ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.3.1.P63

A Reliability Prediction Method for Pipelines with Corrosion Defects Based on Gaussian Mixture Model

Tao Cui[1, a] *, Zhifeng Wu[1, b], Sha Chen[1, c], Chen Shen[1, d], Ailing Wang[1, e], and Yinchao Fang[1, f]

[1] PipeChina Group Southwest Pipeline Co., LTD

[a] qxd_cup@163.com, [b] 2500877249@qq.com, [c] cup_yuan@163.com, [d] 1061225524@qq.com,

[e] 1546135258@qq.com, [f] 18813106190@qq.com

Abstract. The assessment of pipeline corrosion rate is a critical component in the prediction of pipeline reliability. Traditional methods for predicting the reliability of corroded pipelines often suffer from excessive conservatism and insufficient accuracy in corrosion rate estimation. To address these issues, this paper proposes a reliability prediction methodology for pipelines with corrosion defects based on the Gaussian Mixture Model (GMM). First, lever-aging pipeline corrosion inspection data, the GMM is employed to fit the probability distribution of pipeline corrosion rates, thereby establishing a model that captures the distribution patterns of corrosion rates. Subsequently, a pipeline reliability assessment model is developed based on this distribution pattern, with the 95th percentile of the distribution model selected as the predicted corrosion rate. Finally, by integrating the reliability assessment model with the corrosion rate, the future reliability of the pipeline is predicted. Experimental results demonstrate that the deep Gaussian Mixture Model effectively fits the distribution of pipeline corrosion rates, with a K-S test value as low as 0.1295, outperforming traditional probability distribution models. Comparative analysis of prediction results from typical engineering cases indicates that the proposed methodology aids in formulating more scientific and rational maintenance strategies, thereby enhancing the operational efficiency of pipelines. This approach not only improves the accuracy of corrosion rate estimation but also provides a robust framework for reliability prediction, contributing to the optimization of pipeline management practices.

Keywords: corrosion rate statistics; corrosion rate distribution; Gaussian Mixture Model; reliability prediction;

1. Introduction

Oil and gas pipelines serve as the "lifelines" of national energy transportation. However, during long-term operation, pipelines are subjected to the coupled effects of multiple factors, including internal transported media (such as CO_2, H_2S), external complex environments (soil, seawater corrosion), and alternating stresses, making corrosion one of the critical factors affecting pipeline safety and reliability. Corrosion not only leads to wall thinning and reduced load-bearing capacity but can also cause leakage incidents in severe cases, resulting in significant economic, environmental, and social safety losses[1]. In recent years, with the continuous expansion of China's energy pipeline network, the total mileage of long-distance pipelines has increased substantially, making corrosion prevention and pipe-line integrity management increasingly prominent. Therefore, accurately assessing pipeline corrosion rates and their impact on pipeline reliability has become crucial for ensuring safe pipeline operation and optimizing maintenance strategies.

Traditional corrosion reliability assessment methods often rely on empirical formulas or statistical regression models, assuming that corrosion defect parameters follow specific distributions such as nor-mal or log-normal[2][3]. However, these methods have significant limitations: empirical models struggle to characterize the interactions of complex environmental factors, leading to overly conservative pre-dictions and unnecessary maintenance costs; assumptions about the distribution types of corrosion defects (e.g., depth, length) lack a unified basis, with most studies relying on artificial assumptions or small sample data, directly affecting the accuracy of reliability analysis.

ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.3.1.P63

In recent years, machine learning techniques have been extensively applied in engineering prediction and analysis[4][5]. Particularly, probabilistic modeling-based machine learning methods, such as the Gaussian Mixture Model (GMM), have provided new insights for reliability analysis due to their advantages in modeling complex distributions and hierarchical feature extraction. The GMM combines the nonlinear representation capabilities of deep neural networks with the probabilistic modeling characteristics of Gaussian Mixture Models, enabling adaptive learning of the latent distribution features of corrosion rates from high-dimensional, unstructured corrosion data[6]. This approach overcomes the limitations of traditional parametric distribution assumptions (e.g., exponential, Weibull distributions, significantly enhancing the flexibility and accuracy of distribution fitting. This paper proposes a novel reliability prediction methodology for pipelines with corrosion defects based on the GMM: First, leveraging multi-source corrosion data obtained from internal inspections by a Chinese pipeline company, the GMM is used to achieve high-precision fitting of the multimodal distribution of corrosion rates without relying on prior distribution assumptions. Second, based on the probability density function output by the model, the 95th percentile is extracted as the predicted corrosion rate to construct a time-dependent reliability assessment framework. Finally, by integrating pipeline degradation mechanisms with dynamic reliability models, multi-scenario probabilistic predictions of future reliability are achieved. This methodology, through the synergy of deep probabilistic modeling and engineering mechanisms, provides a more robust solution for pipeline lifespan assessment under complex operating conditions.

2. Methodology

2.1 Corrosion Rate Prediction Based on Gaussian Mixture Model

This paper employs the GMM, a probabilistic framework that represents complex data distributions as a weighted sum of multiple Gaussian distributions. The GMM is particularly advantageous for corrosion rate modeling due to its ability to capture multimodal and non-linear characteristics inherent in corrosion data, which traditional parametric models often fail to represent accurately.

Let x represent the corrosion rate, and assume that the corrosion rate follows a mixture of K Gauss-ian distributions. The probability density function (PDF) of the GMM[7] is given by:

$$p(x\mid\theta)=\sum_{k=1}^{K}\pi_k N(x\mid\mu_k,\sigma_k^2) \tag{1}$$

In this formula, $\theta=\left\{\pi_k,\mu_k,\sigma_k^2\right\}_{k=1}^{K}$ represents the set of the model parameters. π_k is the mixing coefficient (weight) of the k-th Gaussian component, satisfying $\sum_{k=1}^{K}\pi_k=1$ and $\pi_k\geq 0$ for all k. $N(x\mid\mu_k,\sigma_k^2)$ is the probability density function of the k-th Gaussian component, with mean μ_k and variance σ_k^2:

$$N(x\mid\mu_k,\sigma_k^2)=\frac{1}{\sqrt{2\pi\sigma_k^2}}\exp(-\frac{(x-\mu_k)}{2\sigma_k^2}) \tag{2}$$

The GMM is trained using the Expectation-Maximization (EM) algorithm to iteratively estimate parameters that maximize the likelihood of observed corrosion rate data, enabling adaptive learning of underlying distribution patterns without restrictive assumptions. Constructed from pipeline inspection data (e.g., corrosion depth, length), the GMM captures the probabilistic nature of corrosion rates, offering flexibility for complex, multimodal distributions, adaptability to diverse scenarios, and interpretability through distinct Gaussian components. In pipeline corrosion evaluation, the maximum corrosion rate often leads to overly conservative maintenance, while the

ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.3.1.P63
local rate underestimates risks. To balance conservatism and accuracy, this study analyzes corrosion distribution patterns from internal inspection data and adopts the 95th percentile corrosion rate for assessment, providing a reliable representation of pipeline conditions to support informed maintenance decisions.

2.2 Reliability Evaluation

Firstly, define the limit state function of the pipeline. Apply the modified ASME-B31G method to calculate the remaining strength of the corroded area. This calculation method mainly considers the influence of the length and depth of the corrosion point on the remaining strength of the pipeline. The modified ASME-B31G[8] formula for calculating the remaining strength is as follows:

$$p_f = \frac{2t\sigma_{flow}}{D} \frac{1-0.85\dfrac{d}{t}}{1-0.85\dfrac{d}{Mt}} \tag{4}$$

when $\frac{l^2}{Dt} \le 50$:

$$M = \sqrt{1+0.6275\frac{l^2}{Dt} - 0.003375\left(\frac{l^2}{Dt}\right)^2} \tag{5}$$

when $\frac{l^2}{Dt} > 50$:

$$M = 0.032\frac{l^2}{Dt} + 3.3 \tag{6}$$

In the formula, P_f represents the failure pressure of the pipeline, measured in MPa; M represents the bulging coefficient; l represents the defect length of the pipeline, measured in mm; D represents the outer diameter of the pipeline, measured in mm; t represents the wall thickness of the pipeline, measured in mm; d represents the defect depth of the pipeline, measured in mm; σ_f represents the rheological stress, measured in MPa:

$$\sigma_f = 1/2 \times (SMYS + SMTS) \tag{7}$$

In the formula, $SMYS$ represents the specified minimum yield limit, measured in MPa; $SMTS$ represents the tensile limit, measured in MPa.

The local corrosion rate is estimated by a linear growth model based on the data from two rounds of magnetic flux leakage internal detection. This model assumes that the corrosion grows uniformly over time. The corrosion rate is calculated by comparing the changes in the local corrosion depth at the same location found in the two rounds of magnetic flux leakage internal detection. That is:

$$d(T) = d_0 + \upsilon_r (T - T_0) \tag{8}$$

$$l(T) = l_0 + \upsilon_a (T - T_0) \tag{9}$$

In the formula, $d(T)$ represents the depth of the corrosion point at time T, in mm; $l(T)$ represents the length of the corrosion point at time T, in mm; υ_r represents the radial corrosion rate, in mm/s; υ_a represents the axial corrosion rate, in mm/s; d_0 represents the corrosion depth at time T_0, in mm; l_0 represents the corrosion length at time T_0, in mm.

It is worth noting that in the linear growth model, since it is assumed that the corrosion rate of each corrosion point is constant within the analysis time interval, the actual nonlinear nature of corrosion growth is simplified. If the service time of the pipeline is relatively long, this excessive simplification may lead to conservative calculation results. Finnaly, the failure pressure of the pipeline at time T is:

ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.3.1.P63

$$p_f\left(T\right)=\frac{2t\sigma_{flow}}{D}\frac{1-0.85\dfrac{d_0+\upsilon_r\left(T-T_0\right)}{t}}{1-0.85\dfrac{d_0+\upsilon_r\left(T-T_0\right)}{Mt}} \tag{10}$$

Based on the reliability theory, the limit state function Z is the difference between the failure pressure of the pipeline and the maximum allowable working pressure of the pipeline:

$$Z=p_f-P_0 \tag{11}$$

In this formula, P_0 is the max allowed pressure of the pipeline, MPa.

When $Z>0$, the pipe section is in a safe state. When $Z<0$, the pipe section is in a failed state. When $Z=0$, the pipe section is in an extreme state. Therefore, the failure probability of the pipe section is actually the probability of $Z<0$ under the influence of random variables.

The failure probability of the pipe section can be obtained by solving the limit state, and then the reliability can be derived. The failure probability of the pipe section is difficult to be calculated by using traditional calculation methods based on the limit state equation. However, the Monte Carlo method does not require the analytical solution of the limit state equation. Instead, it only needs to repeatedly conduct simple random sampling to generate a series of sample numbers $\left(x_1,x_2,\cdots,x_n\right)$, and then substitute these sample numbers into the limit state equation $Z=g\left(x_1,x_2,\cdots,x_k\right)$ to obtain the failure probability distribution. Finally, the failure probability of the pipe section can be derived based on the reliability theory and statistical analysis methods. This method is widely applied in the reliability analysis of pipelines[9].

By combining with the corrosion rate prediction in 2.1, the reliability of the pipeline in the future can be predicted.

3. Case Study

3.1 Data Foundation

In this paper, we conduct our case study based on the actual corrosion internal inspection data of a certain pipeline. The basic parameters of the pipeline are shown in Table 1.

Table 1. The basic parameters of the pipeline

Parameters	Description	Distribution	Mean	Standard Deviation
t	Wall thickness, (mm)	Normal	16.0	0.5
D	Diameter, (mm)	Normal	711	0
σ_y	Yield strength, (MPa)	Normal	467	32.9

3.2 Comparison of Corrosion Rate Fitting Models

In order to calculate the corrosion rate of pipeline assessment and predict the reliability of pipelines, it is necessary to determine the distribution patterns of the radial and axial corrosion rates of pipelines. The Gaussian mixture model is adopted to fit the corrosion data, and normal distribution, logarithmic distribution, generalized extreme value distribution and exponential distribution are selected as comparisons. The fitting results are shown in Fig. 1.

ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.3.1.P63

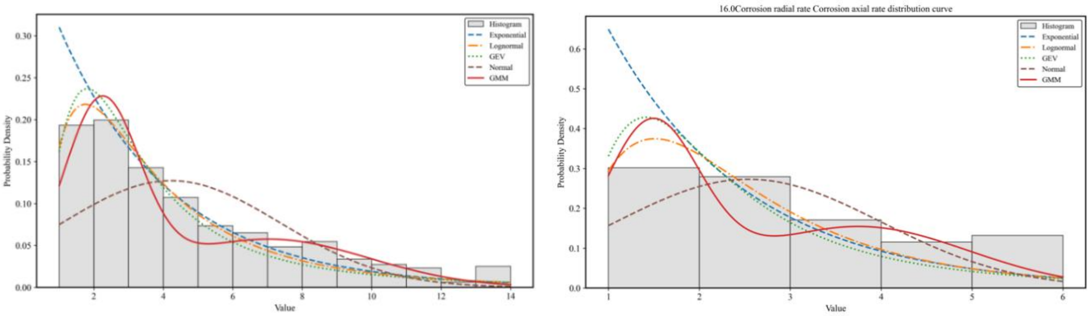

Fig.1 Probability density curve of radial and axial corrosion rate

As shown in Fig. 1, all four models can fit the corrosion detection data well. Among them, the Gaussian Mixture Model is closer to the original data. To further verify the optimal distribution law that the pipeline data conforms to, the K-S test[10] is adopted for further verification. The results are shown in Table 2.

Table 2. K-S test values of different models

Model	K-S test value	
	Corrosion Axial Rate	Corrosion Radial Rate
Exponential	0.3019	0.1933
Lognormal	0.2028	0.1297
GEV	0.2089	0.1302
Normal	0.2255	0.1877
GMM	0.1963	0.1295

As can be seen from the table, the GMM model K-S test values of the radial corrosion rate and the axial corrosion rate are the smallest, being 0.1963 and 0.1295 respectively.

3.3 Prediction of the Reliability of the Pipeline

The determination of pipeline evaluation corrosion rate is crucial to the evaluation of the reliability of the pipeline. When selecting the evaluation corrosion rate of the pipeline, both the conservative type of pipeline evaluation should be considered and should not be too conservative. Therefore, the axial and radial corrosion rates of the pipeline with 95% quantile are selected as the evaluation corrosion rate in this paper.

By calculating the cdf function of GMM, it can be seen that the corrosion rate corresponding to the radial 95% quantile is 0.6577 mm/a, and the axial 95% quantile corresponds to the corrosion rate is 1.3049mm/a. It can be found that the corrosion rates of 95% quantile of the pipe in radial and axial direction are inconsistent and greatly different, because adjacent corrosion with axial distance less than 3 times the wall thickness is regarded as mutually influencing corrosion in ASME-B31G. Therefore, the adjacent corrosion point whose axial or circumferential distance from the corrosion point boundary is less than 3 times the wall thickness is generally labeled as a corrosion during the analysis of the detection data in MFL. In addition, axial corrosion growth can develop to both ends at the same time, while radial corrosion can only develop along the radial, so the evaluated axial corrosion rate is larger than the radial corrosion rate.

The radial and axial 95% quantile corrosion rates of the pipeline based on the probability distribution law were put into equation (10) respectively, and the future failure probability change curve of the pipeline was obtained through Monte Carlo sampling, as shown in Fig 2.

ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.3.1.P63

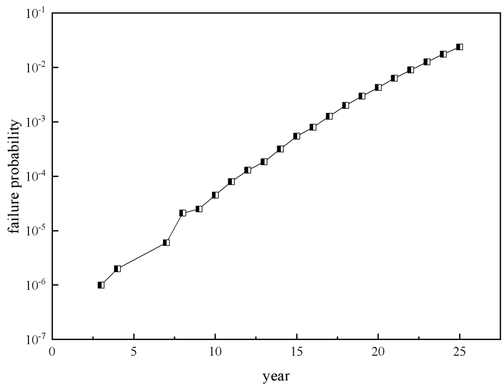

Fig.2 Failure probability change curve of pipeline failure

It can be seen from the figure that the reliability of the pipeline remains basically unchanged at the initial stage of operation. When the pipeline has been in service for XX years, the degree of corrosion has intensified, and the reliability of the pipeline has begun to decrease. After 100 years, the reliability is reduced to XX. By setting a reliability-based maintenance threshold, maintenance can be performed before pipeline reliability falls below the threshold, reducing the risk of pipeline leakage due to corrosion and perforation.

4. Summary

(1) The two-round MFL internal test data of a pipeline was analyzed as an example, and the GMM model was introduced to fit the MFL internal test data. Through K-S test, it is found that GMM is more consistent with the radial and axial corrosion rate distribution than the traditional probability distribution model.

(2) Considering the need for conservative and comprehensive calculation of pipeline corrosion rate, the corrosion rate based on the pipeline probability distribution law is used as the pipeline evaluation corrosion rate. According to the calculation in this paper, the radial and axial evaluation corrosion rates are 0.6577mm/a and 1.3049mm/a, respectively. The variation curve of pipeline reliability in the future is obtained by Monte Carlo simulation.

(3) This paper provides a method to analyze the probability distribution model of pipeline MFL internal detection data, which can provide a reference for the study of other pipeline probability distribution laws.

Acknowledgements

This research is supported by “Monitoring·and Sensing·System for Safe and Efficient Operation of Cross-regional Complex·Oil and Gas·Pipeline Network and Its Application(2022YFB3207600)".

References

[1] Li C , Yang F , Jia W ,et al.Pipelines reliability assessment considering corrosion-related failure modes and probability distributions characteristic using subset simulation[J].Process Safety & Environmental Protection: Transactions of the Institution of Chemical Engineers Part B, 2023, 178.DOI:10.1016/j.psep.2023.08.013.

[2] Zhang X, Naining C A O, Yayun L I. Residual life prediction of buried oil and gas pipelines based on gumbel extreme value type i distribution[J]. Journal of Chinese Society for Corrosion and protection, 2016, 36(4): 370-374.

[3] LIU Chenglei, YANG Jing, ZHU Ming, ZHANG Chang, GUAN Zhe, XIAO Ningjin. Remaining Life Prediction Method of Corroded Pipeline Based on Normal Distribution[J]. Corrosion & Protection, 2023, 44(3): 100-106,118. DOI: 10.11973/fsyfh-202303015

ISSN:3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.3.1.P63
[4] Fang J, Cheng X, Gai H, et al. Development of machine learning algorithms for predicting internal corrosion of crude oil and natural gas pipelines[J]. Computers & Chemical Engineering, 2023, 177: 108358.

[5] Seghier M E A B , Hche D , Zheludkevich M .Prediction of the internal corrosion rate for oil and gas pipeline: Implementation of ensemble learning techniques[J].Journal of Natural Gas Science and Engineering, 2022, 99:104425-.DOI:10.1016/j.jngse.2022.104425.

[6] Wang M , Li X , Chen L ,et al.A deep-based Gaussian mixture model algorithm for large-scale many objective optimization[J].Applied Soft Computing, 2025, 172.DOI:10.1016/j.asoc.2025.112874.

[7] HOSSAIN M. A simple and effective approach to investigate the dominant contaminant sources and accuracy in water quality estimation through Monte Carlo simulation, Gaussian Mixture Models (GMMs), and GIS machine learning methods[J]. Ecological Indicators, 2025,171: 113188.

[8] A S S M , B A S M .Failure pressure estimation error for corroded pipeline using various revisions of ASME B31G - ScienceDirect[J].Engineering Failure Analysis, 109[2025-03-25].DOI:10.1016/j.engfailanal.2019.104284.

[9] Maciej,Witek.Gas transmission pipeline failure probability estimation and defect repairs activities based on in-line inspection data[J].Engineering Failure Analysis, 2016.DOI:10.1016/j.engfailanal.2016.09.001.

[10] Lili F , Mathematics D O , University L .Parameter Estimation and Case Analysis of Generalized Extreme Value Distribution[J].Journal of Capital Normal University(Natural Science Edition), 2016.