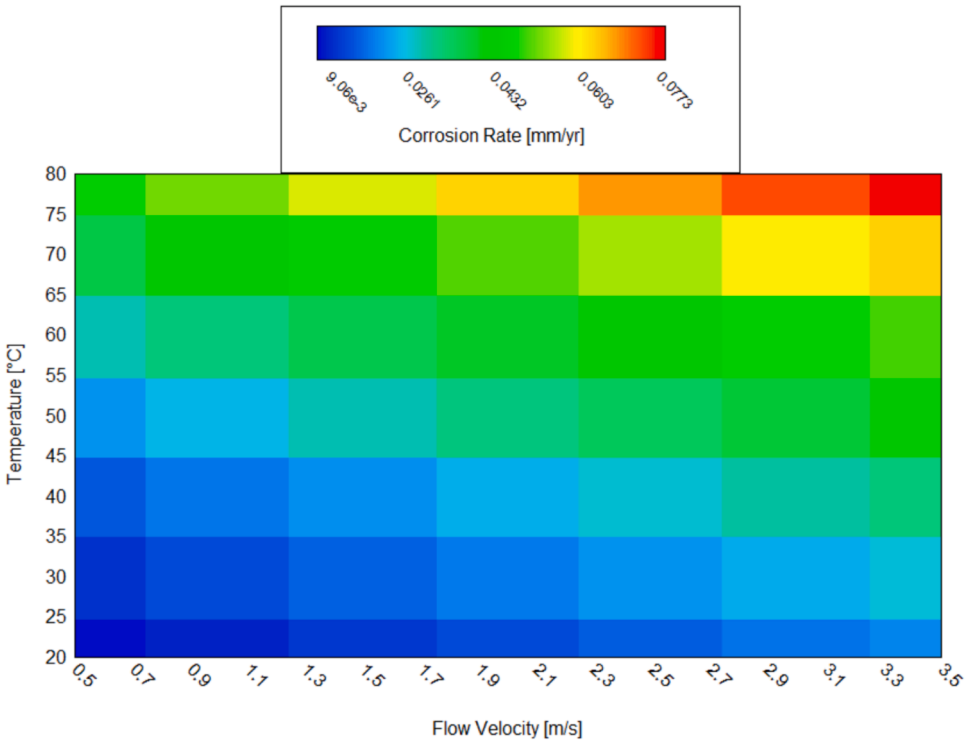


Fig. 3. The heat map of temperature, velocity, and corrosion rate.

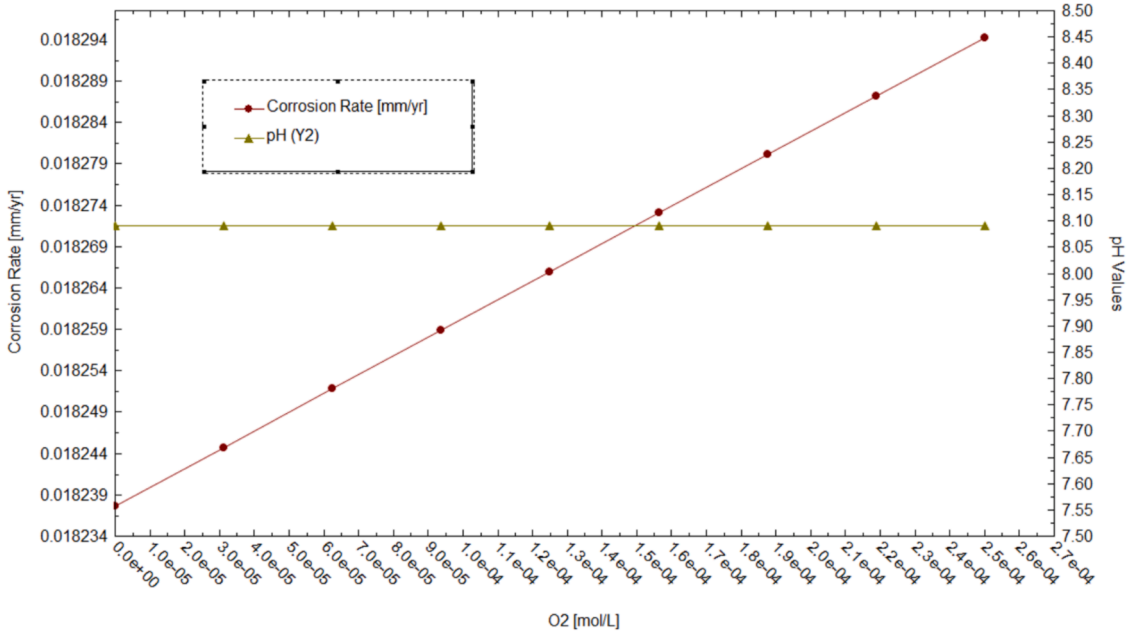


Fig. 4. Corrosion rate and pH value vs. O_2 concentration.

value decreases. This was because CO_2 is an acidic gas.

In the prior research, Song (2010) validated a model by using the experimental data obtained at 90 °C with a solution under both saturated and unsaturated conditions. When either the CO_2 pressure or the presence of CO_2 increased, the corrosion rate increased, as shown in Fig. 8. The literature result was agreeable with the simulation using OLI Studio Corrosion Analyzer.

Fang et al. (2010) investigated the CO_2 corrosion rate of C1018 carbon steel in an aqueous solution with NaCl concentrations 3 – 25 wt% at 20 °C, pH 4.0, 5.0, and 6.0, respectively. They found out that the CO_2 corrosion rate of carbon steel significantly decreased as the salt concentration increased.

Liu et al. (2016) investigated the effects of chloride ion, temperature, pH value, CO_2, and O_2 on the corrosion induced leakage of the inner wall of the crude oil pipelines (AISI 1020 steel). The corrosion rate increased when the concentration of chloride ions increased from 44,000 mg/L to 144,000 mg/L. When the pH values were 6.5, 5.5, 4.5, and 3.5, the corrosion rates were 0.1637 mm/year, 0.1764 mm/year, 0.2607 mm/year, and 0.2767 mm/year, respectively.

Bai et al. (2018) investigated the influence of CO_2 partial pressure on

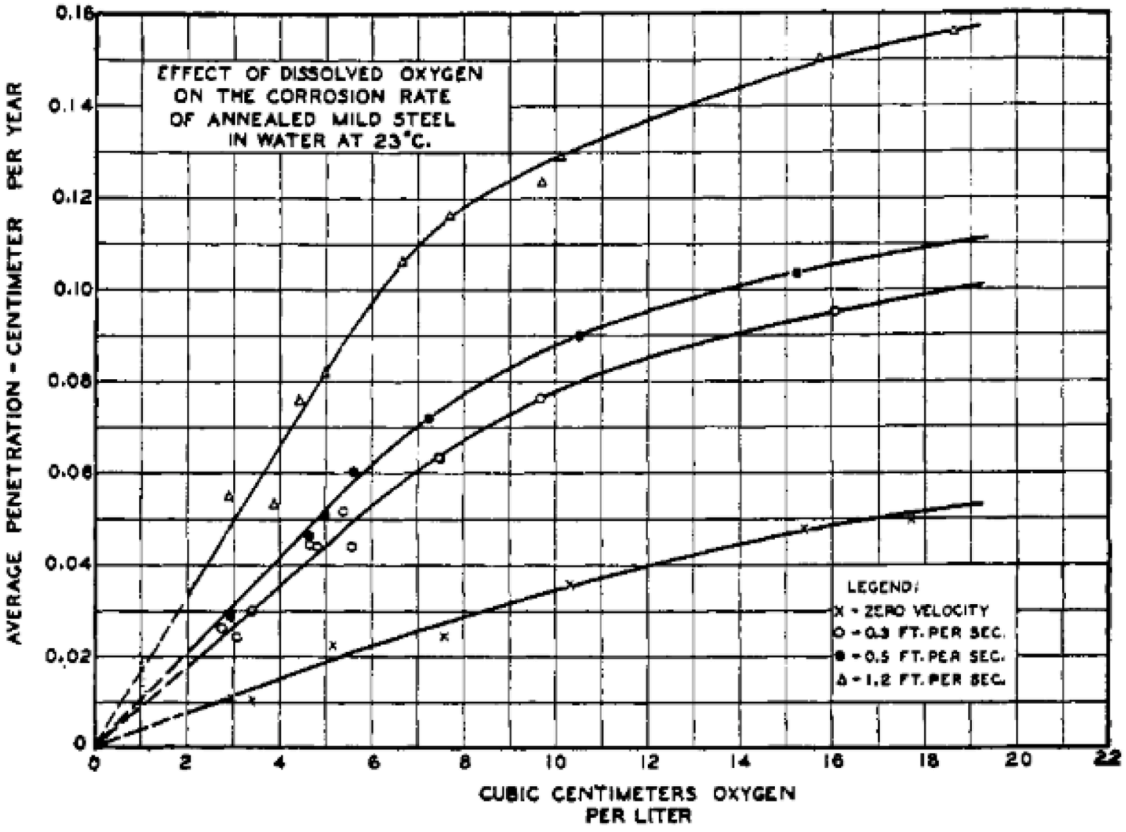


Fig. 5. Effect of dissolved oxygen on the corrosion rate of annealed mild steel in water at 23 °C (Cox and Roetheli, 1931).

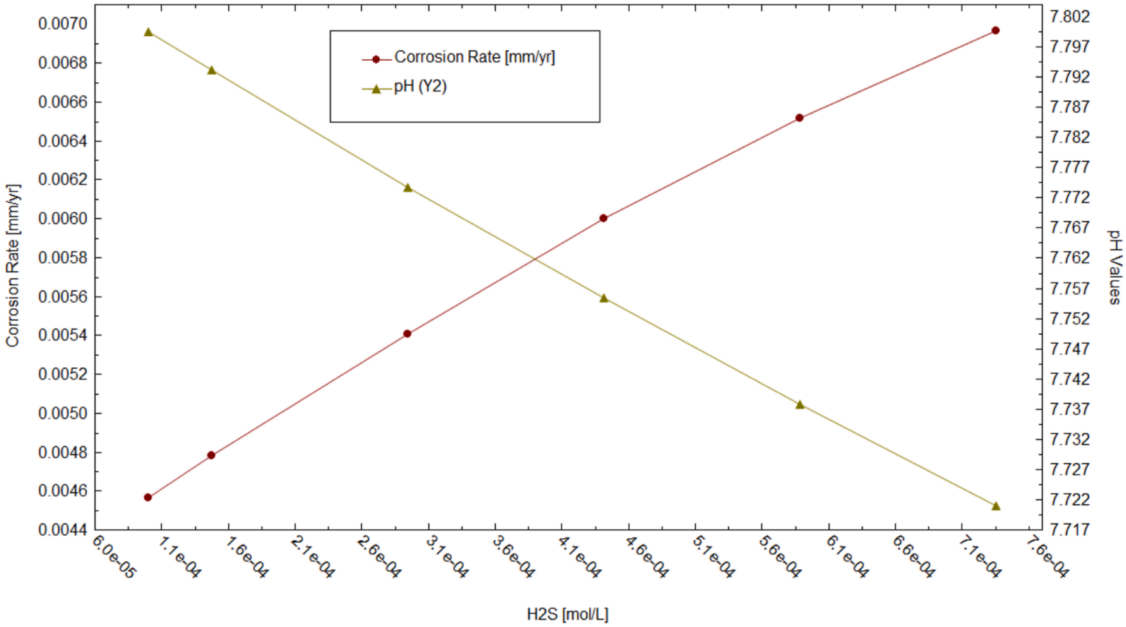


Fig. 6. Corrosion rate and pH Value vs. H_2S concentration.

the corrosion rate of J55 carbon steel in 30% (V/V) crude oil/brine. They reported that the system pH values decreased with the increased CO_2 partial pressure, as shown in Fig. 9. When the CO_2 partial pressure was small, the system pH values decreased significantly with the increased CO_2 partial pressure. This conclusion agreed exactly with this simulation.

4.3. OLI simulation of natural gas pipeline corrosions

In the base case of the simulation, the amount of the natural gas (NG) stream was 10,000 mol. The pipeline material was Carbon steel 1018. The pipe diameter was 36 inches.

The operating parameters of natural gas pipelines and the components of the hydrocarbons in NG were listed in Tables 6 and 7. The most

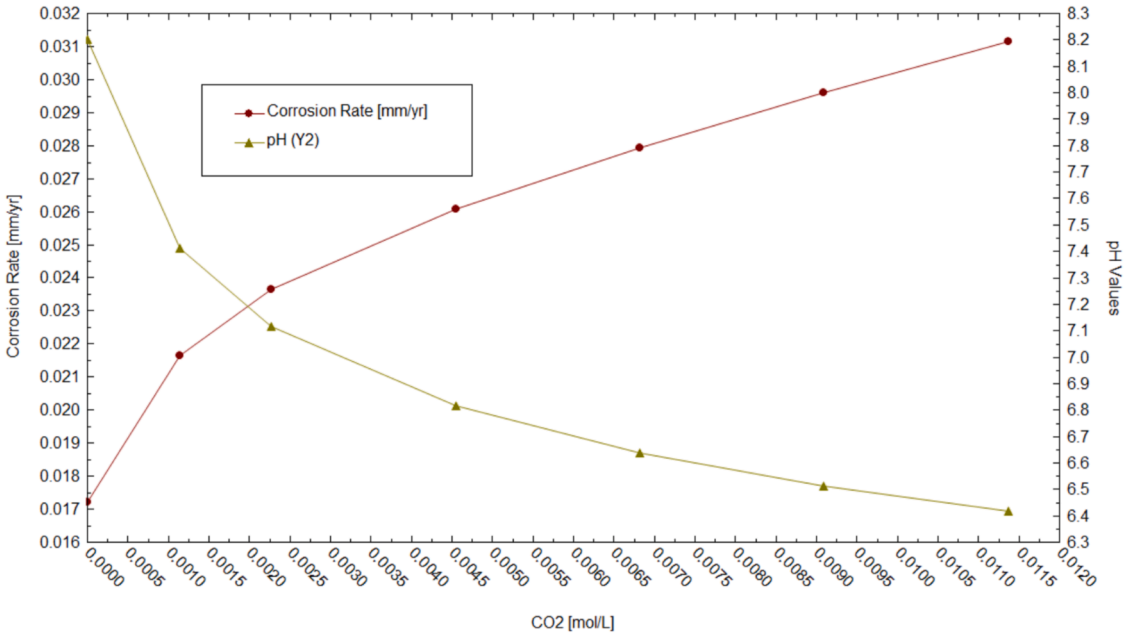


Fig. 7. Corrosion rate and pH value vs. CO_2 concentration.

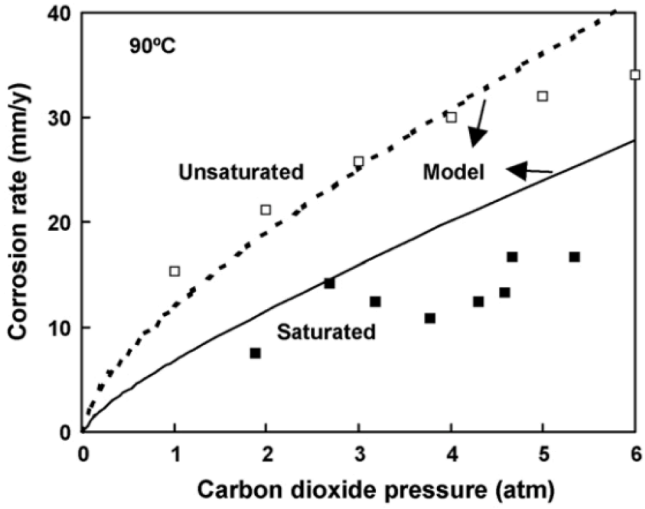


Fig. 8. Model validation with experimental data obtained at 90 °C (Song, 2010).

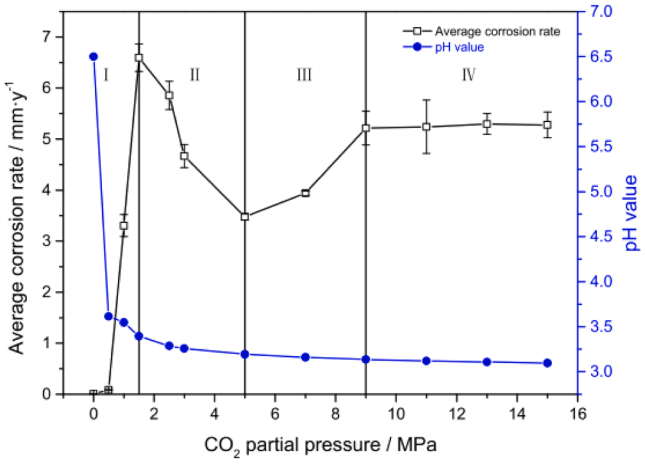


Fig. 9. Corrosion rate and pH value as a function of CO_2 partial pressure in 30% crude oil/brine mixtures (Bai et al., 2018).

important components for corrosion are compositions of water (H_2O), hydrogen sulfide (H_2S), carbon dioxide (CO_2), and oxygen (O_2). In addition, the temperature (*T*), pressure (*P*), and velocity (*v*) also made noticeable impacts. Then OLI Studio Corrosion Analyzer was used to simulate the corrosion rate under the specific conditions.

4.4. Investigation of the impacts of parameters on corrosion rates of natural gas pipelines

4.4.1. Effect of temperature, pressure, and velocity on corrosion rate

Fig. 10(a) showed that the internal corrosion rate increased as *P* increased, while *v* was constant. Similarly, the change of internal corrosion rate and *v* were proportional when *P* stayed constant.

Fig. 10(b) showed that the internal corrosion rate decreased as *T* increased, while *v* was constant. However, the internal corrosion rate increased as *v* increased when *T* was constant. As *v* increased, the internal corrosion rate decreased quickly as *T* increased.

Table 6
The parameters of OLI simulation for natural gas pipelines corrosion.

Parameter	Unit	Range	Base case value
Temperature	*°F*	50 – 70	70
Pressure	*psia*	400 – 1200	1200
Velocity	*ft/s*	10 – 30	30
Concentrations of Sensitive Components			
H_2S	*mol %*	0 – 0.00151	0.00151
CO_2	*mol %*	0.50000 – 2.00000	2.00000
O_2	*mol %*	0 – 0.00010	0.00010
H_2O	*mol %*	0.05834 – 0.13434	0.13434
N_2	*mol %*	3.00000 – 5.00161	3.00000

Table 7
The components of hydrocarbons in natural gas.

Parament	Unit	Mole Fraction
CH_4	*mol%*	90.00000
C_2H_6	*mol%*	3.00000
C_3H_8	*mol%*	1.00000
n-C_4H_{10}	*mol%*	0.37000
i-C_4H_{10}	*mol%*	0.37000
C_5H_{12}	*mol%*	0.05000
i-C_5H_{12}	*mol%*	0.05000
C_6H_{14}	*mol%*	0.01415

Figs. 10(c) showed that internal corrosion rate increased as P increased and it also increased as T increased, when v was 30 ft/s. In Fig. 10(d), the red color represented high corrosion rate, and the blue color represented low corrosion rate. Fig. 10(d) showed that the corrosion rate was low when P and T were low. The corrosion rate was high when P and T were high. In addition, for the pressure range of 450 – 550 psia, when T increased, the internal corrosion rate also increased.

4.4.2. *Effect of H_2S, O_2, and CO_2 on corrosion rate*

Simulation results showed that for pipeline grade natural gas, the corrosion rate actually decreased with H_2S concentration in the certain range. Some representative data were listed in Table 8.

Fig. 11(a) was reported by Anderko and Young (1999), who compared their computer simulation work with experimental data reported by Sardisco et al. (1963). At low concentrations, H_2S can form a protective film on the internal pipeline surface, which can reduce the corrosion rate. This is because the H_2S reacts with the metal surface to form a thin layer of FeS, which acts as a barrier against further corrosion (Anderko and Young, 1999). However, this protective film is not effective at high H_2S concentrations, as the film can break down and lead to accelerated corrosion. At high concentrations, H_2S can react with water to form sulfuric acid (H_2SO_4), which is a highly corrosive compound. This acid can then react with the internal metal surface of the pipeline, leading to severe corrosion.

The decreasing trend of the black line, as shown in Fig. 11(b) was consistent with the downward part of the red line (left side of the black vertical line). The comparison of OLI Studio Corrosion Analyzer simulation data, as shown in Fig. 11(b), with literature data presented in Fig. 11(a), confirmed that corrosion rate prediction using OLI Studio Corrosion Analyzer is reliable and consistent with prior experimental research.

Kermani and Morshed (2003) reported that CO_2 corrosion is indeed ubiquitous at high CO_2 partial pressure (P_{CO_2}) and it progressively disappears at low P_{CO_2}, unless H_2CO_3 is supplemented by free organic acids. Within the temperature range of 20–70 °C, the internal corrosion rate exhibits a direct relationship with temperature, wherein an increase in temperature corresponds to an increase in the internal corrosion rate. Concurrently, an elevation in CO_2 partial pressure signifies a higher concentration of CO_2 within the natural gas, consequently leading to an augmented internal corrosion rate as shown in Fig. 12. In this study, our simulation results consistently demonstrate a positive correlation between temperature and the internal corrosion rate within the temperature range of 50 to 70°F. Furthermore, the observed escalation in corrosion rate resulting from higher CO_2 concentration aligns with prior research outcomes.

Touali (2013) found out that P_{CO_2} has a significant impact on the corrosion of metals. Lower P_{CO_2} results in lower carbon steel weight loss, whereas higher P_{CO_2} leads to higher weight loss as shown in Fig. 13. In the crude oil and natural gas industry, the corrosion damage of CO_2 is primarily caused by the carbonic acid concentration, which is influenced by P_{CO_2} and reservoir water content. Our research result is also consistent with the findings of Touali.

5. Development of corrosion ml algorithms prediction models

5.1. *Crude oil pipelines corrosion rate data*

The crude oil pipelines corrosion rate (CR) data were generated from the simulation using OLI Studio Corrosion Analyzer. The input variables used in OLI Studio Corrosion Analyzer include sodium ion mass concentration ("$C_{Na^{+1}}$"), calcium ion mass concentration ("$C_{Ca^{+2}}$"), magnesium ion mass concentration ("$C_{Mg^{+2}}$"), chlorine anion mass concentration ("$C_{Cl^{-1}}$"), sulfate anion mass concentration ("$C_{SO_4^{-2}}$"), carbon dioxide mass concentration ("C_{CO_2}"), bicarbonate anion mass concentration ("$C_{HCO_3^{-1}}$"), hydrogen sulfite mass concentration ("C_{H_2S}"), bisulfide anion mass concentration ("$C_{HS^{-1}}$"), oxygen mass concentration ("C_{O_2}"), pH value ("pH"), temperature ("T"), pressure ("P"), and pipeline flow velocity ("v"). A total of 8748 sets of the crude oil pipelines corrosion rate data were obtained using OLI Studio Corrosion Analyzer.

As shown in Fig. 14(a), the histogram of pH values showed that the pH value's spread was from 5.67 to 9.39, and it was a normal distribution. 81.24% of the data mainly distributed in the bins of 7.0 – 8.0 and 8.0 – 9.0, which was consistent with the range of 7.0 – 9.0 in the previous study (Lynch et al., 2014).

As shown in Fig. 14(b), the histogram of corrosion rate showed that the corrosion rate spread was from 0.004 to 0.3628, and its distribution skewed left. 82.33% of the data were distributed in the bins of 0.01 – 0.06 and 0.06 – 0.11. The OLI Studio Corrosion Analyzer simulation results were consistent with a previous study (Collier et al., 2012) which reported that the average corrosion rate of carbon steel in brine with crude oil was 2.1 ± 1.9 mpy (0.00508 – 0.1016 mm/yr).

Figs. 6 and 7 showed that the changed concentrations of H_2S, and CO_2 caused the change of pH value, which ultimately affected the corrosion rate. The scatter plot in Fig. 15 showed that the datasets with corrosion rates greater than 0.20 were mainly distributed in the area with pH values less than 7.0. Moreover, the scatter plot in Fig. 15 didn't disclose any obvious zone-based relationship between pH value and corrosion rate.

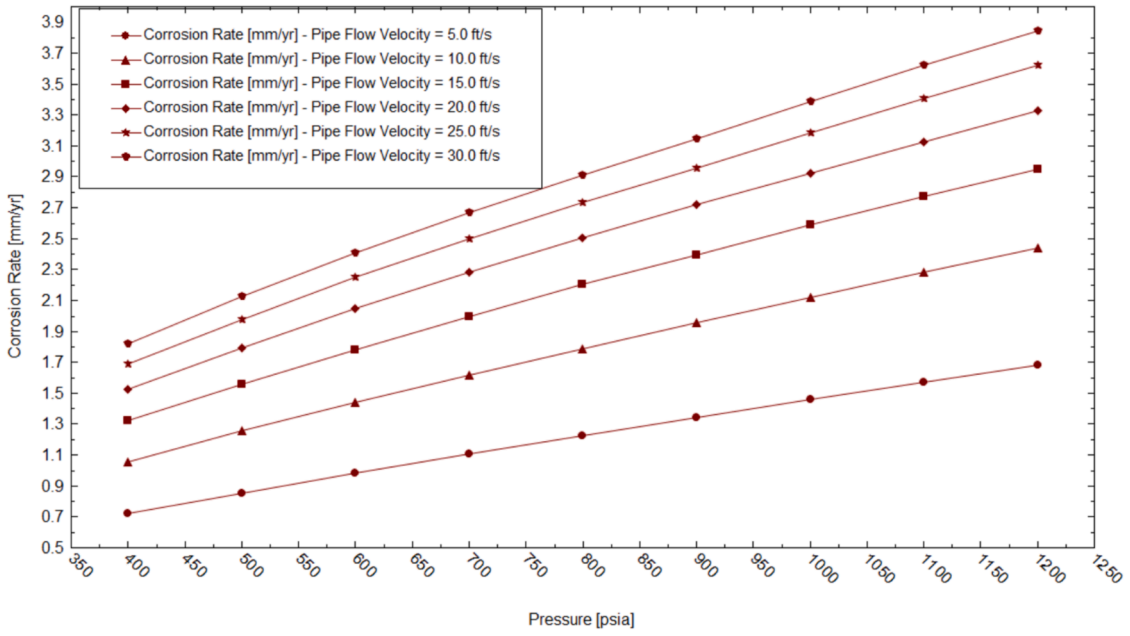
5.2. *Natural gas pipelines corrosion rate data*

The natural gas pipelines corrosion rate data were generated from the simulation using the OLI Studio Corrosion Analyzer. The input variables were used in the OLI Studio Corrosion Analyzer include temperature ("T"), pressure ("P"), pipeline flow velocity ("v"), moles of carbon dioxide ("n_{CO_2}"), methane ("n_{CH_4}"), ethane ("$n_{C_2H_6}$"), propane ("$n_{C_3H_8}$"), butane ("$n_{n-C_4H_{10}}$"), Isobutane ("$n_{i-C_4H_{10}}$"), pentane ("$n_{C_5H_{12}}$"), isopentane ("$n_{i-C_5H_{12}}$"), hexane ("$n_{C_6H_{14}}$"), oxygen ("n_{O_2}"), hydrogen sulfite ("n_{H_2S}"), water ("n_{H_2O}"), and nitrogen ("n_{N_2}"). A total of 3240 datasets of the natural gas pipelines corrosion rate data were obtained.

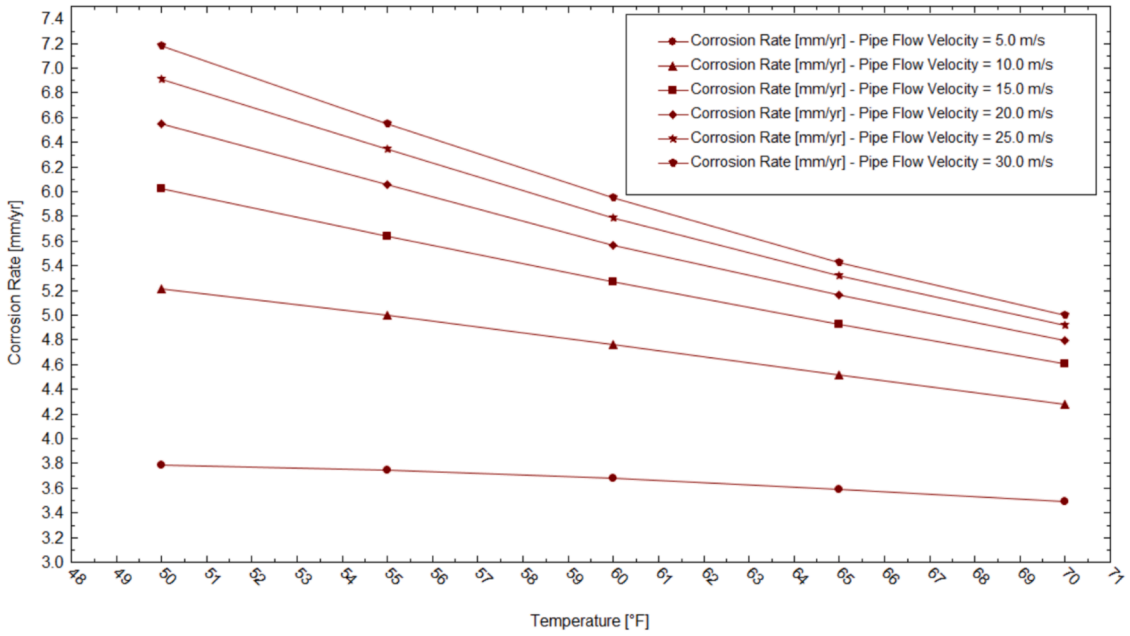
As shown in Fig. 16, the histogram of corrosion rate showed that the corrosion rate spread was from 0.078 to 9.130, and its distribution skewed right. The data were distributed in the bins of 0.078 – 1.586 and 1.586 – 3.095. In addition, it was clear from Table 6 that the histograms of the data indicated the presence of non-normal distributions at various scales. For example, the temperature varied in a range of 50 – 70°F. Similarly, the pressure ranged from 400 to 1200 psia, while the range of the output variable, corrosion rate, was between 0.078 and 9.130 mm/yr.

5.3. *Feature selection*

Feature selection is to examine the strength of the relationship of the feature with the response variable. There are three reasons why feature selection is important. First, the multicollinearity of features may cause over-fitting. Reasonably reducing the number of features can help build a robust model. Subsequently, the resulting predictive model can clearly depict the relationship between features and response variables. Third,



(a) Corrosion Rate vs. Pressure and Velocity at 70 °F



(b) Corrosion Rate vs. Temperature and Velocity at 1200 psia

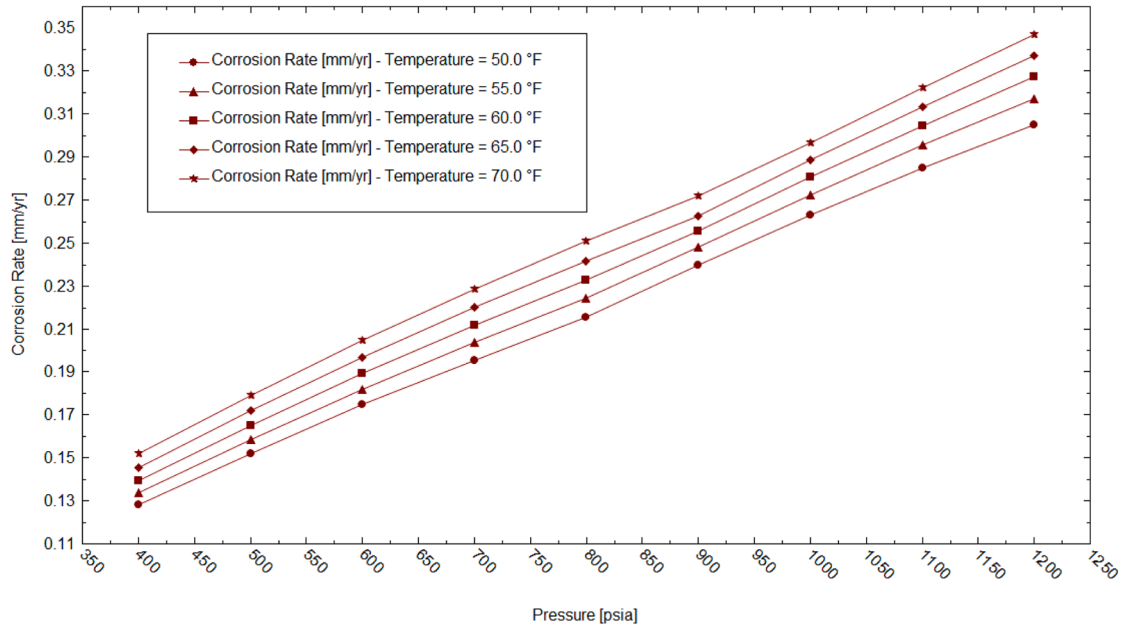
Fig. 10. The relationships of corrosion rate vs. temperature, pressure, and velocity. (a) Corrosion rate vs. pressure and velocity at 70°F. (b) Corrosion rate vs. temperature and velocity at 1200 psia. (c) Corrosion rate vs. Temperature and pressure at 30 ft/s velocity. (d) Heatmap of corrosion rate vs. temperature and pressure at 30 ft/s velocity.

feature selection can reduce model complexity and running time. In this section, three feature selection methods: XGBoost Feature Importance (Brownlee, 2016), Permutation Feature Importance (Brownlee, 2020), and Random Forest Regression Feature Importance (Gupta, 2020) were used to select the important features for the models.

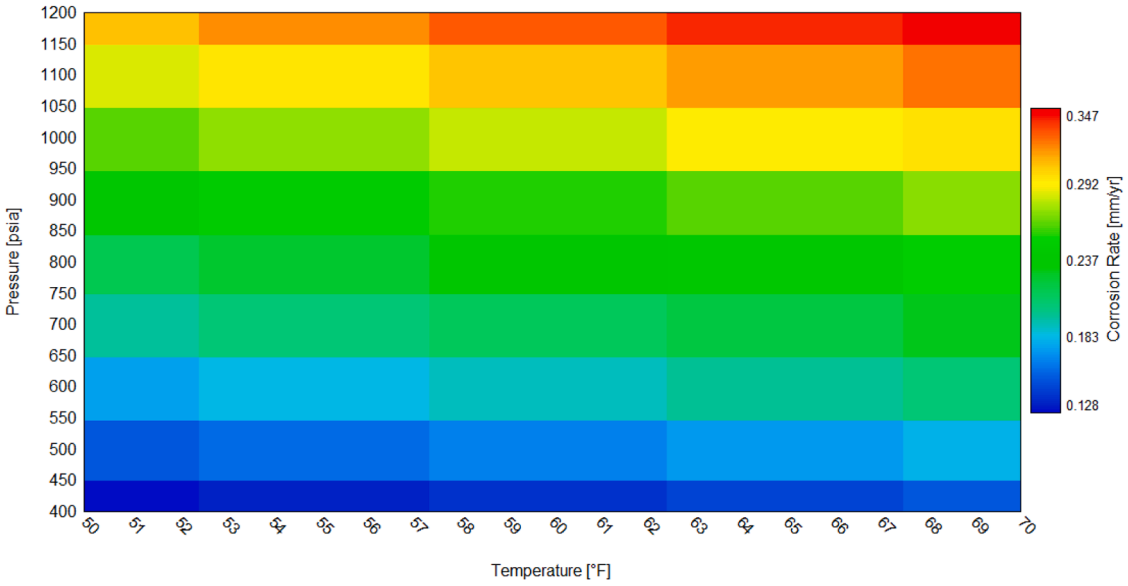
6. Evaluation of ML models

6.1. ML models for crude oil pipelines corrosion rate

Eight thousand seven hundred forty-eight (8748) datasets from OLI Studio Corrosion Analyzer simulation were available. The important features (T, C_{O_2}, $C_{Cl^{-1}}$, v, and pH) were selected as the input features to train the ML models for CR. 70% of data were randomly chosen as the training data, 15% of data were randomly selected as the validation



(c) Corrosion Rate vs. Temperature and Pressure at 30 ft/s Velocity



(d) Heatmap of Corrosion Rate vs. Temperature and Pressure at 30 ft/s Velocity

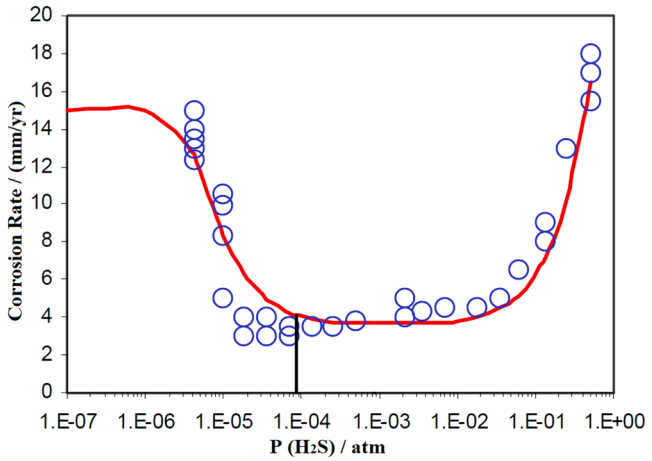
Fig. 10. (*continued*).

Table 8
H_2S composition and the corresponding corrosion rate @ 70°F and 800 psia.

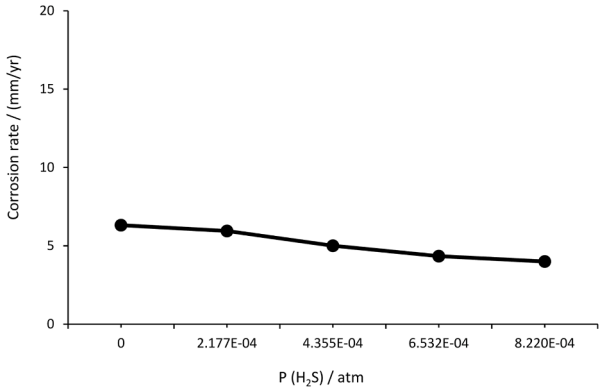
H_2S mole fraction	H_2S partial pressure (atm)	Corrosion rate (mm/yr)
0	0	6.31514
4.000E-06	2.177E-04	5.94669
8.000E-06	4.355E-04	5.00284
1.200E-05	6.532E-04	4.33975
1.510E-05	8.220E-04	3.99673

data, and the remaining 15% of data were used to test the prediction model respectively. The ML *CR* models using RF and CatBoost algorithms were built in Python. The prediction accuracies of the predictive sub-models were evaluated based on R^2, mean squared error (MSE), mean absolute error (MAE), and root mean square error (RMSE).

Five important features, T, C_{O_2}, C_{Cl^-}, v, and pH, were used as input features to train the *CR* models. Two one-zone RF - *CR* and CatBoost - *CR* models and their performances are listed in Table 9, fully utilizing 8748 datasets in total. 70% of data were randomly chosen as the training data, 15% of data were randomly selected as the validation data, and the remaining 15% of data were used to test the prediction model respectively. In addition, the early stopping method was used in the model training to avoid overfitting. As shown in Fig. 17, for validation, the RF - *CR* and CatBoost - *CR* models achieved the R^2 values of 0.996 and 0.999, and the corresponding MSE, MAE, and RMSE values of 0.000006, 0.000688, 0.002463 and 0.000002, 0.000557, 0.001315, respectively. For testing, the RF - *CR* and CatBoost - *CR* models achieved the R^2 values of 0.994 and 0.999, and the corresponding MSE values of 0.000010,



(a) Corrosion Rate vs. H_2S partial pressure (Anderko and Young, 1999)



(b) OLI Simulation Result of Corrosion Rate vs. H_2S Partial Pressure

Fig. 11. Corrosion rate vs. H_2S partial pressure. (a) Corrosion rate vs. H_2S partial pressure (Anderko and Young, 1999). (b) OLI simulation result of corrosion rate vs. H_2S partial pressure.

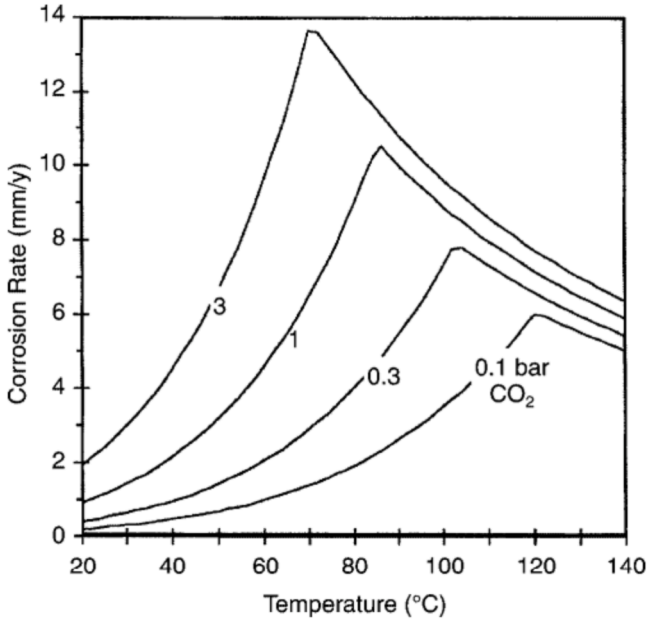


Fig. 12. Effect of temperature on CO_2 corrosion (Kermani and Morshed 2003).

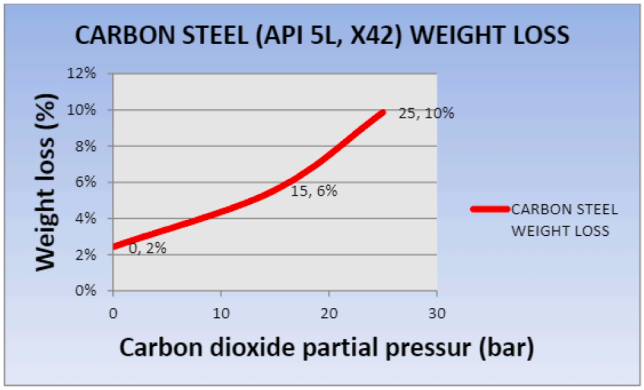


Fig. 13. Carbon steel API 5 L, X42 weight loss (Touali 2013).

0.000752, 0.003125 and 0.000001, 0.000500, 0.001088, respectively. Clearly, the RF - *CR* model was as accurate as the CatBoost - *CR* model.

6.2. ML models for natural gas pipelines corrosion rate

Three thousand two hundred and forty (3240) data were available for machine learning. The important features (n_{O_2}, n_{H_2S}, P, v, n_{CO_2}, and T) were selected as the input features to train the ML models for *CR*. 70% of data were randomly chosen as the training data, 15% of data were randomly selected as the validation data, and the remaining 15% of data were used to test the prediction model respectively. The ML *CR* models using RF and CatBoost algorithms respectively were built in Python. The prediction accuracies of the predictive sub-models were evaluated based on R^2, MSE, MAE, and RMSE.

Six important features, n_{O_2}, n_{H_2S}, P, v, n_{CO_2}, and T, were selected as the input features to train the *CR* models. Two one - zone RF - *CR* and CatBoost - *CR* models were built in Table 10, fully utilizing 3240 datasets in total. 70% of data were randomly chosen as the training data, 15% of data were randomly selected as the validation data, and the remaining 15% of data were used to test the prediction model respectively. In addition, the early stopping method was used in the model training to avoid overfitting. As shown in Fig. 18, for validation, the RF - *CR* and CatBoost - *CR* models achieved the R^2 values of 0.997 and 0.999, and the corresponding MSE values of 0.005658, 0.035013, 0.075220 and 0.001019, 0.008313, 0.031917, respectively. For testing, the RF - *CR* and CatBoost - *CR* models achieved the R^2 values of 0.998 and 0.999, and the corresponding MSE values of 0.005624, 0.037200, 0.074992 and 0.000211, 0.007933, 0.014516, respectively. Clearly, the RF - *CR* model was as accurate as the CatBoost - *CR* model.

7. Conclusion

In this study, a hybrid approach was used to predict corrosion of crude oil and natural gas pipelines. At first, a first principles-based electrochemistry software package, OLI Studio Corrosion Analyzer, was used to simulate various corrosion scenarios to generate corrosion data. Then two ML algorithms, RF and CatBoost were used to predict the internal corrosion rate of crude oil pipelines and natural gas pipelines, respectively.

8748 datasets of crude oil pipeline corrosion rates were generated using OLI Studio Corrosion Analyzer under various scenarios. The models for CR predictions have high predictive accuracies using five important features, T, C_{O_2}, C_{Cl^-1}, v, and pH, and none of the original data were deleted. In addition, both the one-zone RF and CatBoost models for CR achieved superior predictive performance.

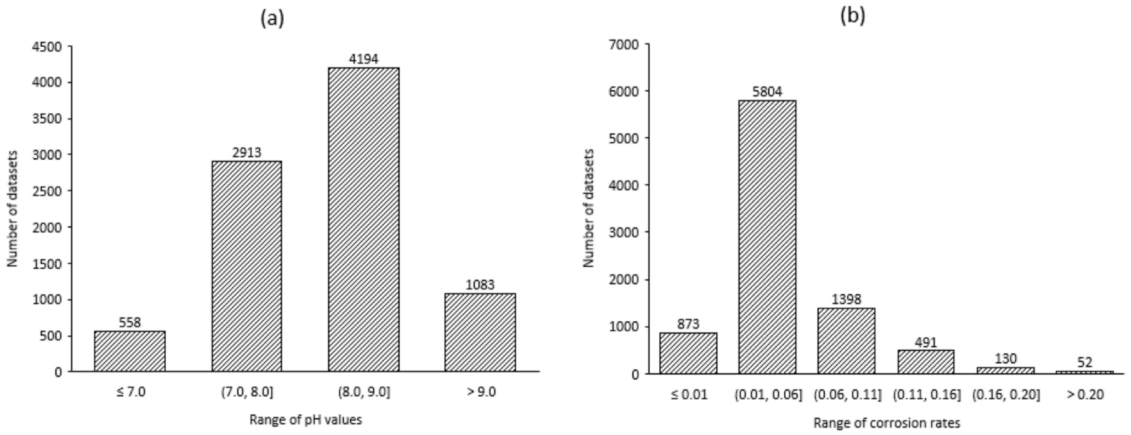


Fig. 14. Frequency histograms of the pH Value (a) and corrosion rate (b).

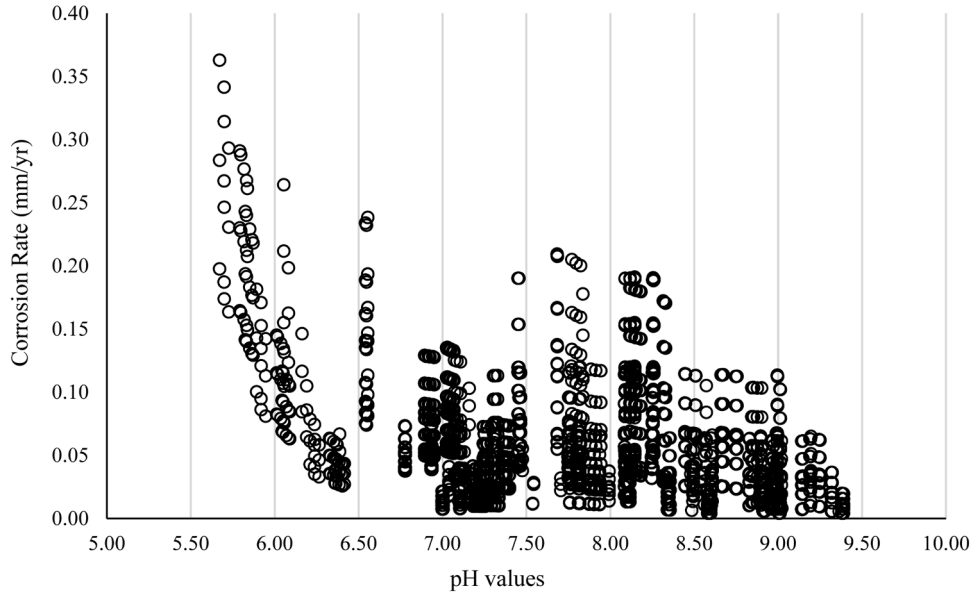


Fig. 15. Scatter-plot of pH Value vs. corrosion rate.

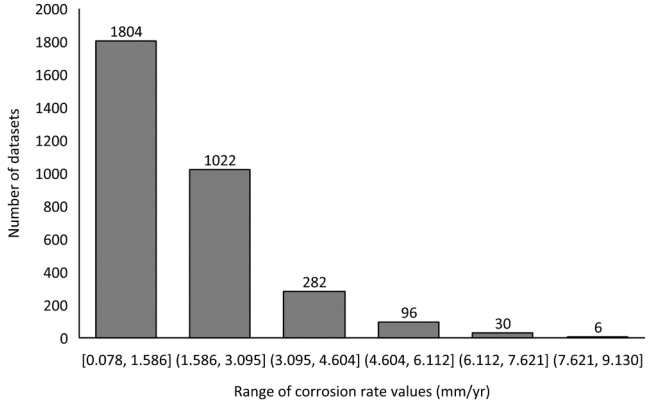


Fig. 16. Distribution histograms of natural gas pipelines corrosion rate.

3645 natural gas pipeline corrosion rate datasets were generated using OLI Studio Corrosion Analyzer under various scenarios. The models for CR predictions have high predictive accuracies using six important features, n_{O_2}, n_{H_2S}, P, v, n_{CO_2}, and T, and none of the original data were deleted. Both the one-zone RF and CatBoost models for CR achieved superior predictive performance.

All these case studies demonstrated the great application potential of using ML for the energy industry. An ML model that featured by simplicity, general applicability, and high reliability will offer tremendous value for operational excellence and asset integrity.

While the RF- and CatBoost-CR models can effectively model the internal corrosion rate of crude oil pipelines and natural gas pipelines, there is still room for improvement in this field. Sulfur compounds are a major cause of corrosion, and addressing this issue is of paramount importance. OLI Studio Corrosion Analyzer has a good corrosion chemistry for inorganic sulfur. However, it cannot simulate the corrosion rate caused by organic sulfur, so the dataset generated in the lab simulation cannot represent the corrosion impact of organic sulfur. Moreover, the impact of microbial corrosion is also significant, but there is not enough open database available. Once the organic sulfur and microbial corrosion data are available, the ML models can be retrained for better prediction.

Table 9
Performance the RF and CatBoost - *CR* models.

Model	Features	Outlier	Validation				Testing			
			R^2	MSE	MAE	RMSE	R^2	MSE	MAE	RMSE
one-zone RF - *CR*	*T*, C_{O_2}, C_{Cl^-}, *v*, and *pH*	0	0.996	0.000006	0.000688	0.002463	0.994	0.000010	0.000752	0.003125
one-zone CatBoost - *CR*	*T*, C_{O_2}, C_{Cl^-}, *v*, and *pH*	0	0.999	0.000002	0.000557	0.001315	0.999	0.000001	0.000500	0.001088

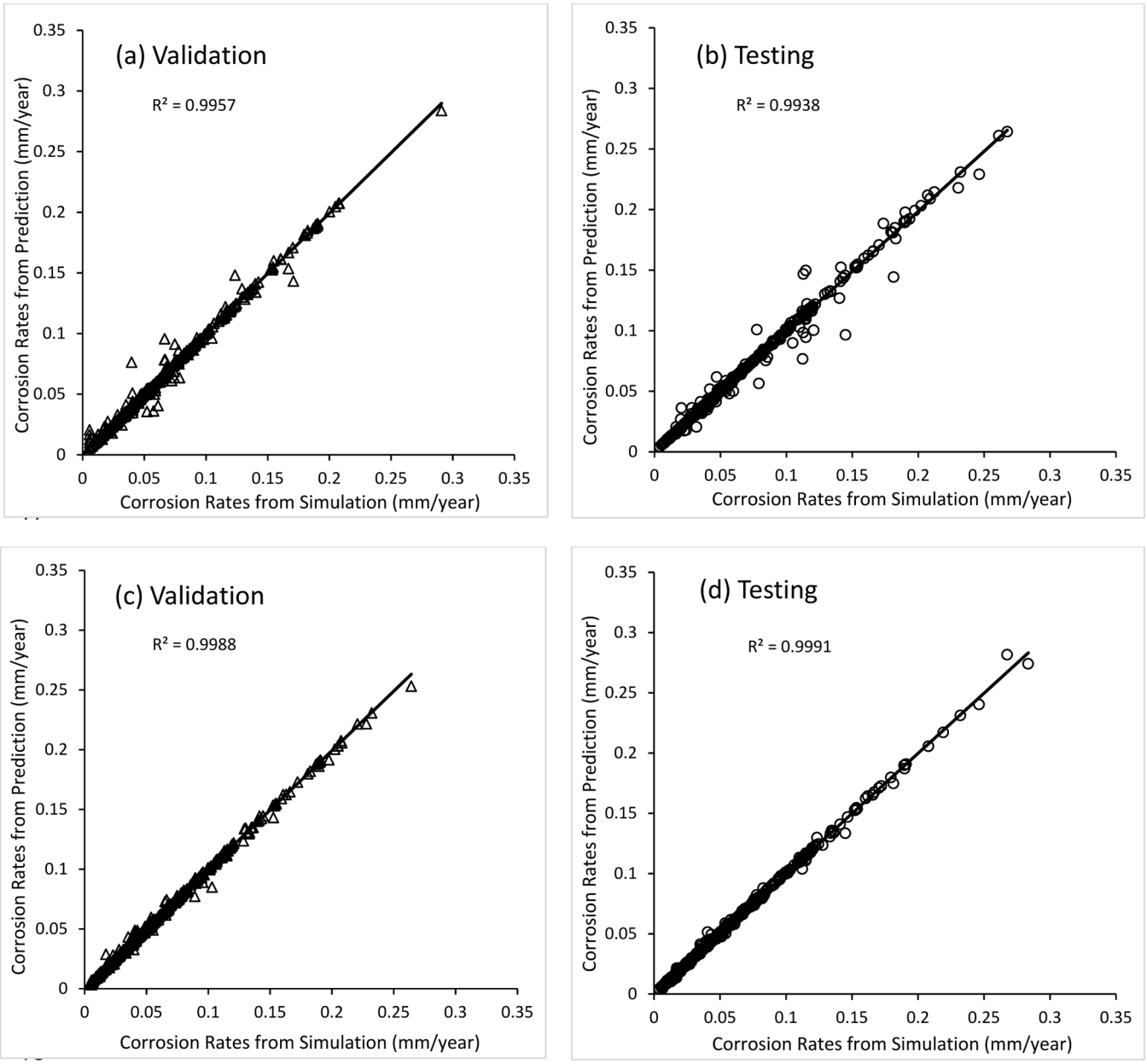


Fig. 17. Predicted corrosion rate (*CR*) vs. simulation *CR* values of crude oil pipelines. (a) RF - *CR* model for validation, (b) RF - *CR* model for testing, (c) CatBoost - *CR* model for validation, and (d) CatBoost - *CR* model for testing.

Table 10
Performance of the RF and CatBoost - *CR* models.

Model	Features	Outlier	Validation				Testing			
			R^2	MSE	MAE	RMSE	R^2	MSE	MAE	RMSE
one-zone RF - *CR*	n_{O_2}, n_{H_2S}, *P*, *v*, n_{CO_2}, and *T*	0	0.997	0.005658	0.035013	0.075220	0.998	0.005624	0.037200	0.074992
one-zone CatBoost - *CR*	n_{O_2}, n_{H_2S}, *P*, *v*, n_{CO_2}, and *T*	0	0.999	0.001019	0.008313	0.031917	0.999	0.000211	0.007933	0.014516

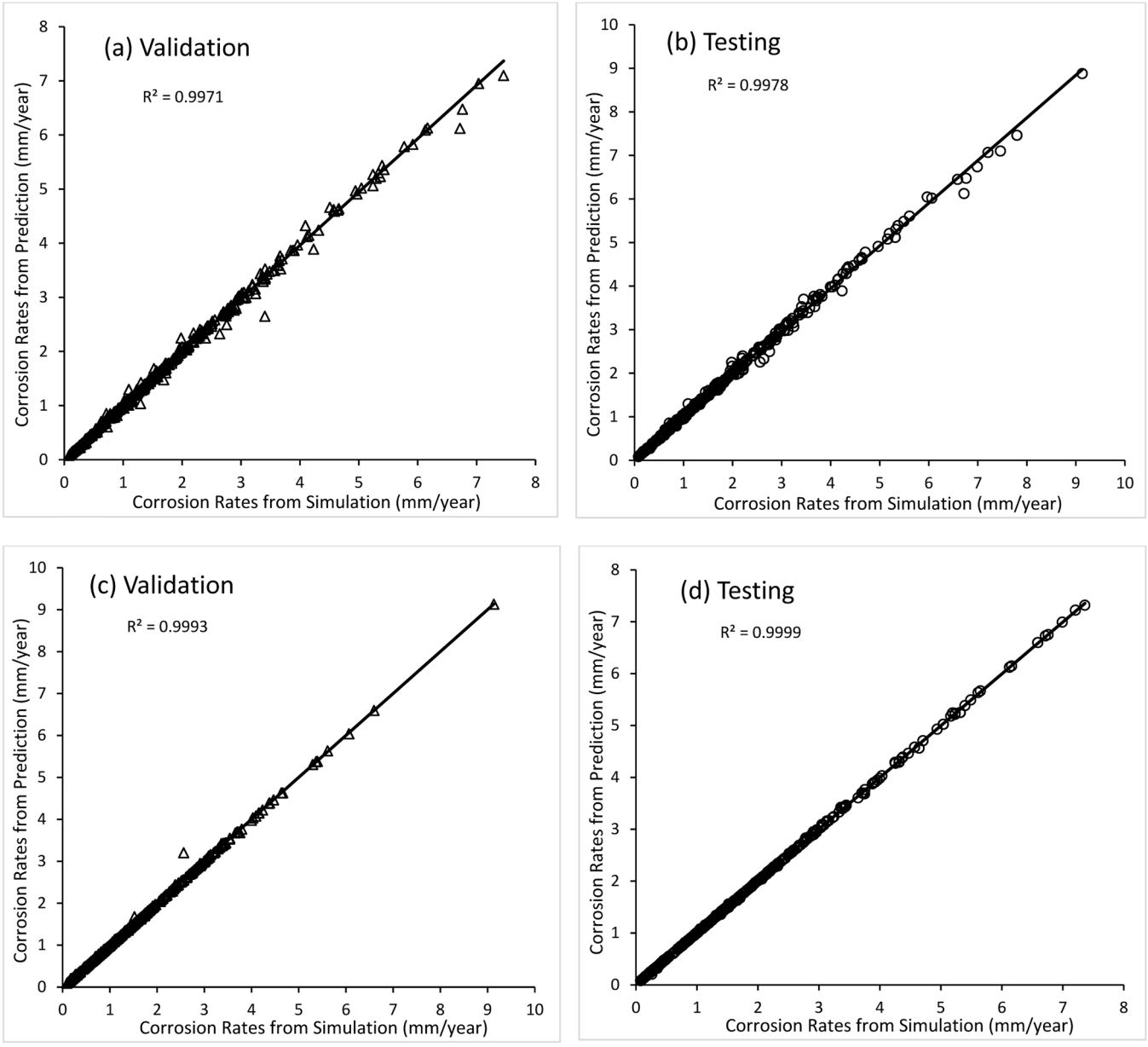


Fig. 18. Predicted *CR* vs. simulation *CR* values of natural gas pipelines. (a) RF - *CR* model for validation, (b) RF - *CR* model for testing, (c) CatBoost - *CR* model for validation, and (d) CatBoost - *CR* model for testing.

Author contribution statement

Jian Fang: Methodology, OLI Simulation, Model development, Model coding, Visualization, Writing – original draft, Writing – review & editing. **Xiao Cheng:** Visualization, Writing – review & editing. **Huilong Gai:** Validation models, Visualization, Writing – review & editing. **Sidney Lin:** Supervision, Conceptualization, Writing – review & editing. **Helen H. Lou:** Supervision, Methodology, Conceptualization, Writing – review & editing, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This research was supported by Lamar University Center for Midstream Management and Science (CMMS) and Lamar University College of Engineering. The authors also wish to thank Dr. Yueqing Li, Dr. Xianchang Li, and Dr. Berna Eren Tokgoz for helpful conversations and supports.

References

Al-Fakih, A.M., Algamal, Z.Y., Lee, M.H., Abdallah, H.H., Maarof, H., Aziz, M, 2016. Quantitative structure–activity relationship model for prediction study of corrosion inhibition efficiency using two-stage sparse multiple linear regression. J. Chemom. 30, 361–368. https://doi.org/10.1002/cem.2800.

Anderko, A., Young, R.D., 1999. Simulation of CO2 /H2S corrosion using thermodynamic and electrochemical models. In: Nace Corrosion 99. San Antonio, Texas. Nace Corrosion, pp. 1–19.
Bai, H., Wang, Y., Ma, Y., Zhang, Q., Zhang, N., 2018. Effect of CO2 partial pressure on the corrosion behavior of J55 carbon steel in 30% crude oil/brine mixture. Materials (Basel) 11, 1–15. https://doi.org/10.3390/ma11091765.
Baker, M. Jr. (2008). Pipeline Corrosion: final Report.
Ben Seghier, M.E.A., Höche, D., Zheludkevich, M., 2022. Prediction of the internal corrosion rate for oil and gas pipeline: Implementation of ensemble learning techniques. J. Nat. Gas Sci. Eng. 99 https://doi.org/10.1016/j.jngse.2022.104425.
Biau, G., Scornet, E., 2016. A random forest guided tour. Test 25, 197–227. https://doi.org/10.1007/s11749-016-0481-7.
bp, 2022. Crude Assays. https://www.bp.com/en/global/bp-trading-and-shipping/documents-and-downloads/technical-downloads/crudes-assays.html.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1201/9780429469275-8.
Brownlee, J., 2016. Feature importance and feature selection with XGBoost in Python. https://machinelearningmastery.com/feature-importance-and-feature-selection-with-xgboost-in-python/.
Brownlee, J., 2020. How to calculate feature importance with Python. https://machinelearningmastery.com/calculate-feature-importance-with-python/.
Chern-Tong, H., Aziz, I.B.A, 2016. A corrosion prediction model for oil and gas pipeline using CMARPGA. In: 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), pp. 403–407. https://doi.org/10.1109/ICCOINS.2016.7783249.
Chrysafis, I., Mallinis, G., Gitas, I., Tsakiri-Strati, M., 2017. Estimating Mediterranean forest parameters using multi seasonal Landsat 8 OLI imagery and an ensemble learning method. Remote Sens. Environ. 199, 154–166. https://doi.org/10.1016/j.rse.2017.07.018.
Collier, J., Papavinasam, S., Li, J., Shi, C., Liu, P., Podlesny, M., 2012. Comparison of corrosivity of crude oils using rotating cage method. In: the Symposium on Crude Oil Corrosivity (Toronto, Canada: NACE 2012 Northern Area Eastern Conference). Available at: https://www.nrcan.gc.ca/maps-tools-and-publications/publications/minerals-mining-publications/comparison-corrosivity-crude-oils-using-rotating-cage-method/8246.
Cox, G.L., Roetheli, B.E., 1931. Effect of oxygen concentration on corrosion rates of steel and composition of corrosion products formed in oxygenated water. Ind. Eng. Chem. 23, 1012–1016. https://doi.org/10.1021/ie50261a011.
Du, J., Liao, K., Li, J., 2005. Prediction the pitting depth growth in oil & gas pipelines with the times series analysis method. Xin Jiang Oil Gas Jiang Oil Gas 1, 80–82.
El-Abbasy, M.S., Senouci, A., Zayed, T., Mirahadi, F., Parvizsedghy, L., 2014. Artificial neural network models for predicting condition of offshore oil and gas pipelines. Autom. Constr. 45, 50–65. https://doi.org/10.1016/j.autcon.2014.05.003.
Fang, H., Brown, B., Technology, M., 2010. High salt concentration effects on CO2 corrosion and H2s corrosion. In: NACE - International Corrosion 2010 Conference Series. NACE International, pp. 1–29.
Gomez, C., Mangeas, M., Petit, M., Corbane, C., Hamon, P., Hamon, S., et al., 2010. Use of high-resolution satellite imagery in an integrated model to predict the distribution of shade coffee tree hybrid zones. Remote Sens. Environ. 114, 2731–2744. https://doi.org/10.1016/j.rse.2010.06.007.
Grabner Instruments, 2019. Vapor pressure in the transport, storage, and bending of crude oil. https://www.azom.com/article.aspx?ArticleID=18157.
Gupta, A., 2020. Feature selection techniques in machine learning. https://www.analyticsvidhya.com/blog/2020/10/feature-selection-techniques-in-machine-learning/.
Haghighatlari, M., Hachmann, J., 2019. Advances of machine learning in molecular modeling and simulation. Curr. Opin. Chem. Eng. 23, 51–57. https://doi.org/10.1016/j.coche.2019.02.009.
Hancock, J.T., Khoshgoftaar, T.M., 2020. CatBoost for big data: an interdisciplinary review. J. Big Data 7, 1–45. https://doi.org/10.1186/s40537-020-00369-8.
Hatami, S., Ghaderi-Ardakani, A., Niknejad-Khomami, M., Karimi-Malekabadi, F., Rasaei, M.R., Mohammadi, A.H., 2016. On the prediction of CO_2 corrosion in petroleum industry. J. Supercrit. Fluids 117, 108–112. https://doi.org/10.1016/j.supflu.2016.05.047.
Ismail, A., Adan, N.H, 2014. Effect of oxygen concentration on corrosion rate of carbon steel in seawater. Am. J. Eng. Res. 03, 64–67.
Javadli, R., Klerk, A.de, 2012. Desulfurization of heavy oil. Appl. Petrochem. Res. 1, 3–19. https://doi.org/10.1007/s13203-012-0006-6.
Jechura, J. (2018). Crude Oil Assay – WTI (from OGJ article). http://inside.mines.edu/~jjechura/Refining/02_Assay_WTI_OGJ.pdf.
Kermani, M.B., Morshed, A., 2003. Carbon dioxide corrosion in oil and gas production—a compendium. Corrosion 59, 659–683. https://doi.org/10.5006/1.3277596.
Knowino, 2010. Crude Oil Desalter. https://www.theochem.ru.nl/~pwormer/Knowino/knowino.org/wiki/Crude_oil_desalter.html.
Legat, A., 2007. Monitoring of steel corrosion in concrete by electrode arrays and electrical resistance probes. Electrochim. Acta 52, 7590–7598. https://doi.org/10.1016/j.electacta.2007.06.060.
Liu, J., Wang, H., Yuan, Z., 2012. Forecast model for inner corrosion rate of oil pipeline based on PSO-SVM. Int. J. Simul. Process Model. 7, 74–80. https://doi.org/10.1504/IJSPM.2012.047863.
Liu, Y., Zhang, B., Zhang, Y., Ma, L., Yang, P, 2016. Electrochemical polarization study on crude oil pipeline corrosion by the produced water with high salinity. Eng. Fail. Anal. 60, 307–315. https://doi.org/10.1016/j.engfailanal.2015.11.049.
Liao, K., Yao, Q., Wu, X., Jia, W., 2012. A numerical corrosion rate prediction method for direct assessment of wet gas gathering pipelines internal corrosion. Energies 5, 3892–3907. https://doi.org/10.3390/en5103892.
Liu, Z.G., Mu, Z.T., 2011. Research on aircraft LY12CZ aluminum alloy corrosion damage prediction based on ARIMA model. Adv. Mater. Res. 308–310, 1016–1022. https://doi.org/10.4028/www.scientific.net/amr.308-310.1016.
López, L., Lo Mónaco, S., 2017. Vanadium, nickel and sulfur in crude oils and source rocks and their relationship with biomarkers: implications for the origin of crude oils in Venezuelan basins. Org. Geochem. 104, 53–68. https://doi.org/10.1016/j.orggeochem.2016.11.007.
Louppe, G. (2014). Understanding Random Forests: from Theory to Practice. doi: 10.13140/2.1.1570.5928.
Luo, M., Wang, Y., Xie, Y., Zhou, L., Qiao, J., Qiu, S., et al., 2021. Combination of feature selection and CatBoost for prediction: the first application to the estimation of aboveground biomass. Forests 12, 1–22. https://doi.org/10.3390/f12020216.
Lynch, J., Sohn, H., Wang, M., 2014. Sensor Technologies for Civil Infrastructures. Elsevier. Woodhead Publishing. ISBN:9781782422440.
Mohammed, Q.Y., 2017. Determination of salt content in crude oil, turbine oil and some refinery products volumetrically. J. Chem. Pharm. Sci. 10, 34–37. Available at: www.jchps.com.
Mosavat, N., Abedini, A., Torabi, F., 2014. Phase Behaviour of CO2-Brine and CO2-oil systems for CO2 storage and enhanced oil recovery: experimental studies. Energy Procedia 63, 5631–5645. https://doi.org/10.1016/j.egypro.2014.11.596.
OLI Studio, 2011. OLI Product Description. http://www.corrscience.com/wp-content/uploads/2011/03/Corrosion-Analyzer-2011.pdf.
Ossai, C.I., 2019. A data-driven machine learning approach for corrosion risk assessment—a comparative study. Big Data Cogn. Comput. 3, 1–22. https://doi.org/10.3390/bdcc3020028.
Pasban, A.A., Nonahal, B., 2018. The Investigation of precision of analytical methods for determination of salt content in Iranian crude oils. Anal. Methods Environ. Chem. J. 1, 23–28. https://doi.org/10.24200/amecj.v1.i01.33.
Peng, S., Zhang, Z., Liu, E., Liu, W., Qiao, W., 2021. A new hybrid algorithm model for prediction of internal corrosion rate of multiphase pipeline. J. Nat. Gas Sci. Eng. 85, 103716 https://doi.org/10.1016/j.jngse.2020.103716.
PetroWiKi, 2015. Desalting. https://petrowiki.spe.org/Desalting.
Poe, W.A., Mokhatab, S., 2017. Chapter 1: Introduction to Natural Gas Processing Plants. Modeling, Control, and Optimization of Natural Gas Processing Plants. https://doi.org/10.1016/b978-0-12-802961-9.00001-2. ISBN 978012802961-9.
Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., and Gulin, A. (2019). CatBoost : unbiased boosting with categorical features.
Royer, R.A., Unz, R.F., 2002. Use of electrical resistance probes for studying microbiologically influences corrosion. Corrosion 58, 863–870. https://doi.org/10.5006/1.3287671.
Sardisco, J.B., Wright, W.B., Greco, E.C., 1963. Corrosion of iron in an H2S-CO2-H2O system: corrosion film properties on pure iron. Corrosion 19, 354t–359t.
Scornet, E., 2015. Random Forests and Kernel Methods. Cornell University arXiv: 1502.03836. https://doi.org/10.1109/TIT.2016.2514489.
Singh, R., 2017. Chapter 5: Hazards and Threats to a Pipeline System. Pipeline Integrity Handbook, 2nd Edn. https://doi.org/10.1016/b978-0-12-813045-2.00005-3 ISBN: 978-0-12-813045-2.
Song, F.M., 2010. A comprehensive model for predicting CO2 corrosion rate in oil and gas production and transportation systems. Electrochim. Acta 55, 689–700. https://doi.org/10.1016/j.electacta.2009.07.087.
Soomro, A.A., Mokhtar, A.A., Kurnia, J.C., Lashari, N., Lu, H., Sambo, C., 2022. Integrity assessment of corroded oil and gas pipelines using machine learning: a systematic review. Eng. Fail. Anal. 131, 105810 https://doi.org/10.1016/j.engfailanal.2021.105810.
Touali, Y.A. (2013). London South Bank University MSc Dissertation Petroleum Engineering.
Trench, C.J., 2001. How Pipelines Make the Oil Market Work—their Networks, Operation and Regulation. Allegro Energy Gr, pp. 1–22.
Velázquez, J.C., Caleyo, F., Valor, A., Hallen, J.M., 2009. Predictive model for pitting corrosion in buried oil and gas pipelines. Corrosion 65, 332–342. https://doi.org/10.5006/1.3319138.
WANG, Q., 2008. Generation mechanism and control measures for H2S in oil wells, Liaohe Oilfield. Pet. Explor. Dev. 35, 349–354. https://doi.org/10.1016/S1876-3804(08)60082-8.
Wang, Z.M., Song, G.L., Zhang, J., 2019. Corrosion control in CO2 enhanced oil recovery from a perspective of multiphase fluids. Front. Mater. 6, 1–16. https://doi.org/10.3389/fmats.2019.00272.
Winkler, D.A., 2020. Role of artificial intelligence and machine learning in nanosafety. Small 16, 1–13. https://doi.org/10.1002/smll.202001883.
Xinru, X., Jingyi, Y., Ying, J., Jinsheng, G, 2006. Effects of process conditions on desalting and demetalization of crude oil. Pet. Sci. Technol. 24, 1307–1321. https://doi.org/10.1081/LFT-200056651.
Xu, X., Yang, J., Gao, J., 2006. Effects of demulsifier structure on desalting efficiency of crude oils. Pet. Sci. Technol. 24, 673–688. https://doi.org/10.1081/LFT-200041172.
Xu, Y., Huang, Y., Wang, X., Lin, X, 2016. Experimental Study on Pipeline Internal Corrosion Based on a New Kind of Electrical Resistance Sensor. Sens. Actuators B: Chem. 224, 37–47. https://doi.org/10.1016/j.snb.2015.10.030.
Zhang, G., Lu, M., Chai, C., Wu, Y., 2006. Effect of HCO3- concentration on CO2 corrosion in oil and gas fields. J. Univ. Sci. Technol. Beijing Miner. Metall. Mater. 13, 44–49. https://doi.org/10.1016/S1005-8850(06)60012-1.
Zhang, H., Lan, H.Q., 2017. A review of internal corrosion mechanism and experimental study for pipelines based on multiphase flow. Corros. Rev. 35, 425–444. https://doi.org/10.1515/corrrev-2017-0064.