Construction of NOx emission states identification method of diesel bus based on judgment matrix: A case study of Nanjing

Zixin Liu
School of Transportation, Southeast University
liuzixin_seu@seu.edu.cn
+86 1885162198

Tiezhu Li (Corresponding Author)
Professor, School of Transportation, Southeast University
litiezhu@seu.edu.cn
+86 18136481719

Haibo Chen
Institute for Transport Studies, University of Leeds, Leeds
H.Chen@leeds.ac.uk

Ying Li
Dynnoteq Limited
International House, 24 Holborn Viaduct, London, EC1A 2BN, UK
ylitransportation@gmail.com

ABSTRACT

As the number of motor vehicles continues to rise worldwide, road motor vehicle emissions are becoming increasingly polluting. How to identify the emission states, analyze and reduce high-emission driving behaviors are the key to controlling road motor vehicle exhaust emission pollution. At present, there is no clear quantitative definition of high emissions for vehicles, and the identification variables of emission states are limited to emission factors or instantaneous emissions. There is limited consideration of the change rate of instantaneous emissions.

In this regard, a method for obtaining instantaneous emission state labels of diesel buses with PEMS as the data source is proposed: Firstly, the data sets of different devices are created, the iForest algorithm is used to automatically identify the outlier data, and a time-delay processing method for the vehicle emission and driving states data is proposed. Then, the models based on decision tree are used to predict the instantaneous emissions and its change rate. In the next step, considering the time series characteristics of emissions, an emission states judgment matrix is constructed to obtain the emission state labels of vehicles.

Compared with the current mainstream cluster recognition algorithm, the method of obtaining emission state labels through the judgment matrix strengthens the consideration of the time series characteristics of emissions. It can provide clear definition of different emission states and identify high-emission states/low-emission states which will establish a good foundation for subsequent real-time identification of emission states and high-emission driving behavior analysis.

Key Words: Diesel buses, Emission states, Decision tree model, Adaptive clustering algorithm, Judgment matrix

INTRODUCTION

With the rapid development of economy and society, the number of motor vehicles in the world continues to rise. In 2021, the total emissions of four pollutants (including CO, HC, NOx and PM) from motor vehicles in China were 15.577 million tons. Among the various models, buses are the main contributors to CO and HC emissions; Trucks are the main contributors to NOx and PM. Motor vehicle exhaust emissions have become the main source of air pollution in medium and large cities in China, and the interaction between pollutants will form secondary pollutants, causing respiratory diseases such as asthma and pneumonia, posing a huge threat to the health of residents. Therefore, the precise control and treatment of urban road motor vehicle pollution emissions has become a top priority of urban pollution control.

The forecast of vehicle emissions needs to be based on a large amount of real data. The Portable Emission Measurement System (PEMS) is widely used in vehicle emissions forecasting because of its high sensitivity and ability to accurately measure the true emissions of the actual road of a motor vehicle. KC Johnson et al. conducted comparative experiments in the mobile emission laboratory and PEMS that comply with European federal regulations, respectively, and verified the effectiveness and accuracy of PEMS for heavy-duty diesel vehicle exhaust emission detection[1]. Artur Jaworski et al. established an emission model based on the enhanced regression tree method by collecting vehicle emission data on roundabouts through PEMS, and the experimental results show that the model can effectively estimate the emission level of exhaust pollution at roundabouts [2].

Motor vehicle emissions belong to a typical time series data, and time series algorithms are widely used in different research fields, such as road traffic flow prediction, environmental pollution prediction, and weather forecasting. Bishop et al. used the Advanced Vehicle Simulator (ADVISOR) [3] and the Quantitative Regression (QR) model [4] to estimate vehicle emissions under real driving conditions. In order to solve the problems of poor prediction effect of linear regression method in highly complex and time-varying nonlinear data and inability to identify exogenous drivers, Liu et al. used an improved and optimized SVM method to predict NOx emissions in diesel engines [5], and Lv et al. used an innovative least squares support vector machine method to predict NOx emissions in coal-fired boilers [6], but these methods are based on a pre-defined nonlinear framework. Does not accurately reflect true nonlinear relationships in time series [7] [8] [9].

At present, there is no clear definition of vehicle high emission in academic studies, and the emission data itself does not carry the category label of emission states. XIE et al. proposed an automatic and rapid identification method for vehicle high emission. Used K-Medoids to cluster and label the 3D training data set. Then used K-NN algorithm in the trained model to realize automatic and rapid identification of high-emission sources [10]. In the case of scarce high-emission category tags, KANG et al. used semi-supervised learning method to amplify labeled data sets, and constructed high-emission source identification models based on single classification support vector machine and semi-supervised single classification support vector machine respectively [11]. Li et al. adopted the weighted extreme learning machine to solve the

problem of sample imbalance between high emission samples and normal emission samples. The active learning method was adopted to solve the problem that the model could not assign the inspection label of the Vehicle administration to the new samples [12]. Kang et al. adopted the method of single classification support vector machine to solve the problem of low reliability of the label of normal emission samples of the Vehicle administration, and improved the identification accuracy of single classification support vector machine by introducing semi-supervised learning [11]. However, this method requires manual adjustment of the hyperparameters in the training process. The degree of automation of the model is low.

At present, the identification of high emissions of motor vehicles mainly considers the instantaneous emissions or emission factors, but does not consider the position of instantaneous emissions in its change cycle in detail, that is, the change rate of instantaneous emissions. Therefore, this paper takes the normal identification process of high emission of motor vehicles as the logical basis and take the change rate of instantaneous emission into the judgment variable group. Taking the NOx emission of diesel buses in Nanjing as the research object, the instantaneous emission and the change rate of the instantaneous emission were predicted respectively. According to the distribution of the two emission variables, the emission states identification matrix was constructed to complete the construction of the diesel bus NOx emission states identification method.

METHODOLOGY

Based on the real-time vehicle driving data and real-time vehicle emission data that can be collected by equipment of OBD and PEMS, this paper processed the time-delay effect of data through a customized process, established the emission related prediction model, and obtained the emission state labels according to the model results.

DATA ACQUISITION AND PREPROCESSING

This part processes the collected vehicle running state data and emission related data. Firstly, a more comprehensive feature data set is constructed through data feature derivation. Then the outlier data are identified and processed by artificial recognition, Isolated Forest algorithm and Three-Point Simple Moving Average algorithm. Finally, the driving cycles are divided according to the requirements of the subsequent prediction model, and the data from different sources is processed by the time-delay effect processing method considering the characteristics of the driving cycles.

Dataset creation and outlier recognition

Take NOX as an example to explore the correlation between vehicle driving states and exhaust emission changes. Considering the data accuracy and driving route characteristics, relevant data items are screened out, including: time, longitude and latitude, driving speed, instantaneous emission of NOX. In order to improve the accuracy of subsequent prediction, data features are derived. It has been shown in the literature that considering the change rate of acceleration can effectively improve the accuracy of emission calculation model. Therefore, the instantaneous acceleration and its changing rate are calculated on the basis of traveling speed, and the change rate of instantaneous emission of NOX in seconds is calculated on the basis of instantaneous emission of NOx emission, constituting a new characteristic dimension.

When data is collected during the actual operation of the buses, the quality of the data cannot be monitored in real time. Due to equipment failure and other problems, the data may be missing and abnormal (outlier), so it is necessary to preprocess the data to avoid the influence of wrong data on the subsequent research.

First, the missing values are identified by calculating the adjacent timestamp difference:

$$TimeGap_i = TimeStamp_i - TimeStamp_{i-1}$$

$TimeGap_i = 1s$ indicates that there are no missing values; $TimeGap_i > 3s$ indicates that the device is disconnected for too long and the data bar needs to be interrupted; When$TimeGap_i \in (1, 3]$, the missing values need to be calculated by the mean method.

Considering the characteristics of Isolation and small amount of abnormal data, Isolation Forest algorithm was used to calculate the degree of isolation of data points to identify outliers.

Considering that the Isolation Forest algorithm is sensitive to global sparse points and adopts an unsupervised anomaly detection method, manual inspection is required for the identified outlier data items. The three-point moving average algorithm is used to replace the data bar that is truly an outlier.

Time delay effect treatment

Considering the difference of data transmission time of different equipment and the time consumed by mechanical transmission, there is a certain lead-lag relationship between emission data and vehicle driving state in the time series, which will reduce the correlation between variables and reduce the accuracy of emission state identification. By moving part of the data along the time series, the negative influence of the lead and lag effect can be eliminated, and the moving time unit is the time delay value between the data.

Since pollutant emissions have a high correlation with the driving state such as real-time speed, acceleration, this paper corrects the TimeLag by considering the micro-drive characteristics on the basis of processing the time-delay effect through data correlation, and establishes a data time-delay effect processing method considering the micro-travel characteristics.

The emission characteristics of buses in the idle state are quite different from normal driving state. In order to accurately identify the emission characteristics of different driving states, it is necessary to exclude the idle state data of the vehicle. Calculate the average velocity $\overline{Speed}$, velocity standard deviation σ_{Speed}, mean positive acceleration $\overline{Acc_+}$, mean negative acceleration $\overline{Acc_-}$, standard deviation of acceleration σ_{Acc}. For the i_{th} driving cycle($i = 1,2, \ldots m$), the characteristic values of each driving cycle are calculated separately as well.

Based on the eigenvalues, the time-delay value of each driving cycle is calculated as:

$$TimeLag = TimaLag_{basic} \times (1 + \alpha_i)$$

Where $TimaLag_{basic}$ refers to the underlying time-delay value, α_i refers to the correction factor for each driving cycle.

For m driving cycles, the $TimaLag_{basic}$ expressed as the average of all driving cycle translation time values:

$$TimaLag_{basic} = \frac{\sum_{i=1}^{m} \tau_i}{m}$$

For the i_{th} driving cycle, set the interval length K of the moving time value to$[-5s, 5s]$, X_t and Y_t represent its velocity and instantaneous emission of NOx sequence data. Calculate its lead lag correlation coefficient series:

$$r_k = \frac{S_k}{S^2} = \begin{cases} \frac{1}{n-k}\sum_{t=1}^{n-t}\left(\frac{X_t - \bar{X}}{S}\right)\left(\frac{Y_{t+k} - \bar{Y}}{S}\right), k \geq 0 \cap k \in K \\ \frac{1}{n+k}\sum_{t=1}^{n-t}\left(\frac{X_t - \bar{X}}{S}\right)\left(\frac{Y_{t+k} - \bar{Y}}{S}\right), k \leq 0 \cap k \in K \end{cases}$$

Where $\bar{X}$ and $\bar{Y}$ are the mean of the time series data, S_k refers to the covariances.
Find the maximum value in the leading lag correlation coefficient series, and its corresponding translation time value k is the translation time value τ_i of the short trip. Assign weights to the travel characteristic terms, and calculate the percentage difference between the i_{th} driving cycle and the total stroke as the correction coefficient α_i the driving cycle:

$$\alpha_i = \frac{\overline{Speed}_i}{\overline{Speed}} \times \omega_{AS} + \frac{\sigma_{Speed_i}}{\sigma_{Speed}} \times \omega_{\sigma S} + \frac{\overline{Acc_{+_i}}}{\overline{Acc_+}} \times \omega_{AA+} + \frac{\overline{Acc_{-_i}}}{\overline{Acc_-}} \times \omega_{AA-} + \frac{\sigma_{Acc_i}}{\sigma_{Acc}} \times \omega_{\sigma A}$$

PREDICTION MODEL CONSTRUCTION

Based on the fixed relationship between vehicle fuel efficiency and mechanical transmission, considering that the existing literature supports the functional relationship between vehicle speed, acceleration, the changing rate of acceleration and instantaneous emissions, a supervised learning algorithm is used to predict the instantaneous emission of NOX and its change rate.
Considering the computational efficiency and model accuracy, the model takes the Decision Tree and its integrated algorithm as the direction, establish and draw the learning curve of the Random Forest Model, the Gradient Boosting Decision Tree Model and the eXtreme Gradient Boosting Model respectively. By observing the influence of the change of a single hyperparameter on the MSE, the approximate range of adjustment is determined. Then the parameters are precisely tuned using the five-fold cross-validation and grid search method, and fuse the established regression model.
Since the grid search method has some effect in suppress overfitting, a NOX instantaneous emission quality prediction model with higher prediction accuracy and stronger generalization ability can be obtained by comparing the validation set of each model with Sum of Squared Error (SSE), Mean Squared Error (MSE) or Root Mean Squared Error (RMSE).

EMISSION ATATE LABELS ACQUISITION

Instantaneous emission of NOx is no longer used as the only influence variable for emission state identification. The change rate of instantaneous emission of NOx is introduced to supplement the description of the change trend of instantaneous emission, and the instantaneous emission state is replaced by continuous emission state, so that the relevant identification of emissions can better adapt to the continuity of the driving state. At the same time, interval is used instead of numerical judgment to reduce emission states identification errors caused by prediction errors.

Influencing factor clustering

The influencing factors of emission states of NOx include instantaneous emissions and its change rate. In order to improve the adaptability of the emission state identification method, the automatic boundary detection method is adopted. Synthesize the most recent dataset

generated by the prediction, and then determine the maximum group number and boundary group number for each influencer.

The Initial K-Center algorithm is used to determine the position of the initial center point. First, the center values of the high-value region and the low-value region of each influencing factor are calculated. Then form matrix containing all the central values, At the end of the algorithm run, the initial K center is automatically generated. Use different adaptive clustering algorithms to cluster the influencing factors separately, With the help of DBI as a metric to determine the best clustering algorithm, DBI calculates the distance to measure the similarity in the same category and the difference between different categories:

$$E_{DBI} = \frac{1}{k}\sum_{i=1}^{k}\max_{j\neq 1}\left[\frac{avg(C_i)+avg(C_j)}{d_{cen}(u_i,u_j)}\right]$$

Where k refers to the number of clusters; $C_\theta(\theta = i,j)$ refers to the sample points of the θ class, $avg(C_\theta)$ refers to the average Euclidean distance between the sample points and the center pointu_θ of the θ class; $d_{cen}(u_i,u_j)$ refers to the Euclidean distance between the center points of different classes.

The smaller the DBI refers to the higher the similarity between the same classes, the greater the differences between different classes and the better the clustering effect. Finally, the high-value interval, normal-value interval and low-value interval of the influencing factors under the optimal clustering algorithm are marked.

Construction of Emission State Labels Judgment Matrix

Define the high emission state of NOx: Instantaneous emission of NOx and its change rate belong to the corresponding high-value interval. Define higher emission state of NOx: One of the instantaneous emissions of NOx or its change rate belongs to its corresponding high-value interval and the other one belongs to its normal-value interval. The same applies to defining low emission state of NOx and lower emission state of NOx. According to the definition of different emission states, the identification rules for different emission states labels of motor vehicles are shown in Table 1.

Table1 Judgment Matrix for NOx Emission States Labels

		Instantaneous emission of NOx		
		Low-value interval	Normal-value interval	High-value interval
Change rate of instantaneous emission of NOx	Low-value interval	Low emission state	Lower emission state	Normal emission state
	Normal-value interval	Lower emission state	Normal emission state	Higher emission state
	High-value interval	Normal emission state	Higher emission state	High emission state

CASE STUDY

ROUTE DISTRIBUTION AND DATA ANALYSIS

The study collected the driving data of Nanjing No. 100 diesel bus for 4 days. The route distribution is shown in the Fig1-a). Data acquisition is carried out by OBD and PEMS. The

acquisition frequency is 1Hz, and the collected items include vehicle position information (longitude, latitude, elevation), exhaust system states information (exhaust pipe temperature, exhaust mass flow rate, exhaust volume flow rate), exhaust emission information (instantaneous fuel consumption of each emission, instantaneous quality), motor states information (current, voltage, motor temperature) and other vehicle operating parameters in a total of 177 items. A total of more than 33,000 valid driving data were collected on 5 complete trips.

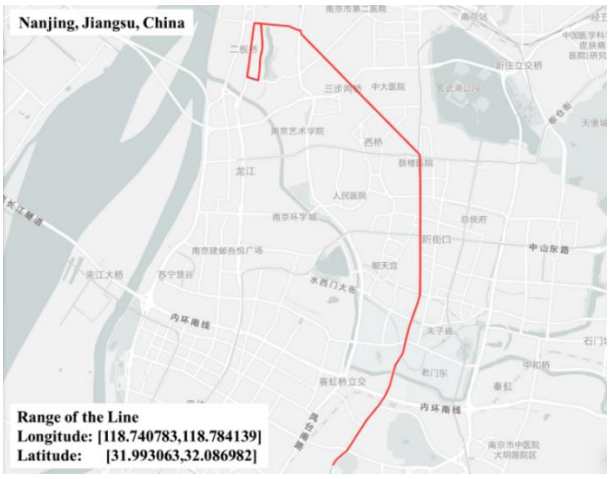
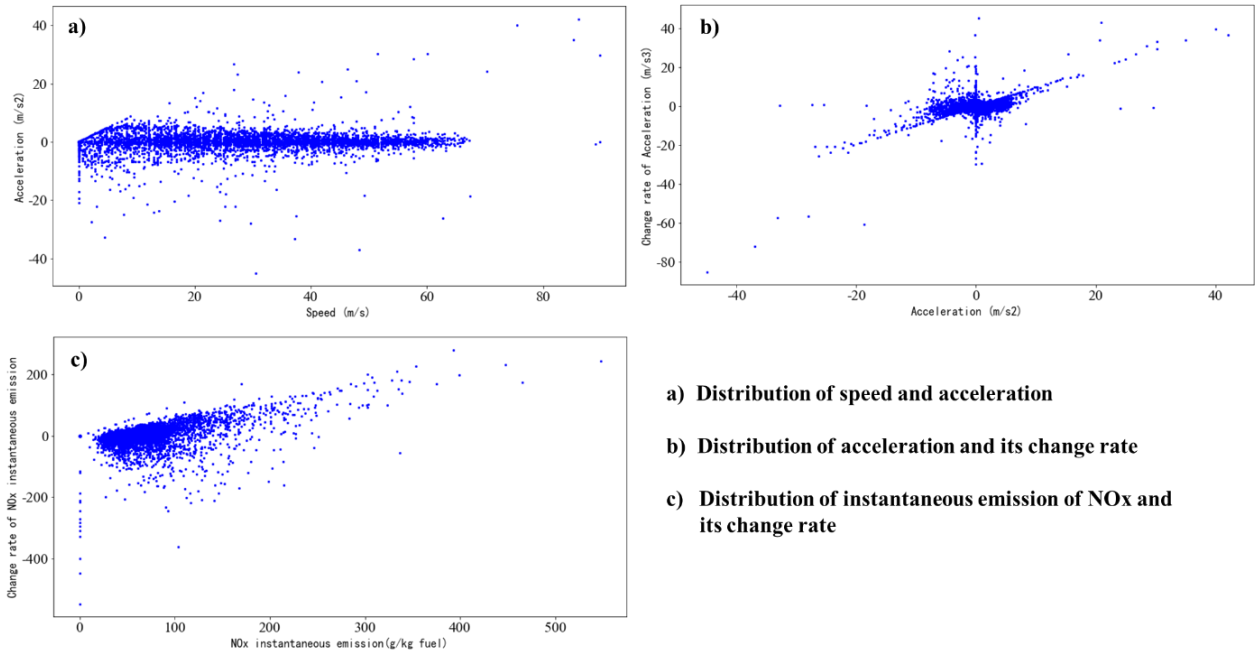


Fig1 Route distribution and data distribution

A new data set is constructed based on the second-by-second speed of the vehicle and the instantaneous emission of NOx to complete the derivation of data characteristics. According to the instantaneous speed data of the vehicle, the instantaneous acceleration and its change rate are calculated. The change rate of instantaneous emission of NOx is calculated based on the data of instantaneous emissions of NOx. Taking Day1 data as an example, it can be seen from Fig2 that without considering whether the relationship between variables is abnormal, due to equipment collection accuracy and other problems, there are numerical unreasonable data points in each variable indicates that the data needs to be processed with outliers.



a) **Distribution of speed and acceleration**

b) **Distribution of acceleration and its change rate**

c) **Distribution of instantaneous emission of NOx and its change rate**

Fig2 Distribution of data by variable

The Isolation Forest algorithm calculates the degree of alienation of data points to identify outliers, and the results are shown in the figure below. It can be seen from Fig 3 that instead of taking the instantaneous emission of NOx as the only identification standard, the outliers identified by the isolation algorithm are distributed in various numerical intervals of instantaneous emission of NOx. By identifying outliers in the driving state related variables and emissions-related variables, the causes of outline points can be effectively explained.

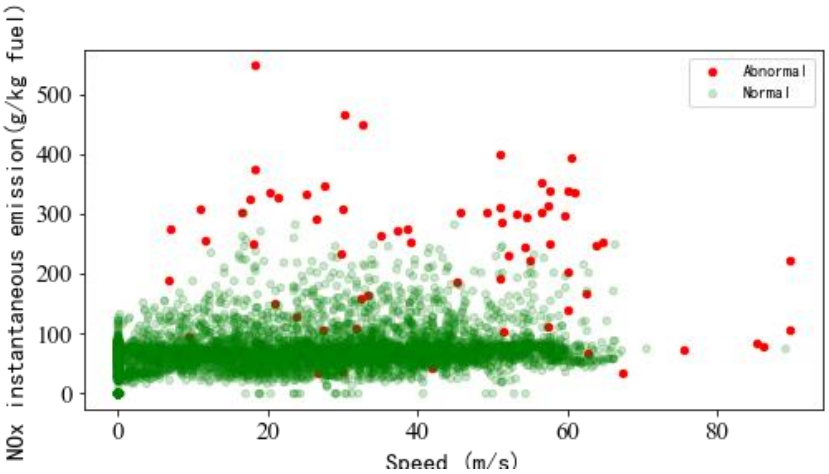


Fig3 Distribution of normal data and abnormal data after Isolation Forest Algorithm

The causes of outline points are mainly divided into four categories, among which the internal anomalies in the category account for more than 89%, and the abnormal relationship between different categories account for about 11%. Taking the driving state related variables as an example, since the acceleration and its change rate are calculated by the speed and acceleration respectively, there will be no abnormalities in the relationship between them, so the anomalies in the same category variables are described as value anomalies, that is, the value of isolated points are outside the normal threshold. Anomalies in the relationship between different categories can be described as the value of abnormal points are within their normal threshold, but the relationship function between categories does not apply to most other points. Since there is a certain functional relationship between instantaneous emission of NOx and driving state related variables, the identification of abnormal data between different categories can effectively screen out hidden abnormal data to improve the accuracy of prediction models and the emission state judgment matrix.

EMISSION PREDICTION MODEL CONSTRUCTION

The prediction of NOx instantaneous emissions and its change rate are based on the decision tree model. Firstly, the learning curves of RF, GBDT, and XGBoost are plotted, the influence of the change of a single hyperparameter on the MSE is observed, and the approximate range of hyperparameter tuning is determined. Since the different hyperparameters will be intrinsically linked, the five-fold cross-validation and grid search method are used for precise parameter tuning. The three established regression models are fused through the decision tree to obtain the fusion model. By calculating MSE for validation sets, the final prediction function is determined. Taking the process of forecasting instantaneous emissions of NOx as an example:

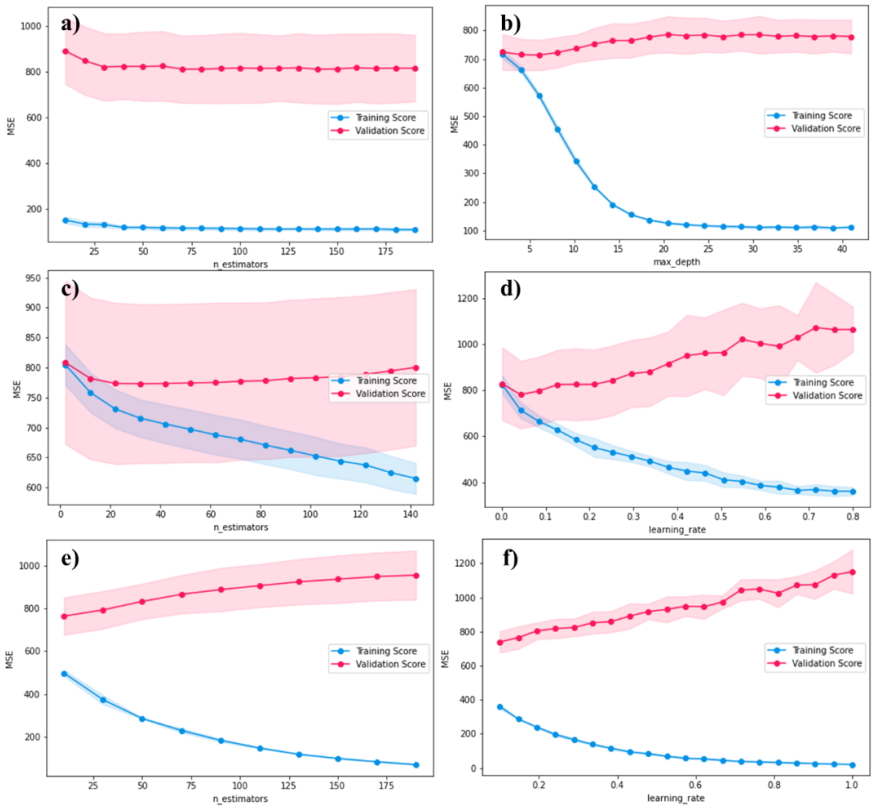
Random Forest Model: $n_{estimators}$ and max_{depth} were chosen as variables that need to be adjusted precisely. By plotting the learning curve, It can be seen From Fig4-a) and b) that the

model is severely overfitting. As the $n_{estimators}$ grows, the MES decreases initially, but after 100, the model plateaus. Continuing to increase the value of the $n_{estimators}$ will increase the operation time significantly. The growth trend of max_{depth} is similar, after reaching 25, the MSE of the model basically no longer changes. Therefore, these two values are used as the center of the grid search for fine-tuning parameters. The final search results are optimal when $n_{estimators} = 130,\ max_{depth} = 26$.

GBDT model: Select $n_{estimators}$ and learning_rate as variables which need to be adjusted precisely. As can be seen from Fig4-c) and d), with the growth of $n_{estimators}$, the MES showed a downward trend at the beginning, but gradually increased after 20. The MSE of the model is minimized when $learning_{rate} = 0.05$, and further growth increases the validation set error. The final search results are optimal when $n_{estimators} = 32,\ learning_{rate} = 0.093$.

XGBoost model: Plotting the learning curve for $n_{estimators}$ and learning_rate. As can be seen from Fig4-e) and f), the MES continues to increase as $n_{estimators}$ grows. The MSE of the model is minimized at the $learning_{rate}$ minimum. Therefore, set the parameter space as $n_{estimators} \in [2,25]$, $learning_{rate} \in [0.01,0.1]$. The final search results are optimal when $n_{estimators} = 16,\ learning_{rate} = 0.079$.

Fusion model: RF, GDBT and XGBoost are used as base learners, and decision trees are fused as meta-learners to establish an ensemble model.



a) n_estimators in RF

b) Max_depth in RF

c) n_estimators in GBDT

d) Learning_rate in GBDT

e) n_estimators in XGBoost

f) Learning_rate in XGBoost

Fig4 MSE changes with parameters in models

A comparison of all algorithm results is shown in Table 2. Through comparison, it can be seen that the grid search of all algorithms suppresses the overfitting case to some extent. Since the MSE of the validation set is more representative of the generalization ability of the model than the test set, it can be seen that the accuracy order of model prediction: GDBT⩾XGBoost ⩾ RF⩾ Fusion Model. Therefore, the accurately adjusted GBDT model is finally used to predict

the instantaneous emissions of NOx.

Table2 Prediction accuracy of instantaneous emission of NOx for each model

	RF	RF grid	GBDT	GBDT grid	XGBoost	XGBoost grid	Fusion	Grid Fusion
	Prediction accuracy of instantaneous emission of NOx							
Train Set	118	118	639	696	157	620	1699	1561
Validation Set	850	811	781	745	960	758	1537	1589
Test Set	723	787	659	710	841	711	1837	1679
	Prediction accuracy of change rate of instantaneous emission of NOx							
Train Set	116	610	631	704	145	652	1580	1540
Validation Set	824	752	763	749	938	756	1708	1679
Test Set	737	694	699	684	825	688	1647	1338

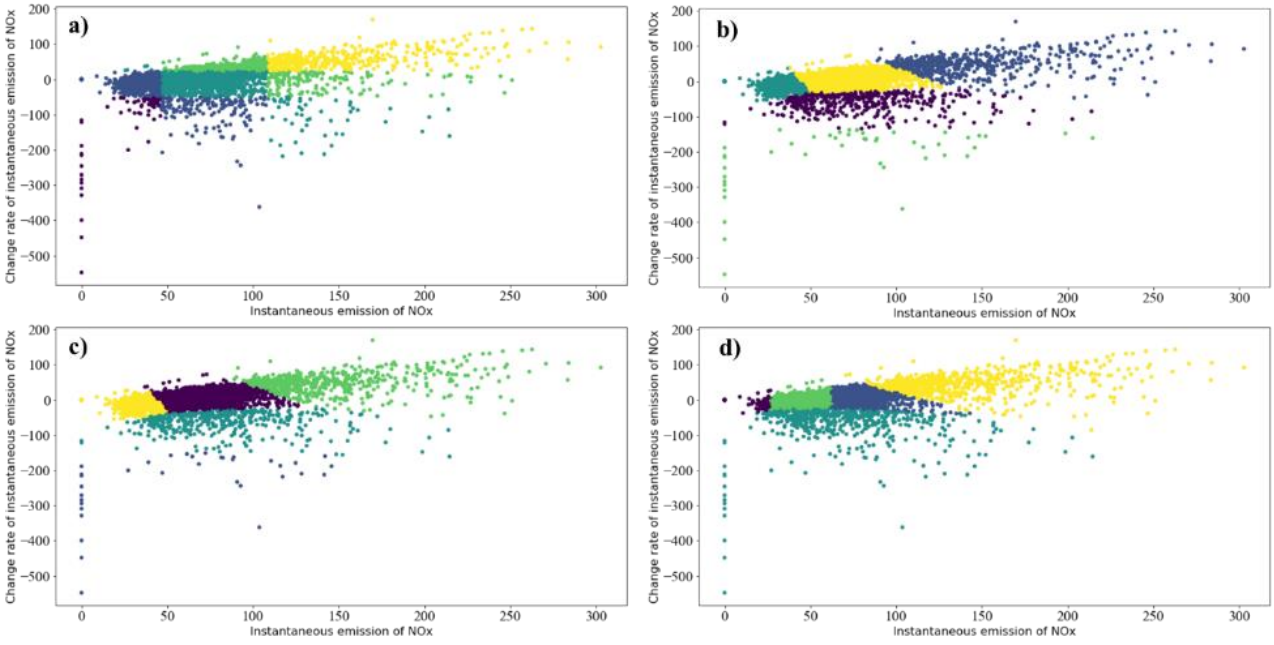
The same process was used for the selection of the prediction model for the change rate of instantaneous emissions of NOx, and all algorithms are compared in Table 2. Therefore, the GBDT model with precisely adjusted parameters is finally used to make the final prediction.

EMISSION STATE LABELS ACQUISITION

The instantaneous emissions of NOx and its change rate was clustered separately. The clustering effect of K-means, K-medoids and K-means++ algorithms was measured by DBI as shown in Table 4. By comparison, the K-Means algorithm was selected for clustering of both two emission variables. Labels are assigned to the instantaneous emission state of the vehicle by the judgment matrix proposed above.

Table4 Clustering effect of emission variables for each model

	K-Means	K-Means++	K-Medoids
Instantaneous emission of NOx	0.52089	0.52099	0.61728
Change rate of instantaneous emission of NOx	0.4957	0.54849	0.58091



a) Judgment matrix **b) K-Means** **c) K-Means++** **d) K-Medoids**

Fig5 Cluster results of emission state labels for each model

Taking instantaneous emissions of NOx and its change rate as variables, the clustering effect of the judgment matrix, K-Means algorithm, K-Means++ algorithm and K-Medoids algorithm on emission state is compared as shown in Fig5. It is clear that the K-Means algorithm, K-Means++ algorithm and K-Medoids algorithm give greater weight to the instantaneous emissions when clustering emission states. The above three algorithms can obtain higher DBI values because they take the data distribution characteristics as the only clustering logic, but the obtained emission states cannot be well explained and defined, and the obtained emission states cannot reflect the time series characteristics of vehicle emissions.

Relatively speaking, the method of obtaining vehicle emission state labels through the judgment matrix can provide the definition and interpretation of different emission states, and the definition process gives the same weight to the instantaneous emission and its change rate, which strengthens the consideration of emission change trend (as time series characteristics of emissions). The judgment matrix is more in line with the characteristics of vehicle emission formation mechanism and its change distribution. At the same time, the acquisition of emission state labels through the judgment matrix can clearly identify the high-emission state/low-emission state, and provide a good theoretical basis for the subsequent real-time identification of emission states and high-emission driving behavior analysis.

CONCLUSION

This paper proposed an emission states identification method based on judgment matrix. Based on PEMS high-precision emission data, the method comprehensively considered the time series characteristics of vehicle instantaneous emissions and its change rate and defined various emission states. By constructing the judgment matrix, effectively solved the problem that the common emission states identification method takes the numerical distribution characteristics as the only clustering condition, and the clustering results cannot be interpreted.

First, the datasets obtained by different acquisition devices are synchronized and the features are derived. Outliers are identified by timestamp distribution and Isolated Forest algorithm and handled by three-point moving average algorithm. Then, based on the decision tree model, the RF model, GBDT model, XGBoost model and fusion model are established. The five-fold cross-validation and grid search method were used to accurately adjust the parameters. The models with the highest prediction accuracy were selected to predict the instantaneous emission of NOX and its change rate. Finally, the adaptive clustering algorithm with better clustering effect is selected to cluster the two emission variables. Based on the emission formation mechanism and the logical relationship of emission variables, the emission states judgment matrix is established, and the emission state is labeled.

The feasibility of the method was evaluated on the dataset of diesel bus emissions and operating in Nanjing. Compared with other traditional clustering methods, The advantages of the method of emission prediction by operating data and emission states labeling by judgment matrix include: 1) effectively define different emission states; 2) quickly identify high emission states; 3) provide a good theoretical basis for subsequent real-time identification of emission states and analysis of high emission driving behavior.

ACKNOWLEDGEMENTS

The research in this paper was jointly supported by the National Key Research and Development Program of China (No. 2021YFE0112700), and the EU-funded project MODALES (grant agreement no. 815189).

REFERENCE

[1] Johnson K C, Durbin T D, Cocker D R, Iii, et al. On-road comparison of a portable emission measurement system with a mobile reference laboratory for a heavy-duty diesel vehicle[J]. Atmospheric Environment, 2009, 43(18): 2877-2883.
[2] Jaworski A, Madziel M, Lejda K. Creating an emission model based on portable emission measurement system for the purpose of a roundabout[J]. Environmental Science and Pollution Research, 2019, 26(21): 21641-21654.
[3] Bishop J D K, Stettler M E J, Molden N, et al. Engine maps of fuel use and emissions from transient driving cycles[J]. Applied Energy, 2016, 183: 202-217.
[4] Bishop J D K, Molden N, Boies A M. Using portable emissions measurement systems (PEMS) to derive more accurate estimates of fuel use and nitrogen oxides emissions from modern Euro 6 passenger cars under real-world driving conditions[J]. Applied Energy, 2019, 242: 942-973.
[5] Liu B, Hu J, Yan F, et al. A novel optimal support vector machine ensemble model for NOX emissions prediction of a diesel engine[J]. Measurement, 2016, 92: 183-192.
[6] Lv Y, Liu J, Yang T, et al. A novel least squares support vector machine ensemble model for NOx emission prediction of a coal-fired boiler[J]. Energy, 2013, 55: 319-329.
[7] Qin Y, Song D, Cheng H, et al. A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction[C]. 26th International Joint Conference on Artificial Intelligence (IJCAI), 2017: 2627-2633.
[8] Liu Y, Gong C, Yang L, et al. DSTP-RNN: A dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction[J]. Expert Systems with Applications, 2020, 143.
[9] Li Y, Zhu Z, Kong D, et al. EA-LSTM: Evolutionary attention-based LSTM for time series prediction[J]. Knowledge-Based Systems, 2019, 181.
[10] Xie H, Zhang Y, He Y, et al. Automatic and Fast Recognition of On-Road High-Emitting Vehicles Using an Optical Remote Sensing System[J]. Sensors, 2019, 19(16).
[11] Kang Y, Li Z, Lv W, et al. High-emitting vehicle identification by on-road emission remote sensing with scarce positive labels[J]. Atmospheric Environment, 2021, 244.
[12] Li Z, Kang Y, Lv W, et al. High-emitter identification model establishment using weighted extreme learning machine and active sampling[J]. Neurocomputing, 2021, 441: 79-91.