

Remote NO_x emission prediction model based on LSTM neural network

Tiantian Wang^{1,*}, Yayu Cheng¹, Jie Hu², Chuan Wan³, and Bifeng Zou³

¹ School of mechanical and electrical engineering and vehicle engineering, East China Jiaotong University, Nanchang 330100, China

² School of automotive engineering, Wuhan University of technology, Wuhan 430070, China

³ Power transmission development department, Jiangling Automobile Co., Ltd., Nanchang 330100, China

Keywords: neural network, model, remote monitoring, no_x emission prediction.

Abstract. Test results from many researchers show that NO_x emission from many on-broad heavy-duty diesel vehicles is higher than which been registered. Therefore, CN_VI emission regulations clearly proposes that the heavy-duty diesel vehicles should be supervised by a T-BOX which can transmit CAN message from vehicle OBD interface to the remote monitoring platform. Based on the formation mechanism of NO_x emission and the variety of OBD data flow, the LSTM (Long Short-Term Memory) neural network model inputs such as engine speed, torque, atmospheric pressure, coolant temperature, fuel consumption rate and intake air mass flow are selected by using partial least square method (PLS). 19877 groups of data from engine test results were used for model training and verification, the root mean square error of training and test are $R_{TR} = 29.7 \times 10^{-6}$ and $R_{TE} = 19.9 \times 10^{-6}$, with a high prediction accuracy which can fully meet the requirements of the SCR system DeNO_x performance diagnosis module in the OBD remote monitoring system.

1 Introduction

With the rapid development of the national economy, the main contribution factors of urban pollution sources have begun to shift from fixed source emissions to mobile source emissions. Emissions of mobile sources in Beijing, Shanghai, Hangzhou, Jinan, Guangzhou and Shenzhen are the primary sources, accounting for 45.0%, 29.2%, 28.0%, 32.6%, 21.7% and 52.1% respectively [1]. By the end of 2018, there were 18.18 million diesel trucks in China, accounting for 7.9% of the vehicle ownership, which were mainly commercial vehicles with high intensity of use, harsh operating environment and relatively high pollution emissions, and emissions of nitrogen oxides (NO_x) and particulate matter (PM) accounted for 60.0% and 84.6% of the national vehicle emissions respectively.

Since the implementation of the CN_IV and CN_V emissions regulations, a number of research and testing institutions found that the real-road emissions of diesel vehicles,

* Corresponding author: wtt31@126.com

especially commercial heavy-duty diesel vehicles, exceeded the standard limit seriously, and the cheating behaviors of on-board fault diagnosis (OBD) system and selective catalytic reduction (SCR) system in the on-road vehicle have not been effectively regulated^[2]. Against this background, the country proposed the "Win the Blue Sky Guard Station Plan" at the end of 2017, and in the GB17691-2018 "Emission Limits and Measurement Methods for Pollutant Emissions of Heavy-Duty Diesel Vehicles (6th Phase of China) " released on June 22, 2018, the demand for remote monitoring of heavy-duty diesel vehicles were proposed for the first time^[3]. In the same year, "Action Plan for the Control of Diesel Trucks" jointly issued by 11 departments such as the Ministry of Ecological Environment, which putted forward the installation requirements of OBD remote online monitoring system for diesel trucks.

OBD remote online monitoring system needs to judge the operation and emission status of the target vehicles through various on-line models. At present, the domestic diesel vehicle in use are basically the vehicles below the national V, with a single downstream NO_x sensor. Therefore, the diagnosis of the DeNO_x performance of the SCR system cannot be directly realized, and a method that can predict the NO_x emissions of the diesel engine through the OBD data flow is urgently needed to facilitate the realization of the above diagnosis function.

Regarding the estimation of NO_x emissions, domestic and foreign scholars have done a lot of research and put forward a variety of methods to build a prediction model, such as Taskinen^[4] using combustion simulation software KIVA-2 to establish the NO_x emission prediction model of medium-speed diesel engines. This method has a strong explanation for NO_x generation, but the modeling method is complicated, the input parameters are more, the requirements for computing resources and data are higher, and it is not suitable for remote monitoring needs. d'Ambrosio, Asprion and Finesso etc.^[5-7] used a semi-experienced approach to build a NO_x emission prediction model of diesel engines based on Zeldovich theory. This model has a simple structure and can explain the generation of NO_x to a certain extent, but it is difficult to use for remote diagnostic due to the need for accurate cylinder pressure signals as input. For diesel engines without in-cylinder pressure sensors, Banerjee, Anand, and Li Nan, etc.^[8-10] used neural networks (NN) to establish a NO_x emission prediction model for diesel engines. The neural network method has good nonlinear fitting and is simple to use, but it requires many training samples and has a large prediction error for the untrained area. The diesel engine NO_x emission prediction model Chang Yingjie etc.^[11] built using the partial least square (PLS) requires less training data and good robustness, but its nonlinear fitting ability is low. Since the method of PLS modeling can only solve nonlinear problems in a very small extent, for many nonlinear practical problems, the PLS method is not very effective^[12]. Hu Jie etc.^[11] use the neural network partial least square (NNPLS) to establish a NO_x emission prediction model of diesel engines, which also has the characteristics of both PLS (requires few training sample point and can fully extract useful information from independent variables) and NN (high degree of nonlinear fitting) ,and it can obtain higher prediction accuracy under the training of a small number of sample points. However, for remote online monitoring systems, the amount of data is often abundant, so this paper selects a recurrent neural network^[14-15] that is more suitable for prediction of sequence and large amounts of data to build a NO_x emission prediction model, which can accurately predict NO_x emissions of the target engine and then realize the diagnosis of DeNO_x performance of SCR system and NO_x sensor signal.

2. OBD remote monitoring system

2.1 Introduction of systems and functions

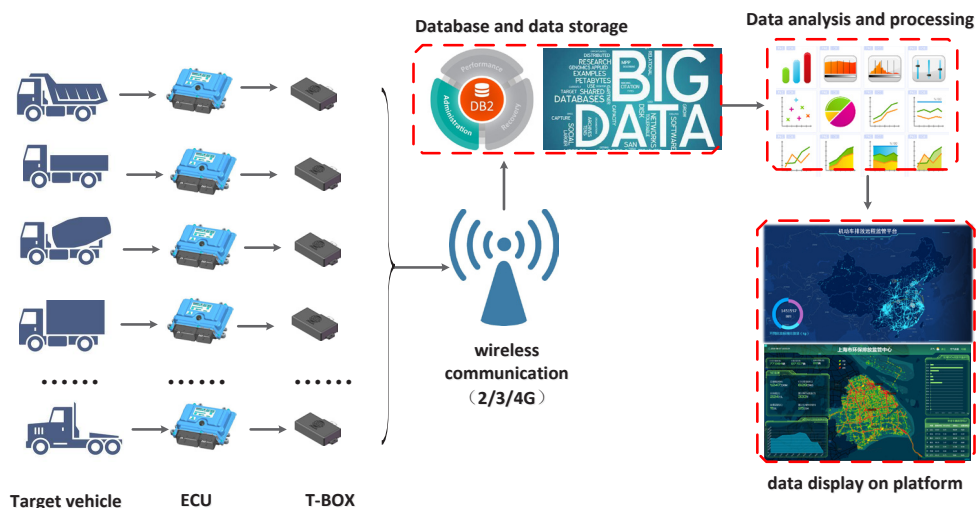


Fig. 1. OBD remote monitoring system.

OBD remote monitoring system is mainly composed of OBD remote monitoring terminal (also known as vehicle terminal/T-BOX) and remote monitoring platform. As shown in Figure 1, OBD remote monitoring terminal is usually installed in the vehicle OBD diagnostic interface, which can collect the main operating parameters of the vehicle during operation through the CAN bus (e.g. engine speed, intake mass flow, fuel consumption rate, engine net output torque, SCR inlet temperature/SCR outlet temperature, urea injection volume, urea level sensor signal, downstream NO_x sensor measurement etc.). Through the wireless communication module of the terminal, the data will be sent to the remote monitoring platform in the format specified by the platform using the 2G/4G network. The platform can accept the data uploaded by the terminals installed on different vehicles, and carry out a series of data pre-processing, storage and calculation operations to achieve the normalized tracking of continuous, real-time and online status monitoring of vehicle operating condition information, I/M status information and pollutant emissions.

After processing a large number of long-term running data from on-line vehicles, many common cheating behavior of vehicles are identified based on temperature sensor signal diagnosis module, NO_x sensor signal diagnosis module, urea injection system working status diagnosis module etc. Based on the feedback results of the real state of the actual vehicle, train the neural network self-learning model that determines the diagnosis result, determine the threshold of core criteria and the weight factor of diagnosis result of sub-model. The learned model can accurately identify the abnormal using behavior of vehicles such as complete or partial pull-out of the temperature sensor, partial dilution of the NO_x sensor, pull-out of the NO_x sensor, failure of the NO_x sensor, the using of temperature/ NO_x sensor signal generator, reflux of the injected urea, the using of poor quality urea and poor performance of SCR catalyts.

This paper focuses on the NO_x emission prediction method used for the diagnosis of DeNO_x performance of the vehicle SCR system to ensure that the accuracy of prediction output by the model meets the needs of remote diagnostics.

2.2 Method of SCR system DeNO_x performance diagnosis

After the platform accepts the OBD data uploaded by the vehicle, the filtered data is used as the input to the DeNO_x performance diagnostics module. Due to the technology configuration of a single downstream NO_x sensor used in diesel vehicles, it is necessary to diagnose the DeNO_x performance of the SCR system through the following methods:

- 1) Predict upstream NO_x emission level through the NO_x emission prediction model of the platform, as shown in figure 2.
- 2) Estimate the maximum conversion capacity of the catalyst through inlet and outlet temperature of the SCR system, estimate the theoretical maximum reaction capacity of the SCR system under the actual injection condition by the reactant real-time injection volume, and obtain the DeNO_x quantity, as shown in Figure 3.
- 3) Predict downstream NO_x emission mass through upstream NO_x primary emission cumulative mass and the theoretical DeNO_x mass. Finally, compare the downstream NO_x emission mass and the actual NO_x emission mass calculated by t NO_x sensor (as shown in Figure 4) to evaluate the real-time conversion capability (performance) of the SCR system.

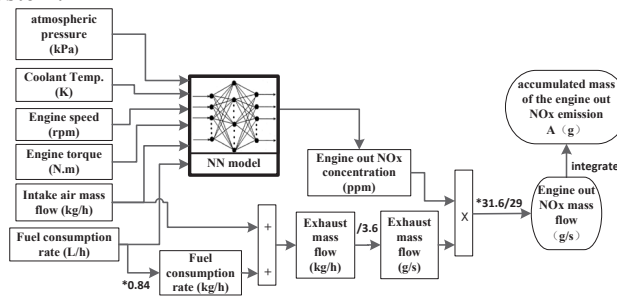


Fig. 2. Prediction of original NO_x levels through neural network model by using of OBD data.

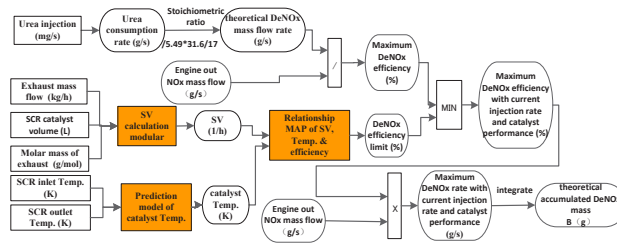


Fig. 3. Prediction of DeNO_x efficiency through dosing rate, exhaust gas temperature and space velocity.

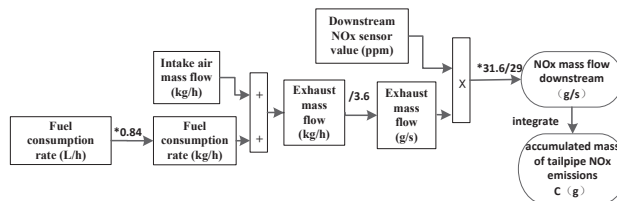


Fig. 4. Calculation of NO_x emission level in tailpipe through the downstream NO_x sensor signal.

The evaluation method of the real-time conversion capacity of the SCR system is as follows. Since the conversion efficiency of normal vanadium based SCR catalyst can generally reach more than 60% under the conditions above the starting temperature,

Condition 1 gives the basic judgment of the catalyst conversion efficiency, Condition 2 is to diagnoses the DeNO_x efficiency of SCR system by checking the logic correspondence among the accumulated mass of the upstream NO_x emission, the accumulated mass of urea injection and the accumulated mass of the downstream NO_x emission. When Condition 2 fulfilled, the conversion performance of SCR system is considered normal.

$$\text{cond. 1: } \frac{\sum_{i=1}^5(Ai-Ci)}{\sum_{i=1}^5 Ai} > 50\% \tag{1}$$

$$\text{cond. 2: } 0.6 \sum_{i=1}^5 Ci < \sum_{i=1}^5(Ai - Bi) < 1.5 \sum_{i=1}^5 Ci \tag{2}$$

In the formulas:

A is the accumulated mass of the engine out NO_x emission during a certain time;

B is the theoretical accumulated DeNO_x mass corresponding to the consumed urea during a certain time;

C is the accumulated mass of tailpipe NO_x emissions during a certain time.

When the downstream NO_x sensor works normally and the signal is reliable, the accuracy of this diagnosis method mainly depends on the accuracy of the upstream NO_x emission prediction model.

This paper uses the LSTM neural network method to predict the engine out NO_x emission of the target vehicles because OBD remote monitoring platform can easily get a lag quantity of real-time data that can be quickly classified in to diffident models and brands.

3 Remote NO_x emission estimation model

3.1 Influence factors of diesel NO_x generation and preliminary confirmation of input parameters

Three conditions are required for the formation of diesel exhaust NO_x: excess oxygen in the combustion chamber, high temperature of the combustion process, and duration of high temperature [17].

As the advance angle of fuel injection is delayed, the amount of fuel involved in the mix combustion decreases, and the initial heat release and peak of the heat release are reduced. Thus, the maximum combustion temperature is reduced. Meanwhile, the duration of high temperature in combustion chamber get shorted and the NO_x emission is significantly reduced.

The diesel engine speed can affect both the high temperature and the duration of the combustion process. When the injection angle is fixed, the higher the speed, the shorter the actual injection time. At the same time, the charge coefficient decreases with increasing speed, resulting in a reduction in NO_x emission concentration.

The load represents the maximum combustion temperature and the duration of the high temperature. Torque, cycle fuel injection quantity (mg/stk), excess air coefficient etc. represent the load factor. And the torque can be obtained by product of the rated torque in the vehicle static information and the percentage of net output torque in the OBD data flow; cycle fuel injection quantity can be calculated by the proportional relationship between the number of engine cylinders in the vehicle static information and the speed and the fuel consumption rate in the OBD data flow; excess air coefficient can be calculated by the intake quantity and the fuel consumption rate in the OBD data flow.

For diesel engines with EGR, the EGR rate is also an important factor that affects the original NO_x emission. However, most of heavy-duty vehicles that meet the implementation of the CN_IV and CN_V Emissions Regulations adopt the SCR-only DeNO_x technology

route, the proportion of vehicles using EGR is small, so the EGR rate is not considered in this paper.

Intake resistance, exhaust resistance, temperature after intercooling etc. also have a great impact on emissions, but these parameters mainly depend on load factors, which has been reflected in the fuel consumption rate factors and will not be considered.

The environmental factors such as the atmospheric pressure, ambient temperature, ambient humidity of the engine-post-processing system also have an impact on the generation of NO_x, and the atmospheric pressure can be obtained through the on-board sensor signal, the ambient temperature and ambient humidity etc. can be confirmed through the input in the platform atmospheric environment module with the GPS signal later. Then correct NO_x emissions.

According to the requirement of GB17691 for remote monitoring of OBD data flow, the types of data that are shown in Table 1 must be read from OBD interface.

Table 1. Requirements for remote monitoring of data flow in GB17691.

Number	Data items
1	Vehicle speed
2	Atmospheric pressure (direct measurement or estimation)
3	Maximum benchmark torque of the engine
4	Engine net output torque
5	Friction torque
6	Engine speed
7	Engine fuel flow
8	NO _x sensor output
9	SCR inlet temperature (if applicable)
10	SCR outlet temperature (if applicable)
11	DPF differential pressure
12	Intake air mass flow read by the air quality flow sensor
13	Urea tank level
14	Fuel tank level
15	Engine coolant temperature
16	Latitude and Longitude

Based on the standard output parameters of OBD interface in Table 1, combined with the parameters related to the formation of NO_x emissions, preliminarily select seven parameters, such as engine speed, engine net output torque, atmospheric pressure, cooling water temperature, fuel consumption, SCR inlet temperature, and intake mass flow, as the initial input of the model. Then filter and confirm relatively independent characteristic parameters as the final input of the neural network by the PLS method.

3.1 PLS method

The PLS method uses the idea of main component analysis and reduce the multiple correlation of the independent variables by extracting the component. However, it is different from principal component analysis. Partial least square method extracts components in groups of independent and dependent variables at the same time. The extracted components satisfy the following two conditions ^[18]:

(1) The extracted component group should contain the variation information in the independent variable and the dependent variable as much as possible;

(2) The correlation between the independent and dependent variable components in the same group is as large as possible, that is, their correlation coefficients are as large as possible.

If t represents the component extracted from the independent variable X and r represents the component extracted from the dependent variable Y , the above two conditions mean

choosing the largest variance $\text{Var}(t)$ and $\text{Var}(r)$ and the largest correlation coefficient $\text{Corr}(t, R)$.

consumption

Assuming that there are w and v that make formula $t = X_w$ and $r = Y_v$ true, the above condition can be expressed the extremum problem in formula (3):

$$\max \sqrt{\text{Var}(t)\text{Var}(r)} \text{Corr}(t, r) = \max \text{Cov}(Xw, Yv) = \max(w^T X^T Y v) \quad (3)$$

Solve w and v using singular value decomposition. A set of eigenvectors $t=X_w$ and $r=Y_v$ can be obtained through decomposition of the equation $X^T Y = \sum e_i f_i g_i^T$, where $w=f_i$ and $v=g_i$. It has been proved in literature [19] that the covariance $\text{Cov}(t, r)$ at this time is the largest.

The load vectors of feature vector are defined as $m = X^T t / t^T t, n = Y^T r / r^T r$; The regression factor of the regression model $\hat{r} = ct$ is $c = t^T t / r^T r$, and the components of the first group extracted are tm^T and ctn^T . The independent variables and dependent variables were decomposed to: $X = X_1 + tm^T$, $Y = Y_1 + ctn^T$, and do the same thing for X_i and Y_i . Assuming A times of decomposition, independent variables and dependent variables can be expressed as:

$$X = t_1 m_1^T + \dots + t_A m_A^T + X_A \quad (4)$$

$$Y = c_1 t_1 q_1^T + \dots + c_A t_A q_A^T + Y_A \quad (5)$$

The relationship between the variables X and Y can be expressed as the following:

$$Y = X_1 b_1 + \dots + X_B b_B + Y_E \quad (6)$$

Where B is the number of independent variables, B_i is the regression factor of the i th independent variable, X_i is the i th independent variable, and Y_E is the residual matrix.

3.1.1 Screening of independent variables

Although the partial least square method extracts the most relevant components from independent variables and dependent variables, the insignificant variables in the independent variables will still have an adverse effect on the model [20], so the independent variables must be filtered. The independent variables are tested by the significance coefficient as follows:

The regression factor b_i is calculated using all samples for the i th independent variable and the regression factor $b_{i(-j)}$ is calculated using the sample except the j th sample point, and the following parameters are defined:

$$\sigma_{b_i} = \sqrt{\sum_{j=1}^n (b_{i(-j)} - b_i)^2} \quad (7)$$

The significance of the independent variable is judged by whether the significance coefficient $|b_{i(j)} / t_{0.05, n-1} \sigma_{b_i}|$ is greater than 1. The greater the significance coefficient, the higher the significance of the independent variable. When the significance coefficient is less than 1, the independent variable is not significant.

However, simply filtering the independent variables by whether the significance coefficient $|b_{i(j)} / t_{0.05, n-1} \sigma_{b_i}|$ is greater than 1 does not result in the best results. Therefore, two parameters in the formula (7) and formula (9) are used to determine the accuracy of the model

to assist in the selection of independent variables. The root mean square error can show the average prediction effect of the model.

In the modeling process, the sample is divided into two parts: the modeling sample and the testing sample. Training samples are used to build the model, and test samples are used to verify the prediction performance of the model.

The root mean square of the training sample prediction error:

$$R_{tr} = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / n} \quad (8)$$

where y_i represents the actual value of the i th training sample in the dependent variable, \hat{y}_i represents the predicted value of the i th training sample and n is the number of training samples.

The root mean square of the test sample prediction error:

$$R_{te} = \sqrt{\sum_{j=1}^m (y_j - \hat{y}_j)^2 / m} \quad (9)$$

where y_j represents the actual value of the j th test sample in the dependent variable, \hat{y}_j represents the predicted value of the j th test sample and m is the number of test samples.

After judging the significance coefficient is less than 1, adopt the method of screening one by one to remove the insignificant variables. First calculate the R_{tr} and R_{te} values of the model before removing an independent variable, and then calculate the R_{tr} and R_{te} values again after removing the independent variable. If both R_{tr} and R_{te} become smaller, it is reasonable to remove the independent variable, otherwise it should not remove the independent variable. If the final R_{tr} and R_{te} values are large, it means that the relevant information has been missed or the nonlinearity was serious, and the information should be re-extracted.

3.1.2 Removal of specific sample points

Because of experimental errors or other reasons, a small number of sample points deviate too much from other sample points, and when they are used to model or validate the model, there is a large deviation in the mode. Therefore, this part of the data needs to be eliminated. There are a statistical test method proposed in the literature [18].

The contribution rate of the j th sample to the h th component t_h is defined as:

$$T_{hj}^2 = \frac{t_{hj}^2}{(n-1)s_h^2} \quad (10)$$

where t_{hj}^2 is the square of the h th component value calculated with the j th sample. s_h^2 is the variance of component t_h .

The cumulative contribution rate of the j th sample is:

$$T_j^2 = \sum_{h=1}^m \frac{t_{hj}^2}{(n-1)s_h^2} \quad (11)$$

where m is the number of extracted components.

When the cumulative contribution is too large, the probability that the component will deviate the result is high. The specific determination conditions are as follows:

$$T_j^2 < F(m, n - m) \frac{m(n^2-1)}{n^2(n-m)} \quad (12)$$

The sample point is retained if the above conditions are met, otherwise it should be eliminated.

3.1.3 Determination of the structure of the PLS model

The number of extracted components will directly affect the effect of partial fitting of neural network. Determine the optimal number of components by increasing the number of components one by one. The method steps are as follows: the number of components increases one by one from 1. We can use the method in Figure 5 to determine the optimal structure of the neural network part with current number of components to obtain the PLS model with different components. As the number of components increases, the prediction error of the model first decreases. When the number of components of the model is 6, the test sample error is minimum. And then the number of components starts to increase, and the error of training sample changes little when the number of components is 4 to 6. Therefore, the optimal number of components of this model is 6.

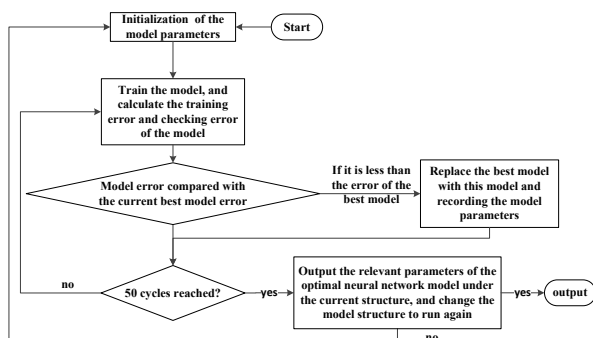


Fig. 5. The process of determining model structure.

This paper finally confirms that the parameters used for the neural network model estimate are as follows: engine speed, torque, atmospheric pressure, cooling water temperature, fuel consumption and intake mass flow. The engine test bench data are used for model training and validation. We combined with the sensor signal of the T-BOX and all environmental factors are used as input of the environmental correction factor MAP when estimating emissions for the vehicle.

3.1.4 Possible error in the PLS model

Based on previous experience in establishing the NO_x emission estimation model in the injection control strategy, it has been found that the deviation of the PLS model is mainly due to the following aspects [14]:

- (1) Error of the test data. There are some errors in the data obtained from the experiment, which will result in the error of the model prediction results. The errors are randomly distributed, and the influence of this error on the model decreases gradually as the training sample increases.
- (2) The deviation caused by fitting. The PLS method is only applicable to linear fitting, Although the accuracy of the model is improved to a certain extent by adding some quadratic terms, the quadratic curve still has a large deviation from the actual curve;
- (3) Error caused by the residual terms. During modeling with the PLS method, there is always a residual Y_E for the dependent variable that cannot be eliminated, and it also have an impact on the model;

(4) Others. The selected independent variables and training samples are not representative, so the model is poorly optimized.

In order to reduce the error of the PLS model, the PLS method of neural network is used. The composition extracted by the PLS method is fitted directly to the dependent variable, eliminating the influence of the residual terms. At the same time, the error caused by fitting can be greatly reduced by using of the neural network method.

3.2 Determination of neural network model

Although the BP (back propagation) neural network is popular and technologically sophisticated in the industry, recurrent neural network is more appropriate for sequence prediction and large amounts of data [15]. Moreover, the LSTM neural network overcomes the problem of gradient disappearance and gradient explosion caused by long-distance memory, and is a more suitable choice for training the network model [21].

2. 2.1.1 Selection of activation functions and learning algorithm

The main function of the activation function is to strengthen the ability of network learning and provide nonlinear modeling ability for the network [22]. Different activation functions have different characteristics and application situation. In the LSTM neural network, the Sigmoid function can be selected as the gate activation function. The Sigmoid function can produce values between 0 and 1. 1 means remember and 0 is for forgot. In addition, activation functions such as Tanh, ReLU, Sigmoid can be selected as state activation functions. Compared with the ReLU function, Tanh function is more robust, and its convergence speed is faster than Sigmoid function, so we choose the Tanh function as the state and output function.

The common learning algorithm of LSTM neural networks is BPTT (back-propagation through time). The BPTT algorithm is a common method for training RNN, in fact, its essence is still the BP algorithm. It calculates the hidden state and output, error and partial derivative of weight of each layer, and then optimize the parameter of vertical propagation between levels and lateral propagation on time using gradient descent [23].

3. 2.2 Determination of the number of nodes in hidden layer

The number of nodes in the hidden layer has a great impact on the performance of the LSTM neural network. If the number of hidden layer nodes is too small, the information processing ability and generalization ability of the network will decrease. Conversely, if there are too many hidden layer nodes, the network will also be complicate, the model training speed will be affected, and the risk of "overfitting" will increase. However, based on the Kolmogorov theorem, it can be seen that this hidden layer has $2n-1$ neurons, where n represents the number of neurons in the input layer. Because the number of input layer nodes of LSTM model created in this paper is 6, the number of hidden layer nodes is 13.

3.2.3 Adam optimization algorithm

The optimization algorithm is essentially a mathematical method. And common optimization algorithms include SGD (Stochastic gradient descent, SGD), Momentum, Nesterov Momentum, Adagrad (Adaptive sub-gradient, Adagrad), RMSprop, Adam etc. [21]. Adam is actually the result of an optimized combination of Momentum and RMSprop. Adam can not only store exponential decay averages of squares of past gradients like RMSprop, but also can store exponential decay averages of past gradients like momentum. In short, Adam can automatically adjust the learning rate of parameters, increase training speed, and improve

network stability. The main parameters of the recurrent neural network based on LSTM are listed in Table 2: :

Table 2. Main parameters of LSTM-based recurrent neural network model.

Training parameters	Modes and values
Number of hidden layer nodes	13
State activation function	Tanh
Gate activation function	Sigmoid
Loss function	RMSE
Optimize iterative algorithms	Adam
Training cycle	400
Gradient threshold	1
Learning rate	0.01

4. Validation of the model

4.1 Experimental acquisition of model data

We used the bench test results of a commercial heavy-duty vehicle diesel engine as the training and validation data for the model are from and the specific parameters of the engine are shown in Table 3: :

Table 3 Test engine specifications

Parameters	Content
Engine model	7H
Configuration	6-cylinder inline
Displacement	6.8L
Ignition sequence	1-5-3-6-2-4
Number of intake valves	2
Number of exhaust doors	2
Compression ratio	17.7
Type of supercharger	Variable Geometry Turbocharger (VGT)
Fuel injector	Double solenoid valve electronic unit injector
Peak torque	1000Nm@1200~1900rpm
Rated power	209kW@2300rpm

Measure parameters including engine data flow and gas emissions using instruments shown in Table 4, and figure 6 shows the bench test site.

Table 4 Parameter measurement scheme

Test instruments	Test parameters
AVL electric dynamometer test bench	engine speed, torque, coolant temperature, inlet temperature and pressure, exhaust pressure, atmospheric humidity, atmospheric temperature, atmospheric pressure
FTIR	Engine out NO _x emission concentration
ES581	Fuel delivery per cycle, engine speed, engine net output torque, atmospheric pressure, coolant temperature, SCR inlet temperature, intake mass flow

19877 groups of data are divided into the data set and training set by hierarchical sampling method, and 14877 groups are used as sample training data and 5000 groups are used as test data. Because the unit of each input data is different and the order of magnitude is different, it is necessary to normalize the data first.

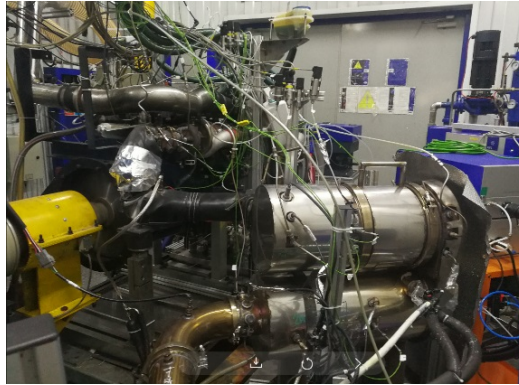


Fig. 6. Engine & after-treatment system test bench.

4.2 Results analysis

LSTM neural network and PLS model were combined to the final neural network model. And the engine speed, torque, atmospheric pressure, cooling water temperature, fuel consumption, intake mass flow were taken as inputs of neural network. The predicted results of training and test sets of the model (the network model output value after denaturalization), the error of training set and the test set, as shown in figure 7 to figure 10 respectively.

As can be seen from figure 7 and figure 8, the predictive model has good learning ability and generalization ability, and with the increase of training times, the mean square error becomes smaller and smaller. The training sample error of the model (R_{TR}) is 29.7×10^{-6} . And the RMSE of the model is 2.27% when the maximum NO_x emission of the training data is 1310×10^{-6} . The test sample error of the model (R_{TE}) is 19.9×10^{-6} . And the RMSE of the model is 2.64% when the maximum NO_x emission of the training data is 753×10^{-6} . And the position with the largest error in the test set is in the 1001 data, with an error -267.2. The test set sample error percentage is shown in Figure 11. It can be seen from Figure 11 that the maximum error percentage of the test set is only 1.18%, which has reached a very high estimation accuracy. The data with the largest error mainly appears in the first few sets of data for in each layer of the training set, which is caused by the basic characteristics of the recurrent neural network and performs well in other cases.

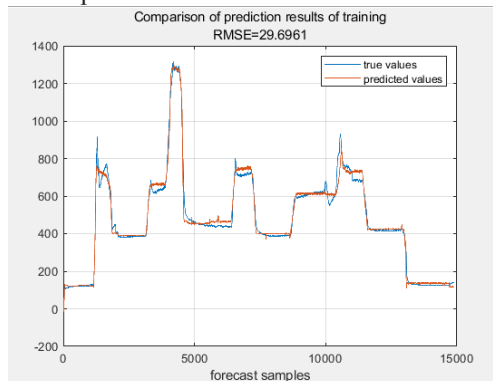


Fig. 7. Comparison of prediction results of training.

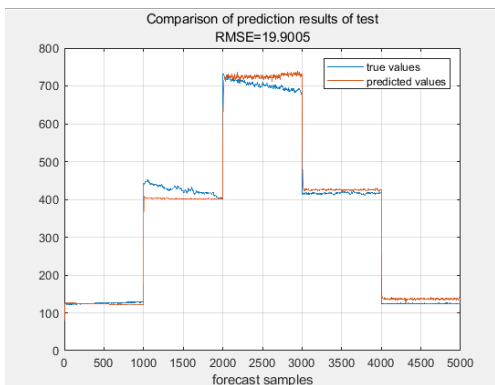


Fig. 8. Comparison of prediction results of test.

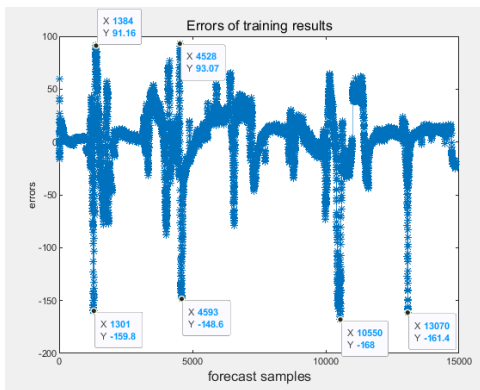


Fig. 9. Errors of training results.

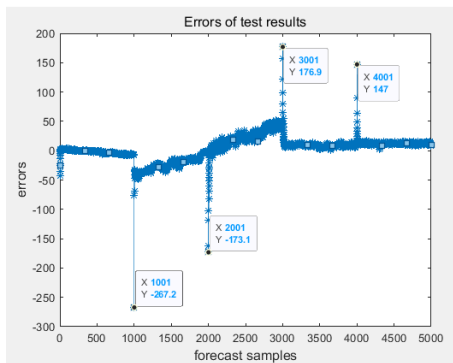


Fig. 10. Errors of test results.

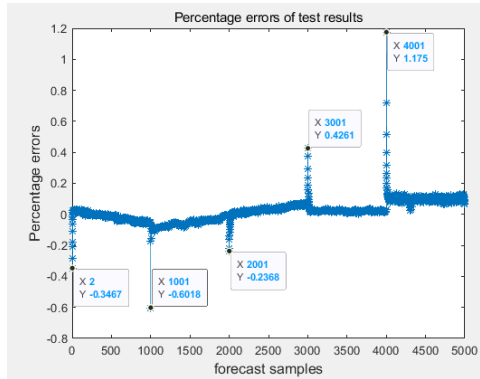


Fig. 11. Percentage errors of test results.

5. Conclusion

(1) The PLS model based on the LSTM neural network model. First of all, select the parameters according to their availability, NO_x emission correlation, and their significance. And then determine the structure of model by incrementing one by one. Finally establish the diesel engine NO_x emissions prediction model and confirm engine speed, torque, atmospheric pressure, cooling water temperature, fuel consumption, mass air flow as the input of the model.

(2) The training sample error of the model (R_{TR}) is 29.7×10^{-6} . And the RMSE of the model is 2.27% when the maximum NO_x emission of the training data is 1310×10^{-6} . The test sample error of the model (R_{TE}) is 19.9×10^{-6} . And the RMSE of the model is 2.64% when the maximum NO_x emission of the training data is 753×10^{-6} .

(3) For this diesel engine, the maximum error percentage of test set is only 1.18%, which can fully meet the requirements for remote diagnosis of DeNO_x performance of SCR system and can be used for reasonable diagnosis of downstream NO_x sensor signal.

Open Fund of the National Engineering Laboratory for Mobile Pollution Source Emission Control Technology (NELMS2017A08)

References

1. Ministry of Ecology and Environment of the People's Republic of China. China Mobile Source Environmental Management Annual Report [M]; 2019
2. Li Mengliang; Li Jingyuan; Xu Yueyun; Research on the emissions and influencing factors of in-use vehicles [C]; Proceedings of the 14th Annual Meeting of the Lubricating Oil Branch; 2010
3. Ministry of Ecology and Environment. GB-17691-2018 Pollutant Emission Limits and Measurement Methods for Heavy-Duty Diesel Vehicles (China Sixth Stage) [S];2018-06-22.
4. Pertti Taskinen. Modelling Medium Speed Diesel Engine Combustion, Soot and NO_x-emission Formations[C]//SAE International, Paris, France, 2000, 2000-01-1886.
5. Stefano d'Ambrosio, Roberto Finesso, Lezhong Fu. A control-oriented real-time semi-empirical model for the prediction of NO_x emissions in diesel engines[J]. Applied Energy, 2014(130): 265–279.

6. Jonas Asprion, Oscar Chinellato, Lino Guzzella. A fast and accurate physics-based model for the NO_x emissions of Diesel engines[J]. *Applied Energy*, 2013(103): 221–233.
7. Roberto Finesso, Ezio Spessa. A real time zero-dimensional diagnostic model for the calculation of in-cylinder temperatures, HRR and nitrogen oxides in diesel engines[J]. *Energy Conversion and Management*, 2014(79): 498–510.
8. Rahul Banerjee, Probir Bose. Development of a Neuro Genetic Algorithm Based Virtual Sensing Platform for the Simultaneous Prediction of NO_x, Opacity and BSFC in a Diesel Engine Operated in Dual Fuel Mode with Hydrogen under Varying EGR Conditions[J]. *SAE Int. Engines*, 2012, 5(2): doi: 10.4271/2011-01-2472.
9. G Anand, S Gopinath, M R Ravi. Artificial Neural Networks for Prediction of Efficiency and NO_x Emission of a Spark Ignition Engine[C]//SAE International, Detroit, Michigan, United States, 2006, 2006-01-1113.
10. Li Nan, Tang Dong, Li Changyuan. Research on prediction of emission characteristics of diesel engine blended biodiesel based on BP neural network [J]. *Automotive Engine*, 2013(5): 48-56.
11. Chang Yingjie, Lu Xianzhong, Wang Shilong. NO_x emission model of EGR diesel engine based on PLS regression [J]. *Journal of Internal Combustion Engine*, 2012, 30(4): 322-329.
12. Sun Fenglin. Research on nonlinear modeling and recursive algorithm of partial least squares regression method [D]. South China University of Technology, 2010.
13. Hu Jie, Lin Feng, Wang Tiantian, et al. NO_x emission prediction model of diesel engine based on neural network partial least squares method[J]. *Journal of Internal Combustion Engine*, 2015, v.33; No.162(06):34-39.
14. Imran F , Yunfei Y . LSTM Algorithm for SCR Outlet NO_x Based Prediction Model[J]. *Kuwait Journal of ence*, 2019, 46.
15. Gers F A , Schmidhuber, Jürgen, Cummins F . Learning to Forget: Continual Prediction with LSTM[J]. *Neural Computation*, 2000, 12(10):2451-2471.
16. Soutner D , Luděk Müller. Application of LSTM Neural Networks in Language Modelling[C]// International Conference on Text. Springer Berlin Heidelberg, 2013.
17. Jonas Asprion, Oscar Chinellato, Lino Guzzella. Optimisation-oriented modelling of the NO_x emissions of a Diesel engine[J]. *Energy Conversion and Management*, 2013(75): 61–73.
18. Wang Huiwen, Wu Zaibin, Meng Jie. Linear and nonlinear methods for partial least squares regression [M]. Beijing: National Defense Industry Press, 2006.
19. Liang Lin. Research on soft sensor modeling method based on nonlinear partial least squares [D]. Tsinghua University, 2000.
20. Roy P P, Roy K. On some aspects of variable selection for partial least squares regression models[J]. *QSAR and Combinatorial Science*, 2008, 27(3):302-313.
21. Hossein Abbasimehr et al., An optimized model using LSTM network for demand forecasting[J]. *Computers & Industrial Engineering*. 2020, (143). <https://doi.org/10.1016/j.cie.2020.106435>
22. Yi Lirong, Wang Shaoyu, Yin Lili et al. Industrial sensor time series data prediction based on multivariable LSTM [J]. *Intelligent Computer and Application*. 2018, vol.8(5): 13-16.
23. Liu Yijuan, Lei Ming, He Yang and others. Overview of artificial neural network [J]. *Network and Information*. 2015, (11): 74-75.