Application of BP neural network models in predicting the desulfurization rate of petroleum coke calcination flue gas

Y W Lv 1,3, Y Zhu 2,3, Y X Jiang 1, L Xue 1 and L Zuo 1

1Faculty of Environmental Science and Engineering, Kunming University of Science and Technology, Yunnan, China, 650093
2Kunming Environmental Monitoring Center, Yunnan, China, 650228
3E-mails:1776281566@qq.com(Y W Lv); zhuyan1029@163.com(Y Zhu)

Abstract. Sulfur-containing prebaked anodes, mainly made from petroleum coke, can give rise to SO2 emission in the production of electrolytic aluminum. In order to study the correlation between the emission volume and desulfurization rate of the flue gas discharged from the calcination of petroleum coke, a traditional double hidden layer BP neural network model was established on the basis of BP neural network by taking the emission data of the petroleum coke calcination flue gas desulfurization system as input parameters. This model was used to predict the desulfurization rate of the petroleum coke calcination flue gas desulfurization system, with an aim to control the SO2 removal rate of flue gas through controlling the flue gas parameters of the desulfurization system, so as to reduce SO2 emission from the flue gas of the electrolytic aluminum industry. The study will be of practical significance for the SO2 removal of the petroleum coke calcination flue gas.

1. Introduction
The carbon industry which produces carbon for the electrolytic aluminum industry is the main consumer of petroleum coke, consuming about 21 million metric tons of petroleum coke annually. In the calcination process of petroleum coke, about 10% - 30% of sulfur contained in petroleum is released and the sulfur dioxide gas generated in the process is discharged in the form of SO2 along with the calcination flue gas [1]. Among the calcined petroleum coke, more than 45% has a sulfur content of more than 3%, and the sulfur-containing flue gas produces heavy pollution to the atmosphere and environment in China [2]. Data show that an increase of 1% in the sulfur content of petroleum coke can lead to an increase of 133mg/m3 in the SO2 concentration of the flue gas discharged from the industrial electrolytic aluminum production process [3]. So it is of great practical significance for China’s environmental protection to strengthen control over the SO2 concentration of the flue gas discharged from the industrial electrolytic aluminum production process and to adopt appropriate methods to predict the SO2 removal rate of the flue gas desulfurization (FGD) system.

Artificial neural network prediction models have been widely used in air pollution control [4]. In 2014, Faris et al. [5] established a BP neural network prediction model for O3 concentration and made quite accurate predictions about the O3 concentration. In 2010, Zhu G.C. et al [6] created the algorithm and network structure model of back propagation artificial neural network to simulate and calculate pollutant concentrations in streets and valleys. They had good generalization ability. In 2015, Yao N. et al. [7] established a BP neural network model between the air pollutant concentration and its influencing factors. Based on a BP neural network GIS system optimized by the AGES algorithm, the
concentration of air pollutants was predicted, and the predicted result was significantly correlated with the actual concentration. In 2015, Ai H.F. et al. [8] built a BP neural network prediction model based on the time sequence and made an accurate prediction about the haze weather based on the model, so the model has a good practicability. In 2015, Ma C.Y et al. [9] established an atmospheric visibility prediction model based on the genetic neural network and accurately predicted the atmospheric visibility based on the model, and the correlation and absolute error of the prediction results are better than those of BP neural network. In 2011, Liu Y.H. [10] created a seasonal air pollutant prediction model based on BP neural network and predicted the concentration of SO2, NO2 and PM10 in Foshan City based on the model. The prediction results show that the seasonal prediction model has good stability. In 2016, Feng J. et al. [11] established an hourly PM2.5 concentration prediction model based on BP neural network and predicted the PM2.5 concentration in Tianjin City based on the model. The results show that the model can make accurate predictions in actual weather conditions with uncertain meteorological parameters. All the results show that the prediction models based on BP neural network are feasible methods to simulate the concentration change of air pollutants. Therefore, it is theoretically feasible to apply neural network models to predict pollutant concentration changes in industrial exhaust gases. We have reason to believe that the model can also be applied to specific industrial production, so we have tried for the first time to apply the model to the prediction of petroleum coke flue gas desulfurization. In this paper, a BP neural network model was established to predict the desulfurization rate of the flue gas desulfurization system, and the data related to the petroleum coke calcination flue gas (including SO2 concentration, flue gas temperature, flue gas pressure, air flow, pH in the regulating tank, etc.) were collected. Based on the existing data, the model adopted two hidden layers. After training, the SO2 removal rate of different FGD systems were predicted with the model under the optimal combination of parameters, with an aim to control the SO2 removal rate of flue gas through controlling the flue gas parameters of the desulfurization system, so as to reduce SO2 emission from the flue gas of the electrolytic aluminum industry, and to control the desulfurization rate quantitatively by changing the parameters of the desulfurization system. This is of practical significance to controlling the SO2 concentration of the flue gas from electrolytic aluminum enterprises.

Like any ordinary neural network, BP neural network consists of an input layer, hidden layers and an output layer. There are several neurons in a hidden layer, and the weights of the neurons in each layer are related. A hidden layer and another hidden layer can be connected by an activation function, which solves the nonlinear correlation among the weights of neurons in the hidden layers of the multilayer feedforward neural network.

The input layer, as the input end of the network, should preprocess the data before inputting the data into the model, such as standardization and normalization; otherwise, losses and drifts may appear in the training process. The hidden layer, as the middle layer of the network, is the most critical. It can be designed as multi-layer or single layer. In case of multi-layer, it is better to add an activation function as the connection between multiple layers, and the last hidden layer should directly connect the output layer. The output layer includes error calculation and prediction result output. In the training process, error calculation directly affects the determination of the final prediction result. The error is processed by a calculation unit and propagated back to the hidden layer, so as to adjust the weights of neurons in the hidden layer. This process should iterate repeatedly until the stop condition is met.

BP neural network [12] inputs data in the input layer, processes the data and then extracts information in the hidden layer, and calculates errors and predicts the final result in the output layer. The BP network structure is shown in Figure 1.

Firstly, the design of BP network structure is mainly the design of the hidden layer and the output layer. In this paper, the hidden layer is divided into two layers. The two layers are connected by an activation function. The output layer includes a prediction function and an error calculation. Secondly, the training mode of BP neural network is forward propagation and backward feedback. The most important basic unit is the error between the real target value and the predicted value. Only after the
error is obtained can we determine whether the parameters need to be updated. At the same time, the
activation function connects the prediction function with the last hidden layer, thus making up for the
lack of a linear model among neurons in the artificial neural network. In this paper, we chose ReLU as
the activation function between hidden layers.

![Neural Network Diagram](image)

**Figure 1. Three-layer BP neural network structure.**

2. Data analysis

2.1. Data source and processing

The data used in this paper is the flue gas parameter data of YL company's carbon petroleum coke
calcination flue gas desulfurization system. The flue gas desulfurization (FGD) rate is affected by
many factors, including SO$_2$ concentration, flue gas temperature, flue gas pressure, oxidation air flow
and the regulating tank’s pH. This paper established a FGD rate prediction model based on BP neural
network to predict the FGD rate. The data was sourced from the flue gas parameters of YL company’s
petroleum coke calcination flue gas desulfurization system and the corresponding SO$_2$ concentration
sample data collected at the outlet of the FGD system. It included 3,200 groups of data in total.

The premise of model training is the preparation of data, because data is the basis of model learning.
Therefore, it is necessary to carry out noise removal, standardization or normalization for the data, so
as to finally get pure and standard data. First, it is needed to eliminate noises from the flue gas
parameter data. The noises are abnormal values appearing in the data set. In the training process, the
noises can cause great interference to the training result. The methods used to remove the noise data
usually include the box chart detection, the clustering detection of discrete points, the detection of
discrete points based on the multivariate normal distribution, etc. The box chart detection method was
adopted to eliminate the noise data in the study. After the abnormal values were eliminated with the
box chart detection method, the sample size of the effective flue gas parameter data was 2804 groups,
of which 2524 groups were used to train the model and 280 groups were used to verify the model.

After the abnormal values were removed, the effective flue gas parameters were standardized with
their respective mean value and standard deviation. The calculation formula of the parameter
standardization is shown in Formula (1):

$$x_i = \frac{x_i - mean_i}{std_i}$$  \hspace{1cm} (1)

Wherein, $x_i$ represents the ith parameter, and $mean_i$ and $std_i$ represent the mean value and
standard deviation of the ith parameter, respectively. Input parameters and output parameters all need
to be standardized. The mean value and the standard deviation are shown in Table 1.
Table 1. Mean values and standard deviations of input parameters and output parameters.

| Parameter                          | Mean value | Standard deviation |
|------------------------------------|------------|--------------------|
| SO₂ concentration (mg/m³)          | 630.4208   | 296.75             |
| Inlet temperature (°C)             | 141.3813   | 25.5448            |
| Outlet temperature (°C)            | 50.02610   | 7.9540             |
| Inlet pressure (Pa)                | 494.0219   | 147.5468           |
| Outlet pressure (Pa)               | 19.9179    | 13.2958            |
| Oxidation air flow (m³/h)          | 2528.5589  | 877.5316           |
| Regulating tank’s pH               | 11.2019    | 167.3029           |
| FGD rate (%)                       | 76.48      | 23.73              |

2.2. Debugging of model parameters

The super parameters involved in the model were designed. The target function is a regression function and adopts the square loss function. The calculation formula is Formula (2)

\[ L(Y, f(X)) = (Y - f(X))^2 \] (2)

Wherein, \( X \) represents the input sample, \( Y - f(X) \) represents the difference between the real target value and the predicted value, i.e. the residual value. The purpose of debugging is to minimize this function value. The mean square error (\( MSE \)) is used as the measurement index:

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} (Y - f(X))^2 \] (3)

The training is carried out by changing the number of iterations, the number of neurons in the hidden layer, the number of randomly selected samples in a single iteration as well as the learning rate. The training results are shown in Table 2.

Table 2. Predicted target loss values under different training parameters.

| Number of iterations | Number of neurons in hidden layer 1 | Number of neurons in hidden layer 2 | Sample size of a single iteration | Learning rate \( \eta \) | Loss in the desulfurization rate |
|----------------------|-------------------------------------|-------------------------------------|-----------------------------------|----------------------|---------------------------------|
| 2500                 | 16                                  | 32                                  | 32                                | 1e-3                 | 2.565                           |
| 2500                 | 32                                  | 64                                  | 32                                | 1e-3                 | 21.452                          |
| 2500                 | 64                                  | 128                                 | 32                                | 1e-3                 | 23.124                          |
| 2500                 | 128                                 | 256                                 | 32                                | 1e-3                 | 189.154                         |
| 5000                 | 16                                  | 32                                  | 64                                | 1e-2                 | 0.786                           |
| 5000                 | 32                                  | 64                                  | 64                                | 1e-2                 | 0.898                           |
| 5000                 | 64                                  | 128                                 | 64                                | 1e-2                 | 7.89                            |
| 5000                 | 128                                 | 256                                 | 64                                | 1e-2                 | 25.461                          |
| 5000                 | 256                                 | 512                                 | 64                                | 1e-2                 | 48.435                          |
| 10000                | 16                                  | 32                                  | 128                               | 1e-4                 | 1.214                           |
| 10000                | 32                                  | 64                                  | 128                               | 1e-4                 | 23.454                          |
| 10000                | 64                                  | 128                                 | 128                               | 1e-4                 | 25.735                          |
| 10000                | 128                                 | 256                                 | 128                               | 1e-4                 | 123.47                          |
The model was trained under the condition that the number of neurons in hidden layer 1 was 16, the number of neurons in hidden layer 2 was 32 and the learning rate was 0.01. The training results are shown in Figure 2.

![Figure 2](image1.png)  ![Figure 3](image2.png)

**Figure 2.** Training loss value trend of BP model.  **Figure 3.** Training loss value trend of the improved combined BP model.

Figure 3 shows that, when the number of iterations reaches 8800, the training loss value is the smallest, that is, 0.7371. Based on this, the optimal parameter combination is obtained, as shown in Table 3.

| Number of Iterations | Number of neurons in hidden layer 1 | Number of neurons in hidden layer 2 | Batch size | Learning rate η | Loss   |
|----------------------|-------------------------------------|-------------------------------------|------------|----------------|--------|
| 8800                 | 16                                  | 32                                  | 128        | 0.01           | 0.7371 |

The traditional BP neural network may lose some feature information in the process of feature extraction. In order to retain this part of feature information, improvements have been made to the traditional BP neural network structure, making the information of hidden layer 1 directly extended to the full connection layer while transferred to hidden layer 2, that is, making hidden layer 1 and hidden layer 2 connected in parallel to form a hidden layer combined BP model while connected in series.

The improved combined BP neural network model was trained again. The training results are as shown in Figure 3.

From the comparison between Fig. 2 and Fig. 3, it can be seen that the training convergence of the combined BP neural network model is significantly faster than that of the traditional BP neural network model. When the number of iterations is 5,800, the test loss value of the combined BP neural network model is the smallest, that is, 0.3998. The optimal parameter combination obtained according to this is shown in Table 4.

| Number of Iteration | Number of neurons in hidden layer 1 | Number of neurons in hidden layer 2 | Batch size | Learning rate η | Loss   |
|---------------------|-------------------------------------|-------------------------------------|------------|----------------|--------|
| 5800                | 16                                  | 32                                  | 128        | 0.01           | 0.3998 |

2.3. *Verification of BP neural network model*

The prediction results were evaluated by means of indicators, such as standardized mean deviation (NMB), standardized mean error (NME), residual standard deviation (RMSE) and correlation coefficient (R).  

5
The traditional BP model is used to predict the removal rate of SO2 in the petroleum coke calcination flue gas. According to the prediction results obtained by the model, NMB value is 3.474, NME value is 36.748, RMSE value is 0.065, R value is 0.633, NMB is less than 10, NME is less than 40, RMSE is less than 0.1, and R is greater than 0.6. The prediction results are valid. The hidden layer combined with BP model is used to predict the removal rate of SO2 in the petroleum coke calcination flue gas. Model prediction results: NMB value is 8.907, NME value is 26.318, RMSE value is 0.020, R value is 0.656; NMB is less than 10, NME is less than 40, RMSE is less than 0.1, and R is greater than 0.6. These indicators are all within a reasonable range. This indicates that the prediction results are valid. For visual display, the simulation results of predicted values and measured values are shown in Figure 4.

The traditional BP model is used to predict the removal rate of SO2 in the petroleum coke calcination flue gas. According to the prediction results obtained by the model, NMB value is 3.474, NME value is 36.748, RMSE value is 0.065, R value is 0.633, NMB is less than 10, NME is less than 40, RMSE is less than 0.1, and R is greater than 0.6. The prediction results are valid. The hidden layer combined with BP model is used to predict the removal rate of SO2 in the petroleum coke calcination flue gas. Model prediction results: NMB value is 8.907, NME value is 26.318, RMSE value is 0.020, R value is 0.656; NMB is less than 10, NME is less than 40, RMSE is less than 0.1, and R is greater than 0.6. These indicators are all within a reasonable range. This indicates that the prediction results are valid. For visual display, the simulation results of predicted values and measured values are shown in Figure 4.

### Table 5. Comparison of indicators of the two BP models.

| BP network type | NMB  | NME  | RMSE | R   |
|-----------------|------|------|------|-----|
| Traditional BP  | 3.474| 36.748| 0.065| 0.633|
| Combined BP     | 8.907| 26.318| 0.020| 0.656|

3. **Results and discussions**

In summary, the NMB values of the two BP models are less than 10, the RMSE values are less than 0.1, and the correlation coefficient R values are greater than 0.6. Therefore, both BP models have an ideal prediction effect on the removal rate of SO2. The oxygen content and flow rate of the flue gas are negatively correlated with the SO2 concentration in the flue gas. The correlation between the temperature and pressure of the flue gas and the SO2 concentration in the flue gas is not significant. However, changes in the concentration of NO and other pollutants produced during the calcination of petroleum coke have a significant effect on the concentration of SO2 in the flue gas. There is a negative correlation between NOx concentration and SO2 concentration in the flue gas. The increase of NOx concentration in the flue gas can lead to the decrease of SO2 concentration to a certain extent. The concentration of dust, another pollutant, in the flue gas has no significant correlation with the SO2.
concentration in the flue gas. Therefore, based on the BP models, it can be concluded that increasing the oxygen concentration and flow rate of the petroleum coke calcination flue gas can reduce SO$_2$ emissions. Based on the fitting comparison of the measured and predicted values, it can be seen that both the traditional BP model and the hidden layer combined BP model have a good effect in predicting the desulfurization rate of the petroleum coke calcination flue gas. The combined BP model has better performance than the traditional BP model to a certain extent. The combined BP model is superior to the traditional BP model. This also proves the feasibility of improving the SO$_2$ removal rate by controlling the flue gas parameters of the desulfurization system. The results show that increasing the oxygen concentration and flow rate of the petroleum coke calcination flue gas can reduce SO$_2$ emissions.

![Figure 4](image_url)

**Figure 4.** Comparison between the predicted and measured values of the desulfurization rate.

**References**

[1] Lv Y H 2018 Technical analysis on the desulfurization of petroleum coke calcination flue gas *J. Chem. Mgmt.* 2 128-9

[2] Wei P W 2019 Discussions on the desulfurization technology of petroleum coke calcination flue gas *J. Pet. Technol.* 26 243-52

[3] Zhao B, Luo Y T, Su Z W, Huang H and Guo R M 2011 Advances in the research of petroleum coke desulfurization technology *J. Carbon. Technol.* 30 30-2

[4] Guo Q C 2015 Prediction of air pollution based on neural network *J. Electron. Test.* 18 5-6

[5] Faris H, Alkasassbeh M and Rodan A 2014 Artificial Neural Networks for Surface Ozone Prediction: Models and Analysis *J. Pol. J. Environ. Stud.* 23 341-8

[6] Zhu G C, Fang M J and Zheng X X 2010 Study on the numerical model of NOx concentrations in streets and valleys based on artificial neural network *J. Environ. Eng.* 4 875-80

[7] Yao N, Ma Q L and Zhang J 2015 Prediction of air pollutant concentration by means of the
optimized BP neural network and GIS system based on the AGNES algorithm *J. China Environ. Moni.* **31** 113-7

[8] Ai H F and Shi Y 2015 Prediction of haze based on BP artificial neural network *J. Comput. Simu.* **32** 402-5

[9] Ma C Y, Zu J and Fu Q F 2015 Prediction of air visibility based on the genetic neural network model *J. Environ. Eng.* **9** 1905-10

[10] Liu Y H, Xie M and Cai M 2011 Study on the prediction model of air quality in Foshan City based on BP neural network *J. Safe. Environ.* **11** 125-30

[11] Feng J, Liu G and Huang Y 2016 Prediction of PM$_{2.5}$ concentration in Tianjin based on BP neural network *J. Environ. Sci. Mana.* **41** 121-5

[12] Rumelhart D E, Hinton G E and Williams R J 1988 Learning representations by back-propagating errors *J. Read. Cogn. Sci.* **323** 399-421