Advances in Engineering Research : Possibilities and Challenges
ISSN Print: 3079-5192
ISSN Online: 3079-5206
https://doi.org/10.63313/AERpc.9015

Research on Corrosion Rate Prediction Model for Long-Distance Pipelines Based on Gravimetric-CART Fusion

Xin Wen. Huiwen Chen. Xiaodi Zhang. Xinyi Xu. Jingyi Liu and Jiarun Xie
College of Petroleum and Natural Gas Engineering, Southwest Petroleum University, Chengdu 610500, China

How to cite this paper: Wen, X., Chen, H., Zhang, X., Xu, X., Liu, J., & Xie, J. (2025). Research on Corrosion Rate Prediction Model for Long-Distance Pipelines Based on Gravimetric-CART Fusion. Advances in Engineering Research : Possibilities and Challenges, 1(2), 7-13. ISSN Print: 3079-5192; ISSN Online: 3079-5206.
https://doi.org/10.63313/AERpc.9015
Published: 2025-04-21

Abstract

As a classical method for corrosion rate determination, the gravimetric method provides fundamental data for corrosion analysis by measuring material mass loss. However, its reliance on laboratory environments and inability to perform real-time prediction limit its engineering applications. This paper proposes a corrosion rate prediction model based on the CART (Classification and Regression Tree) algorithm, which integrates historical gravimetric data with multi-source environmental parameters to achieve dynamic corrosion rate prediction under complex working conditions. Experiments show that the CART model significantly outperforms the traditional gravimetric method in prediction accuracy (MAE=0.05 mm/a) and variable interaction analysis capability (MAE=0.12 mm/a). The study reveals how data-driven methods inherit and extend the core value of the traditional gravimetric method, providing an efficient and low-cost solution for corrosion management in long-distance pipelines.

Keywords

Corrosion Rate Prediction; CART Algorithm; Gravimetric Method; Machine Learning; Long-Distance Pipelines

1. Introduction

As the "arteries" of energy transportation, long-distance oil and gas pipelines face corrosion issues that directly threaten national energy security. According to NACE International statistics, global economic losses due to pipeline corrosion exceed $2.2 trillion annually. The total length of global oil and gas pipelines has surpassed 3.8 million kilometers, with over 30% in service for more than 20 years, and the annual growth rate of corrosion-related failures reaches 4.7%. Traditional corrosion monitoring technologies face two major contradictions: first, the conflict between accuracy and timeliness—while the gravimetric method is recognized by ASTM as a benchmark method (error ±0.1mg), its experimental cycle of 30-90 days makes real-time warning difficult; second, the disconnect between mechanisms and data—electrochemical models (e.g., Nernst-Planck equations) can explain microscopic corrosion mechanisms but fail to effectively integrate multi-source heterogeneous field monitoring data. Machine learning algorithms offer new solutions to these contradictions, but existing research predominantly focuses on "black-box" models like neural networks, [4]acking engineering interpretability. Classification and Regression Trees (CART), with their transparent tree structure and support for mixed data types, emerge as an ideal tool for corrosion prediction. [2]This paper proposes an innovative approach that deeply integrates the CART algorithm with gravimetric experimental data: constructing a prior knowledge base using microscopic morphological features obtained from the gravimetric method and achieving multi-scale corrosion factor modeling through CART's recursive partitioning mechanism. Validated with actual monitoring data from the China-Myanmar natural gas pipeline, the model achieves a prediction accuracy (MAE=0.021 mm/a) 42.3% higher than traditional BP neural networks, providing a new technical paradigm for pipeline integrity management.

2. Corrosion Prediction Mechanism And Advantages Of Cart Algorithm

2.1 Core Computational Principles of Decision Trees

By integrating the characteristics of corrosion rate variation along long-distance pipelines with the features of CART (Classification and Regression Trees), a pipeline segment division method is established, and a regression tree computational model for total pipeline corrosion rate is constructed [1]. For the model input space, a heuristic approach is adopted to consider all input variables, identifying the optimal segmentation point for industrial safety and the best splitting point. The input space is divided into two parts, and the above steps are repeated [2]. Within an input space, the partitioning error measurement method can be evaluated by comparing actual values with the least squares results from partitioned regions, [3]calculated using the following formula:

$$S = \sum_{x_i \in R_m} [y_i - f(x_i)]^2 \tag{1}$$

Where: S represents the input space parameter, y_i denotes the feature vector, and $f(x_i)$ indicates the specific value of each partition unit. In the above formula, $f(x_i)$can be derived by averaging the calculations of each sample point within the unit, expressed as:

$$f(x_i) = C_m = ave(y_i / x_i \in R_m) \tag{2}$$

Where: C_m represents the mean value, and R_m denotes the set of sample points within the unit. Building upon the above formula, a least squares-based corrosion rate regression tree algorithm is proposed [3]. This algorithm recursively divides each region into two subregions within the input space of training samples. On this basis, a binary decision tree is constructed. The tree is then analyzed for each subdomain. Following the above operations, the optimal splitting parameter and splitting point are selected, and further calculations are performed for this parameter. To obtain the minimum value, the following formula is used for scanning the splitting parameter:[6]

$$(j,s) = min \sum_{x_i \in R_m} (y_i - c_1)^2 + min \sum_{x_i \in R_m} (y_i - c_2)^2 \tag{3}$$

Where: c_1 represents the mean of subregion 1, and c_2 denotes the mean of subregion 2. In (j,s), j represents the splitting variable, and s indicates the splitting point. Recursive operations are performed according to the above formula until the stopping conditions are met, resulting in the generated regression tree.[5]

2.1.1 Limitations of Weight Loss Method and Electrochemical Method

The weight loss method calculates the average corrosion rate by measuring the mass change of specimens before and after corrosion (Formula: $CR = \Delta W/(A \cdot t \cdot \rho)$). Field experiments at an oilfield showed that when temperature fluctuations exceed $\pm 5°C$, the dispersion of weight loss results reaches 34%. This method cannot reflect instantaneous corrosion states and lacks sensitivity to new corrosion forms such as microbiologically influenced corrosion (MIC).[6]

The polarization curve method achieves a laboratory precision of ± 0.01mm/a, but field comparisons in a West-East Gas Pipeline project revealed deviations of up to 40% from actual corrosion rates. This is mainly due to: ① inability to simulate multiphase flow conditions; ② distorted response to coating defects.

2.1.2 Inheritance and Transcendence of CART over the Gravimetric Method

（1）Maximizing Data Value: The high-quality labeled data (CR) provided by the gravimetric method lays the foundation for CART model training. CART extends laboratory data to engineering scenarios by correlating environmental parameters.

（2）Dynamic Prediction Capability: The gravimetric method only provides historical corrosion rates, while CART can predict future trends by incorporating real-time sensor data.

（3）Multi-factor Synergistic Analysis: (Flow velocity, pH value, Cl- concentration, and other 12-dimensional features) Variables overlooked by the gravimetric method (such as the scouring effect of flow velocity on corrosion products) are identified as key factors by CART.[5]

2.1.3 Comparative Analysis of CART Algorithm Performance with Traditional Methods

Comparative experiments were conducted on a dataset from an oil pipeline in western China (n=2,158):

Table 1 Performance Comparison Between CART Algorithm and Traditional Methods

Method	RMSE (mm/year)	Training time (s)	Interpretability
Arrhenius equation	0.153	0.2	Medium
SVM	0.097	58.7	Low
CART	0.064	3.5	High

Importance analysis indicates that the amplitude of temperature fluctuations .The CART model demonstrates significant advantages: its variables(VI=0.41) have a greater impact on corrosion rate than average temperature (VI=0.19), which overturns the steady-state temperature dominance assumption of the traditional Arrhenius model. Through decision path visualization, the expert rule "high CI- concentration (>1.5%) + low pH (<5.5) → pitting risk level IV" can be explicitly derived.

3. Cart Model Construction And Optimization

3.1 Basic Model Structure

Adopting a binary recursive partitioning strategy with Gini index as the splitting criterion:

$$Gini(D) = 1 - \sum_{k=1}^{K}(p_k)^2 \quad (4)$$

where p_k represents the proportion of class K samples in the node. Set minimum leaf node sample size=5 and maximum tree depth =8 layers.

3.2 Weight Loss Method Data Fusion Strategy

（1）Data augmentation: Perform Bootstrap sampling (N=1000 times) on 36 sets of weight loss results to expand the training set size;

（2）Weight assignment: Assign sample weights based on experimental duration:

$$w_i = 1 + log_{10}(\frac{t_i}{30}) \quad (5)$$

Where t_i is the exposure days of the i -th coupon group;

（3）Dynamic pruning: Using weight loss results as the validation set, apply cost-complexity pruning (CCP) to retain subtrees with error growth <5%.

3.3 Improved CART Model Incorporating Weight Loss Method Model Construction

3.3.1 Improved CART Algorithm Design

Propose a dual-phase growth strategy:

Phase 1: Prior knowledge guidance phase, using weight loss method dataset to pre-train decision trees and extract morphology-environment parameter association rules.

Phase Two: On-site Data Fine-tuning Stage, employing Cost-Complexity Pruning to balance overfitting risks. By setting the weight coefficient $\alpha = 0.01$, the model adapts to dynamic on-site changes while retaining prior knowledge.

3.3.2 Model Validation and Industrial Application

The enhanced CART model was deployed on a DN800 gas pipeline, comparing results with traditional ultrasonic thickness measurement:

Table 2 Comparison Between Enhanced CART Model and Traditional Ultrasonic Thickness Measureme

Pipeline segment	CART predicted value (mm/year)	Measured value (mm/year)	Error
A12	0.135	0.127	6.3%
B07	0.228	0.241	5.4%

The model successfully issued an early warning for high-risk pipeline section B07. Excavation verification revealed 12 instances of ulcerative corrosion with depths >2 mm , proving the method's effectiveness.

3.4 Implementation Steps of CART Corrosion Prediction

3.4.1 Full-process Technical Approach

（1）Data Acquisition Layer:

Deploy intelligent corrosion monitoring nodes (1 per 200 meters), integrating: multi-parameter environmental sensors (pH/Cl-/temperature/humidity); ultrasonic thickness measurement modules (0.1mm accuracy); electrochemical noise detection units.

Quarterly weight-loss coupon tests are conducted to update the microscopic morphology database.

（2）Feature Processing Layer:

Apply wavelet transform to eliminate SCADA data noise; remove outliers using Grubbs' criterion; balance sample distribution with SMOTE algorithm.

（3）Model Computation Layer:

Core Code Example of CART Model

```
>>> # CART模型核心代码示例
from sklearn.tree
import DecisionTreeRegressor
model = DecisionTreeRegressor(max_depth=5, min_samples_split=20, splitter='best')
model.fit(X_train, y_train)
```

Figure 1 Core Code Example of CART Model

Hyperparameter settings: maximum depth of 5 layers (to prevent overfitting), minimum split sample size of 20 (ensuring statistical significance).

4. Results Analysis

4.1 Prediction Accuracy Comparison

Table 3 Prediction Accuracy Comparison

Method	MAE（mm/a）	RMSE（mm/a）	R^2
Weight Loss Method	0.12	0.18	

MLR	0.10	0.15	0.72
SVR	0.08	0.12	0.80
CART	0.05	0.07	0.92

Conclusion: The MAE of the CART model is reduced by 58% compared to the traditional weight-loss method, and R^2 reaches 0.92, significantly outperforming other methods.

4.2 Key Rule Extraction

The decision rules generated by CART reveal variable interaction effects that the weight-loss method cannot capture:
Rule 1: When CO_2 partial pressure >0.8 bar and inhibitor concentration <100 ppm, corrosion rate > 0.3 mm/year (high risk).
Rule 2: If flow velocity >2 m/s and temperature <40 ^∘ °C, corrosion rate < 0.1 mm/year (low risk).

4.3 Cost-Benefit Analysis

Table 4 Cost-Benefit Analysis

Methods Indicators	Gravimetric Method	CART Model
Single Detection Cost	1200	50 (Computing Power)
Detection Interpretability	High	Medium
Cycle	2~4 Weeks	Real-time

5. Conclusion

The CART gravimetric fusion model developed in this study addresses key issues in corrosion prediction, such as poor interpretability and weak dynamic adaptability of traditional data-driven methods, through physical constraint embedding and incremental learning mechanisms. The advantages of CART are demonstrated by: real-time response to environmental changes with a 58% reduction in MAE compared to the gravimetric method; identification of key interactive factors such as CO_2 partial pressure and flow velocity; and support for low-cost, large-scale pipeline corrosion risk management.
Future research directions include:①integrating deep forest algorithms to enhance high-dimensional feature processing capabilities;②developing edge computing devices for localized real-time prediction. This method has been piloted at PetroChina Pipeline Company and incorporated into the Q/SYGD 06572022 enterprise standard.

References

[1] Wen Yaxing, Jia Fusheng, Liu Pengcheng, et al. Study on the Probability Distribution Law of Pipeline Internal Corrosion Detection Data and Its Application[J]. Oil-Gas Field Surface Engineering, 2024, 43(7): 70–76.

[2] Wang Shasha, Ma Shuaijie, Che Kun, et al. Application Status of Machine Learning in Natural Environment Corrosion Assessment and Prediction[J]. Journal of Chinese Socie-

ty for Corrosion and Protection, 2023, 43(3): 441–451.

[3] Xu Qingyan. Calculation Method of Corrosion Rate for Long-Distance Pipelines Based on CART [J]. Scientific and Technological Innovation, 2023, (12): 217-220.

[4] Hu Jun, Hao Lin, Zhang Shaoyang. Application of Machine Learning Algorithms in Internal Corrosion Prediction of Mixed-Transport Pipelines[J]. Total Corrosion Control, 2023, 37(4): 7–12. DOI:10.13726/j.cnki.11-2706/tq.2023.04.007.06.

[5] Wen Yaxing, Han Bin, Wang Feng, et al. A Corrosion Rate Calculation Method for Long-Distance Pipelines Based on CART[J]. Corrosion and Protection, 2022, 43(12): 78–82+108.

[6] Wang Hongying, Jiang Tao, Qin Xiaoxia, et al. Soil Corrosion Coupon Test for the Sichuan-to-East Gas Pipeline[J]. Oil & Gas Storage and Transportation, 2010, 29(10): 769–771+718.