Research on SRB Corrosion Test Optimization and Internal Corrosion Rate Prediction Model for Shale Gas Gathering Pipeline

Qiwen Gong and Xingyu Peng*

Southwest Petroleum University, Chengdu, Si Chuan, China
Email: 2158140050@qq.com (Q.G.); pengxy1949@163.com (X.Y.P.)
*Corresponding author
Manuscript received April 23, 2025; accepted June 3, 2025; published August 8, 2025

***Abstract*—Internal corrosion caused by Sulfate-Reducing Bacteria (SRB) poses a critical challenge to shale gas gathering pipelines. This study combines experimental analysis and a Bayesian network-based corrosion probability model to explore SRB corrosion mechanisms and predict corrosion rates. Experimental results demonstrate a direct correlation between SRB concentration and accelerated corrosion rates. The proposed Bayesian model integrates factors such as SRB concentration, temperature, flow rate, and pH, validated against field data with relative errors <35%. This research provides theoretical support for pipeline maintenance strategies.**

***Keywords*—shale gas pipeline, SRB detection, microbial corrosion, bayesian network**

I. Introduction

Sulfate-reducing bacteria are widely distributed parthenogenetic anaerobic bacteria that use organic matter or iron as an electron donor and sulfate as an electron acceptor. It reduces sulfate to H_2S and induces chemical, electrochemical and other chemical reactions that ultimately cause corrosion of metals [1].

Sulfate-Reducing Bacteria (SRB)-induced internal corrosion represents a significant threat to the integrity and longevity of shale gas gathering pipelines, driven by complex microbial-electrochemical interactions under anaerobic conditions. While existing studies have highlighted SRB's role in accelerating corrosion, limitations persist in quantitatively linking microbial activity to corrosion rates and integrating multifactorial influences (e.g., temperature, pH, flow dynamics) into predictive frameworks. This study addresses these gaps by synergizing experimental SRB concentration analysis with a Bayesian network-based probabilistic model, which holistically evaluates critical corrosion drivers and predicts failure risks. Validated against field data with relative errors below 35%, the proposed model offers a robust tool for optimizing pipeline maintenance and enhancing operational safety in shale gas production systems.

II. Literature Review

The development of shale gas abroad is relatively early for domestic. According to the research, more literature has been studied and found that the common types of shale gas pipeline corrosion are CO_2 corrosion and SRB corrosion. Ye [2] studied the synthetic behavior of N80 steel in CO_2 saturated formation water by weightlessness test, electrochemical test and surface characterization. The results showed that the corrosion rate of N80 steel showed a trend of slowing down and then slowing down with increasing temperature, reaching a maximum around 60 °C. Fatah [3] and others developed an empirical equation for SRB corrosion based on metabolizing species, and used the equation to predict the depth of pitting corrosion caused by SRBs. Yang [4] applied water sample analysis and XRD analysis to the buried pipeline from the sink skid of the Jiao Shibei. A gas gathering station to the production separator, bacterial culture and other methods, and found that the main cause of pipeline perforation is microbial corrosion.

According to the research results, it is known that different operating conditions and pipeline materials will produce different types of corrosion, therefore, in order to effectively guarantee the safe operation of shale gas gathering and transportation system, the use of corrosion failure probability prediction method, so as to effectively grasp the corrosion state and parameters of the gathering and transportation pipeline, is the most important to guarantee the shale gas gathering and transportation pipeline and even the whole gas field's safe, economic and efficient operation.

III. Bacterial Counting Experiments

A. Experimental Materials

The water collected at the separator of A# and B# gathering pipeline of a shale gas field in Sichuan Province was used as the experimental water samples, which were separated, purified, enriched and cultured to obtain the experimental strains. Part of the experimental water samples and reagent bottles required for the experiment are shown in Figs. 1 and 2.

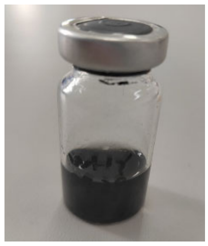

Fig. 1. Experimental water samples.

Fig. 2. Reagent bottles required for the experiment.

B. Experimental Apparatus

The main instruments required for the bacterial counting experiment mainly include high temperature and high pressure sterilizer (XFS-280CB), standard two-person single-side purification workbench (SW-CJ-2FD), and electrically heated constant temperature incubator (303-00AB), and the diagrams of the experimental instrumentation are shown in Figs. 3–5.

Fig. 3. Autoclave sterilizer.

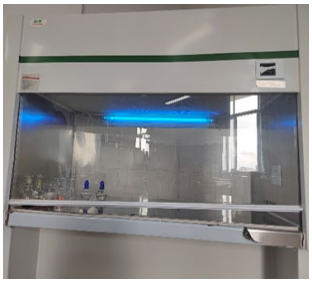

Fig. 4. Ultra-clean bench.

Fig. 5. Electrothermal constant temperature incubator.

C. Experimental Methods

1) Dilute the sample

SRB (Sulfate Reducing Bacteria) Bacterial strain determination was performed using the national standard of the People's Republic of China GB/T14643.5-2009 [5] "Determination of Bacterial Algae in Industrial Circulating Cooling Water Part 5: Determination of Sulfate Reducing Bacteria MPN Method". Dissolve the above solvents in 1L of water, adjust the pH to between 7.0 and 7.4 with NaOH solution or NaCl solution, and dispense in 500 mL graduated triangular bottles of no more than 350 mL each, with the mouths of the bottles stuffed with cotton and wrapped in kraft paper, and sterilize them with an autoclave at 120 °C for 15 min. (Pay attention to the sterilization of the graduated pipettes, test tubes, and sampling bottles.), The medium composition is shown in Tables 1 to 3:

Table 1. SRB liquid medium components

Reagent Name	Content (g/L)	Reagent name	Content (g/L)
Dipotassium hydrogen phosphate	0.5	Magnesium sulfate	2.0
Sodium sulfate	0.5	Sodium lactate	3.5
Ammonium chloride	1.0	Yeast juice	1.0
Calcium Chloride	0.1	-	-

2) Preparation of IOB (Iron Oxidizing Bacteria) liquid medium

Table 2. IOB liquid medium composition

Reagent Name	Content (g/L)	Reagent Name	Content (g/L)
Magnesium sulfate	0.5	calcium chloride	0.2
Ammonium sulfate	0.5	Sodium nitrate	0.5
Dipotassium hydrogen phosphate	0.5	Ammonium iron citrate	10.0

3) Preparation of TGB (Putrefactive Bacteria) liquid medium

Table 3. TGB liquid medium composition

Reagent Name	Content (g/L)	Reagent Name	Content (g/L)
Sodium chloride	5	peptone (biochemistry)	5
Beef paste	3	-	-

D. Analysis of Experimental Results

The presence of the strain is indicated by the production of a black precipitate accompanied by the odor of hydrogen sulfide, indicated by "+" (positive), and the rest of the test tubes are indicated by "–" (negative). It should be noted that if the blank sample shows a positive reaction, it indicates that there is contamination in the assay process and the assay is invalid. Figs. 6–8 show the pictures of enrichment of bacteria after 14 days of culturing the extracted liquid from A# gathering pipeline.

The Table 4 below shows the statistics of the three bacterial counts after 14 days of incubation.

Fig. 6. Results after 14 d of 10 TGB enrichment.

Fig. 7. Results after 14 d of IOB enrichment.

Fig. 8. Results after 14 d of 12 SRB enrichment.

Table 4. Statistical results of each bacterial count

Pipeline number	SRB (Sulfate Reducing Bacteria)	IOB (Iron Oxidizing Bacteria)	TGB (saprophytic bacteria)
A#	+ + + 2.5×10^5 (pcs/mL)	− − − 3.0×10^2 (pcs/mL)	+ − − 1.2×10^2 (pcs/mL)
B#	+ + + 1.5×10^5 (pcs/mL)	+ − − 2.8×10^2 (pcs/mL)	+ + − 4.0×10^3 (pcs/mL)

According to the results of bacterial counting, three kinds of bacteria, SRB, IOB and TGB, exist in A# and B# gathering pipelines, of which SRB has the highest number and IOB has the lowest number. SRB is anaerobic bacteria, and IOB and TGB are aerobic bacteria, of which IOB and TGB consume O2 when metabolizing, thus providing an anaerobic environment for SRB, and it can be speculated that microbial corrosion of A# and B# gathering pipelines is mainly dominated by SRB, which is the main cause of microbial corrosion.

IV. A Bayesian Network-based Probabilistic Model for Corrosion

A. Bayesian Network Overview

Bayes' theorem relates the posterior probability of an event (i.e., the probability of an event after observation) to the prior probability of the event (i.e., the probability before observation), the probability of the observation event, and the conditional probability of the observation event given the occurrence of the event [6]. This can be expressed mathematically as shown in Eq. (1):

$$P\left(A_i \middle| B\right) = \frac{P\left(B \middle| A_i\right) P\left(A_i\right)}{\sum_{j=1}^{n} P\left(B \middle| A_j\right) \cdot P\left(A_j\right)} \tag{1}$$

where $P\ (A_i|\ B)$ is the posterior probability of the event A_i given the observation B; $P\ (A_i)$ is the prior probability of the event A_i before observation B; $P\ (B|\ A_i)$ is called the likelihood function and is the probability of observing B given the occurrence of event A_i.

In Eq. (1), the denominator represents the probability of the observation event and is obtained by summing the conditional probabilities of the event given by B. The prior probability and posterior probability can be thought of as the "cause" and "effect" of a process, respectively. Bayesian Networks are often depicted graphically, with multiple random variables connected by causal relationships.

B. Constructing Bayesian Network Model

In this study, the 1# gathering pipeline of a shale gas field in Sichuan is used as an engineering example to develop a probabilistic model. To construct the Bayesian Network structure, a causal relationship between variables is considered, making it simpler to evaluate the probability distributions among variables. Therefore, the directed edges between variables in the Bayesian Network constructed in this study represent causal relationships [7].

1) Node selection and variable set setting

Based on the investigation of the 1# gathering and transportation pipeline, this study found that pipeline corrosion mainly includes uniform corrosion (including CO_2 corrosion), local corrosion, and SRB corrosion. Table 5 presents the main corrosion influencing factors for each corrosion mode. Each influencing factor corresponds to a node in the Bayesian Network.

Table 5. Main influencing factors for various corrosion modes

Corrosion mode	Main influencing factors
Uniform corrosion	T, CO_2, O_2
Local corrosion	Gas velocity, gas density, fluid velocity, liquid holdup, pipe inclination
SRB corrosion	SRB quantity, T, CO_2 Partial pressure, pH, Cl^-, SO_4^{2-}, Ca^{2+}, HCO_3^-

2) Bayesian network learning

Bayesian Network learning consists of two main types: structure learning and parameter learning. Structure learning determines the most suitable network structure based on existing data, while parameter learning focuses on estimating the conditional probability table for each node after the structure is established [8].

a) Structure learning

This process aims to identify the appropriate network structure based on the study's nature. It can be performed with either complete or incomplete historical data. With complete data, structure learning can be achieved through methods like Matlab programming or the K2 algorithm. In cases of incomplete data, the missing values must be imputed before structure learning can proceed. For this study on the 1# gathering pipeline, the structure learning process involved three steps: (a) initializing incomplete data by random generation, (b) refining data and estimating network structure using parameter methods, and (c) iterating between structure learning and data refinement until the optimal structure is found [9].

b) Parameter learning

Parameter learning involves estimating the probability distribution of the Bayesian Network parameters from available data. With comple102te data, the Maximum Likelihood Estimation (MLE) is used. For incomplete data, methods like the Monte Carlo, Gaussian approximation, or Expectation-Maximization (EM) algorithms are employed. This study utilized the EM algorithm for parameter learning. The log-likelihood function of the estimated parameter θ [10]:

$$L(\theta)=\sum_{m}\log\sum_{n}P\left(X=x_i,Y=D\right) \tag{2}$$

This can be expressed mathematically as shown in Eq. (2): where X represents the unknown hidden variable, Y represents the observed variable, and D is the known training set.

The definition q (X=x| Y) is a probability function taking any value of [0,1], representing the X=x probability of observing the value of Y, so Σ xxq(X=x| Y) = 1.

Rewrite L as a function of q (X| Y), suppose P (X=x, Y=D) is a convex function with extreme values, then according to Jensen's inequality, as shown in Eq. (3):

$$L=\sum_{m}\log\sum_{n}q(X|\,Y)\times\frac{P(X=x,Y=D)}{q(X|\,Y)}\geq\sum_{m}\sum_{n}q(X|\,Y)\times\log\frac{P(X=x,Y=D)}{q(X|\,Y)} \tag{3}$$

It follows that, first of all, a lower bound on L can be given for a particular q (X| Y), so when the function q (X| Y) is known, the maximum value of L can be obtained theoretically. Second, in order for L to reach its expected maximum value, which is the lower bound, it is necessary to adjust q (X| Y), in effect, the parameters θ of L itself, so that:

$$L=\sum_{m}\sum_{n}q(X|\,Y)\times\log\frac{P(X=x,Y=D)}{q(X|\,Y)} \tag{4}$$

As shown in Eq. (4), the EM algorithm corresponds to the above two processes through E and M steps. Expectation Step denotes the Expectation of finding the maximum value of L when we know. q (X| Y). Maximization Step, for a particular parameter θ, is modified q (X| Y) so that L exists.

The EM algorithm is a continuous iterative process, getting a lower bound on the function based on the current parameter in step E, updating the parameter in step M, thus modifying q (X| Y) continuously to update the lower bound and approximating the maximum value of L itself.

The function q (X| Y) is actually a process of repairing the hidden variables through the observed values. Although X cannot be obtained directly from the training set, the function Q can give the probability of a set of hidden variables obtained from the current samples. In the iterative process of EM algorithm, if θ_t represents the parameters of the current iteration, and θ_{t+1} represents the parameters of the next iteration, because the observed value Y will not change, the essence of solving the lower bound of L is to solve the parameters of the next iteration when the current parameters are given. Thus noting q (X| Y) =P (X| θ_t), the lower bound of function L is Q (θ_{t+1}| θ_t), called the expected likelihood log function, and the transformation can be obtained as follows:

$$Q\left(\theta_{t+1}|\,\theta_t\right)=\sum_{m}\sum_{n}P\left(X|\,\theta_t\right)\times\log P\left(X,D|\,\theta_{t+1}\right) \tag{5}$$

This can be expressed mathematically as shown in Eq. (5): For different models, EM algorithm needs to select different expected likelihood functions. For the Bayesian Network model, the same logical likelihood function proposed in MLE is used to obtain the expected logical likelihood function.

$$Q\left(\theta_{t+1}|\,\theta_t\right)=\sum_{i}^{N}\sum_{j}\sum_{k}^{M}P\left(X|\,\theta_t\right)\log\theta_{ijk}\chi\left(x,j,k,D_l\right) \tag{6}$$

As shown in Eq. (6), the process of calculation Q (θ_{t+1}| θ_t) corresponds to step E in EM. EM algorithm is an iterative convergence process, will stop when θ_{t+1} and θ_t are equal, but in practice, computers use floating-point representation for computation, which can lead to slight variations in results due to limitations in memory precision. Therefore, when using similarity algorithms that converge, a threshold must be set to determine when the difference between the two values is small enough to stop the algorithm. Additionally, iterative algorithms are often influenced by the initial values of the parameters, so it is necessary to provide appropriate initial values to ensure optimal results.

3) Building bayesian networks

In this paper, a causal graph consisting of 38 nodes is constructed based on existing experience and all Bayesian network models are modelled and manipulated. The modeling process employed a two-stage method, and the Bayesian Network inference was performed using the variable elimination method. The Bayesian Network models were constructed using expert knowledge and experimentally obtained data, shown in Fig. 9.

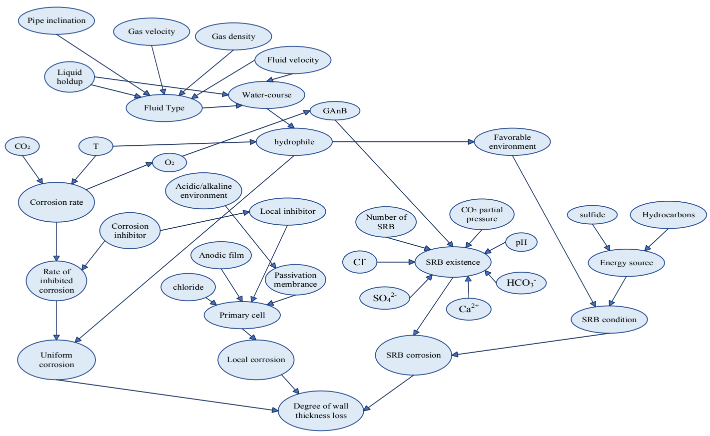

Fig. 9. The established Bayesian Network model.

4) Determine the prior probability of the root node

In a Bayesian Network, the prior probability of the root node refers lity derived from historical statistics or expert analysis. This probability is only influenced by factors directly related to the root node and is not affected by external factors. When a large amount of statistical data is available, the prior probability of the root node can be determined using machine learning. There are two methods for determining the prior probability of the root node: manual establishment based on expert knowledge and experience, and establishment based on data structure. In this paper, a Bayesian Network was constructed by combining expert knowledge and data structure [9].

This paper classifies the root node states into five levels: very high (VH), high (H), medium (M), low (L), and very low (VL). The criteria for these occurrence probability levels are provided in Table 6.

Table 6. Criteria of levels of root node's occurrence probability

Probability level	Probability interval	Center of probability	State description
1	(0,0.2]	0.1	The probability of corrosion during the service life of the pipeline is very low (VL)
2	(0.2, 0.4]	0.3	The probability of corrosion during the service life of the pipeline is low (L)
3	(0.4, 0.6]	0.5	The probability of corrosion during the service life of the pipeline is medium (M)
4	(0.6, 0.8]	0.7	The probability of corrosion during the service life of the pipeline is high (H)
5	(0.8, 1]	0.9	The probability of corrosion during the service life of the pipeline is very high (VH)

Table 7. Prior probabilities of some variables based on historical data

Node variables	VL	L	M	H	VH
T	0.81094258	0.11584678	0.06225149	0.01090552	0.00005363
CO_2	0.82307614	0.11428207	0.04538675	0.01640020	0.00085484
Number of SRB	0.82672667	0.11026451	0.02004984	0.04241256	0.00054643
CO_2 partial pressure	0.75998897	0.19264816	0.02712414	0.02001586	0.00022287
PH	0.77763944	0.14985359	0.05720090	0.01530010	0.00000596
HCO_3^-	0.80329444	0.10381341	0.07930264	0.01351052	0.00007898
Ca^{2+}	0.80216057	0.10006297	0.06883110	0.02893526	0.00001010
SO_4^{2-}	0.81127319	0.11844734	0.05266886	0.01674582	0.00086479
Cl^-	0.82761467	0.13836611	0.02354764	0.01025681	0.00021477

C. *Predicting and Analyzing Pipeline Corrosion Probability*

Now that the nodes, network structure and inference mode have been determined, the constructed conditional probability Table 7, the causal reasoning results of Bayesian Network, along with the probability distribution of wall thickness loss, are shown in Figs. 10–12, respectively for illustration.

Fig. 10 shows the conditional probabilities obtained from Eq. (6) using Bayesian inference.

Fig. 11 shows the model prediction built by Bayesian Network to calculate the interface of the probability value of the degree of node wall thickness loss.

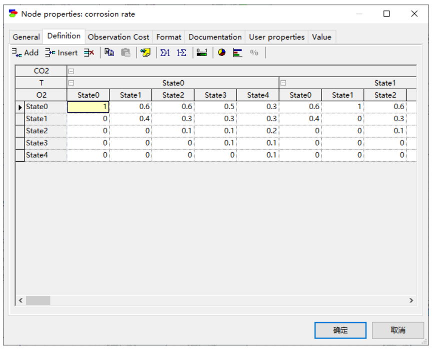

Fig. 10. Construction of conditional probability table of corrosion rate nodes.

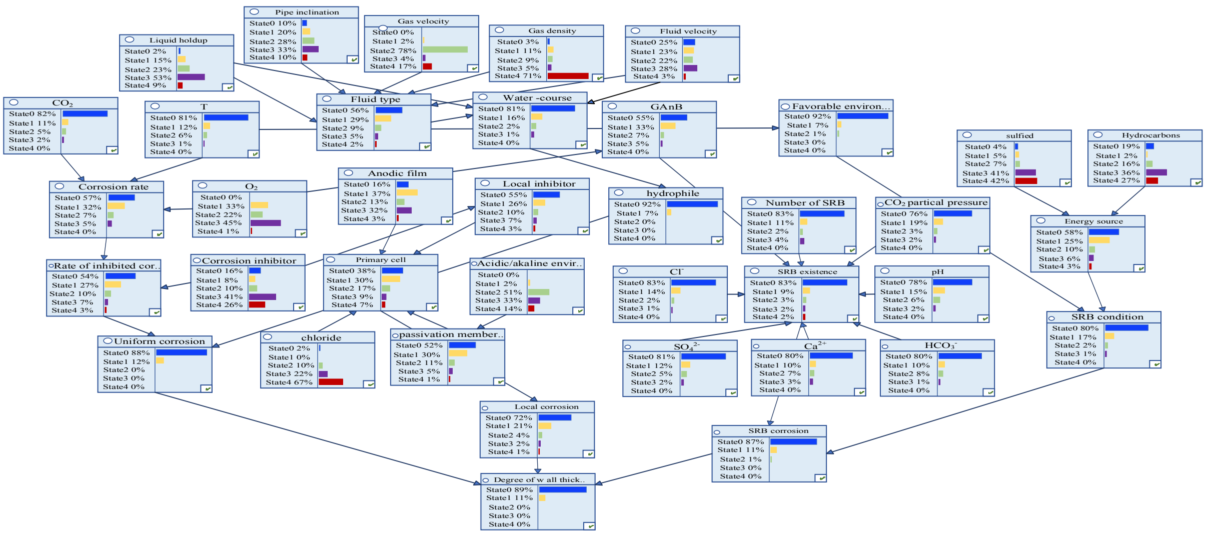

Fig. 11. The causal reasoning results of Bayesian Network.

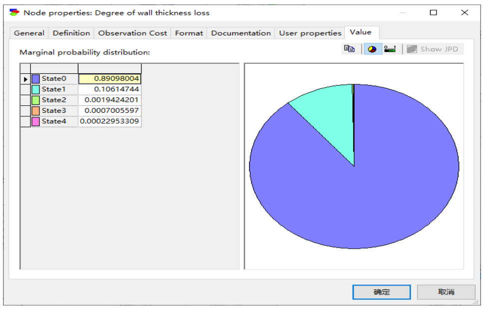

Fig. 12. The probability distribution of wall thickness loss.

Fig. 12 shows the detailed calculation of the probability of corrosion failure corresponding to the degree of wall thickness loss, as predicted by the model built using Bayesian Network.

D. Case Study

Through the constructed Bayesian network to get A #, B # two gathering pipeline node wall thickness loss degree of corrosion degree calculation results, to get each pipe section corresponding to the probability of corrosion failure, and with the field situation of corrosion failure probability comparison, the results are shown in Figs. 13 and 14.

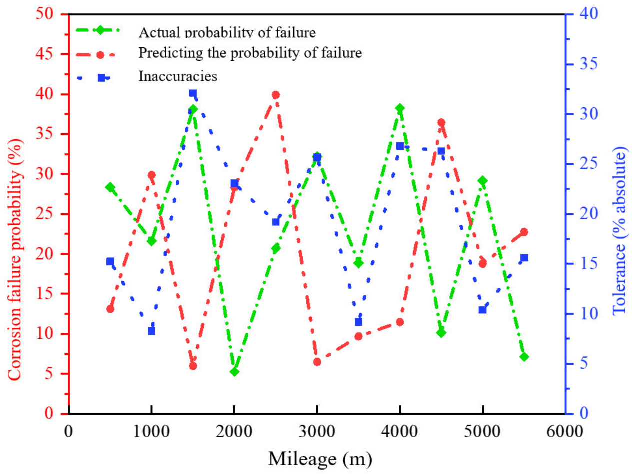

Fig. 13. Comparison analysis results between predicted and measured values of pipeline A#.

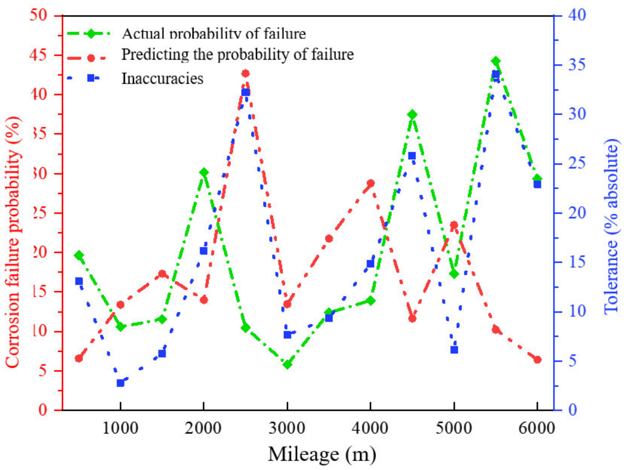

Fig. 14. Comparative analysis results of predicted and measured values of pipeline B#.

The model predicts corrosion failure probabilities for A# and B# gathering pipelines with small errors compared to field data. For A#, 6 segments have errors below 20%, while 5 segments range from 20% to 35%, with the largest error at 32.15% in segment No. 3. For B#, 8 segments have errors below 20%, and 4 segments range from 20% to 35%, with the largest error at 34.08% in segment No. 22. Overall, the Bayesian network-based model predicts pipeline corrosion with good accuracy.

V. Conclusion

Bacterial analysis revealed three types of bacteria—SRB, IOB, and TGB—in both A# and B# pipelines. SRB, being anaerobic, was the most abundant and likely the main cause of microbial corrosion, while IOB and TGB, aerobic bacteria, create an oxygen-deprived environment that supports SRB growth. Thus, microbial corrosion is primarily driven by SRB.

A Bayesian network causality diagram was developed using GeNIe software, combining expert knowledge, historical data, and corrosion factors identified through orthogonal experiments. The model used GMM and input parameters to generate conditional probability tables, creating uniform and pitting corrosion failure probability models for the pipelines.

Corrosion failure probabilities were predicted for 23 pipe sections of A# and B# pipelines. Comparing with field data, 6 segments in A# had errors below 20%, 5 segments had errors between 20% and 30%, with the largest error (32.15%) in segment No. 3. For B#, 8 segments had errors below 20%, 4 segments had errors between 20% and 30%, and the largest error (34.08%) was in segment No. 22. Despite some larger errors, the overall accuracy supports the Bayesian network model's effectiveness in predicting pipeline corrosion failure.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Qiwen Gong completed the following two works and articles:

1. Experimental research: through experimental analysis, a direct correlation between SRB concentration and corrosion rate acceleration was found.

2. Bayesian network model construction: combined with the experimental data, the paper proposes a corrosion probability prediction model based on Bayesian network.

The rest of the content is completed by Peng Xingyu for model validation and application: the model is validated by comparing with the field data, and the relative error obtained is less than 35%, which proves that its prediction effect is good, and provides theoretical support for pipeline maintenance strategy. All authors had approved the final version.

References

[1] G. Yun, W. Yiping, H. Meandan *et al.*, "Review of China's natural gas development in 2022 and outlook for 2023," *Natural Gas Technology and Economy*, vol. 17, no. 1, pp. 1–10, 2023.

[2] Z. R. Ye, Z. C. Qiu, R. Y. Sui *et al.*, "Effect of temperature on corrosion behaviour of N80 steel in saturated formation water," in *IOP Conference Series: Materials Science and Engineering*, vol. 504, no. 1, 2019.

[3] M. C. Fatah, M. C. Ismail, and B. Wahjoedi, "Empirical equation of Sulphate Reducing Bacteria (SRB) corrosion based on abiotic

chemistry approach," *Anti-corrosion Methods and Materials*, vol. 60, no. 4, pp. 206–212, 2013.
[4] C. Yang, "Risk analysis and preventive measures of pipeline leakage in Fuling shale gas field," *China Petroleum and Chemical Standards and Quality*, vol. 39, no. 2, pp. 70–72, 2019.
[5] *Method for the Determination of Bacteria and Algae in Industrial Circulating Cooling Water Part 5: Determination of Sulfate-Reducing Bacteria MPN Method*, GB/T 14643.5-2009 [S].
[6] S. Wang, S. Lin, and Y. Lu, "Causal analysis with Bayesian networks," *Computer Science*, vol. 27, no. 5, pp. 1–3, 2000.
[7] W. Chen and Q. Liao, "Research on Bayesian network modeling based on genetic algorithm," *Computer Engineering and Design*, vol. 30, no. 1, pp. 188–190, 2009.
[8] K. Cheng, J. Yao, Z. Cheng *et al.*, "Prediction method of pipeline corrosion depth based on correlation and Bayesian inference," *Oil and Gas Storage and Transportation*, vol. 40, no. 8, pp. 854–859, 2021.
[9] Q. Hu, T. Sun, F. Wang *et al.*, "Bayesian-based corrosion depth prediction method for gas pipelines," *Oil and Gas Storage and Transportation*, vol. 38, no. 11, pp. 1219–1226, 2019.
[10] Z. Li and C. Xiong, "Dynamic assessment of corrosion damage in submarine pipelines based on Bayesian inference," *Engineering Mechanics*, pp. 1–10, 2025. http://www.engineeringmechanics.cn/CN/10.6052/j.issn.1000-4750.2023.12.0999