

Quantifying the influence of corrosion defects on the failure prediction of natural gas pipelines using generalized extreme value distribution (GEVD) model and Copula function with a case study

Changzhou Wang, Lijing Zhang* and Gang Tao

College of Safety Science and Engineering, Nanjing Tech University, Nanjing 210009, China

* Corresponding author, E-mail: zhanglj@njtech.edu.cn

Abstract

Since the corrosion defects in gas pipelines have similar corrosion characteristics to the surrounding soil, the random growth of these defects may be correlated, so we can not simply treat the corrosion defects as completely correlated or independent. Therefore, this paper proposes a method that can accurately calculate the failure probability of the pipeline system considering the correlation of corrosion defects: using MATLAB software to fit the parameters of the GEV model to select an appropriate distribution model; using Monte Carlo simulation (MC), considering different correlation coefficients and quantities, the system failure probability of pipeline corrosion defects is calculated; the results show that the system failure probability of the pipeline and the correlation coefficient are basically linear; when the correlation coefficient is increasing, the pipeline is regarded as an independent. There is a large error between the calculated failure probability of the series system and the actual result; the system failure probability of the pipeline increases with the increase of the assumed number of corrosion defects. When the correlation coefficient is greater than or equal to 0.6, the system failure probability of the pipeline increases significantly. An increase and the system failure probability of the pipeline decreases significantly.

Citation: Wang C, Zhang L, Tao G. 2024. Quantifying the influence of corrosion defects on the failure prediction of natural gas pipelines using generalized extreme value distribution (GEVD) model and Copula function with a case study. *Emergency Management Science and Technology* 4: e002 <https://doi.org/10.48130/emst-0024-0002>

Introduction

Affected by the external environment such as soil medium, stray current, and microbial corrosion, external corrosion (mostly localized corrosion) often occurs in gas pipelines during service; accidents such as fire and explosion occur. Therefore, characterizing the uncertainty distribution of corrosion defects and propagating this uncertainty into the failure probability prediction model is crucial for formulating the maintenance cycle and improving the reliability of the pipeline.

Jia^[1], Xi^[2] and other scholars studied the application of the generalized extreme value distribution model in the characteristics of corrosion defect probability distribution. Taking the soil corrosion test data of Tarim and Dagang oil fields as an example, Weng^[3] demonstrated that the overall probability distribution of the regional corrosion test data conforms to the normal random function and the maximum data conforms to the Gumbel function. The statistical results show that the Gumbel function can be used to describe the probability distribution of the maximum corrosion depth of pipelines in the short term. The probability distribution of the maximum corrosion depth of the pipeline changes with different service time. However, in most pipeline failure probability calculation studies, each corrosion defect of the pipeline is regarded as an independent event affecting pipeline failure, and the events are independent of each other, and then the series model is used to calculate the failure probability of the pipeline. Since the soil and other external environment around the adjacent defects have similar

corrosion characteristics with the pipeline, the random growth of these defects may be correlated, and the results obtained by using the series model are different from the actual situation. This correlation should be taken into account when evaluating the system reliability of natural gas pipelines with multiple corrosion defects and developing appropriate failure prevention measures. Yu^[4] proved that the binary copula function can be constructed and applied to practical problems based on the correlation between the data. Wang et al. and Zhang et al.^[5,6] considered copula function to predict pipeline corrosion defects. Based on this, Zeng^[7] proposed a calculation method of pipeline failure probability based on correlation coefficient theory to simulate the correlation between corrosion defects, which can accurately evaluate the reliability of pipeline system. In order to evaluate the reliability of a pipeline system with multiple corrosion defects, the generalized extreme value distribution model (GEV) can be used as a suitable model to describe the maximum depth of corrosion, while the Copula-based method can reasonably simulate the random correlation between the growth of different defects.

GEV model and Copula function

GEV distribution

The research shows that the probability distribution of the maximum corrosion depth of the pipeline varies with the actual situation. The limitation of a single extreme value distribution

type will lead to the low prediction accuracy of the built model, while the GEV distribution can adaptively optimize the extreme value distribution according to the simulated parameters. GEV distribution is the probability distribution of the maximum value (minimum value) set in probability theory, let $y = \frac{x-\mu}{\sigma}$, then the GEV distribution can be divided into the following three forms:

Gumbel distribution ($\eta = 0$):

$$F_1(y) = \exp(-\exp(-y)), \forall y \quad (1)$$

Frechet distribution ($\eta > 0$):

$$F_2(y) = \begin{cases} 0, & y \leq 0 \\ \exp(-y^{-\eta}), & y > 0 \end{cases} \quad (2)$$

Weibull distribution ($\eta < 0$):

$$F_3(y) = \begin{cases} \exp(-y^{-\eta}), & y \leq 0 \\ 1, & y > 0 \end{cases} \quad (3)$$

x is an extreme variable, μ is the position parameter, σ is the scale parameter, η is the threshold parameter.

From Eqns (1) to (3), the cumulative probability distribution function of the GEV distribution can be deduced:

$$\text{GEV}(x) = \exp(-(1 + \eta y)^{-1/\eta}), \quad 1 + \eta y > 0 \quad (4)$$

Copula function

The Copula function is the joint distribution function of n ($n \geq 2$) standard uniformly distributed variables U_i ($i = 1, \dots, n$):

$$C(u_1, u_2, \dots, u_n) = P(U_1 < u_1, U_2 < u_2, \dots, U_n < u_n) \quad (5)$$

$C(u_1, u_2, \dots, u_n)$ representing the copula function, u_i is the value of U_i ($i = 1, \dots, n$). An n -membered probability distribution function $F(x_1, x_2, \dots, x_n)$ can be represented by the copula function as follows:

$$C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) = F(x_1, x_2, \dots, x_n) \quad (6)$$

In the formula, $F_i(x_i)$, $i = 1, 2, \dots, n$ is the marginal probability distribution function of $F(x_1, x_2, \dots, x_n)$. In addition, with the parameter $F_i(x_i)$, the copula is a function with marginal distribution $F_i(x_i)$ multivariate distribution function. Copula function set includes Gaussian Copula function, T-Copula function and Gumbel Copula function. In this study, Gaussian copula was selected to characterize the correlation between different corrosion defects.

$$C(u_1, u_2, \dots, u_n) = \Phi_n(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \dots, \Phi^{-1}(u_n); R) \quad (7)$$

$\Phi_n(\cdot; R)$ is an n -variable standard normal distribution function with a correlation coefficient R , and $\Phi^{-1}(\cdot)$ is the inverse of the standard normal distribution function. R , which can also be expressed as ρ refers to the Pearson linear correlation coefficient between corrosion depths. If the correlation matrix R is known, the Gaussian copula is fully defined^[4].

Pipeline failure probability calculation

Probability model of pipeline corrosion depth distribution based on GEV distribution

Since it is difficult to obtain the data of the entire pipeline, this paper adopts the method of estimating a large sample with a small sample, and conducts sampling detection to obtain the data in the most severely corroded area.

If the distribution law of pipeline corrosion depth is to be verified by imaging, it may be assumed that n samples of corroded pipelines are drawn, and the maximum corrosion depth x_i ($i = 1, 2, \dots, n$) of each area is used as a statistical

variable, and then sorted from big to small, and then the average permutation method of sequential statistics is used to calculate the cumulative probability, which is:

$$\text{GEV}(x_i) = \frac{i}{n+1}, \quad n = 1, 2, \dots, n \quad (8)$$

$n = 1, 2, \dots, n$, Fitting x_i and $\ln(1/\text{GEV}(x_i))$, if the image is exponentially distributed, it proves that the probability distribution of the maximum corrosion depth of the pipeline conforms to the GEV distribution.

Based on the MATLAB program, the GEV distribution parameters were fitted to obtain the corresponding μ, σ, η .

Failure modes and limit state equations of pipelines

The limit state function (LSF) can be used to define the corresponding failure mode, on the basis of which appropriate maintenance measures can be taken. This article uses two limit state functions to define small leaks and bursts. A small leak is defined as a failure event in which a defect (i.e., corrosion pit) penetrates a pipe wall to its wall thickness threshold percentage. Based on industry practice, a wall thickness threshold of 80% is recommended. Therefore, the first limit state function is:

$$\text{LSF}_1 = 0.2t - d \quad (9)$$

Where, d is the corrosion depth of the pipeline.

For bursting, its limit state function is:

$$\text{LSF}_2 = P_b - P_{op} \quad (10)$$

Where is the pipeline operating pressure, its commonly used models include improved B31G, PCORRC, DNV-RPF10 and other methods. In this paper, we use the improved B31G model to calculate the burst pressure, and its calculation formula is as follows:

$$P_b = \lambda \frac{2t}{D} \sigma \left[\frac{1 - 0.85 \frac{d}{t}}{1 - 0.85 \frac{d}{t} \frac{1}{M}} \right] \quad (11)$$

$$\begin{cases} M = \sqrt{1 + 0.6275 \left(\frac{2L}{\sqrt{Dt}} \right)^2 - 0.003375 \left(\frac{2L}{\sqrt{Dt}} \right)^4}, & L \leq \sqrt{50Dt} \\ M = 3.3 + 0.032 \left(\frac{L}{\sqrt{Dt}} \right)^2, & L > \sqrt{50Dt} \end{cases}$$

(12)

P_b is the pipe failure pressure (Mpa); L is the axial corrosion length, mm; λ is the error factor of multiplication model; D is the pipe outer diameter, mm; σ is the yield strength (Mpa); d is the corrosion depth of the pipeline, mm; t is the pipe wall thickness, mm; M is Folias expansion coefficients.

By $\text{LSF}_1, \text{LSF}_2$, small leakage can be defined as $\text{LSF}_1 \leq 0$ $\text{LSF}_2 > 0$, define breakthrough as $\text{LSF}_1 > 0 \cap \text{LSF}_2 \leq 0$.

Calculation of system failure probability of the pipeline

The Stevnsen-Moses method is a typical point estimation method. Its core idea is to assume that the mutual relationship between the failure modes in the series system is two ideal states, such as complete positive correlation or mutual independence. The failure modes of the structural system are not completely positively correlated with each other, nor are they completely independent, so the results obtained by the Stevnsen-Moses algorithm tend to be conservative or unsafe. Suppose the system failure probability of a pipeline with corrosion defects is P_f , then there are:

Pipeline corrosion dependent failure probability

$$\max P_{fi} \leq P_f \leq 1 - \prod_{i=1}^i (1 - P_{fi}) \tag{13}$$

P_{fi} is the failure probability of the pipeline at each defect.

It can be seen that the Steverson-Moses method can only describe the range of system failure probability. This paper proposes a system failure probability analysis method based on Monte Carlo simulation, which can accurately calculate the actual system failure of pipelines containing multiple corrosion defects probability, the calculation steps are as follows:

- 1) Let $N_0 = 0$, $SL = 0$, SL represents the number of leaks in the pipeline;
- 2) According to the distribution model of corresponding variables, generate random variables or fixed value samples such as wall thickness, pipe outer diameter and tensile strength;
- 3) n ($n = 2, 4, 6, 8, 10$) pieces of corrosion depth data with correlation coefficients of 0, 0.1, ..., 0.9 are respectively generated by the Copula function;
- 4) The value of LSF_1, LSF_2 is calculated according to the corresponding variable,
- 5) If $LSF_1 \leq 0$, so that $SL = SL + 1$;
- 6) Let $N_0 = N_0 + 1$; If $N_0 < N$, stop the cycle.

For the total number of N simulation tests, the failure probability PSL of small leakage and rupture of the pipeline can be calculated by the following formula:

$$P_{SL} \approx \frac{1}{N} SL \tag{14}$$

Example analysis

Data preprocessing

The Nanjing gas transmission pipeline has a history of more than 40 years since the 1970s, and a large number of gas pipelines in the main urban area have been in service for more than 10 years. When the pipelines were excavated, it was found that the sample pipelines had multiple corrosion phenomena, as shown in Fig. 1. The surface of the pipeline has been corroded by soil, microorganisms and other environmental factors over a long period, and the surface pitting corrosion phenomenon is serious. The main form was local corrosion on the outside of the pipelines.

In this paper, a certain section of pipeline was selected for analysis, and the material is Q235 steel. The measured data are shown in Table 1.

Fourteen groups of pipeline sections with serious corrosion degree are selected, and the results of residual wall thickness of pipeline are the average values of five measurements, as shown in Table 2.

Determination of GEV model parameters

Select the maximum corrosion depth of each section of pipeline, and calculate the GEV distribution probability of corrosion depth by the average arrangement method of



Fig. 1 Pipeline corrosion sample.

sequential statistics. As shown in Fig. 2, the local maximum corrosion depth of pipeline x_i and $\ln(1/GEV(x_i))$. There is a negative exponential distribution relationship, so the above maximum corrosion depth obeys Frechet distribution.

The model parameters are fitted by MATLAB, $\eta = 0.2294$, $\mu = 0.1588$, $\sigma = 0.46$. $\eta > 0$, which proves that the model obeys Frechet distribution, which is the same as the verification result in Fig. 2. The prediction probability calculated by the average arrangement method of sequential statistics for pipeline corrosion depth and the result of the fitted GEV distribution curve are shown in Fig. 3.

Simulation of pipeline corrosion depth under different correlation coefficients

The Copula function can be used to fit the depth data of pipeline corrosion defects with different correlation coefficients. To illustrate this method, 1,000 correlation coefficients (ρ) are randomly generated. Two-dimensional pipeline corrosion depth data of 0, 0.3, 0.6 and 0.9, respectively, and their distributions are shown in Fig. 4.

It can be seen from Fig. 4 that as the correlation coefficient (ρ) between the data increases, the corrosion depth distribution becomes more regular. When the correlation coefficient is 1, the two sets of corrosion depth data will be completely linear.

Calculation results and analysis of pipeline failure probability

Because the internal pressure of the pipeline is less than 25% yield strength, the pipeline is operated under low circumferential stress conditions, and the failure mode is perforation failure caused by corrosion. Therefore, the failure probability of pipeline system in this paper is the probability of small leakage.

Table 1. Pipeline parameters and probability distribution form.

Parameter	Average value	Variable coefficient	Unit	Distribution form
Axial length of corrosion defect (L)	30	35%	mm	Logarithmic normal distribution
Outer diameter (d)	60	1%	mm	Normal distribution
Maximum annual internal pressure (p)	0.259	—	Mpa	Definite value
Thickness (t)	4	1.5%	mm	Normal distribution
Yield strength (σ_u)	235	3%	Mpa	Normal distribution
Accuracy coefficient of model (x_m)	0.97	10.5%	—	Normal distribution

As is shown in Fig. 5, The corrosion depth decreases with the increase of correlation coefficient. The Fig. 6 shows the change trend of the system failure probability value of the pipeline

Table 2. Residual wall thickness of pipeline.

Sample number	Residual pipe wall thickness (mm)	Sample number	Residual pipe wall thickness (mm)
1	3.12	8	3.56
2	3.48	9	3.43
3	3.67	10	2.58
4	3.60	11	3.28
5	3.46	12	2.98
6	3.74	13	3.37
7	3.68	14	3.44

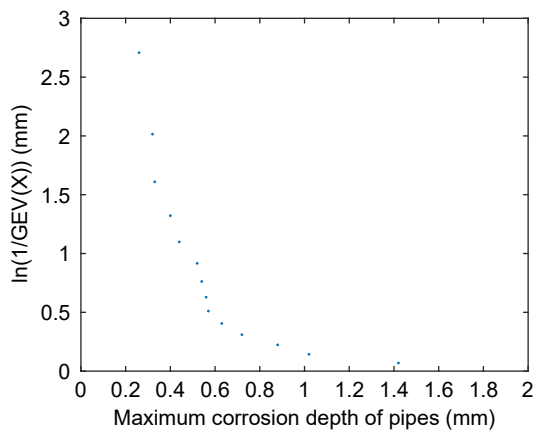


Fig. 2 Cumulative probability distribution law of local maximum corrosion depth of the pipeline.

with the increase of the correlation coefficient between corrosion defects when the number of pipeline corrosion defects is set as 2, 4, 6, and 8 respectively. As can be seen from the following figure, the system failure probability of the pipeline is within the upper and lower limits. Therefore, the failure probability calculation method adopted in this paper is effective.

Generally speaking, the system failure probability of a pipeline decreases with the increase of correlation coefficient, because the larger the correlation coefficient, the greater the joint failure probability of corrosion defects. For example, for a pipeline with two corrosion defects, the system failure probability is $P_f = P_{f_1} \cup P_{f_2} = P_{f_1} + P_{f_2} - P_{f_1} \cap P_{f_2}$. When the correlation degree is greater, the corrosion depth distribution trend of

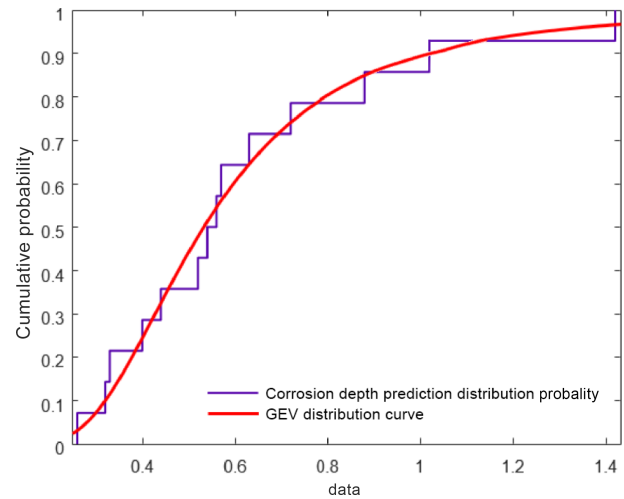


Fig. 3 Probability distribution curve of corrosion depth.

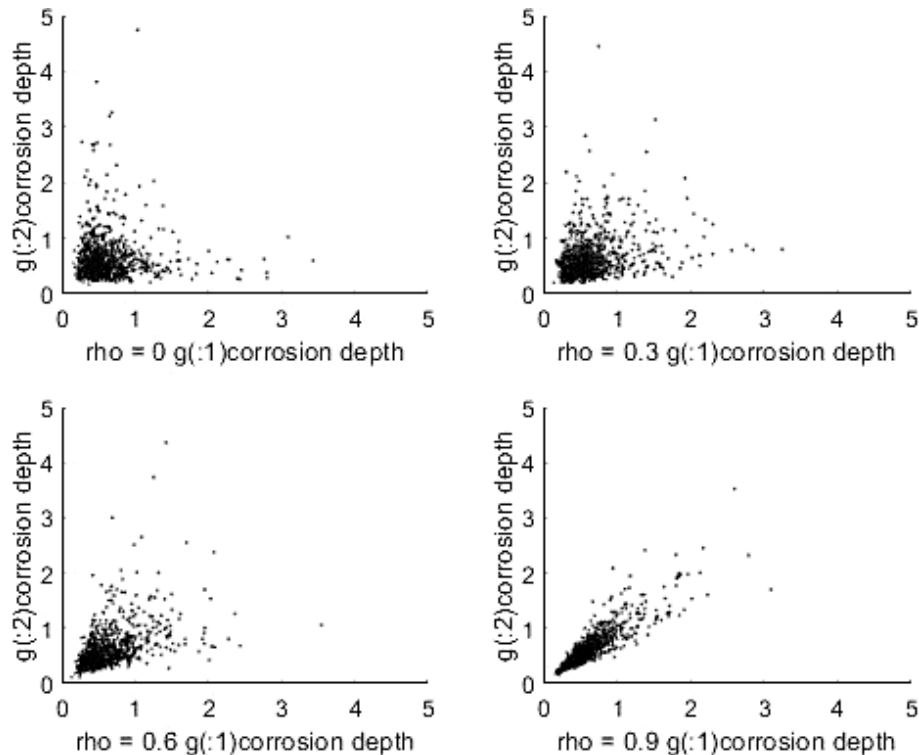


Fig. 4 Distribution of corrosion depth under different correlation coefficients.

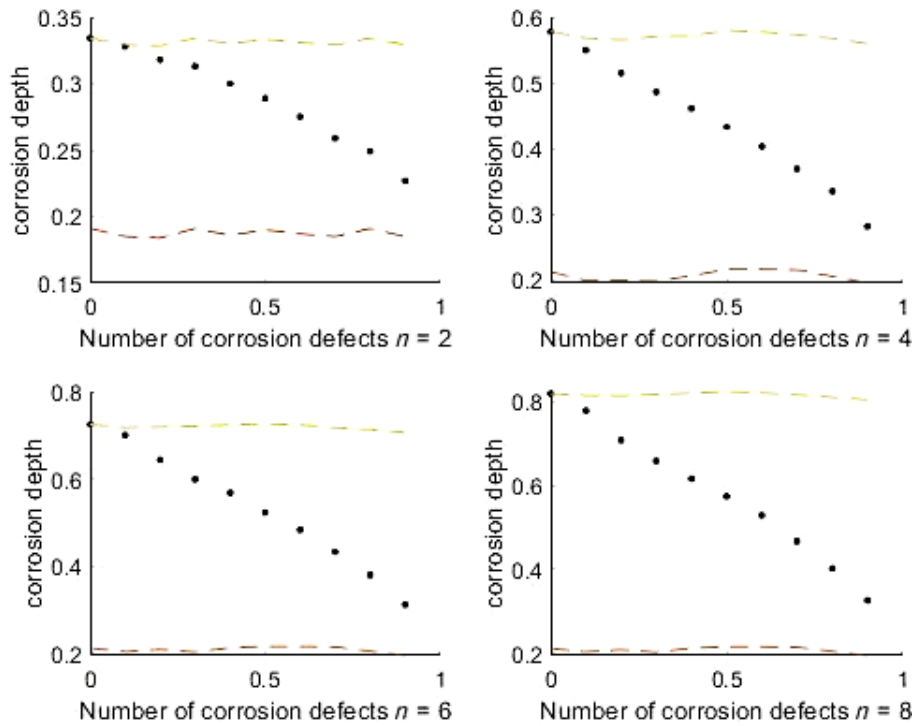


Fig. 5 Trend diagram of system failure probability with correlation coefficient of pipeline corrosion defects.

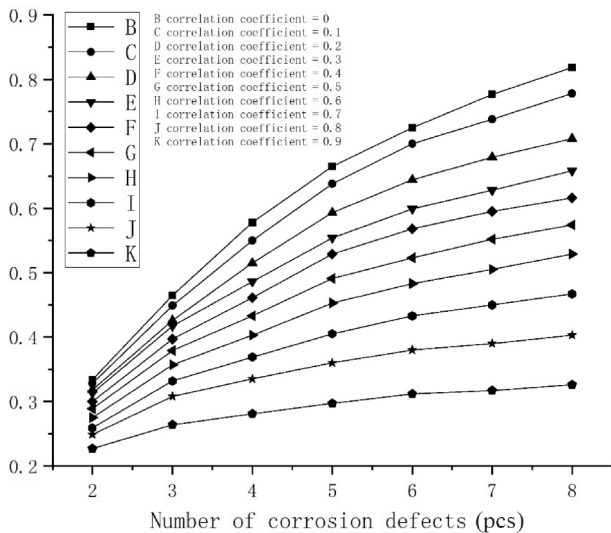


Fig. 6 Trend diagram of system failure probability with pipeline corrosion defects.

pipeline is the same, $P_{f_1} \cap P_{f_2}$. The larger the value of, P_{f_1} and P_{f_2} . The value of is basically unchanged, so the smaller P_f is. When the correlation coefficient is less than 0.3, the system failure probability is not far from the upper limit, so it can be used.

$1 - \prod_{i=1}^i (1 - P_{f_i})$ The formula is used to estimate the system failure probability of a pipeline. When the correlation coefficient increases with the number of defects, using $1 - \prod_{i=1}^i (1 - P_{f_i})$ will make the difference between the calculated failure probability

and the system failure probability of partial correlation getting to increase. Therefore, when the correlation coefficient is increasing, there is a big error in calculating the failure probability of a pipeline according to the series system with independent defects compared with the actual situation.

The influence of the number of defects on the probability of pipeline failure is shown in Fig. 6. When the correlation coefficient is less than 0.6, the failure probability of the pipeline system fluctuates greatly with the increase of the number of corrosion defects. When the number of defects increases from 2 to 8, the pipeline failure probability increases by more than 100%. As the number of corrosion defects increases with the correlation of corrosion defects, the impact on the system failure probability becomes smaller.

Conclusions

In this paper, the GEV model is adopted, and the corresponding extreme value distribution model is selected according to the simulated parameters. Through graphic verification and MATLAB software data fitting, it can be concluded that the threshold parameters $\eta = 0.2294 > 0$, the corrosion depth of pipeline in this paper obeys Frechet distribution.

The method used in this paper calculates that the pipeline failure probability falls within the upper and lower limits of Stevenson-Moses method, indicating that the pipeline corrosion probability calculation method adopted in this paper is effective and reliable when defects are correlated.

The system failure probability of the pipeline decreases with the increase of the correlation coefficient, and the failure probability of the pipeline is always greater than that of the defect correlation when the defects are completely independent. A simple assumption that the defects are independent or

completely correlated can lead to conservative or unsafe estimates. When the correlation coefficient is large, assuming that the pipeline is an independent series system to calculate the failure probability of the pipeline has a large error compared with the actual situation. The system failure probability of the pipeline increases with the increase of the number of assumed corrosion defects. When the correlation coefficient is less than 0.6, the failure probability of the pipeline system is greatly affected by the change in the number of corrosion defects, and the impact of the number of corrosion defects on the system failure probability is less with the increase of the correlation of corrosion defects.

In this paper, the Copula function is used to generate data sets based on existing pipeline corrosion samples to simulate and calculate the failure probability of pipeline system, laying a foundation for further research on pipeline corrosion defects and the correlation between the surrounding soil environment and other factors, and putting forward a new idea and calculation method.

Author contributions

The authors confirm their contribution to the paper as follows: study conception and design: Wang C; data collection: Wang C, Zhang L, Gang Tao; analysis and interpretation of results: Wang C; draft manuscript preparation: Wang C. All authors reviewed the results and approved the final version of the manuscript.

Data availability

All data generated or analyzed during this study are included in this published article.

Acknowledgments

Thanks to the two teachers for their advice and help in the process of writing this paper, thanks to each person who provided help in publishing this paper.

Conflict of interest

The authors declare that they have no conflict of interest.

Dates

Received 15 November 2023; Accepted 7 February 2024;
Published online 21 March 2024

References

1. Jia ZY. 2023. Residual life prediction and case analysis of pitting pipeline based on generalized extreme value distribution. *Petroleum Engineering Construction* 49(3):54–59,81 (in Chinese)
2. Xi ZS. 2020. *Residual life prediction of oil and gas pipeline based on Bayesian survival analysis*. Thesis. Xi'an Building University of Science and Technology, China.
3. Weng YJ. 2002. Distribution characteristics of corrosion test data of regional soil tests -- I characterization of probability distribution. *Corrosion Science and Protection Technology* 14(5):249–52 (in Chinese)
4. Yu X. 2020. *Study on the Construction and Application of binary Copula function*. Thesis. Sichuan Normal University, China.
5. Wang YF, Su C, Xie MJ. 2020. Remaining useful life prediction of corroded oil pipelines based on binary inverse Gaussian process. *Journal of Southeast University (Natural Science Edition)* 50(6):1038–44 (in Chinese)
6. Zhang J, Zhang PY, Yu WC, Diao F, Wang SJ, et al. 2020. Structural reliability evaluation method of corroded pipelines with failure data. *Safety and Environmental Engineering* 27(3):185–92 (in Chinese)
7. Zeng HL. 2011. *Reliability Evaluation of corroded Pipeline based on defect Correlation*. Thesis. Lanzhou University of Technology, China.



Copyright: © 2024 by the author(s). Published by Maximum Academic Press on behalf of Nanjing Tech University. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.