A failure prediction model for corrosion in gas transmission pipelines

Kimiya Zakikhani

Graduate Student, Dept. of Building, Civil and Environmental Engineering, Concordia Univ., Montréal, QC, Canada
E-mail: kimiya.zakikhani@gmail.com

Fuzhan Nasiri

Associate professor, Department of Building, Civil and Environmental Engineering (BCEE), Concordia University, Montreal, QC, Canada
H3G 1M8, email: fuzhan.nasiri@concordia.ca

Tarek Zayed

Professor, Department of Building and Real Estate (BRE), The Hong Kong Polytechnic University, Hong Kong
E-mail: tarek.zayed@polyu.edu.hk

A failure prediction model for corrosion in gas transmission pipelines

Transmission pipelines comprise a major part of a gas network, conveying natural gas within jurisdictions and across international boundaries. In the United States, more than 10,000 failure incidents have been reported in gas transmission pipelines in a 20-year period from 1996 to 2016 leading to a cumulative property damage of more than $748 Million. Among different failure sources, corrosion is ranked as the most frequent one, corresponding to approximately a quarter of total failures. Though in-line inspection is counted as the most frequently applied corrosion monitoring technique for oil and gas pipelines, it imposes considerable costs due to the necessity of implementing frequent inspections using smart devices. For this reason, several failure prediction models have been developed to estimate the corrosion failure. However, the majorities of these prediction models rely solely on experimental tests or limited historical records which undermine the extent of their applicability and ignore pipeline environmental and geographical circumstances. The objective of this research is to develop failure prediction models for external corrosion in underground gas transmission pipelines by considering both conventional and environmental/geographical variables. For this objective, multiple regression analysis was performed on the accessible historical data reported for gas transmission pipelines. Two main climate regions of Great Plains and South East in the US were selected, and their corresponding failure prediction models were developed. Such development was based on a step by step procedure analyzing different scenarios. Considering diagnostic measures, null hypothesis and residual analysis, scenario 3 was selected as satisfactory. The validation tests of the developed models present a root mean square error (RMSE) of 0.04 and 0.07 and R-Sq of 0.93 and 0.75, respectively. The results of this research can be applied in maintenance planning of gas transmission pipeline to estimate the critical time in which a pipeline may encounter external corrosion failure, and to accordingly schedule the maintenance activities.

Keywords: Pipelines, gas transmission, external corrosion, time of failure, multiple regression, petroleum

1. Introduction

Transmission pipelines are considered as a major part of gas network, conveying invaluable natural gas across countries and provinces. These facilities convey the petroleum and gas products from a storage facility or a gathering line to another storage facility, distribution center or a large volume customer (1). According to pipeline and hazardous safety administration (PHMSA) (1), more than 1,700 failures have been reported in gas transmission pipelines since 1984. Among these failures, corrosion, as a time-dependent failure source, is ranked as the most frequent one corresponding to approximately one quarter of the total failures and leading to more than $189 Million of property damages.

In order to ensure pipeline safety, corrosion monitoring of pipeline is performed through different methods of in-line inspection (2). This technique requires frequent assessment of pipeline condition through application of high-tech devices such as magnetic flux and ultrasonic tools to report anomalies such as metal loss, dents and gouges. However, due to the required high frequency of in-line inspections and resolution perquisites, this method is considered as excessively expensive and time consuming. For this reason, in the recent years more focus has been contributed towards development of models which can estimate corrosion failure to avoid performing unnecessary expensive in-line inspections. These models are usually based on the failure pressure models obtained from theory and experimental tests (3-6). These tests require information on design variables of the pipeline and often ignore effective environmental/geographical factors of corrosion failure such as soil temperature.

The objective of this research is to develop failure prediction models to estimate the time of corrosion of gas transmission pipelines by considering environmental/geographical attributes in addition to the conventional variables. To attain this objective, first the effective parameters on pipeline corrosion failure are identified from a review of state of the arts. Then, the corresponding data are collected with respect to both historical data on pipeline failure as well as climatological data bases. Finally, a multiple regression analysis is employed to exploit the collected data and generate prediction models for two selected climatological regions in the US. This method was applied since regression analysis provides a practical approach, presenting the relationship between response and explanatory variables in form of equations. Such equations can be simply applied in practice and industry. Through this research, pipeline operators can estimate time of corrosion failure by considering both conventional design variables in addition to environmental/geographical conditions of the pipelines. In this research a state-of-the-art review reveals the efforts contributed to oil and gas pipeline failure and highlights the remaining gaps. Data is collected from accessible historical records and automated best subset regression is carried out to estimate time of failure due to corrosion in gas transmission pipelines. Finally, the developed models are tested through null hypothesis procedure and a residual analysis and the models are validated.

2. Background

Different sources can lead to a failure in oil and gas pipelines. These failure sources include manufacturing/material/weld failures, natural force damage, equipment failure, excavation failure, incorrect operation, corrosion and other outside force damage (1,7,8). Among these, corrosion corresponds to approximately one quarter of failures in gas transmission pipelines. Pipeline corrosion is defined as deterioration metal over time due to interaction of pipe with the environment and can be classified as internal, external or stress corrosion cracking (9,10). The effective parameters on corrosion failure are summarized below.

External corrosion: In external corrosion the environment is considered as water or moist soil for onshore pipelines and seawater for offshore pipelines (11). In general, external corrosion can be categorized into three groups including differential cell, microbiologically influenced corrosion and stray current corrosion (12). Differential cell corrosion takes place when parts of a pipe are over-exposed to different oxygen concentrations leading to generation of cells. Different parameters can lead to this type of corrosion including differential aeration cell, soil properties such as PH value, temperature, soil type, moisture content, resistivity, redox potential, cover, galvanic corrosion, surface films and relative size of anodic and cathodic areas (12-14).

Microbiologically influenced corrosion (MIC) corresponds to activity of microorganisms including bacteria, archaea and fungi to promote corrosion by converting the metal oxide to a less protective layer (15). Stray current corrosion corresponds to traveling of stray current along the pipe to other areas and its returning to the power source. In this type of corrosion, the extent of metal loss is proportional to the current leaking from the pipe. Examples of direct current include foreign pipelines not properly bonded to the intended pipeline in direct current sources such as rail roads and mining operations (12,16).

Internal corrosion: This corrosion type takes place when an electrolyte is available and completes the corrosion cell. Natural gas is prone to internal corrosion due to presence of contaminants such as water, carbon dioxide and hydrogen sulfide and the possibility of reaction with condensed water (11).

Stress corrosion cracking (SCC): SCC leads to cracking of material due to combined actions of corrosion and tensile stresses (7). In this type of corrosion failure, colonies of longitudinal surface cracks form in the pipe and link up to join long-shallow flaws (12).

Due to the inefficiency associated with performing in-line inspection on oil and gas pipelines, more attention has recently been targeted towards developing corrosion failure prediction models. These models can predict various corrosion failure parameters such as probability and rate of failure, cause and risk of failure as well as consequence of corrosion failure or its rate. The developed models may be either based on available in-line inspection data, empirical equations, numerical finite element modeling or reported historical data on pipeline corrosion failures (17). As one of these parameters, probability of corrosion failure was estimated by Li et al. (18) and Witek (6) by referring to empirical equations for burst pressure and application of a Monte Carlo simulation. Similarly, wen et al. (19) estimated the probability of failure using Monte Carlo simulation from limit state functions for several types of failure including corrosion, equipment impact, and weld defect. In addition, Dundulis et al. (20) integrated Bayesian method with hoop stress values obtained from Monte Carlo simulation and finite element analysis to estimate the probability of failure in gas pipelines due to corrosion.

Regarding to reliability analysis, compared to different failure sources in petroleum pipelines, corrosion is the most highlighted. Through iterative numerical analysis, Weiguo et al. (21) predicted remaining life of buried gas pipelines under corrosion and cyclic loads. Similarly, Kucheryavyi et al. (22) performed reliability assessment towards corrosion failure in a defective gas pipeline based on mechanical equations on pipeline strength. Through application of a Monte Carlo simulation and principals of reliability analysis, Teixeira et al. (23) and Ossai et al. (24) proposed two methodologies to assess reliability of oil and gas pipelines. Teixeira et al. (23)

performed a first-order reliability analysis on burst pressure results from numerical and experimental analysis. By considering maximum corrosion pit depth, Ossai et al. (24) applied a Monte Carlo simulation in conjunction with Weibull probability to assess pipeline reliability. In addition to reliability and probability analysis, Liao et al. (25), Caleyo et al. (26) and Papavinasam et al. (27) predicted corrosion rate parameter. This failure parameter was obtained through application of a back propagation neural network, Monte Carlo simulation and empirical equations respectively.

Reviewing the state-of-the-art reveals that different design and environmental parameters are effective on external corrosion failure (fig. 1). However, the majority of these studies are merely based on empirical equations. Such methods are a function of design parameters without considering environmental factors such as the impact of soil temperature and geographical location on corrosion failure. In addition, the applicability of these models may be limited since they are solely based on either experimental tests or in-line inspections without considering historical failure records reflecting geo-environmental impacts.

2. Methodology

This research aims at developing a corrosion failure prediction model for gas transmission pipelines through consideration of both conventional and geo-environmental parameters which are often ignored. This objective is fulfilled through application of multiple linear regression on the highlighted variables. For this objective, the accessible experimental equations were reviewed to extract the conventional corrosion failure parameters. Then, the corresponding failure and geo-environmental data were collected. Finally, through a step by step procedure, the model with the most satisfactory diagnostic measures and validation results is determined as the final model. The proposed research methodology is presented in figure 2.

Different corrosion assessment methods have been developed based on experimental testing and numerical studies such as ASME B31G (28), modified B31G (29), SHELL 92 (30), and SAFE (31). These methods are based on estimates for burst pressure leading to pipeline failure. As one of these assessment methods SHELL 92 is based on a rectangular shape assumption for corrosion defect and empirical factors to consider nonlinearity of pipeline material (32). In addition, application of this method leads to the highest probability of failure and expected number of repair actions. According to this method, the burst pressure of a pipeline is considered as a function of pipeline design parameters (diameter, thickness, ultimate tensile strength) and Folias empirical factor. Equations 1 to 5 are derived from the literature and correspond to SHELL 92 corrosion assessment method (18,33). According to these equations, failure pressure of a pipeline is a function of the hoop stress and corrosion defect size, assuming a rectangular defect shape.

$$p_f = \frac{1.8UTS.THK}{DIAM}\left(\frac{1-\frac{d(T)}{t}}{1-\frac{d(t)}{t}M^{-1}}\right) \tag{1}$$

$$M = \sqrt{1 + 0.805\,\frac{L(T)^2}{DIAM.THK}} \tag{2}$$

Assuming a quasi-steady corrosion process,

$$d(T) = d_0 + V_r(T - T_0) \tag{3}$$

$$L(T) = L_0 + V_a(T - T_0) \tag{4}$$

Finally, Eq. 1 is converted to the following equation as,

$$p_f = \frac{1.8UTS.THK}{DIAM}\left(\frac{1-\frac{d_0+V_r(T-T_0)}{t}}{1-\frac{d_0+V_r(T-T_0)}{t}M^{-1}}\right) \tag{5}$$

In the equations above *UTS*, p_f, *DIAM*, *THK*, *d(T)*, *T*, *M*, *L(T)*, V_r, V_a,d_0,L_0 and T_0 correspond to ultimate tensile strength, failure pressure, pipe diameter, wall thickness, time-dependent depth of the defect, elapsed time, Folias (bulging) factor, the axial length of the defect projected on longitudinal axis, the radial corrosion rate, the axial corrosion rate, the measured depth and length of a defect at time T0 (the time of last inspection) respectively. Therefore, pipeline corrosion failure occurs met when the operating pressure goes beyond the failure pressure.

3.1. Multiple regression analysis

Predictive models used in data mining are divided into two main categories, i.e. classification (discrimination) and predictive (or regression). The objective of both operations is to estimate the value of the dependent (target, response, explained or endogenous) variable as a function of a certain number of other variables (explanatory, control or exogenous variables) (34). For predictive models, if the dependent variable is quantitative or qualitative, the technique will be categorized as prediction or classification respectively. Regression analysis is a statistical method that utilizes the relationship between two or more variables to predict a dependent variable from independent ones (7). In the simplest form, a simple linear regression is expressed as Eq. 7 where a continuous dependent variable Y relates to a continuous independent variable X (35).

$$y_i = \alpha + \beta x_i + \varepsilon_i \tag{6}$$

It is assumed that the values of x_1, x_2, …, x_n of X are controlled and the corresponding values of $y_1, y_2, \ldots, y_n$ of Y are observed. In this equation, y_i is the value of the response variable in the "ith" trial, α and β are the regression parameters, x_i is the value of the predictor variable in the ith trial and ε_i is the random error. In a more general form of multiple regression, a multiple linear regression is applied to the case of several independent variables of X_i to obtain the dependent variable of Y as (34);

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon \tag{7}$$

3.2.Data collection

In order to develop predictive models for external corrosion of gas transmission pipelines, incident data reported by PHMSA -pipeline and hazardous materials safety administration (1)- and climatological data reported by national climatic data center (36) were collected. Among the accessible failure records for oil and gas pipelines, PHMSA is considered as one of the most comprehensive failure record databases for oil and gas pipelines that covers incidents in US since 1970. According to PHMSA, an incident is defined as an event resulting to leakage and has led to one or more of the following criteria:

- “A death, or personal injury necessitating in-patient hospitalization; or”
- “Property damage, including the product loss cost of 50,000 USD or more”,
- “An event that is significant even though if it does not meet the above criteria” (1).

Depending on the type of the pipeline and the material transferred, the database comprises four categories of petroleum pipelines including gas transmission (GT), gas distribution (GD), hazardous liquid (HL) and liquefied natural gas (LNG). Though PHMSA database is considered as a comprehensive database on petroleum pipelines failures, the data reported in different time intervals for each facility type do not follow an identical presentation. As an example, for GT pipelines, the records reported for 1970-1984 do not provide information on the thickness of pipelines while other intervals do. This problem leads to devoting more time into exploring and interpreting the reported data.

For more efficient identification of the reported parameters, they are categorized into three main categories including system, consequence and incident parameters as presented in Figure 3. System parameters correspond to the physical properties of the system (e.g., installation year, depth of cover), design characteristics (e.g. diameter, thickness, maximum operating pressure, yield strength) and the environmental attributes in which the system is located (e.g. soil type, soil temperature, class). Consequence parameters correspond to the consequences that take place after the failure including safety damages (injuries and death) and monetary damages (e.g. loss, operator and total damages). Finally, incident parameters correspond to the ones that are related to the incident itself rather than what happens afterward (e.g. date, pressure, the part that failed source of failure etc.).

Since the concentration of this research is on development of predictive models for time of failure in underground gas transmission pipelines, the deployed parameters correspond to system parameters. In GT pipelines corrosion failure is a time-dependent failure resource and is considered as the most frequent type of failure. This failure source corresponds to approximately a quarter of total failures records (Fig. 4).

The collected data were then compared with the identified efficient factors in experimental equations for corrosion failure (equations 1 to 5) and the important parameters were extracted accordingly. These parameters include pipe wall thickness (*THK*), diameter (*DIAM*), and pressure at time of incident (*INCP*), maximum allowable pressure (*MAOP*), specified minimum yield strength (*SMYS*) and depth of cover (*cov*) as presented in table 1.

Table 1. Variable thresholds

Variable	Minimum	Maximum
TIME (years)	4.00	85
TEMP (cent.)	35.00	86
INCP (psi)	35.0	1200
MAOP(psi)	120.0	1337
DIAM (in)	3.50	36

THK (in)	0.1300	0.5
SMYS (psi)	1050	1200
COV (in)	12.00	99.6

In addition, through referring to other databases, two geo-environmental parameters were extracted from national climatic data center (36). As one of soil corrosively measures, average monthly soil temperatures for each failure record were extracted according to the location and the time of the incident. In addition, to take account of possible climatological effects on time of failure, the failure records in US were classified into eight different regions including Northeast, Southeast, Midwest, Great Plains, Northwest, Southwest, Alaska, and Hawai‘i/Pacific Islands as presented in figure 5. Such classification is based on historical climate trends and climate scenarios of the future. The climate factors that are considered in this analysis include regional floods, thunderstorms, drought, heat waves, water levels and winter storms. On the other hand, the considered climate trends include temperature, precipitation, extreme heat and cold, extreme precipitation, wind, freeze-free season, snowfall, water levels, ice cover and humidity (36).

3.Development of time-based failure prediction models

After selection of the corresponding important variables in failure prediction of oil and gas pipelines, data preparation and processing were performed prior to model development. For this objective, first, the aberrant values are detected and considered as missing. Aberrant value is an enormous value resulting from an incorrect measurement or input error. In the next step, to handle the missing values through SPSS software, an automatic multiple imputation procedure was performed to interpret the patterns of missingness and to replace missing values with plausible estimates according to variable constraints. The automatic imputation mode selects the most suitable imputation method (monotone or fully conditional specification) based on data characterization. Through imputation, the software provides a number of imputed datasets for different imputation iterations which are then integrated with mean values to be replaced in the dataset for missing values.

After collecting the corresponding data on external corrosion failure of gas transmission pipelines, the procedures to develop a time of failure prediction model are pursued. For this objective, the selected effective parameters on external corrosion of petroleum pipelines obtained from the literature review and the databases (PHMSA and NCDC) are fed into the models. Since the objective of the prediction model is to estimate the time of failure, the subtraction between pipeline installation and incident date is considered as the output (time of failure). Considering the empirical/ experimental assessment methods discussed earlier (Eqs. 1 to 5), corrosion failure is considered as a function of pipe wall thickness, diameter, pressure at time of incident, maximum allowable pressure and tensile strength (a function of specified minimum yield strength). The values of these parameters were collected from PHMSA database (1). Soil temperature and depth of cover are also considered as input variables (soil property parameters) obtained from the literature (12-14). These data are extracted from national climatic and PHMSA data center. To investigate whether corrosion failure time prediction can be attributed to climatological measures, the climate regions obtained from national climatic data center for each data record is considered as another explanatory variable. In summary, these variables are presented in Table 2.

Table 2 . Model variables and their descriptions

Variable	Depth of cover (in)	Soil temp. (cent.)	Incident pressure (psi)	Max operating pressure (psi)	Diameter (in)	Thickness (in)	Min yield strength (psi)	Climate region	Failure time (year)
Type	quan.	quan.	quan.	quan.	quan.	quan.	quan.	qual.	quan.

Abbrev.	COV	TEMP	INCP	MAOP	DIAM	THK	SMYS	REG	TIME
Prediction role	exp.	exp.	exp.	exp.	exp.	exp.	exp.	exp.	resp.
Minimum	12	35	35	120	3.50	0.13	1050	-	4
Maximum	99.6	86	1200	1337	36	0.5	1200	-	85

4. Results analysis

After selection of explanatory and response variables, multiple linear regression analysis is deployed to generate the prediction model. In this research, these models are based on steel, underground offshore pipelines with cathodic protection and coating. It should be noted that the process of model development is not a straightforward process and requires numerous trials and error. For the sake of clarification, three different scenarios are highlighted to present the model development procedure. Such scenarios are based upon classification of failure records and consideration of empirical ratios in the corresponding models.

Prior to model development for each scenario, data preparation was performed through detecting the outliers, normalizing input and output values and processing data through mutual imputation as previously discussed. For each scenario, data were randomly divided into training and validation dataset consisting of 80% and 20% of the total data set respectively. To develop a regression-based model for each scenario, the regression model is first developed for the training dataset through an automated best-subset procedure using Minitab software. This method was selected since this it facilitates modeling and exploring datasets in which many potential predictors are available. In this method, according to Mallow's Cp, the model that presents the testing database more efficiently is selected and its corresponding diagnostic measures are extracted. In best-subset analysis, the best subset is selected according to its Mallow's Cp value which shall be close to the number of variables plus one. If the diagnostic measures of the selected subset are not satisfactory, the subsequent scenario is implemented until satisfactory diagnostic measures are obtained. As diagnostic measures for the training phase, R-Sq, R-Sq adjusted, mean square error and residual analysis are examined. In addition, the model is also tested for the validation dataset and accordingly the model that best describes the relation between input and output variables is selected.

4.1 Scenario1. Best subset or stepwise regression on all failure records

As the first attempt to predict the time of failure for gas transmission pipelines, the data was entered into the model as is, without any prior filtering and classification and the analysis was performed through considering linear and nonlinear terms (with second order and interactions) in the model through stepwise analysis respectively. The explanatory variables are either quantitative or qualitative. The qualitative data is transformed into quantitative values to facilitate their input in the model. Between different subsets presented the one with four variables is selected since it has a Mallow's Cp value of 4.9 which is close to 5 (as highlighted in table 3). However, the proficiency of this scenario in prediction of time of failure is rejected due to weak diagnostic measures of the training dataset for both linear and nonlinear analysis R-sq and R-sq (adj) of 16.7% and 15.6% respectively as presented in table 3). Therefore, the analysis was prolonged to scenario 2 as follows.

Table 3. Model development based on all records

Vars	R-Sq (%)	R-Sq (adj) (%)	Mallows Cp	S	TEMP	INCP	MAOP	DIAM	THK	SMYS	REG
1	12.6	12.4	14.5	0.16			X				
1	5.4	5.1	42.4	0.17		X					
2	14.8	14.3	8.2	0.16			X				X
2	14.0	13.5	11.3	0.16			X		X		
3	15.6	14.8	6.9	0.16			X		X		X
3	15.6	14.8	7.1	0.16		X	X				X
4	**16.7**	**15.6**	**4.9**	**0.16**		**X**	**X**		**X**		**X**
4	16.0	15.0	7.5	0.16		X	X	X	X		
5	17.4	16.1	4.1	0.16		X	X	X	X		X
5	16.8	15.5	6.3	0.16		X	X		X	X	X
6	17.4	15.9	6.0	0.16	X	X	X	X	X		X
6	17.4	15.9	6.1	0.16		X	X	X	X	X	X
7	17.4	15.6	8.0	0.16	X	X	X	X	X	X	X

4.2 Scenario2. Geographical classification of data

For scenario 2, data was classified according to the geographical variable and best-subset regression was performed for each regional dataset. According to the diagnostic measures obtained from these analyses, it is concluded that such classification improves prediction model efficiency. Due to limited number of records for some regions, two test data groups were generated for model implementation including,

- Underground pipelines located at Great Plains region with external corrosion failures,
- Underground pipelines located South-East region with external corrosion failures.

Best subset linear multiple regression analyses were performed on these data classifications and the subset which can best describe the output was selected (as highlighted in Tables 4 and 5). The results show that the diagnostic measure of R-sq for training data set has improved compared to the first scenario (41% and 60% versus 15%). However, the results are not yet satisfactory. Therefore, another scenario is considered to verify the results.

Table 4. Best subset model development based for Great Plains region (Scen. 2)

Vars	R-Sq (%)	R-Sq (adj) (%)	Mallows Cp	S	TEMP	INCP	MAOP	DIAM	THK	SMYS	COV
1	32.2	31.1	10.9	0.17			X				
1	19.5	18.3	25.1	0.19							X
2	43.1	41.3	0.7	0.16			X				X
2	34.3	32.3	10.6	0.17		X	X				
3	44.9	42.4	0.6	0.16			X		X		X
3	44.2	41.6	1.5	0.16		X	X				X
4	45.4	42	2.1	0.16			X	X	X		X
4	45.4	42	2.2	0.16		X	X		X		X
5	**45.5**	**41.2**	**4**	**0.16**		**X**	**X**	**X**	**X**		**X**
5	45.4	41.1	4.1	0.16	X		X	X	X		X
6	45.5	40.2	6	0.16	X	X	X	X	X		X
6	45.5	40.2	6	0.16		X	X	X	X	X	X
7	45.5	39.3	8	0.16	X	X	X	X	X	X	X

Table 5. Best subset model development for South East region (Scen. 2)

Vars	R-Sq (%)	R-Sq (adj) (%)	Mallows Cp	S	TEMP	INCP	MAOP	DIAM	THK	SMYS	COV
1	24.9	23.3	48.4	0.19			X				
1	19.9	18.2	54.7	0.19							X
2	49.3	47.2	19.4	0.15			X				X
2	41.4	38.9	29.4	0.17		X					X
3	54.1	51.2	15.2	0.15			X		X		X
3	53.9	50.9	15.6	0.15			X			X	X

4	61.8	58.5	7.5	0.14			X		X	X	X
4	60.6	57.2	9	0.14			X	X		X	X
5	**64.5**	**60.5**	**6.1**	**0.13**			**X**	**X**	**X**	**X**	**X**
5	63.8	59.7	7	0.138		X	X		X	X	X
6	66	61.4	6.1	0.13		X	X	X	X	X	X
6	64.8	60	7.7	0.13	X		X	X	X	X	X
7	66.1	60.6	8	0.13	X	X	X	X	X	X	X

4.3 Scenario 3. Parametric ratios from empirical equations

Though the diagnostic measures obtained from scenario 2 were improved compared to the previous, further effort was dedicated to improving diagnostic measures of the training dataset model. The results obtained from multiple linear regression using nonlinear terms proved that in general, adding nonlinear terms improves model efficiency by improving the diagnostic measures. In this step other terms were added to the model by going through the experimental equations (equations 1 to 5). For this objective, linear variables, their corresponding second order terms and, nonlinear terms and their corresponding second order terms were fed into the model. These terms are presented in Table 6 which are based on the first term of equation 1 and by assuming the ratio of incident pressure and specified minimum yield strength with maximum operation pressure as new variables.

Table 6. Nonlinear terms fed into the model

Nonlinear variables entered in the model				
DIAM/THK	MAOP/SMYS	INCP/MAOP	SMYS*THK/DIAM	MAOP*THK/DIAM

After selection of the variables to be fed into the model, an automated best subset regression was performed and according to Mallow's Cp and adjusted R^2, the subset which can best describe the output was selected (as highlighted in tables 7 and 8). Tables 7 and 8 present the results corresponding to performing the automated best subset regression for Great Plains and South East regional classifications. In the next step, the variables of the selected subset were fed into a multiple linear regression analysis to present a mathematical model. Equation 8 corresponds to the model developed for Great Plains region with diagnostic measures of S=0.12, R2= 73% and R^2 (adj)=65%. In addition, equation 9 corresponds to the models developed for South East region respectively with diagnostic measures of S=0.08, R2=89% and R^2 (adj)=84%. Compared to the previous scenarios, these measures have improved considerably. However, these cannot be claimed as final models since the null hypothesis tests and normality shall also be investigated.

Table 7. Best subset model development based for Great Plains region (Scen. 3)

Vars	R-Sq (%)	R-Sq (adj)(%)	Mallows Cp	S	TEMP	INCP	MAOP	DIAM	THK	SMYS	COV	DIAM/THK	MAOP/SMYS	INCP/MAOP	SMYS*THK/DIAM	MAOP*THK/DIAM	$(DIAM/THK)^2$	$(MAOP/SMYS)^2$	$(INCP/MAOP)^2$	$(SMYS*THK/DIAM)2)^2$	$(MAOP*THK/DIAM)^2$	$TEMP^2$	$INCP^2$	$MAOP^2$	$DIAM^2$	THK^2	$SMYS^2$	COV^2
1	26.6	25.2	71.1	0.173																				X				
1	24.1	22.7	75.2	0.176			X																					
2	34.8	32.3	59.5	0.165							X													X				
2	33.5	30.9	61.7	0.167																				X				X
3	40.1	36.5	52.8	0.160							X		X														X	
3	39.5	35.8	53.9	0.161									X														X	X
4	51.6	47.7	35.8	0.145									X							X	X			X				
4	49.2	45.1	39.8	0.149									X			X				X				X				
5	60.4	56.3	23.4	0.133									X		X	X		X						X				
5	57.5	53.1	28.1	0.137									X		X			X			X			X				
6	63.6	58.9	20.1	0.128		X							X		X	X		X						X				
6	63.5	58.9	20.2	0.129							X		X		X	X		X						X				
7	65.9	60.7	18.3	0.126							X		X	X	X	X		X						X				
7	65.8	60.5	18.5	0.126		X					X		X		X	X		X						X				
8	67.3	61.5	18.0	0.124							X		X	X	X	X		X				X		X				

8	67.1	61.2	18.3	0.125	X						X		X	X	X	X		X						X				
9	68.5	62.0	18.0	0.123		X			X		X		X		X	X		X						X		X		
9	68.3	61.9	18.2	0.124		X			X				X		X	X		X						X		X		X
10	70.2	63.3	17.2	0.121		X			X		X		X		X	X		X				X		X		X		
10	70.0	63.0	17.5	0.122		X			X				X		X	X		X				X		X		X		X
11	71.8	64.5	16.5	0.119	X	X			X		X		X		X	X		X				X		X		X		
11	71.6	64.2	16.9	0.120					X		X		X	X	X	X		X	X			X		X		X		
12	**73.1**	**65.3**	**16.3**	**0.118**	**X**				**X**		**X**		**X**	**X**	**X**	**X**		**X**	**X**			**X**		**X**		**X**		
12	72.8	64.8	16.9	0.119	X				X				X	X	X	X		X	X			X		X		X		X
13	74.1	65.7	16.7	0.117	X	X			X		X	X	X		X			X			X	X	X	X		X		
13	73.9	65.4	17.1	0.118	X	X			X		X		X		X		X	X			X	X	X	X		X		
14	74.8	65.8	17.5	0.117	X		X		X		X	X	X	X	X			X	X		X	X		X		X		
14	74.5	65.3	18.1	0.118	X		X		X			X	X	X	X			X	X		X	X		X		X		X
15	75.6	65.9	18.3	0.117	X	X		X	X		X	X	X		X			X			X	X	X	X	X	X		
15	75.4	65.8	18.5	0.117		X		X	X		X	X	X		X		X	X			X	X	X	X	X	X		
16	76.4	66.2	18.9	0.116	X	X		X	X		X	X	X		X		X	X			X	X	X	X	X	X		
16	76.3	66.1	19.1	0.117			X	X	X		X	X	X	X	X		X	X	X		X	X		X	X	X		
17	77.1	66.3	19.8	0.116	X	X		X	X	X	X	X	X		X			X			X	X	X	X	X	X	X	
17	77.1	66.2	19.8	0.116	X		X	X	X		X	X	X	X	X		X	X	X		X	X		X	X	X		
18	78.0	66.7	20.3	0.116			X	X	X	X	X	X	X	X	X		X	X	X		X	X		X	X	X	X	
18	77.9	66.6	20.4	0.116	X		X	X	X		X	X	X	X	X	X	X	X	X		X	X		X	X	X		
19	78.8	66.9	21.0	0.115	X		X	X	X	X	X	X	X	X	X		X	X	X		X	X		X	X	X	X	
19	78.6	66.6	21.3	0.116		X	X	X	X	X		X	X	X	X		X	X	X		X	X	X	X	X	X	X	
20	79.7	67.4	21.5	0.114		X	X	X	X	X	X	X	X	X	X		X	X	X		X	X	X	X	X	X	X	
20	79.5	67.0	21.9	0.115	X	X	X	X	X	X	X	X	X	X	X			X	X		X	X	X	X	X	X	X	
21	80.4	67.6	22.3	0.114	X	X	X	X	X	X	X	X	X	X	X		X	X	X		X	X	X	X	X	X	X	
21	80.1	67.1	22.8	0.115		X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	
22	81.0	67.6	23.3	0.114	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	
22	81.0	67.5	23.4	0.114	X	X	X	X	X	X	X	X	X	X	X		X	X	X		X	X	X	X	X	X	X	X
23	81.8	67.9	24.0	0.114	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X

23	81.6	67.6	24.3	0.114	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
24	82.4	67.9	25.0	0.113	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

$$\text{TIME} = 0.985 + 1.389\ \text{TEMP} - 1.669\ \text{THK} - 0.1746\ \text{COV} - 4.308\ \text{MAOP/SMYS} + 1.194\ \text{INCP/MAOP} - 1.629\ \text{SMYS*THK/DIAM} + 1.304\ \text{MAOP*THK/DIAM} + 3.817\ (\text{MAOP/SMYS})^2 - 0.847(\text{INCP/MAOP})^2 - 1.488\ \text{TEMP}^2 - 0.5963\ \text{MAOP}^2 + 1.856\ \text{THK}^2 \quad (8)$$

Table 8. Best subset model development based for South East region (Scen. 3)

Vars	R-Sq (%)	R-Sq (adj)(%)	Mallows Cp	S	TEMP	INPRS	MAOP	DIAM	THK	SMYS	COV	DIAM/THK	MAOP/SMYS	INCP/MAOP	SMYS*THK/DIA	MAOP*THK/DIA	(DIAM/THK)^2	(MAOP/SMYS)^2	(INCP/MAOP)^2	(SMYS*THK/DIA)^2	(MAOP*THK/DIA)^2	TEMP^2	INCP^2	MAOP^2	DIAM^2	THK^2	SMYS^2	COV^2
1	27.5	25.4	90.1	0.182																								X
1	26.5	24.5	91.8	0.183							X																	
2	50.4	47.6	52.9	0.152			X																					X
2	49.9	47.0	53.7	0.153																				X				X
3	66.3	63.3	27.7	0.127				X																			X	X
3	65.8	62.8	28.5	0.128				X			X																X	
4	71.5	68.1	20.7	0.119				X										X									X	X
4	70.9	67.4	21.8	0.120				X					X														X	X
5	76.0	72.2	15.1	0.111							X				X	X		X									X	
5	75.7	71.9	15.6	0.111											X	X		X									X	X
6	79.5	75.5	11.1	0.104								X				X	X	X									X	X
6	79.0	74.9	12.0	0.105				X			X				X	X		X									X	
7	81.4	77.1	9.8	0.100					X									X		X	X					X	X	X
7	81.2	76.9	10.1	0.101				X			X				X	X		X					X				X	
8	83.4	78.9	8.4	0.097		X							X		X	X			X					X			X	X
8	83.0	78.3	9.1	0.098		X		X					X			X			X	X							X	X
9	85.7	81.2	6.4	0.091		X			X				X			X			X	X				X			X	X
9	85.7	81.1	6.5	0.091		X			X				X		X	X			X					X			X	X
10	88.0	83.5	4.6	0.085		X			X				X			X			X	X				X		X	X	X
10	87.6	83.0	5.3	0.087		X			X				X		X	X			X					X		X	X	X

11	88.7	83.9	5.4	0.084		X			X				X		X	X			X		X			X		X	X	X
11	88.5	83.6	5.7	0.085	X	X			X				X			X			X	X				X		X	X	X
12	**89.3**	**84.1**	**6.3**	**0.084**	**X**	**X**			**X**				**X**		**X**	**X**			**X**		**X**			**X**		**X**	**X**	**X**
12	89.2	84.0	6.5	0.084		X			X				X		X	X			X		X	X		X		X	X	X
13	89.7	84.2	7.6	0.083	X	X			X				X	X	X	X			X	X	X			X			X	X
13	89.7	84.1	7.6	0.084	X	X			X				X		X	X			X		X	X		X		X	X	X
14	90.2	84.2	8.8	0.083	X	X			X				X	X	X	X		X	X	X	X			X			X	X
14	90.1	84.1	8.9	0.084		X			X				X	X	X	X		X	X	X	X	X		X			X	X
15	90.7	84.4	9.8	0.083	X	X			X			X	X	X	X	X		X	X	X	X			X			X	X
15	90.7	84.4	9.9	0.083		X			X			X	X	X	X	X		X	X	X	X	X		X			X	X
16	91.2	84.5	11.0	0.082	X	X	X		X	X		X	X	X	X	X		X	X	X	X			X				X
16	91.2	84.4	11.1	0.083		X	X		X	X		X	X	X	X	X		X	X	X	X	X		X				X
17	91.6	84.5	12.3	0.083	X	X	X		X			X	X	X	X	X	X	X	X	X	X			X			X	X
17	91.6	84.5	12.3	0.083	X	X	X		X	X		X	X	X	X	X	X	X	X	X	X			X				X
18	91.9	84.1	13.9	0.084	X	X	X		X		X	X	X	X	X	X	X	X	X	X	X			X			X	X
18	91.8	84.1	14.0	0.084	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X			X				X
19	92.1	83.7	15.5	0.085	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X		X				X
19	92.1	83.7	15.5	0.085	X	X			X		X	X	X		X	X	X	X	X	X	X	X	X	X		X	X	X
20	92.3	83.3	17.1	0.086	X	X	X		X		X	X	X		X	X	X	X	X	X	X	X	X	X		X	X	X
20	92.3	83.1	17.3	0.086	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X			X
21	92.4	82.3	19.1	0.088	X	X	X		X		X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X
21	92.3	82.3	19.1	0.088	X	X	X		X	X	X	X	X		X	X	X	X	X	X	X	X	X	X		X	X	X
22	92.4	81.2	21.0	0.091	X	X	X	X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
22	92.4	81.2	21.0	0.091	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X
23	92.4	79.9	23.0	0.094	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
23	92.4	79.9	23.0	0.094	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X
24	92.4	78.4	25.0	0.098	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

TIME=1.785 - 0.0750 TEMP + 1.980 INCP + 0.834 THK - 2.417 MAOP/SMYS

- 3.530 SMYS*THK/D + 6.36 MAOP*THK/D - 0.925 $(\mathrm{INCP/MAOP})^2$ - 3.25 $(\mathrm{MAOP*THK/D})^2$

$- 0.919\ MAOP^2 - 0.668\ THK^2 - 0.882\ SMYS^2 - 0.3599\ COV^2$ (9)

5.3.1 Null hypothesis test

The diagnostic measures of Mallow's Cp, R-sq and R-sq (adj) were considered as the preliminary model selection criteria. However, implementation of null hypothesis testing is necessary to determine whether a relationship between the response and explanatory variables indeed exists. For this objective, P-value (statistical significance) is a measure to interpret the results of regression analysis and tests the null hypothesis for each term. For a multiple linear regression model with the general form below as hypothesis we have,

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon \quad (7)$$

- null hypothesis: H_0: $\beta_1 = \beta_2 = \ldots = \beta_p = 0$
- against the alternative hypothesis: H_0: $\beta_1 \neq \beta_2 \neq \ldots \neq \beta_p \neq 0$

In the analysis of variance table (ANOVA), a low p-value (< α=significance level) indicates that the null hypothesis can be rejected, where the coefficient of that term is not equal to zero. On the other hand, a large p-value indicates that term is not significant. The significance level of α is used as a probability cutoff for making decisions about the null hypothesis were is assumed as 5%. P-value can be obtained from F-test which is a function of mean square regression (MSR) and mean square error (MSE).

$$F^* = \frac{MSR}{MSE} \quad (10)$$

For a multiple regression with *n* number of observations and *p* number of independent variables;

$$MSR = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{p} \quad (11)$$

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-p-1} \quad (12)$$

Where $\hat{y}_i$, $\bar{y}$ and y_i are estimated, mean and actual values of response variable respectively.

P-value is defined as the probability of obtaining a value of F^{*}through referring to an F-distribution with p numerator degree of freedom and n-p-1 denominator degrees of freedom. According to tables 9 and 10, ANOVA results illustrate that for a significance level of 5% the developed models are valid, and the null hypothesis is rejected. This is an indication of robust results.

Table 9. ANOVA results for Great Plains region model

Term	Coef	SE Coef	T-Value	P-Value
Constant	0.985	0.181	5.43	0.000
TIME	1.389	0.905	2.23	0.03
THK	-1.669	0.686	-2.43	0.019
COV	-0.1746	0.0651	-2.68	0.010
MAOP/SMYS	-4.308	0.721	-5.98	0.000
INCP/MAOP	1.194	0.583	2.05	0.047
SMYS*THK/DIAM	-1.629	0.259	-6.29	0.000
MAOP*THK/DIAM	1.304	0.206	6.32	0.000
$(MAOP/SYMS)^2$	3.817	0.714	5.34	0.000
$(INCP/MAOP)^2$	-0.847	0.541	-2.57	0.013
$TEMP^2$	-1.488	0.866	-3.43	0.001
$MAOP^2$	-0.5963	0.0731	-8.16	0.000
THK^2	1.856	0.722	2.57	0.014

Table 10. ANOVA results for South East region model

Term	Coef	SE Coef	T-Value	P-Value
Constant	1.785	0.223	8.01	0.000
TIME	-0.0750	0.0629	-2.12	0.040
INCP	1.980	0.459	4.31	0.000
THK	0.834	0.412	2.08	0.044
MAOP/SMYS	-2.417	0.480	-5.03	0.000

SMYS*THK/DIAM	-3.530	0.735	-4.80	0.000
MAOP*THK/DIAM	6.36	1.78	3.56	0.002
$(INCP/MAOP)^2$	-0.925	0.238	-3.90	0.001
$(MAOP*THK/DIAM)^2$	-3.25	1.91	-2.09	0.043
$MAOP^2$	-0.919	0.307	-3.00	0.006
THK^2	-0.668	0.409	-2.36	0.023
$SMYS^2$	-0.882	0.144	-6.11	0.000
COV^2	-0.3599	0.0736	-4.89	0.000

5.3.2 Residual analysis

Another criterion that should be fulfilled in model implementation is to verify that model error is unpredictable which implies that explanatory information is not present in the error. On the other hand, a satisfactory model shall have a normal probability plot which implies normal distribution of the residuals that is an assumption in regression analysis. Both developed models are in accordance with these criteria resulting from a visual residual analysis (figure 6).

5. Validations and sensitivity analysis

After examination of the diagnostic measures, null hypothesis and residual analysis, scenario 3 was selected as the final model for time of failure prediction in gas transmission pipelines of the two regions. To verify application of such models, they were tested by feeding the remaining 20% data (validation data set) into the developed models and some mathematical validation procedures were conducted on the estimated values obtained from validation dataset. These mathematical validation procedures include average validity percentage (AVP), root mean square error (RMSE), mean absolute error (MAE) and R-Sq as obtained from Eq. 13 to Eq. 17. RMSE and MAE are applied as statistical procedures to test model performance.

$$AIP = \sum_{i=1}^{n} \left|1 - \left(\frac{\hat{y}_i}{y_i}\right)\right| \times \frac{100}{n} \tag{13}$$

$$AVP = 100 - AIP \quad (14)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (15)$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \quad (16)$$

$$R\_Sq = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (17)$$

Table 11 summarizes the results obtained from testing the validation of the dataset for each regional classification. According to this table, the validation measures are not optimum yet satisfactory. This can be due to impact of numerous factors on corrosion failure, while in the framework of this research, only certain parameters are considered, reflecting the limited access to data on all effective failure parameters. Among different validation measures, RMSE is considered as it penalizes any variance by assigning higher weights to errors with larger absolute values than to those with smaller values while MAE and AVP consider equal weights for all errors.

Table 11. Validation outputs for developed failure prediction models

Validation measure	Underground Great Plains	Underground South East
AVP	0.73	0.70
RMSE	0.04	0.07
MAE	0.12	0.11
R-Sq	0.93	0.75

A sensitivity analysis on the contributing variables was implemented to highlight the most sensitive variables in the model and their corresponding influence on the model output. For this objective, regarding to the threshold of each variable in the prediction model, for one record, each variable is adjusted between its corresponding minimum and maximum values in uniform discernments while others are kept as constant and is then fed into the model. In this process, other input variables and the output results shall also satisfy the minimum and maximum thresholds of the prediction model, otherwise the corresponding sensitivity point is removed from the analysis. Figure 7 presents the results corresponding to sensitivity analysis of the developed prediction models. Accordingly, both models are evidently sensitive to all input variables. However, the output results are showing higher sensitivity to diameter, incident pressure and depth of cover for South East regional classification. On the other hand, for Great Plains regional classification, the most sensitive variables include; incident pressure, specified minimum yield strength and maximum operating pressure. It should be also noted that some sensitivity results are interpreted from fewer number of records in cases where the minimum and maximum thresholds for either input or output variables were not respected. In addition, the augmentation or diminishment trend of sensitivity results for each input variable is similar for both models. However, as dimeter increases, the Great Plains regional model outputs get less sensitive compared to the South East regional model. In addition, South East model outputs get more sensitive with augmentation of incident pressure compared to the other model. On the other hand, in case of maximum operating pressure, the sensitivity level of Great Plains model is more than that of South East model. Also, for both geographical regions, time of failure decreases linearly as soil temperature increases.

6. Conclusions

This study provided a first attempt to consider geo-environmental parameters in corrosion failure prediction of gas transmission pipelines. This research aimed at developing corrosion failure time prediction models in gas transmission pipelines through consideration of both experimental equations reported in the literature with incorporation of geo-environmental parameters. For this objective, a literature review of effective parameters on corrosion failure in gas pipelines was provided. A comprehensive record of failure incidents in gas transmission pipelines was also collected. On that basis, the explanatory and response variables of the prediction models were determined. Following a step by step procedure, time of failure prediction models were established and evolved using automated best subset and multiple regression analysis supported by diagnostic measures and statistical tests. Such application is tied to the advantage of this method in presenting comprehensible and practical mathematical equations which represent the relationships between input and output variables. In addition, a validation/sensitivity analysis procedure was adopted. Finally, the results obtained from analysis of scenario 3 were accepted leading to MAE and RSME of 0.12 and 0.04, for Great Plains, and 0.11 and 0.07, for South East regional classifications, respectively. The results point to that both models were mostly sensitive to incident pressure and specified minimum yield strength for South East and Great Plains regional specifications, respectively.

The proposed models will be capable of predicting failure for records that fall within specified input and output variables thresholds in the training phase. Therefore, for a case study pipeline, after collecting data on the explanatory variables and applying them to the developed regression models, time of failure can be estimated. The findings of this study could be beneficial for researchers and practitioners in pipeline operations and maintenance management, providing them with a number of benchmark models to predict the time of corrosion failure based on the geo-environmental classifications. Such models can help operators to avoid performing excessive in-line inspections. This research can be extended into development of models representing other geographical regions through collection of related data. The proposed methodology can also be used as a basis for establishing reliability-based maintenance plans for such pipelines.

REFERENCES

(1) US DOT. Distribution, Transmission & Gathering, LNG, and Liquid Accident and Incident Data. Pipeline and hazardous materials safety administration. US Department of Transportation. 2016; Available at:

http://www.phmsa.dot.gov/portal/site/PHMSA/menuitem.6f23687cf7b00b0f22e4c6962d9c8789/?vgnextoid=fdd2dfa122a1d110VgnVCM1000009ed07898RCRD&. Accessed June, 2016.

(2) Baker M. Pipeline corrosion. final report submitted to the US department of transportation pipeline and hazardous materials safety administration. Washington, DC, USA, 2008.

(3) Xie M, Tian Z. A review on pipeline integrity management utilizing in-line inspection data. Engineering Failure Analysis 2018; 92:222-239.

(4) Prediction of Pipe Wrinkling Using Artificial Neural Network. 2010 8th International Pipeline Conference: American Society of Mechanical Engineers; 2010.

(5) Probabilistic Analysis of the Collapse Pressure of Corroded Pipelines. 35th International Conference on Ocean, Offshore and Arctic Engineering Busan, South Korea: American Society of Mechanical Engineers, 2016.

(6) Witek M. Gas transmission pipeline failure probability estimation and defect repairs activities based on in-line inspection data. Engineering Failure Analysis 2016, 70:255-272.

(7) Senouci A, Elabbasy M, Elwakil E, Abdrabou B, Zayed T. A model for predicting failure of oil pipelines. Structure and Infrastructure Engineering 2014, 10(3):375-387.

(8) EGIG. European Gas Pipeline Incident Data Group, Gas pipeline incidents, 9th Report of the European Gas Pipeline Incident Data Group (period 1970 – 2013), 2015.

(9) Bersani C, Citro L, Gagliardi R, Sacile R, Tomasoni A. Accident occurrance evaluation in the pipeline transport of dangerous goods. Chemical Engineering Transactions, 2010:249-254.

(10) Davis M, Dubois J, Gambardella F, Uhlig F. Performance of European cross-country oil pipelines: Statistical summary of reported spillages in 2008 and since 1971, 2010.

(11) Fessler RR. Pipeline corrosion. Report, US Department of Transportation Pipeline and Hazardous Materials Safety Administration, Baker, Evanston, IL, 2008.

(12) Beavers JA, Thompson NG. External corrosion of oil and natural gas pipelines. Ohio, USA: ASM International Materials Park; 2006, p. 1015-1025.

(13) Ismail A, El-Shamy A. Engineering behaviour of soil materials on the corrosion of mild steel. Journal of Applied Clay Science 2009, 42(3-4):356-362.

(14) Orazem M editor. Underground pipeline corrosion, 2014.

(15) Little B, Lee J. Microbiologically influenced corrosion: an update. Journal International Materials Reviews, 2014, 59(7):384-393.

(16) Zhu Q, Cao A, Zaifend W, Song J, Shengli C. Stray current corrosion in buried pipeline. Anti-Corrosion Methods and Materials, 2011, 58(5):234-237.

(17) Zakikhani K, Nasiri F, Zayed T. A Review on Failure Prediction Models for Oil and Gas Pipelines. journal of pipeline systems engineering and practice, ASCE 2019.

(18) Li X, Yu R, Zeng L, Li H, Liang R. Predicting corrosion remaining life of underground pipelines with a mechanically-based probabilistic model. Journal of Petroleum Science and Engineering 2009, 65(3), 162-166.

(19) Wen K, Gong J, Chen F, Liu Y. A Framework for Calculating the Failure Probability of Natural Gas Pipeline. Journal of Computer Science Technology Updates, 2014, 1(1):1-8.

(20) Dundulis G, Žutautaitė I, Janulionis R, Ušpuras E, Rimkevičius S, Eid M. Integrated failure probability estimation based on structural integrity analysis and failure data: Natural gas pipeline case. Reliability Engineering and System Safety 2016, 156:195-202.

(21) Weiguo Z, Dongjing L, Hai W, Xinxin P. Remaining-life prediction and reliability assessment of buried gas pipelines under corrosion and alternating loads. Journal of Pipeline Systems Engineering and Practice 2014, 6(1):05014002-1-05014002-6.

(22) Kucheryavyi V, Mil'kov S. Reliability analysis of a compression section of a gas pipeline with the presence of longitudinal cracks. Journal of Machinery Manufacture and Reliability 2011, 40(3):290-293.

(23) Teixeira AP, Guedes Soares C, Netto TA, Estefen SF. Reliability of pipelines with corrosion defects. IPVP International Journal of Pressure Vessels and Piping, 2008, 85(4):228-237.

(24) Ossai CI, Boswell B, Davies IJ. Estimation of internal pit depth growth and reliability of aged oil and gas pipelines-a Monte Carlo simulation approach. Corrosion, 2015, 71(8):977-991.

(25) Liao K, Yao Q, Wu X, Jia W. A numerical corrosion rate prediction method for direct assessment of wet gas gathering pipelines internal corrosion. Energies, 2012, 5(10):3892-3907.

(26) Caleyo F, Velázquez J, Valor A, Hallen J. Probability distribution of pitting corrosion depth and rate in underground pipelines: A Monte Carlo study. Corrosion Science 2009, 51(9):1925-1934.

(27) Papavinasam S, Doiron A, Revie RW. Model to predict internal pitting corrosion of oil and gas pipelines. Corrosion, 2010;66(3):035006-035006-11.

(28) Manual for determining the remaining strength of corroded pipelines. ASME New York, NY, 2009.

(29) Kiefner F, Vieth H. A modified criterion for evaluating the remaining strength of corroded pipe. A modified criterion for evaluating the remaining strength of corroded pipe, 1989, PR-3-805.

(30) Burst criteria of corroded pipelines-defect acceptance criteria. Proceedings of the EPRG/PRC 10th Biennial Joint Technical Meeting on Line Pipe Research Cambridge, UK: British Steel, 1995.

(31) A new rupture prediction model for corroded pipelines under combined loadings. 1998 2nd International Pipeline Conference Calgary, Alberta, Canada: American Society of Mechanical Engineers, 1998.

(32) The assessment of corrosion in pipelines–Guidance in the pipeline defect assessment manual (PDAM). International Colloquium Reliability of High Pressure Steel Pipelines, 2003.

(33) New developments in burst strength predictions for locally corroded pipelines. International conference on offshore mechanics and arctic engineering, Copenhagen (Denmark): American Society of Mechanical Engineers, 1995.

(34) Tuffery S. Data mining and statistics for decision making. 1st ed. Chichester, West Sussex; Hoboken, NJ.: John Wiley and Sons, Ltd, 2011.

(35) Ledolter J, Hogg RV. Applied statistics for engineers and physical scientists. Newyork, US: Macmillan, 1992.

(36) NCDC. 2017, Available at: https://www.ncdc.noaa.gov/IPS/cd/cd.html. Accessed 2017, Sep.