

## Research Article

# Principal Component Analysis Based Dynamic Fuzzy Neural Network for Internal Corrosion Rate Prediction of Gas Pipelines

Xiaoxu Chen,<sup>1</sup> Linyuan Wang,<sup>1</sup> and Zhiyu Huang <sup>1,2</sup>

<sup>1</sup>College of Chemistry and Chemical Engineering, Southwest Petroleum University, Chengdu 610500, China

<sup>2</sup>Oil & Gas Field Applied Chemistry Key Laboratory of Sichuan Province, Chengdu 610500, China

Correspondence should be addressed to Zhiyu Huang; 623086967@qq.com

Received 7 March 2020; Revised 14 August 2020; Accepted 7 September 2020; Published 17 September 2020

Academic Editor: Bhabani S. Dandapat

Copyright © 2020 Xiaoxu Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the characteristics of the nonlinear changes in the internal corrosion rate in gas pipelines, and artificial neural networks easily fall into a local optimum. This paper proposes a model that combines a principal component analysis (PCA) algorithm and a dynamic fuzzy neural network (D-FNN) to address the problems above. The principal component analysis algorithm is used for dimensional reduction and feature extraction, and a dynamic fuzzy neural network model is utilized to perform the prediction. The study implementing the PCA-D-FNN is further accomplished with the corrosion data from a real pipeline, and the results are compared among the artificial neural networks, fuzzy neural networks, and D-FNN models. The results verify the effectiveness of the model and algorithm for inner corrosion rate prediction.

## 1. Introduction

Due to the influence of the medium composition, temperature, terrain, and other factors, corrosive substances are easily produced in steel gas pipelines, which can lead to internal corrosion. Internal corrosion is one of the causes of aging in natural gas pipeline systems. Corrosion will cause a thinning of the inner wall of the pipeline and reduce its structural strength, which will lead to natural gas leakage and seriously threaten the safety, integrity and economy of the whole gas transmission system [1–3]. To prevent these phenomena, some in-line inspection instruments and internal detection instruments have been developed. However, these methods are not only complex but also costly. It has been reported that less than 50% of the worldwide existing pipelines can be inspected with in-line inspection instruments [4]. For small-diameter pipelines, it is difficult to carry out internal detection with commonly used internal detection instruments [5]. Therefore, it is important to establish a reliable prediction model of the internal corrosion rate based on easily measurable parameters for studying the rules of internal corrosion.

A number of modeling approaches have been used for the corrosion rate prediction. Chou et al. compared the prediction accuracy of the carbon steel corrosion rate in marine environments based on an artificial neural network (ANN), support vector machine (SVM), classification and regression tree (CART), linear regression (LR) and hybrid metaheuristic regression models, and the results showed that the hybrid metaheuristic regression model had superior prediction accuracy in this case [6]. Jain et al. proposed a quantitative evaluation model based on Bayesian networks for the external corrosion rate of oil and gas pipelines [7]. Rocabruno-Valdés et al. developed an ANN model with three layers to predict the corrosion rate of metals in different biodiesel [8]. Wen et al. combined support vector regression (SVR) with particle swarm optimization (PSO) to establish a model for prediction of the corrosion rate of 3C steel under five different seawater environment factors [9]. Abbas et al. developed neural networks (NN) to predict CO<sub>2</sub> corrosion on pipelines at high partial pressures and assessed the degree of suitability for CO<sub>2</sub> corrosion rate prediction [10]. Hu et al. combined the design of experiment (DOE) approach with ANN to discuss the effects of environmental factors in the deep sea on the Ni-Cr-Mo-V high strength

steel corrosion behavior [11]. Although these works have showed great advantages and potential in solving highly nonlinear problems, these models have shortcomings, such as poor fuzzy logic inference ability when completing the “black box” nonlinear mapping from the input to output. Improved applicability, robustness, and generalizability of the prediction model are still the most important targets.

Fuzzy logic, introduced by Zadeh, contains three features: modeling of nonlinear processes by using IF-THEN rules, employing linguistic variables instead of or in addition to numerical variables, and using approximate reasoning algorithms to formulate complex relationships [12, 13]. A model that combines neural networks with fuzzy logic systems was first proposed by Takagi and Hayashi [14]. This model realized fuzzy reasoning so that the weights of traditional neural networks without an explicit physical meaning are assigned to the physical meaning of the reasoning parameters in fuzzy logic to more accurately describe the relationship between the input and output [15]. In the research of fuzzy neural networks for corrosion rate prediction, Biezma et al. proposed a method based on a fuzzy neural network (FNN) to predict the corrosion rate of buried pipelines with limited detection data, considering the factors that affected soil corrosion [16]. Najjaran et al. proposed two FNN models with different input numbers to describe the soil corrosiveness of buried pipelines [17].

In the present study, the FNN only learns and optimizes the parameters in the fuzzy system and adaptive adjustments based on a preset neural network, which is time consuming and leads to low-accuracy structure identification [18–20]. For neural networks, the core indicator for evaluating model performance is the generalization ability, which is mainly affected by the selection of the structure. Too few nodes will result in large learning errors, while too many nodes can cause overfitting. An unreasonable structure can lead to overfitting or overtraining, which directly affects the generalization ability of the neural network system. To overcome the above mentioned problems, this paper proposes an internal corrosion rate forecasting model using the dynamic fuzzy neural (D-FNN).

The internal corrosion rate prediction effect of the model is largely determined by the correlation between the input data and the output data of the model. In addition to temperature, pH, and flow rate, the internal corrosion rate prediction is affected by oxygen content, pressure, and so on. Consequently, we introduce other factors to augment the prediction. However, if these factors are directly used as D-FNN model input, the redundant information of these factors will cause the inaccurate prediction results. At present, methods for dealing with this problem mainly include principal component analysis (PCA), which is a statistical method used for dimensional reduction and feature extraction. This method is particularly suitable for dealing with situations where such factors are highly interrelated [21–24]. Therefore, this paper proposes a method whose parameters are optimized by a PCA algorithm, which is called PCA-D-FNN, to forecast the internal corrosion rate. The model generates fuzzy rules during the dynamic learning process, which grow exponentially instead of increasing with variables, thus improving the generalization ability of the network [25–27].

This paper is structured as follows. In Section 2, the basic concepts of PCA are described. In Section 3, the D-FNN modeling method is introduced in detail, and describes the proposed hybrid model. In Section 4, an application of the proposed method is presented. Finally, the conclusions are stated in Section 5.

## 2. The Method of Principal Component Analysis

PCA is a common multivariable statistical method used for feature extraction and dimensional reduction in analysis. The method uses a linear projection to map high-dimensional data to a representation in a low-dimensional space that maximizes the variance of the data in the projected dimension by using fewer data dimensions and retaining more original data points, thus realizing the dimensionality reduction process [28–30]. In fact, compared to the univariate approach, this explorative method allows to analyze together all the variables acting on a process and to isolate only the relevant information, minimizing redundant data [31]. There are many factors influencing the corrosion in pipelines, and the relationships among them are complex. Therefore, the PCA algorithm can effectively screen corrosion factors, reduce unnecessary analysis, and provide reasonable initial values for the subsequent construction of D-FNNs.

Assume there is a  $P$ -dimensional random vector  $X = (X_1, X_2, \dots, X_p)^T$ , and normalize the transformation of the sample array elements to obtain a new matrix  $Z$ . By the characteristic equation  $|R - \lambda_j E| = 0$ , the eigenvalues of the matrix  $R$  are obtained. Then, select the principal component by formula

$$G = \frac{\sum_{j=1}^k \lambda_j}{\sum_{i=1}^m \lambda_i}, \quad (1)$$

where  $G$  is the cumulative contribution rate, such that when  $G \geq 85\%$ , we select the corresponding components.

Calculate principal component loads:

$$l(Z_i, X_j) = \sqrt{\lambda_i} r_{ij}. \quad (2)$$

## 3. The PCA-D-FNN Prediction Model

**3.1. The structure of the D-FNN.** The D-FNN combines the advantages of fuzzy systems and neural networks. The D-FNN is based on the extended radial basis function (RBF) neural networks and its essence is a fuzzy system based on the Takagi–Sugeno–Kang (TSK) model [32]. The D-FNN model consists of five information processing unit layers: the input layer, fuzzification layer, fuzzy reasoning layer, defuzzification layer, and output layer; these layers are described in detail in the following. The topological structure is shown in Figure 1.

Layer 1 (input layer):  $x_1, x_2, \dots, x_m$  represent the input variables, where  $m$  is the number of input variables.

Layer 2 (membership function layer): each node represents a membership function. The membership function can be denoted as the Gaussian function:

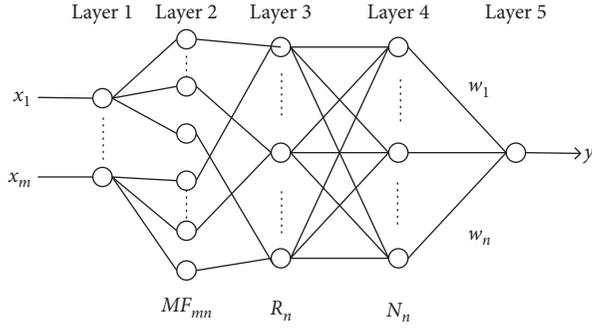


FIGURE 1: The architecture of a dynamic fuzzy neural network [33].

$$\mu_{ij}(x_i) = \exp\left[-\frac{(x_i - c_{ij})^2}{\sigma_j^2}\right], \quad i = 1, 2, \dots, m, j = 1, 2, \dots, n, \quad (3)$$

where  $\mu_{ij}$  is the  $j$ th membership function of the input variable  $x_i$ ,  $n$  is the number of membership functions, and  $c_{ij}$  and  $\sigma_j$  are the center and width of the  $j$ th membership function of the input variable  $x_i$ , respectively.

Layer 3 ( $T$ norm layer): each node represents a fuzzy rule, which is equivalent to the IF part of a possible fuzzy rule. Therefore, the number of nodes in the layer also reflects the total number of fuzzy rules of the system, and the output of the  $j$ th rule is as follows:

$$\phi_j = \exp\left[-\frac{\sum_{i=1}^m (x_i - c_{ij})^2}{\sigma_j^2}\right] = \exp\left[-\frac{\|X - C_j\|^2}{\sigma_j^2}\right], \quad j = 1, 2, \dots, n, \quad (4)$$

where  $X = (x_1, x_2, \dots, x_m)^T$  and  $C_j = (c_{1j}, c_{2j}, \dots, c_{mj})^T$  is the center of the  $j$ th RBF unit. The RBF unit is used in this layer because the network structure of the RBF can be adaptively adjusted during the training phase according to the specific scene without having to be determined before training. This structure simulates the characteristics of the local adjustment and interaction of the human brain and makes the system's approximation ability better. This kind of neural network can well establish the corrosion rate prediction model in this paper and avoid the complicated problems of membership function selection, rule selection, and weight distribution.

Layer 4 (defuzzification layer, also known as the normalized layer): this layer achieves a normalized calculation, and the number of nodes in this layer is equal to the number of fuzzy rules. The output of the  $j$ th node is

$$\phi_j = \frac{\varphi_j}{\sum_{f=1}^n \varphi_f}, \quad j = 1, 2, \dots, n, \quad (5)$$

Layer 5 (output layer): each node in this layer represents an output variable, and the output is the accumulation of all the input signals:

$$y = \sum_{f=1}^n w_f \cdot \phi_f, \quad (6)$$

where  $y$  is the output of the variables,  $w_j$  is the THEN-part or the connected weight of the  $f$ th rule, and  $n$  is the number of total fuzzy rules.

The weights are a linear structure and can be expressed as follows:

$$w_f = a_{f0} + a_{f1}x_1 + \dots + a_{fm}x_m, \quad f = 1, 2, \dots, n, \quad (7)$$

where  $a_{fm}$  are the real-valued parameters.

Substituting equations (12), (13), and (15) into equation (14), the following model is obtained:

$$y = \frac{\sum_{j=1}^n \left[ (a_{j0} + a_{j1}x_1 + \dots + a_{jm}x_m) \exp\left(-\|X - C_j\|^2 / \sigma_j^2\right) \right]}{\sum_{j=1}^n \exp\left(-\|X - C_j\|^2 / \sigma_j^2\right)}. \quad (8)$$

**3.2. Learning Optimization Process of D-FNN.** The structure of the D-FNN is not preset but is formed according to the gradual increase in the learning process. Therefore, the learning algorithm of the system mainly includes the generation of fuzzy rules, the determination of premise parameters, the determination of weights, and the pruning technique of rules, to achieve the specific performance required by the system [34, 35].

**3.2.1. The Generation of Fuzzy Rules.** Determining the structure of the network is one of the main purposes of the training algorithm. To determine whether to add a new rule, it mainly depends on two judging indicators: the accommodating boundary and the system errors. The containable boundary characterizes the coverage of a membership function; multiple existing membership functions have the characteristics of dividing the entire input space. Therefore, if a new sample appears in the coverage of a Gaussian membership function that already exists, it means that this sample can be represented by an existing Gaussian function, so there is no need to add new rules or RBF units to accommodate this new sample. The description of the basis for obtaining rules based on the accommodating boundary is as follows.

For the  $i$ th observed data  $(X_i, t_i)$ , the distance between the input variable  $X_i$  and the current center of the RBF unit is

$$d_i(j) = \|X_i - C_j\| \quad j = 1, 2, \dots, n_c, \quad (9)$$

where  $n_c$  is the number of current fuzzy rules. Define  $kd$  as the effective radius of the accommodating boundary; if  $d_{\min} = \arg \min(d_i(j)) > kd$ , then there is no Gaussian function to represent this new sample and then the fuzzy rules should be increased.

In addition to judging based on the accommodating boundary, system errors need to be considered. If there are too many or too few rules, the unnecessary complexity will be increased, which will worsen the system performance and reduce the generalization ability of the system. Thus, the system error is a vital factor in ensuring the new rules.

For the  $i$ th observed data  $(X_i, t_i)$ , where  $X_i$  is the input vector and  $t_i$  is the expected output, we define the output from the D-FNN as  $y_i$ , and the system error is  $e_i$ :

$$\|e_i\| = \|t_i - y_i\|, \quad (10)$$

when  $\|e_i\| > k_e$ , the fuzzy rules can be increased.

$k_d$  and  $k_e$  of each RBF unit are not fixed during training; with continuous learning, the values of  $k_d$  and  $k_e$  begin to gradually decrease, and local detailed learning is performed. The  $k_d$  and  $k_e$  are defined as follows:

$$\begin{aligned} k_d &= \max[d_{\max} \times \gamma^i, d_{\min}], \\ k_e &= \max[e_{\max} \times \beta^i, e_{\min}], \end{aligned} \quad (11)$$

where  $d_{\max}$  is the largest length of the input space,  $d_{\min}$  is the smallest length expected in the experiment,  $\gamma$  ( $0 < \gamma < 1$ ) is the attenuation coefficient,  $\beta$  ( $0 < \beta < 1$ ) is the convergence coefficient,  $e_{\max}$  is the maximum error of the pre-determined system, and  $e_{\min}$  is the expected accuracy of the system.

The width of the RBF unit can affect the generalization ability of the system. Therefore, the newly generated rules, that is, the width and center of the RBF unit, need to be adjusted. The adjustment method is as follows:

$$C_i = X_i, \quad (12)$$

$$\sigma_i = k \cdot d_{\min}, k > 1, \quad (13)$$

where  $k$  is the overlap factor. When the first sample  $(X_1, t_1)$  is obtained, the network has not yet been established, so the first fuzzy rule is set to

$$C_1 = X_1, \sigma_1 = \sigma_0, \quad (14)$$

where  $\sigma_0$  is the predetermined constant.

**3.2.2. Generation of Weights.** Assuming that  $p$  observed data generate  $n$  fuzzy rules, the output of  $p$  nodes is defined as follows according to the production criterion of rules:

$$\phi = \begin{bmatrix} \phi_{11} & \cdots & \phi_{1p} \\ \vdots & \ddots & \vdots \\ \phi_{n1} & \cdots & \phi_{np} \end{bmatrix}, \quad f = 1, 2, \dots, n. \quad (15)$$

For any input  $(x_{f1}, x_{f2}, \dots, x_{fm})^T$ , the system output  $y_f$  can be represented as

$$Y = W\Psi. \quad (16)$$

Rewrite equation (14) in a matrix form:

$$W = (a_{10} \cdots a_{n0} \quad a_{11} \cdots a_{n1} \quad \cdots \quad a_{jm} \cdots a_{nm}), \quad (17)$$

$$\Psi = \begin{bmatrix} \phi_{11} & \cdots & \phi_{1p} \\ \vdots & \vdots & \vdots \\ \phi_{n1} & \cdots & \phi_{np} \\ \phi_{11}x_{11} & \cdots & \phi_{1p}x_{1p} \\ \vdots & \vdots & \vdots \\ \phi_{n1}x_{n1} & \cdots & \phi_{np}x_{np} \\ \vdots & \vdots & \vdots \\ \phi_{11}x_{m1} & \cdots & \phi_{1p}x_{mp} \\ \vdots & \vdots & \vdots \\ \phi_{n1}x_{m1} & \cdots & \phi_{1p}x_{mp} \end{bmatrix}. \quad (18)$$

The relationship between the expected output  $T = (t_1, t_2, t_3, \dots, t_p)$  and  $\Psi$  is

$$E = \|T - Y\|. \quad (19)$$

Find an optimal parameter coefficient vector  $W_i$  that makes  $E^T E$  the smallest. Select the regression least squares algorithm to solve this problem:

$$W_i = W_{i-1} + S_i \Psi_i^T (T_i - \Psi_i W_{i-1}), \quad (20)$$

$$S_i = S_{i-1} - \frac{S_{i-1} \Psi_i^T \Psi_i S_{i-1}}{1 + \Psi_i S_{i-1} \Psi_i^T}, \quad i = 1, 2, \dots, p, \quad (21)$$

where  $S_i$  is the error covariance matrix of the  $i$ th training data,  $\psi_i$  represents the  $i$ th column of  $\psi$ , and  $W_i$  is the weight matrix obtained after the  $i$ th iterations. The initial parameters are set as  $W_0 = 0$  and  $S_0 = \chi I$ , where  $\chi$  is a sufficiently large positive number and  $I$  is a unit matrix.

**3.2.3. Pruning Algorithm.** In this paper, we trim the number of fuzzy rules in the third layer with the error reduction rate (ERR). This algorithm decomposes the output of the fourth layer into an orthogonal base matrix and an upper triangular matrix by QR decomposition. Then, the ERR is calculated by the orthogonal basis matrix. Using the pruning algorithm, significant neurons are selected so that a parsimonious structure with high performance can be achieved [36]. The  $\eta_i$  reflects the importance of the  $i$ th fuzzy rule; if  $\eta_i$  is larger, then the RBF unit has a greater influence on the entire network. In contrast, if  $\eta_i$  is smaller, then the RBF unit has less impact on the entire network, that is, if  $\eta_i < k_{\text{err}}$ , where  $k_{\text{err}}$  is the preset threshold value, then delete the  $i$ th rule.

**3.2.4. Proposed Hybrid Model.** The proposed hybrid model inherits the merits of the independent models and enhances the performance of the internal corrosion rate prediction compared with previous models. The complexity of the algorithm mainly includes two aspects: PCA and D-FNN. The flow chart of the PCA-D-FNN is shown in Figure 2. The

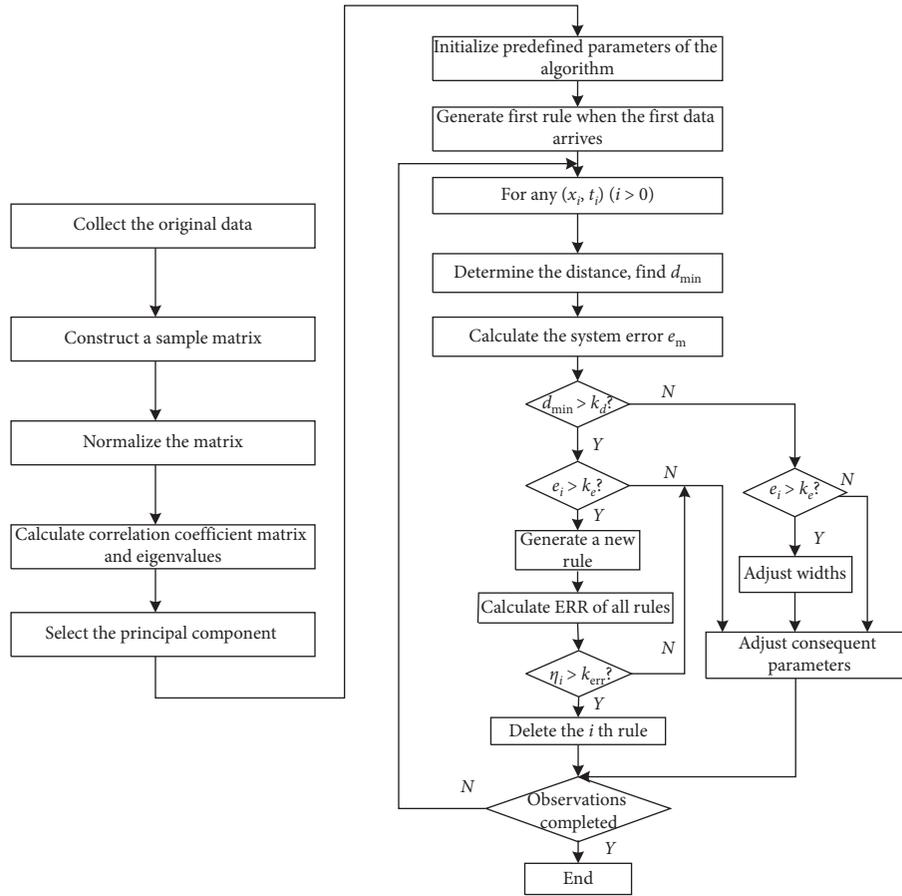


FIGURE 2: The flow chart of PCA-D-FNN.

PCA method is used to analyze the input variables, and a few principal components that can represent all the information are extracted, which will reduce the input dimension of the model and improve the prediction accuracy. The D-FNN, with a compact structure and high performance, is used as the prediction model for internal corrosion rate. The input variables of the D-FNN are screened by PCA, and the model generates fuzzy rules during the dynamic learning process, which grow exponentially instead of increasing with variables, and thus the model has lower computational complexity.

#### 4. Application

**4.1. Dataset.** Natural gas should be purified to remove impurities, such as H<sub>2</sub>O and H<sub>2</sub>S, before entering the pipeline. However, it is difficult to remove these impurities completely. Therefore, the inner wall of the pipeline will be corroded during long-term operation or under special working conditions. Corrosion in pipelines is affected by many factors, and its impact process is complex. Qiao et al. used computational fluid dynamics (CFD) simulation analysis to conclude that the solid particles in the natural gas flow were the main cause of corrosion in the elbow of the gas pipeline [37]. Pfennig et al. found that the presence of CO<sub>2</sub> had a greater corrosive effect on steel pipes at high temperatures (40°C–60°C) [38]. Mansoori et al. used scanning

electron microscopy (SEM) and X-ray diffraction (XRD) to characterize the corrosion products near the damaged part of the gathering pipeline, and they believed that calcium carbonate easily precipitated on the inner surface of the pipeline when the Ca<sup>+</sup> concentration and pH value were high [39]. Javidi et al. believed that pH, temperature, flow rate, CO<sub>2</sub>, corrosion products, and H<sub>2</sub>S had a great influence on the corrosion of gas pipelines [40]. To develop a prediction model of the internal corrosion rate, a total of 9 variables (CO<sub>2</sub> content, H<sub>2</sub>S content, Cl<sup>-</sup> content, moisture content, pH, flow rate, temperature, pressure, and oxygen content) are chosen according to the workers' experience. The corrosion rate is derived from an online monitoring system which is shown in Figure 3. The nine natural gas parameters measured in the 34 samples are introduced into the following models.

#### 4.2. Proposed Hybrid Model

**4.2.1. The PCA Method.** The PCA method is used to analyze the above features, and a few principal components that can represent all the information are extracted, which will reduce the input dimension of the model and improve the prediction accuracy. Using PCA algorithm proposed in Section 2, the nine natural gas parameters measured in the 34 samples were analyzed, and the results are shown in Table 1.

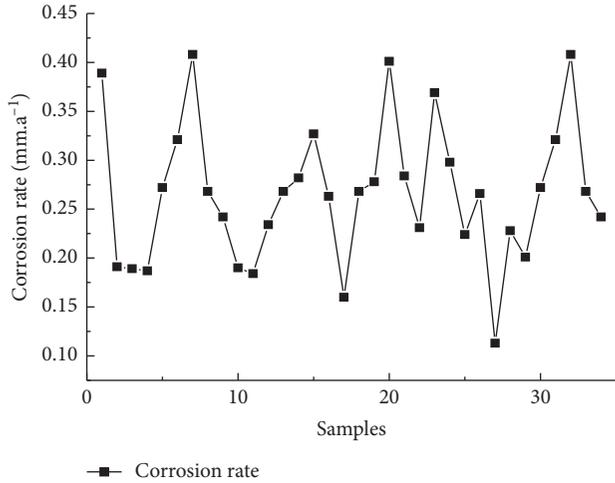


FIGURE 3: Rates of corrosion.

Table 1 shows that the cumulative contribution rate of the first four principal components is 86.62%, which contains most of the internal corrosion information. Among them,  $H_2S$  content has a higher value on the first principal component,  $CO_2$  content has a higher value on the second principal component, moisture content has higher values on the third principal component, and the flow rate is higher on the fourth principal component. Therefore, we choose  $H_2S$  content,  $CO_2$  content, moisture content, and the flow rate as the input of the D-FNN prediction model.

#### 4.2.2. PCA-D-FNN Parameter Setting and Result Analysis.

In this paper, the D-FNN model is established to predict the inner corrosion of the pipeline. There are four input nodes screened out by PCA algorithm, 34 pairs of input and output data are used in this research, while 24 pairs are used as the training dataset and the rest are the test dataset. The precision of the model is set to 0.05. When the accuracy of the training error is less than 0.05, or the maximum iteration number is 80, the training is terminated. The initial parameters of D-FNN are  $d_{max}=4$ ,  $d_{min}=0.2$ ,  $\gamma=0.955$ ,  $e_{max}=1.1$ ,  $e_{min}=0.02$ ,  $\beta=0.5$ ,  $\sigma_0=1.1$ ,  $k=1.1$ ,  $k_d=1$ , and  $k_{err}=0.0015$ . The rule number of the D-FNN is 6, and its mean square error gradually decreases with the training process, which indicates that the structure of the network is basically stable. The D-FNN is trained by 24 training samples, and the network converges after 20 iterations. The results are shown in Figure 4.

To study the prediction accuracy of the proposed model, the root mean square error (RMSE), the mean absolute percentage error (MAPE), and Theil's inequality coefficient (TIC) are employed to evaluate the model performance in this paper. The RMSE is employed to evaluate the difference between the observed values and the actual values, the MAPE is a commonly accepted metric, and the TIC indicates a good level of agreement between the studied process and the proposed model [41]. The calculation methods are defined in Table 2.

The ANN, FNN, and D-FNN models have also been chosen in comparison with the PCA-D-FNN model. In the

TABLE 1: Eigenvalue and cumulative variance contribution.

Principal component	Eigenvalue	Cumulative variance contribution rate
1	7.973	36.52
2	5.222	59.86
3	2.987	80.35
4	1.502	86.62
5	0.412	89.44
6	0.398	92.38
7	0.214	95.95
8	0.177	97.72
9	0.098	99.76
10	0.028	99.89

contrastive experiment, all models were trained using 24 pairs' dataset with the remaining 10 pairs as test dataset. The architecture of the ANN consist of four input nodes, one hidden layer, and one output layer, and the hidden layer contains seven nodes, while the transfer function is *tansig*. In the FNN experiment, each of the four input nodes has four grades of membership of the fuzzy sets, and the membership generation layer consist of 16 nodes, with the model showing stability after generating 6 rules. The architecture of the D-FNN consists of nine input nodes, and the remaining calculation steps are consistent with PCA-D-FNN. The accuracy of the training error was set to 0.05, and the prediction results of the different models on the testing dataset are shown in Table 3. All models use leave-one-out cross-validation (LOOCV) method to investigate the generalization ability of each algorithm, and the RMSE and root MAPE are used to characterize the LOOCV results. The results are shown in Table 4.

From Table 3, the prediction results of the PCA-D-FNN model and algorithm established in this paper for the testing dataset are shown to perform much better than those of the other three models. In detail, the ANN model achieves an RMSE of 0.6863, an MAPE of 0.1244, and a TIC of 0.3471. The FNN model performs better than the ANN model; the model obtains an RMSE of 0.6273, an MAPE of 0.0926, and a TIC of 0.3248. The RMSE, MAPE, and TIC of the D-FNN model are 0.5464, 0.0711, and 0.2784, respectively, which further shows the better performance achieved. The PCA-D-FNN obtains an RMSE of 0.4232, an MAPE of 0.0591, and a TIC of 0.2352; therefore, PCA-D-FNN has the best MAE, RMSE, and MAPE on testing dataset among the four models. The recognition rate of the PCA-D-FNN, D-FNN, and FNN models established in this paper for the testing set is much better than that of the ANN model. From Table 4, the ANN model achieves an RMSE of 0.7324, an MAPE of 8.56%, the FNN obtains an RMSE of 0.6121, an MAPE of 7.82%, the D-FNN model obtains an RMSE of 0.4931, and an MAPE of 6.01%. The PCA-D-FNN obtains an RMSE of 0.4133 and an MAPE of 5.32%. The LOOCV results of different models also show that the results of PCA-D-FNN are better than other algorithms. The result comparison all proves that the PCA-D-FNN model can achieve a good performance in the internal corrosion rate prediction problem. This is due to the fuzziness, as well as the multiple solutions of the relationship

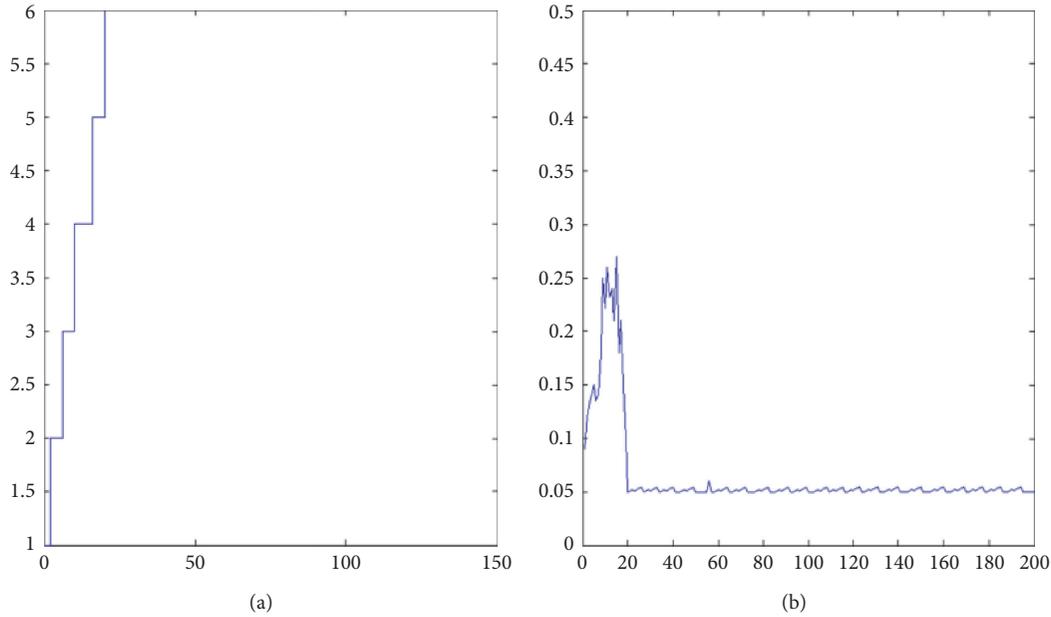


FIGURE 4: (a) The change of fuzzy rules during model training. (b) Root mean square error during training.

TABLE 2: Three metric rules.

Metric	Equation	No.
RMSE	$RMSE = \sqrt{(1/n) \sum_{i=1}^n (O_d - O_r)^2}$	30
MAPE	$MAPE = (1/n) \sum_{i=1}^n  O_d - O_r  / O_d 100\%$	31
TIC	$TIC = \sqrt{(1/n) \sum_{i=1}^n (O_d - O_r)^2} / \sqrt{\sum_{i=1}^n O_d^2 + \sum_{i=1}^n O_r^2}$	32

$O_d$  is the desired output, and  $O_r$  is the result of the algorithm.

TABLE 3: The RMSE, MAPE and TIC of the prediction values for the internal corrosion rate performance of models on the testing dataset.

Evaluation metrics	ANN	FNN	D-FNN	PCA-D-FNN
RMSE	0.6863	0.6273	0.5464	0.4232
MAPE (%)	12.44	9.26	7.11	5.91
TIC	0.3471	0.3248	0.2784	0.2352

TABLE 4: The result of LOOCV with different algorithm.

Evaluation metrics	ANN	FNN	D-FNN	PCA-D-FNN
RMSE	0.7324	0.6121	0.4931	0.4133
MAPE (%)	8.56	7.82	6.01	5.32

between the internal corrosion rate and the influencing factors. Given that the ANN model is the neural network model, the ability and generalization stability of this model are inferior to those of PCA-D-FNN, D-FNN, and FNN when dealing with small samples. The FNN model first needs to convert the variables into grades of membership of the fuzzy sets, a task greatly affected by a researcher’s experience, which affects the accuracy of the model. The D-FNN model that does not implement input dimension reduction has too

many variables that can affect the generalizability of the model. The PCA-D-FNN shows advantages in modeling that employs internal corrosion rate sample sets, which greatly improves the robustness and generalizability of the model and achieves a more accurate result.

The computation time of the ANN, FNN, D-FNN, and PCA-D-FNN models are 1.923s, 2.341s, 2.571s, and 1.621s, respectively. The proposed method can be used for internal corrosion rate prediction of gas pipeline.

### 5. Conclusion

The internal corrosion rate of gas pipeline is affected by many factors, and the reliability of the pipeline will be affected greatly by internal corrosion. Thus, conducting accurate forecasting of the internal corrosion rate appears to be especially important. Therefore, a hybrid model called the PCA-D-FNN is proposed in this paper. PCA is an effective method that is used to extract features and reduce the dimensions of the original sample, and four factors, including 86.62% of the original information, are extracted. Then, the D-FNN is used to conduct the prediction and is shown to take advantage of the fuzzy rules and ANNs to overcome the drawbacks of the single methods. This method generates fuzzy rules in the dynamic learning process, which grow exponentially instead of increasing with variables, thus improving the generalization ability of the network. The experimental results prove the effectiveness of the hybrid model through testing the proposed model by using the collected corrosion data. Through a comparison of PCA-D-FNN with ANN, FNN, and D-FNN models, the PCA-D-FNN model is shown to predict the internal corrosion rate with an RMSE of 0.4232, an MAPE of 5.91%, and a TIC of 0.2352 on testing dataset, which is more accurate than other

models. The LOOCV results of different models also show that the results of PCA-D-FNN are better than other algorithms. It can also be determined that PCA-D-FNN obtains the best forecasting performance with a fast convergence rate and a high ability to search for global optimum. Therefore, the proposed model demonstrates great potential in applications concerned with the internal corrosion rate of pipelines.

### Data Availability

The data used to support the findings of this study have not been made available because they are currently under embargo while the research findings are commercialized. Requests for data, 10 months after publication of this article, will be considered by the corresponding author.

### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

### Acknowledgments

This work was supported by National Science Foundation of China (no. 51874255).

### References

- [1] S. Nešić, "Key issues related to modelling of internal corrosion of oil and gas pipelines—a review," *Corrosion Science*, vol. 49, no. 12, pp. 4308–4338, 2007.
- [2] C. I. Ossai, B. Boswell, and I. J. Davies, "Pipeline failures in corrosive environments—a conceptual analysis of trends and effects," *Engineering Failure Analysis*, vol. 53, pp. 36–58, 2015.
- [3] M. Askari, M. Aliofkhaezrai, S. Ghaffari, and A. Hajizadeh, "Film former corrosion inhibitors for oil and gas pipelines - a technical review," *Journal of Natural Gas Science and Engineering*, vol. 58, pp. 92–114, 2018.
- [4] H. A. Kishawy and H. A. Gabbar, "Review of pipeline integrity management practices," *International Journal of Pressure Vessels and Piping*, vol. 87, no. 7, pp. 373–380, 2007.
- [5] H. R. Vanaei, A. Eslami, and A. Egbewande, "A review on pipeline corrosion, in-line inspection (ILI), and corrosion growth rate models," *International Journal of Pressure Vessels and Piping*, vol. 149, pp. 43–54, 2017.
- [6] J.-S. Chou, N.-T. Ngo, and W. K. Chong, "The use of artificial intelligence combiners for modeling steel pitting risk and corrosion rate," *Engineering Applications of Artificial Intelligence*, vol. 65, pp. 471–483, 2017.
- [7] S. Jain, A. N. Sánchez, and S. Guan, *Probabilistic Assessment of External Corrosion Rates in Buried Oil and Gas Pipelines*, NACE, Houston, TX, USA, 2015.
- [8] C. I. Rocabrano-Valdés, J. G. González-Rodríguez, Y. Díaz-Blanco et al., "Corrosion rate prediction for metals in biodiesel using artificial neural networks," *Renewable Energy*, vol. 140, pp. 592–601, 2019.
- [9] Y. F. Wen, C. Z. Cai, X. H. Liu, J. F. Pei, X. J. Zhu, and T. T. Xiao, "Corrosion rate prediction of 3C steel under different seawater environment by using support vector regression," *Corrosion Science*, vol. 51, no. 2, pp. 349–355, 2009.
- [10] M. H. Abbs, R. Norman, and A. Charles, "Neural network modelling of high pressure CO<sub>2</sub> corrosion in pipeline steels," *Process Safety and Environmental Protection*, vol. 119, pp. 36–45, 2018.
- [11] Q. Hu, Y. Liu, T. Zhang, S. Geng, and F. Wang, "Modeling the corrosion behavior of Ni-Cr-Mo-V high strength steel in the simulated deep sea environments using design of experiment and artificial neural network," *Journal of Materials Science & Technology*, vol. 35, no. 1, pp. 168–175, 2019.
- [12] L. A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 1, pp. 28–44, 1973.
- [13] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-I," *Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975.
- [14] H. Takagi and I. Hayashi, "NN-driven fuzzy reasoning," *International Journal of Approximate Reasoning*, vol. 5, no. 3, pp. 191–212, 1991.
- [15] H. Ishibuchi, R. Fujioka, and H. Tanaka, "An architecture of neural networks for input vectors of fuzzy numbers," in *Proceedings of the IEEE International Conference on Fuzzy Systems*, pp. 1293–1300, San Diego, CA, USA, August 2005.
- [16] M. V. Biezma, D. Agudo, and G. Barron, "A Fuzzy Logic method: predicting pipeline external corrosion rate," *International Journal of Pressure Vessels and Piping*, vol. 163, pp. 55–62, 2018.
- [17] H. Najjaran, R. Sadiq, and B. Rajani, "Modeling pipe deterioration using soil properties - an application of fuzzy logic expert system," *Pipeline Engineering and Construction*, vol. 8, pp. 1–10, 2004.
- [18] C.-F. Hsu, P.-Z. Lin, T.-T. Lee, and C.-H. Wang, "Adaptive asymmetric fuzzy neural network controller design via network structuring adaptation," *Fuzzy Sets and Systems*, vol. 159, no. 20, pp. 2627–2649, 2008.
- [19] H.-C. Lu and H.-K. Liu, "Ant colony fuzzy neural network controller for cruising vessel on river," *Applied Ocean Research*, vol. 42, pp. 43–54, 2013.
- [20] H.-K. Wu, J.-G. Hsieh, Y.-L. Lin, and J.-H. Jeng, "On maximum likelihood fuzzy neural networks," *Fuzzy Sets and Systems*, vol. 161, no. 21, pp. 2795–2807, 2010.
- [21] C. Skittides and W.-G. Früh, "Wind forecasting using principal component analysis," *Renewable Energy*, vol. 69, pp. 365–374, 2014.
- [22] H. Yang, W. Zhou, H.-R. Wen, H.-W. Zhao, and W.-S. Zhang, "The application of principal component analysis (PCA) method in evaluation of soil corrosivity," *Oil & Gas Science and Technology-Revue D IFP Energies Nouvelles*, vol. 504, pp. 43–46+62+1, 2007.
- [23] S. Xu and X. Liu, "Melt index prediction by fuzzy functions with dynamic fuzzy neural networks," *Neurocomputing*, vol. 142, pp. 291–298, 2014.
- [24] H. Huang, Y.-Y. Wu, M.-F. Peng et al., "Application of principal component analysis method in comprehensive evaluation of soil corrosion," *Materials Science and Engineering*, vol. 504, 2019.
- [25] A. R. Khan, A. Mahmood, A. Safdar, Z. A. Khan, and N. A. Khan, "Load forecasting, dynamic pricing and DSM in smart grid: a review," *Renewable and Sustainable Energy Reviews*, vol. 54, pp. 1311–1322, 2016.
- [26] A. T. Jahromi, M. J. Er, X. Li, and B.-S. Lim, "Sequential fuzzy clustering based dynamic fuzzy neural network for fault diagnosis and prognosis," *Neurocomputing*, vol. 196, pp. 31–41, 2016.
- [27] X. Deng and X. Wang, "Incremental learning of dynamic fuzzy neural networks for accurate system modeling," *Fuzzy Sets and Systems*, vol. 160, no. 7, pp. 972–987, 2009.

- [28] N. M. Tran, P. Burdejová, M. Ospienko, and W. K. Härdle, "Principal component analysis in an asymmetric norm," *Journal of Multivariate Analysis*, vol. 171, pp. 1–21, 2019.
- [29] A. Gupta and A. Barbu, "Parameterized principal component analysis," *Pattern Recognition*, vol. 78, pp. 215–227, 2018.
- [30] A. X. Tao, B. R. Chang, C. C. Li, D. R. Wang, and E. R. Liu, "Density-sensitive robust fuzzy kernel principal component analysis technique," *Neurocomputing*, vol. 329, pp. 210–226, 2019.
- [31] E. Bernardi, I. Vassura, S. Raffo, L. Nobili, and F. Passarini, "Influence of inorganic anions from atmospheric depositions on weathering steel corrosion and metal release," *Construction and Building Materials*, vol. 236, pp. 515–526, 2020.
- [32] R.-E. Precup, M.-C. Sabau, and E. M. Petriu, "Nature-inspired optimal tuning of input membership functions of Takagi-Sugeno-Kang fuzzy models for Anti-lock braking systems," *Applied Soft Computing*, vol. 27, pp. 575–589, 2015.
- [33] S. Wu, M. J. Er, and Y. Gao, "A fast approach for automatic generation of fuzzy rules by generalized dynamic fuzzy neural networks," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 578–594, 2001.
- [34] M. Liu, X. Liu, M. Wu, L. Li, and L. Xiu, "Integrating spectral indices with environmental parameters for estimating heavy metal concentrations in rice using a dynamic fuzzy neural-network model," *Computers & Geosciences*, vol. 37, no. 10, pp. 1642–1652, 2011.
- [35] Y. Pan, M. J. Er, X. Li, H. Yu, and R. Gouriveau, "Machine health condition prediction via online dynamic fuzzy neural networks," *Engineering Applications of Artificial Intelligence*, vol. 35, pp. 105–113, 2014.
- [36] S. Wu and M. J. Er, "Dynamic fuzzy neural networks: a novel approach to function approximation," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 30, pp. 358–364, 2000.
- [37] Q. Qiao, G. Cheng, Y. Li, W. Wu, H. Hu, and H. Huang, "Corrosion failure analyses of an elbow and an elbow-to-pipe weld in a natural gas gathering pipeline," *Engineering Failure Analysis*, vol. 82, no. 12, pp. 599–616, 2017.
- [38] A. Pfennig and A. Kranzmann, "Effects of saline aquifere water on the corrosion behaviour of injection pipe steels 1.4034 and 1.7225 during exposure to CO<sub>2</sub> environment," *Energy Procedia*, vol. 1, no. 1, pp. 3023–3029, 2009.
- [39] H. Mansoori, R. Mirzaee, F. Esmaeilzadeh, A. Vojood, and A. S. Dowrani, "Pitting corrosion failure analysis of a wet gas pipeline," *Engineering Failure Analysis*, vol. 82, pp. 16–25, 2017.
- [40] M. Javidi, R. Chamanfar, and S. Bekhrad, "Investigation on the efficiency of corrosion inhibitor in CO<sub>2</sub> corrosion of carbon steel in the presence of iron carbonate scale," *Journal of Natural Gas Science and Engineering*, vol. 61, pp. 197–205, 2019.
- [41] D. J. Murray\_smith, "Methods for the external validation of continuous system simulation models: a review," *Mathematical and Computer Modelling of Dynamical Systems*, vol. 4, no. 1, pp. 5–31, 1998.