



A comprehensive OBD data analysis framework: Identification and factor analysis of high-emission heavy-duty vehicles[☆]

Zeping Cao^a, Kai Shi^b, Hao Qin^a, Zhou Xu^a, Xiaoyang Zhao^a, Jiawei Yin^a, Zhenyu Jia^a, Yanjie Zhang^c, Hailiang Liu^b, Qijun Zhang^{a,*}, Hongjun Mao^a

^a Tianjin Key Laboratory of Urban Transport Emission Research & State Environmental Protection Key Laboratory of Urban Ambient Air Particulate Matter Pollution Prevention and Control, College of Environmental Science and Engineering, Nankai University, Tianjin, 300071, China

^b Tianjin Ecological and Environmental Protection Comprehensive Administrative Law Enforcement Team, Tianjin, 300113, China

^c Tianjin Youmei Environment Technology, Ltd., Tianjin, 300380, China



ARTICLE INFO

Keywords:

On-board diagnostics
On-board emission monitoring
NOx emission
Heavy-duty diesel vehicle
Machine learning

ABSTRACT

On-Board Diagnostic (OBD) systems enable real-time monitoring of NOx emissions from heavy-duty diesel vehicles (HDDVs). However, few studies have focused on the root cause analysis of these emissions using OBD data. To address this gap, this study proposes an integrated analysis framework for HDDV NOx emissions that combines data processing, high-emission vehicle identification, and emission cause analysis. The framework employs a fuel-based window method to identify high-emission vehicles, while binning and machine learning techniques trace the causes of NOx emissions. A case study is conducted using data from 32 vehicles sourced from Tianjin On-Board Diagnostic Platform. Of these, five vehicles were identified as high emitters. A machine learning model was trained for each vehicle, with a detailed analysis conducted on three of them. The analysis involves a preliminary investigation of vehicle emissions status, followed by bin analysis to initially identify the causes of emissions. Finally, machine learning analysis is conducted, including the generation of individual conditional expectation (ICE) plots and multivariable partial dependence plots (PDPs), serving as a supplement to bin analysis when it cannot effectively pinpoint the causes of high emissions. This approach effectively uncovers the underlying factors within OBD big data. Using the analysis framework, we discover the identified causes of high NOx emissions were uneven heating of the Selective Catalytic Reduction (SCR) system and prolonged idling and high-power operation, catalyst degradation at 200–250 °C, and SCR system failure before 425 °C. The proposed framework offers a clear approach for identifying the causes of NOx emissions, aiding policymakers in implementing effective NOx control strategies for HDDVs.

1. Introduction

NOx is one of the primary pollutants emitted by vehicles, with significant implications for human health (Kowalska et al., 2020; Wei et al., 2019), ecosystems (Salmón et al., 2018; Gren et al., 2021; Xu et al., 2019), and the atmospheric environment (Bray et al., 2021; Akimoto and Tanimoto, 2022). It is an air pollutant that necessitates immediate regulatory control. According to a report by the United Nations Environment Programme (Programme and UNE, 2024), NOx emissions from mobile sources account for 40% of anthropogenic emissions. Therefore, controlling NOx emissions from mobile sources is essential to effectively reducing overall NOx emissions.

Stricter vehicle emission standards and the push towards vehicle electrification have led to a substantial reduction in NOx emissions from mobile sources (Mulholland et al., 2022; Ke et al., 2017; McCaffery et al., 2021). Nonetheless, heavy-duty vehicles have considerable weight and operate over long transport distances. Their predominant reliance on diesel fuel presents significant challenges to the widespread adoption of electrification (Davis et al., 2018). Although some studies have focused on replacing diesel with gasoline, JP-8, and other alternatives to reduce NOx emissions (Pandey et al., 2021; Pandey et al., 2020; Pandey et al., 2022; Pandey et al., 2023), diesel is expected to remain the primary fuel for heavy-duty vehicles in the foreseeable future. The “China Mobile Source Environmental Management Annual Report” shows that NOx

* This paper has been recommended for acceptance by Pavlos Kassomenos.

* Corresponding author.

E-mail address: zhangqijun@nankai.edu.cn (Q. Zhang).

emissions from diesel vehicles constitute 80% of the total NOx emissions from mobile sources in China (Ministry of Ecology and, 2023). Similarly, in California, USA, this proportion exceeds 70% (Badshah et al., 2019). To effectively reduce NOx emissions from heavy-duty diesel vehicles (HDDVs), it is essential to investigate their emissions. Some of current research relies on methods such as annual inspections (China MoEaEoPsRo, 2018), remote sensing (Yang et al., 2022; Grange et al., 2020), and smoke plume tracking (Tong et al., 2022). However, the limited duration of data collection for these methods results in reduced representativeness. Other research employs methods such as chassis dynamometer tests, engine dynamometer tests, and real-world driving tests. These approaches provide a more accurate representation of vehicle emissions over extended periods and varying conditions (Agarwal and Mustafi, 2021). Yu et al. examined NOx emissions from diesel vehicles under both certification and in-use testing conditions, revealing that in most instances, NOx emissions exceeded the certified limit of 0.2 g/bhp-hr (Jiang et al., 2022). Li et al. conducted on-road portable emissions measurement system (PEMS) tests to examine the combined impact of vehicle speed, Selective Catalytic Reduction (SCR) system status, and air-fuel ratio on NOx emissions under real-world conditions (Li et al., 2022). Tan et al. used bench tests to evaluate the degradation rate of V₂O₅ catalysts (Tan et al., 2019a). Vishal Kumbhar et al. conducted bench tests to evaluate the NOx reduction effects of using tamarind seed oil methyl ester (Kumbhar et al., 2023). However, these methods are costly and involve complex procedures (Ho et al., 2023; Wei et al., 2022a). They often examine the emissions of only a few vehicles over a short period, which makes it challenging to capture the full picture of real-world emissions.

The On-Board Diagnostic (OBD) system is an automated diagnostic tool integrated into vehicles to monitor and report engine performance and emission control systems. Vehicle operating status and emissions information transmitted via the CAN bus can be accessed through the OBD interface (Chen et al., 2024), enabling direct and continuous monitoring of vehicle emissions (Barbier et al., 2024). With the advancement of internet technology, the NOx online monitoring device OBD-III, which incorporates microcontroller units (MCUs) and wireless communication modules, has been tested in regions such as California and China (Pivetta et al., 2024). This presents an opportunity for large-scale remote monitoring of NOx emissions in real-world scenarios. In California, Tan et al. investigated the emission status of HDDVs and found instances of excess emissions and catalyst degradation, demonstrating the potential benefits of remote monitoring (Tan et al., 2019b). Under the China VI standards, newly sold vehicles are required to be equipped with OBD remote emission management systems that transmit data to local data centers. Through data analysis and real drive emission tests, Zhang et al. confirmed that the on-board emission monitoring (OBM) platform exhibits enough data quality and accuracy, showing a strong level of consistency with PEMS devices. Notably, data from China VI vehicles demonstrate particularly high quality (Zhang et al., 2020).

Building on the reliability of OBM data, current research is concentrated on policy recommendations (Zhao et al., 2024), inventory development (Deng et al., 2020; Lv et al., 2023), and the assessment of emission levels (Wang et al., 2022; Yang et al., 2024; Zhang et al., 2023). Zhao et al. conducted a detailed analysis of OBM data from in-use HDDVs in Beijing and identified a significant proportion of vehicles at risk of exceeding emission standards (Pavlovic et al., 2021). They recommended establishing a long-term observation protocol to enable extended monitoring of in-use vehicles. Lv et al. developed a street-level NOx emission inventory for Tangshan, which has significantly advanced the feasibility of real-time monitoring of NOx emissions from HDDVs (Wei et al., 2022b). Zhang et al. established a method similar to the PEMS evaluation approach, known as the fuel-consumption based window method (FWM), enabling precise identification of high-emission vehicles (Jia et al., 2024a). Current research has thoroughly demonstrated the real-world benefits of using large-scale OBM data. This has established a research chain that includes real-time

monitoring and high-emission vehicle identification. However, effectively identifying the causes of high emissions in vehicles and subsequently implementing targeted control measures remains an unresolved challenge. Moreover, the existing related research is intricate and lacks systematic organization and synthesis. Therefore, a standardized data analysis framework is needed to achieve a comprehensive process that includes data cleaning, high-emission vehicle identification, and influencing factor identification. This allows for an in-depth exploration of the key factors affecting NOx emissions from individual vehicles, thereby enabling scientific and precise control of high-emission vehicles.

While the accuracy of OBD data is relatively low, its large volume compensates for this limitation (Jia et al., 2024b). Large-scale historical data accurately reflects the emissions performance of vehicles under real-world operating conditions. Calculating NOx emission factors (EFs) and SCR conversion efficiency through binning can provide a reasonable indication of a vehicle's emission performance. However, strict binning may result in insufficient data within some bins, leading to considerable uncertainty. Furthermore, such methods struggle to account for the influence of other variables, which may lead to imprecise conclusions. They are also inadequate for addressing complex issues such as the interactions between multiple variables and NOx emissions. Machine learning is a powerful data mining technique capable of handling large-scale datasets and has been widely applied in the environmental field in recent years (Moradi et al., 2020). Data-driven models can effectively uncover the underlying patterns in historical data, and regularization techniques can help the model adaptively reduce the impact of inaccurate measurements (Moradi et al., 2020). Model interpretation methods, such as partial dependence plots (PDP) and individual conditional expectation (ICE) plots, can effectively visualize the captured patterns, illustrating the impact of both single and multiple variables on NOx emissions (Wei et al., 2022c). Therefore, we propose a comprehensive data analysis framework, which includes data cleaning, high-emission vehicle identification, and NOx emission analysis. Impact analysis includes both binning analysis and machine learning analysis, with different methods applied depending on the specific circumstances. Based on the framework, this paper conducts a case study using data from 32 heavy-duty diesel trucks downloaded from Tianjin Ecology and Environment Bureau On-Board Diagnostic Platform. The study provides a detailed analysis of the emission status of each vehicle and investigates the factors contributing to high NOx emissions. The proposed framework can be applied to large-scale analysis of high-emission vehicle data and will offer valuable insights for the future remote and refined management of diesel vehicles.

2. Material and method

2.1. Analysis framework

The proposed NOx analysis framework comprises the following steps. Firstly, collect a substantial amount of vehicle OBM data and conduct data cleaning process. Perform timestamp processing, boundary exceedance handling, and constant value processing. Secondly, assess vehicle emission status. Compute vehicle fuel-based EFs and SCR conversion efficiency, followed by the identification of high-emission vehicles. This study employs the fuel window method to identify high-emission vehicles. This method is convenient and enables efficient identification of high-emission vehicles from the entire fleet. Thirdly, NOx analysis, consists of two parts: bin analysis and machine learning analysis. The first step involves binning analysis, where data is categorized into different bins based on SCR inlet temperature and engine net output torque. Plotting the driving time proportion and NOx emissions in each bin provides a comprehensive view of vehicle driving behavior. Comparing these plots can also reveal the causes of high NOx emissions. Secondly, machine learning analysis involves training an XGBOOST machine learning model for the vehicles, with model parameters tuned using Bayesian optimization. Finally, the trained model is interpreted

using ICE and PDP plots. The method can precisely diagnose the factors influencing NOx emissions and serves as a more robust supplementary method when binning analysis is insufficient. Fig. 1 illustrates the detailed workflow of this analysis framework.

2.2. Data acquisition and cleaning

This study utilizes data from 32 HDDVs downloaded from Tianjin Ecology and Environment Bureau On-Board Diagnostic Platform, all compliant with China VI standards and equipped with OEM-performed OBD systems. The vehicles have identical engine and model specifications. Data from 9 vehicles cover two months, while the remaining 23 vehicles are from one month. Detailed information on these 32 vehicles is provided in Table S1. The specific OBD data items are shown in Table 1. The raw data downloaded from the platform is of low quality, necessitating thorough data processing and cleaning. The data-cleaning process used in this paper is similar to that of Zhang et al. but with some modifications (Zhang et al., 2020). This process includes time series errors, exceeding boundaries errors, and constant value errors.

Time series errors refer to both time series discontinuity errors and time duplication errors. For time series discontinuity errors, the validity of a travel segment and the surrounding data are assessed. If

discontinuous data within a single trip account for more than 20%, that trip is considered invalid. There are two types of time duplication errors. There are two types of time duplication errors. The first type is when identical data appears multiple times, which can be effectively addressed by simply removing duplicate entries. The second type is when identical timestamps are associated with different measurements. For this situation, if a time duplication error is followed by a time series discontinuity error with the same timing as the duplication, the repeated entries are filled into the discontinuity; otherwise, the duplicate values are averaged and resampled at the timestamp.

Handling exceeding boundaries errors is relatively straightforward: we simply remove values that fall outside the instrument's observational range. Our approach is to discard an entire row of data if any parameter within that row is out of range. Notably, this step also removes the fixed feedback value of 65.535 km/h for speed. The specific out-of-bounds ranges are detailed in Table S2.

After performing the two aforementioned steps, the data still contains constant value error. Occasionally, an entire row of data may be repeated dozens of times, which is unlikely given the sensitivity of the sensors. We set a threshold of 15 rows for such repetitions; if the number of duplicate rows exceeds this threshold, we remove the repeated values; otherwise, we retain them. This threshold balances data integrity with

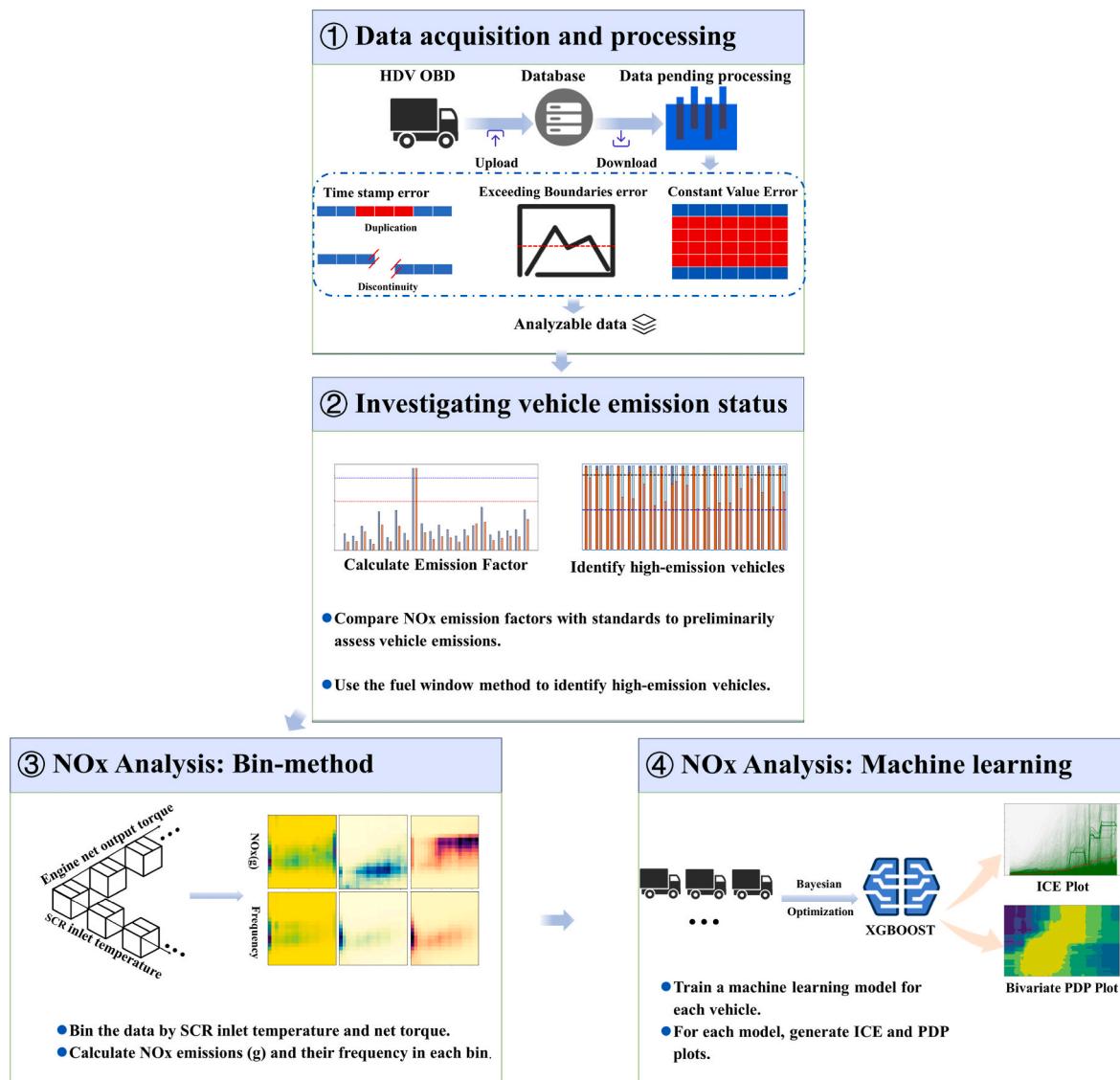


Fig. 1. OBD-based NOx analysis framework for HDDVs.

Table 1
Variables and information recorded by OBD devices.

Variant name	Unit	Increment	Lower Limit	Upper Limit
Data collection time	seconds	1		
Atmospheric pressure	kPa	0.5	0	125
Longitude	Degree	10–5	0	180
Latitude	Degree	10–5	0	90
Vehicle speed	km/h	1/256	0	250.996
Relative engine torque	%	1	–125	125
Relative friction torque	%	1	–125	125
Engine speed	rpm	0.125	0	8031.875
Fuel rate	L/h	0.05	0	3212.75
Intake flow rate	kg/h	0.05	0	3212.75
Fuel tank level	%	0.4	0	100
Engine coolant temperature	°C	1	–40	210
Cumulative range	km	0.1		
SCR upstream NOx concentration	ppm	0.05	–200	3012.75
SCR upstream temperature	°C	1/32	–273	1734.96875
SCR downstream NOx concentration	PPM	0.5	–200	3012.75
SCR downstream temperature	°C	1/32	–273	1734.96875
Urea reagent level	%	0.4	0	100
DPF pressure difference	kPa	0.1	0	6425.5

quality assurance. Examples of time series errors, exceeding boundaries errors, and constant value errors are illustrated in Fig. S1. Fig. S2 shows the error proportions of different columns.

2.3. Vehicle emission status investigation

This study calculates EFs using two methods: NO_xdistance-based, which divides the total NO_x mass by the distance traveled, and NO_xfuel-based, which divides it by the fuel consumed. The instantaneous value of NO_x emissions at time t, denoted as NO_{x,mass,t} (g/s), is calculated using the following formula:

$$NO_{x,mass,t} = \frac{0.001 \times 46}{3600 \times 22.4} \times \frac{1}{\rho_e} \times NO_{x,down-stream,t} \\ \times (FuelFlow_t \times \rho_{fuel} + MAF)$$

where ρ_e is the density of the exhaust gas, which is taken as the density of air, i.e., 1.29 kg/m³. NO_{x,down-stream,t} is the instantaneous NO_x concentration at time t downstream, ppm. FuelFlow_t is the fuel flow rate at time t, g/h. ρ_{fuel} is the density of the fuel, taken as 0.85 kg/L, and MAF is the instantaneous air flow mass at time t, kg/h.

In a segment with a total duration of T, the EFs based on fuel and distance are NO_xfuel-based and NO_xdistance-based, measured in g/kg-fuel and g/km, respectively. The calculation formulas are as follows:

$$NO_{x,fuel-based} = \frac{3600 \times \sum_{t=1}^T NO_{x,mass,t}}{\rho_{fuel} \times \sum_{t=1}^T FuelFlow_t}$$

$$NO_{x,distance-based} = \frac{3600 \times \sum_{t=1}^T NO_{x,mass,t}}{\sum_{t=1}^T v_t}$$

This study employs the FWM proposed by Zhang et al. for high-emission vehicle identification (Zhang et al., 2023). In this method, fuel consumption during a C-WTVC cycle is used instead of power-based window segmentation, allowing for window division based on fuel consumption. Additionally, by linking fuel consumption with power, this method facilitates the conversion between fuel consumption-based and power-based EFs, allowing for comparison with PEMS limits. A

vehicle is classified as high-emission if less than 90% of the windows meet the standard. The Text S1 provides detailed process used in this study.

2.4. Analysis of factors contributing to high-emission vehicles

2.4.1. Bin-method analysis

The OBD measurement values from a single pass are not precise enough, and a certain amount of data is needed to balance this deviation. Distributing the data into different bins based on various variables is an effective method for revealing their impact on NO_x emissions.

We applied two methods to divide total dataset into different bins. First, we segmented the data into different bins based on the inlet temperature and calculated the NO_x conversion efficiency η_{SCR} (%) to each bin. In a segment with a total duration of T, the formula for calculating the η_{SCR} (%) is as follows:

$$\eta_{SCR} = \frac{1}{T} \left(\frac{\sum_{t=1}^T NO_{x,up-stream,t} - \sum_{t=1}^T NO_{x,down-stream,t}}{\sum_{t=1}^T NO_{x,up-stream,t}} \right) \times 100\%$$

where NO_{x,up-stream,t} is the instantaneous NO_x concentration upstream at time t, ppm.

Second, SCR inlet temperature and net output torque were selected as the dividing variables. The horizontal and vertical axes were evenly divided into n parts from the minimum to the maximum values, generating an n × n grid (where the specific value of n is determined by the dataset size. For example, with a dataset size of 1,000,000, n was set to 20). The data was evenly distributed into the grid, and the proportion of the total data in each grid (Frequency) and the total NO_x emission mass in each grid (NO_{x,mass} (g)) were calculated.

The formula for Frequency in a given bin:

$$Frequency = \frac{f}{N}$$

where f is the number of data points in the bin, and N is the total number of data points in the dataset.

The formula for calculating the total NO_x mass (NO_{x,mass}, g) in a bin of total duration is:

$$NO_{x,mass} = \sum_{t=1}^T NO_{x,mass,t}$$

2.4.2. Machine learning analysis

To balance the accuracy and interpretability of the machine learning model, this paper selects vehicle speed (Speed), engine Speed (RPM), engine net output torque (Torque), SCR inlet temperature (Inlet temp), and SCR outlet temperature (Output temp) as the model's input features. The target value of the model is the NO_{x,mass,t} calculated above. Speed, RPM, and Torque can effectively describe the operating state of the vehicle, while the Inlet temp and Outlet temp describe the overall operational status of the SCR system. NO_{x,mass,t} reflects the transient NO_x emissions of the vehicle.

We choose the XGBoost model and use Bayesian optimization to tune the hyperparameters (Chen and Guestrin, 2016; Frazier, 2018). XGBoost supports GPU acceleration and the histogram algorithm, which can significantly reduce the time required for parameter tuning and model building. Bayesian optimization is a hyperparameter tuning algorithm, and unlike grid search, it does not simply evaluate every parameter combination in the parameter space. Instead, it first constructs an objective function and observes the value of the objective function in the parameter space, then uses a probabilistic model to adjust the parameters. Finally, through an iterative optimization process, it selects the optimal parameters. The specific process for building the model is as follows:

First, the data is randomly split into a training set and a test set in a

7:3 ratio. The test set data is never involved in training to prevent data leakage. In the second step, the objective function for Bayesian optimization and the parameter space of the model are determined. Five-fold cross-validation is performed on the training set, and the mean of the negative root mean square error from the five cross-validations is used as the objective function. The selected parameter space is shown in Table S3, with the initial observation points set to 30 and the number of iterations set to 120. In the third step, Bayesian optimization is executed to obtain the optimal parameters. This set of optimal parameters is then used to train the model on the training set, and the corresponding MSE and R^2 are calculated on the test set. Table S4 shows the formulas for calculating the statistical parameters in this process.

This paper chooses ICE plots to explain the model's behavior for single variables and uses PDP to show the interaction of multiple variables on the target value. Since Sklearn's native plotting interface does not support XGBoost, this paper uses a custom Python script to generate ICE plots and multivariate PDP plots. The text S2 provides detailed descriptions of the corresponding formulas and procedures. Additionally, due to the large size of the original dataset, the full dataset was not used. Instead, 1% of the data was randomly sampled for plotting.

3. Result and discussion

3.1. Data cleaning results and vehicle emission status investigation

First, the quality of OBD data from 32 vehicles was examined. Fig. S2 illustrate the proportions of time series errors, exceeding boundaries errors, constant value errors, and analyzable data within the total dataset for each vehicle. It can be observed that vehicles #6, #8, #10, #13, #27, and #28 contain no analyzable data, which might be due to sensor malfunctions. Among the remaining vehicles, except for #14, all have over 60% analyzable data. For OEM-performed China VI heavy-duty trucks, sensor malfunctions in certain vehicles may render part of the OBD data unanalyzable, while errors caused by data transmission issues, positioning errors, and sensor misalignment are generally acceptable, preserving a sufficient amount of analyzable data. Fig. 2 presents the fuel-based and distance-based EFs for the remaining 26 vehicles, with all vehicles renumbered for clarity in the plot. For China VI vehicles, the diesel engine limit and the in-use PEMS limit are set at 1.94 g/kg-fuel and 2.9 g/kg-fuel, respectively (Zhao et al., 2024). Vehicle #9 slightly exceeds the standards (1.13 times for PEMS limit), while vehicles #23–26 show NOx emissions significantly higher than both limits, ranging from 9.2 to 20.3 times the diesel engine limit and 6.1 to 13.6 times the in-use PEMS limit. The FWM was employed to further verify high-emission vehicles. For the vehicle selected in this study, the vehicle weight $M = 25000\text{kg}$, maximum output power $P_{\max} = 346\text{kw}$, certified fuel consumption per 100 km $FC_{c-WTVC} = 39.8\text{L}/100\text{km}$, and the transmission efficiency $\eta_{\text{Trans}} = 0.85$. Based on these parameters, the cycle fuel consumption FC_{WTVC} is calculated to be

6.875 L. The conversion factor Coeff for the EF from fuel-based (mg/kg-fuel) to power-based (mg/kWh) is 5.343, with the max fuel consumption rate (FCR_{\max}) at 64.70 L/h. Using the FC_{WTVC} , the data is divided into multiple windows. For each window, the valid power limit is set at 20% of the FCR_{\max} , which is 12.93 L/h, and the emission limit is set at 690 mg/kWh. If the valid window ratio is less than 50%, the valid power limit is gradually reduced by 1% of the FCR_{\max} until the valid window ratio exceeds 50%. However, if the valid power limit is reduced to 10% of the FCR_{\max} and the ratio still does not exceed 50%, the data is deemed invalid for verification using this method. Fig. 3 illustrates the valid window ratio, compliant emission window ratio, and the final valid fuel threshold used for each vehicle.

In this study, there is a high degree of consistency between vehicles that exceed the two thresholds in the total dataset and those certified as high-emission vehicles using the FWM. Vehicles #9, #23 to #26 not only have fuel-based total EFs exceeding the limits but are also certified as high-emission vehicles. This suggests a correlation between the EFs at the monthly level and the ratio of compliant windows. Fig. S3 plots the EFs of the vehicles on the x-axis and the ratio of compliant windows on the y-axis. The figure shows that as the EFs increase, the proportion of compliant windows decreases significantly. In other words, a simple calculation of the vehicle's overall EFs can provide a significant indication of the outcomes derived from the FWM. Therefore, in the future, more vehicle data could be used to establish this correlation, enabling preliminary and rapid screening of high-emission vehicles, thereby reducing the computational resources required by monitoring platforms.

To assess the validity of fuel-based and distance-based EFs, this study applied the MOVES binning method (specific binning rules can be found in Table S5) to segment each vehicle's data into 23 bins for vehicles that did not exceed the emission limits. The fuel-based and distance-based EFs for each bin were calculated and presented in boxplots, as shown in Fig. S4. To compare the data's dispersion, the coefficient of variation (CV) for each bin was also calculated, with the results provided in Table S6. The figure and CV results reveal that at very low speeds (<1.6 km/h), the distance-based EFs exhibit significant dispersion (bin 1), making it difficult to objectively reflect the vehicle's emissions level. At moderate speeds (1.6–80 km/h), under low VSP conditions, distance-based EFs are more dispersed compared to fuel-based EFs (bins 11, 12, 21, 22). Under higher VSP conditions, the dispersion levels of fuel-based and distance-based EFs are relatively close (bins 13, 14, 15, 16, 23, 24, 25, 27, 28). At very high speeds (>80 km/h), the dispersion of both factors is similar. It is noteworthy that both fuel-based and distance-based EFs exhibit high dispersion in bins 15, 16, 29, and 30, likely due to the limited amount of data within these bins, increasing the uncertainty of the overall dataset. Overall, fuel-based EFs are more stable than mileage-based EFs, less susceptible to the influence of low speeds, and easier to convert into relevant standards, making them more suitable for assessing the emission performance of diesel vehicles.

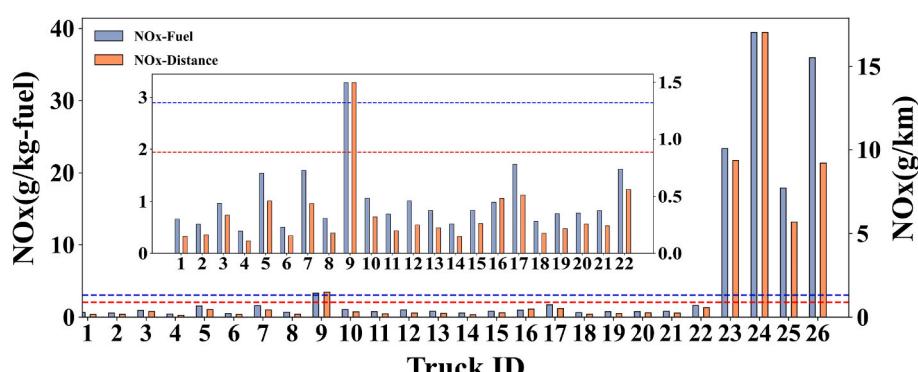


Fig. 2. NOx emission levels for different vehicles. Blue line represents the PEMS limit (2.9 g/kg-fuel) and red line represents the diesel engine limit (1.94 g/kg-fuel).

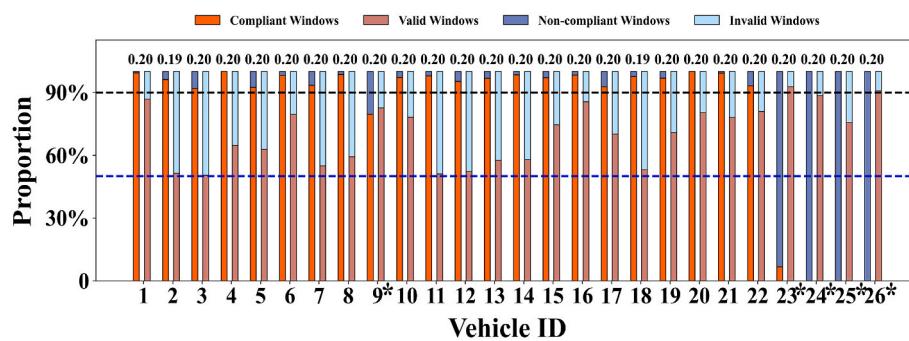


Fig. 3. The proportion of valid windows and compliant windows. An asterisk (*) indicates that the vehicle is certified as a high-emission vehicle. The blue line represents the 50% valid window limit. The black line represents the 90% compliant window limit. The black numbers above the bars refer to the final valid fuel threshold, expressed as a ratio of the maximum fuel consumption.

3.2. NOx analysis: bin-method

To evaluate the efficiency of the SCR system at different temperatures for different vehicles, this study calculated the NOx conversion efficiency η_{SCR} for all vehicles within each 5 °C interval between SCR inlet temperatures of 150 °C and 580 °C. The corresponding 95% confidence intervals were also calculated. For high-emission vehicles (#9, #23–#26), the calculation was refined to use 25 °C intervals to ensure accuracy, with η_{SCR} computed between each two points. Fig. 4 illustrates the results of these calculations. From the figure, it can be observed that at low temperatures (below approximately 250 °C) and high temperatures (above approximately 450 °C), the confidence intervals for η_{SCR} are relatively large. Moreover, as the temperature approaches the mid-range (250°C–450 °C), the confidence intervals become smaller. Additionally, at low temperatures, η_{SCR} gradually increases with rising temperatures, reaching 95% around 230 °C. At mid-range temperatures, the η_{SCR} stabilizes at over 95%. However, at high temperatures, the η_{SCR} decreases with rising temperatures, dropping below 95% at around 460 °C.

The SCR systems of high-emission vehicles have shown varying degrees of catalyst deterioration. For vehicle #9, its η_{SCR} in the 200–250 °C and 400–450 °C ranges is slightly below the average conversion efficiency. For vehicles like #24 and #26, their SCR systems are almost non-functional, with R^2 values calculated from NOx emissions upstream and downstream of the SCR device at 0.80 and 0.79, respectively. Considering the instrument drift, which could decrease the R^2 , the NOx emitted by these vehicles is virtually untreated before being released into the atmosphere. Correspondingly, their fuel-based NOx EFs are the highest, at 39.4 g/kg-fuel and 35.9 g/kg-fuel, respectively. The SCR systems of

vehicles #23 and #25 also show significant deterioration. For vehicle #23, the SCR system is almost ineffective before 375 °C, and for vehicle #25, it is ineffective before 300 °C. Once these temperature thresholds are surpassed, their η_{SCR} increases significantly as the temperature rises, resulting in a marked reduction in NOx emissions, with EFs dropping to 23.3 g/kg-fuel and 17.9 g/kg-fuel, respectively.

The mainstream SCR catalysts in the Chinese market are primarily iron-based, copper-based, and vanadium-based. Copper-based catalysts perform optimally at around 300 °C, achieving a conversion efficiency of approximately 90% when the inlet temperature reaches 400 °C. Iron-based catalysts, in contrast, exhibit peak efficiency at temperatures above 400 °C. Vanadium-based catalysts, which likely correspond to those used in this study, operate most effectively within a temperature range of 300–450 °C, consistent with the trends observed in Fig. 4. However, unlike copper-based and iron-based catalysts, vanadium-based SCR catalysts have relatively low thermal stability. These catalysts are particularly vulnerable to deactivation when exposed to the high temperatures required for active regeneration of diesel particulate filters (DPFs), leading to significant reductions in conversion efficiency. Other forms of degradation, such as sulfur poisoning, hydrocarbon poisoning, Pt-Pd poisoning, and contamination by alkali metals like potassium (K) and sodium (Na), generally have a less pronounced impact on SCR performance (Guan et al., 2014). To address these challenges, we recommend that vehicle maintenance and repair facilities pay close attention to the thermal stability issues associated with vanadium-based SCR systems. Proper maintenance protocols and careful monitoring of regeneration cycles are crucial to minimizing catalyst degradation and ensuring sustained emission control performance.

The deterioration of the SCR system can lead to NOx emissions several times higher, or even tenfold the normal level. For vehicles #23 to #26, this SCR failure is unquestionably the main cause of their excessive emissions. However, for vehicle #9, the current information makes it difficult to determine the main factor contributing to its emissions, as other elements such as driver behavior and operating conditions could also have an impact. To further investigate the factors affecting NOx emissions, this study selected three vehicles: #5 (a normal vehicle), #9 (with uncertain causes), and #23 (with SCR failure). For each vehicle, 1,000,000 random data rows were sampled to ensure comparable results. Two key parameters were selected: net output torque (0–87%) and SCR inlet temperature (150–550 °C). The horizontal and vertical axes were evenly divided into 20 parts, creating a 20 × 20 grid. Data were evenly distributed into the grid, and for each bin, the proportion of data points relative to the total dataset (Frequency) and the total NOx mass (NOx (g)) was calculated. The results are shown in Fig. 5. The analysis of the three vehicles reveals similar frequency distributions, characterized by a substantial proportion of low output power (idle) states. Instances of extreme undercooling (below 210 °C) and overheating (above 410 °C) are relatively rare. For vehicle #5, elevated NOx emissions are observed during idle states and at high

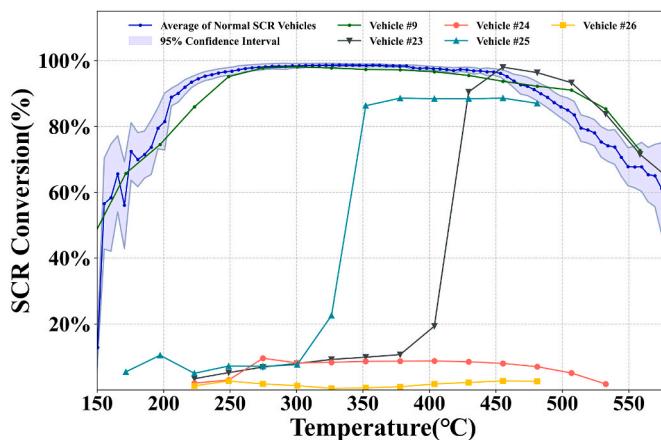


Fig. 4. The average SCR conversion efficiency for different vehicles. The blue line and shaded area represent the average SCR conversion efficiency and the 95% confidence intervals for all normal vehicles. The other lines indicate the average SCR conversion efficiency in 25 °C intervals for high-emission vehicles.

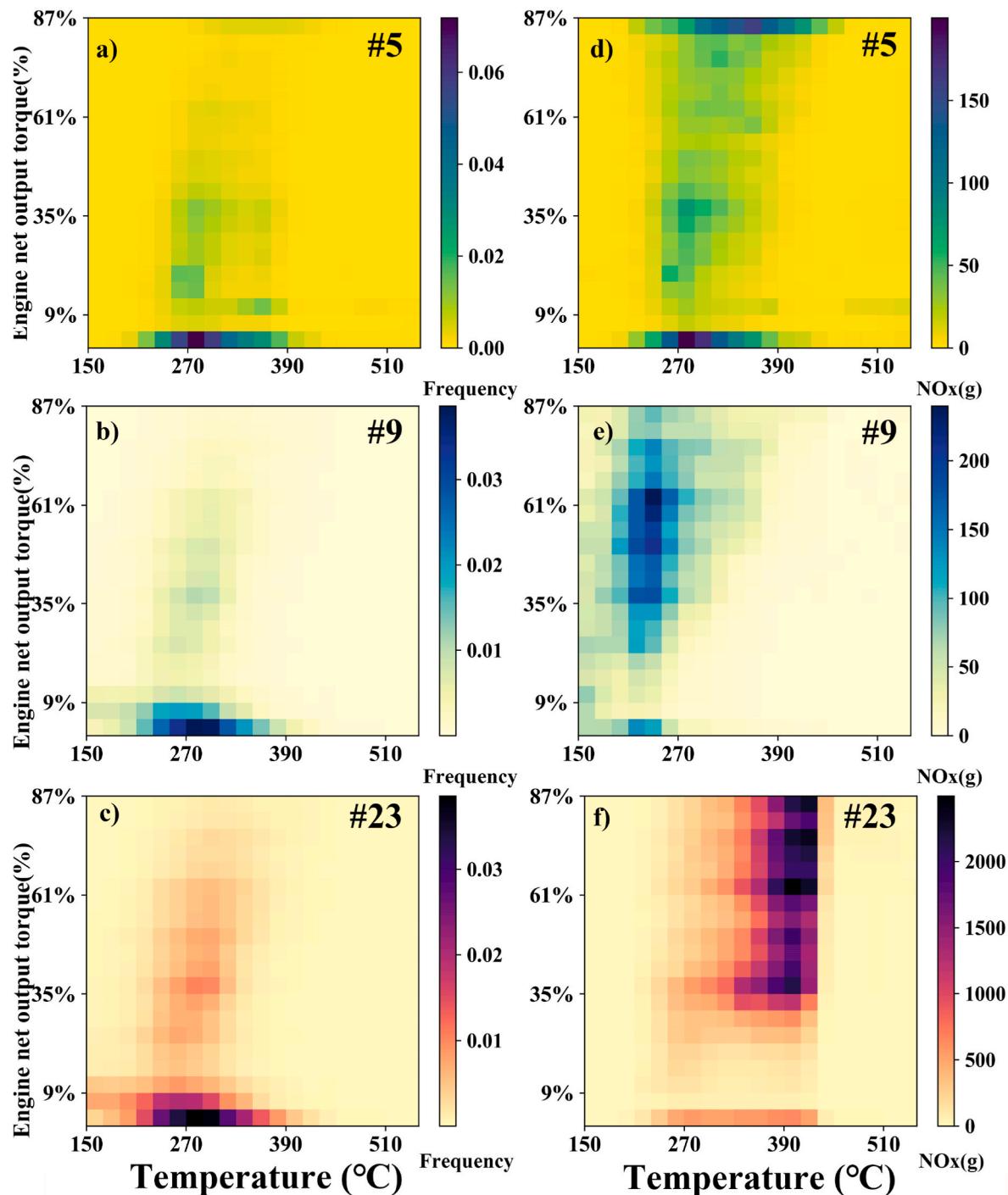


Fig. 5. The bin analysis results for vehicles #5, #9, and #23. The left column (a,b,c) shows the frequency of each bin, while the right column (d,e,f) presents the NOx emissions in each bin. From top to bottom are the results for #5, #9, and #23, respectively.

output powers (83–87%) when the SCR inlet temperature ranges from 230 °C to 450 °C, accounting for a significant portion of total emissions. For vehicle #9, higher emissions are noted between 210 °C and 250 °C, aligning with previous findings. Notably, emissions between 400 °C and 450 °C are not elevated, which can be attributed to the lower frequency of occurrences in this temperature range. Vehicle #23 exhibits consistently high emissions up to 430 °C, beyond which there is a marked decrease. However, even after this reduction, emissions remain significantly higher than those of normal vehicles.

In summary, for vehicle #5, low-power operation and high-power operation are the primary causes of real-world emissions. The former

corresponds to idling, while the latter is associated with prolonged, long-distance driving. Both states are common in real-world operations. On one hand, idling is related to local traffic conditions and operational activities such as loading and unloading. On the other hand, long-distance transport is one of the main functions of heavy-duty vehicles. Engine manufacturers should carefully consider both scenarios to further reduce NOx emissions. For vehicle #9, the elevated emissions are due to a decrease in SCR conversion efficiency in the 210°C–250 °C temperature range. In contrast, for vehicle #23, it is the near failure of the SCR system below 430 °C. In both vehicles' degraded temperature ranges, total NOx emissions increase as net Torque rises. This effect is so

pronounced that it somewhat outweighs the influence of frequency. Therefore, the bin analysis takes into account both operational conditions and the degradation of the SCR system. However, to ensure that each bin contains sufficient data for analysis and to eliminate the interference of frequency, a method capable of enabling continuous analysis is needed.

3.3. NO_x analysis: machine learning

For all vehicles, machine learning models were trained using vehicle Speed, RPM, Torque, Inlet Temp, and Outlet Temp as features, with NO_{x, mass,t} (g/s) as the target. Fig. S5 displays the relationship between fuel-based EFs (x-axis) and the R² values (y-axis) on the test sets of the models. The results suggest a positive correlation between fuel-based NO_x EFs and R². The higher the EF of the vehicle, the better the machine learning model can capture the relationship between vehicle operating conditions, SCR system status, and NO_x emissions. Given that the datasets are sufficiently large and cover most of the vehicles' operating conditions, the differences in test set R² values are more likely attributable to the complexity of modeling each dataset. We propose that vehicles with lower NO_x emissions likely have more active catalysts in their SCR systems, causing the SCR inlet and outlet temperatures to be less indicative of the system's true state. As NO_x emissions increase, catalyst activity deteriorates, simplifying the system and resulting in higher R² values in the models.

This study selected three vehicles (#9, #10, and #23) for machine learning analysis. Fig. 6 presents the Pearson correlation coefficients for the training data, the model's R² values and the MSE on the test set. In each of the three datasets, the correlations between RPM and Speed, as well as between the SCR Inlet and Outlet Temp, are consistently high. The correlation between Torque and NO_x is generally high; however, vehicle #23 exhibits an exceptionally high correlation. This observation suggests significant deterioration of the SCR system. The R² values for the three datasets progressively increase (0.58, 0.66, 0.85), while the MSE also rises (0.011, 0.032, 0.077), reflecting the presence of more high-emission data. All three models performed well on the test sets, indicating that the models successfully captured the key patterns in NO_x emissions. Fig. S6 shows the feature importance based on feature gain, feature cover, and feature weight. In the first model, the most important

feature is Inlet Temp, while for the other two models, Torque becomes the most important feature.

Fig. S7 presents the ICE plots for Inlet temp, Outlet temp, and Torque in relation to transient NO_x emissions. For vehicle #5, as Inlet temp increases, predicted NO_x emissions first decrease and then rise. With increasing Outlet temp, the predicted values consistently rise. In the case of vehicle #9, predictions initially decrease and then increase with rising Inlet temp, while they consistently decrease as Outlet temp rises. For vehicle #23, predicted values remain stable up to 420 °C, after which they drop sharply, and Outlet temp appears unrelated to the predicted emissions. For all three vehicles, the predicted values increase with Torque, with the degree of increase progressively larger. These results indicate that Inlet temp and Outlet temp have different effects on NO_x emissions. The SCR system operates as a whole, with Inlet temp and Outlet temp each reflecting part of the system's condition. Thus, using only Inlet temp to describe the SCR system's state has limitations.

To further explore the combined effect of Inlet temp and Outlet temp on NO_x emissions, a bivariate PDP plot was generated to show the relationship between these temps and predicted NO_x, as seen in Fig. 7. For #5, discrepancies between Inlet and Outlet temps lead to higher NO_x predictions, particularly when the Inlet temp is low and the Outlet temp is high. When the temps are closer, the predicted NO_x values are minimal. For #9, anomalies appear between 200 and 250 °C, and high NO_x predictions are seen when the Inlet temp ranges from 370 to 450 °C and the Outlet temp from 260 to 280 °C, indicating possible SCR degradation. For #23, low predictions only occur around 425 °C Inlet temp and 350 °C Outlet temp, with high predictions elsewhere. Fig. S8 shows PDP plots for Inlet temp-Torque and Outlet temp-Torque, highlighting that the increase in predicted NO_x with Torque varies across temp ranges, further pinpointing areas of low SCR efficiency, consistent with the previous PDP results.

From the above results, it can be observed that the Inlet Temp and Outlet Temp contain different information and can not simply be replaced by one another. There appears to be a certain lag effect between the Inlet Temp and Outlet Temp. Appendix Figure S9 illustrates this lag effect. This study attempted to align the two columns data by minimizing the R² between the Outlet Temp and Inlet Temp, revealing that the lag durations for samples #5, #9, and #23 are 53 s, 67 s, and 57 s, respectively. This indicates that significant differences in the Inlet Temp

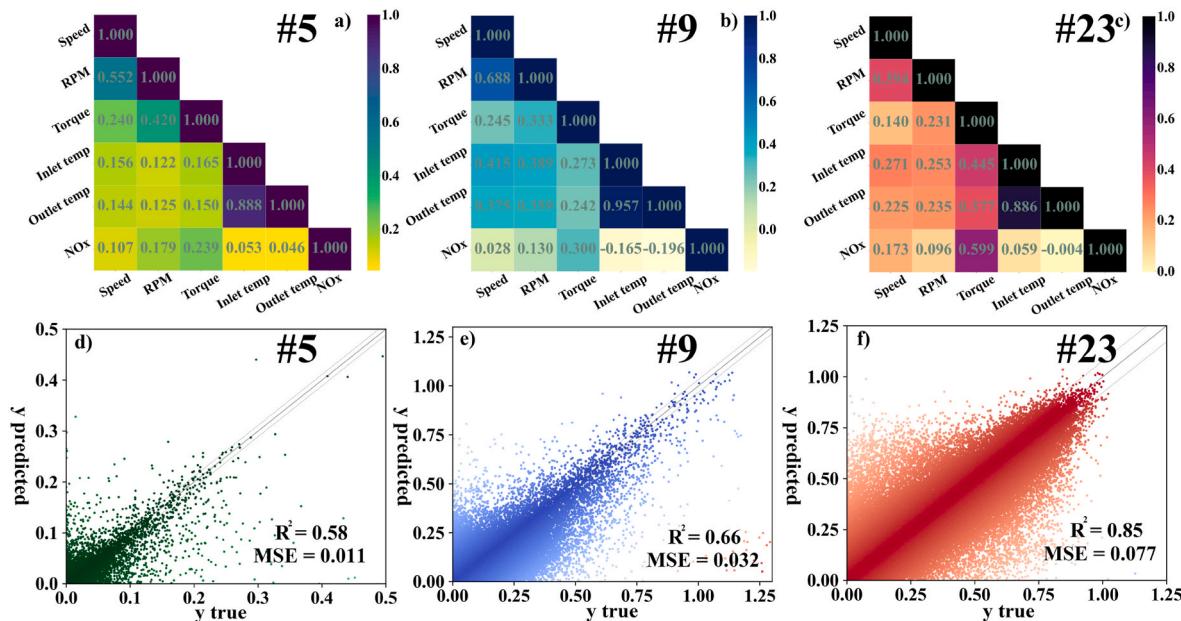


Fig. 6. Variable correlations and corresponding machine learning model results for vehicles #5, #9, and #23. Subfigures a), b), and c) show the correlations between feature values and target values used for machine learning training for vehicles #5, #9, and #23. Subfigures d), e), and f) present scatter plots of the predicted values versus the actual values from the machine learning models.

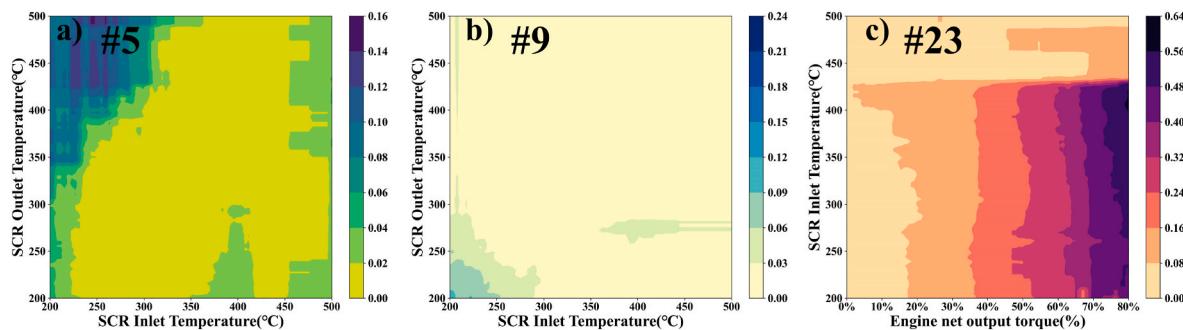


Fig. 7. Bivariate partial dependence plot of NOx predictions and SCR Outlet Temperature-Inlet Temperature for #5, #9, and #23.

and Outlet Temp not only exist in the real world but occur frequently. Although the temperature difference between the inlet and outlet typically does not reach levels that would cause excessive emissions, it may still contribute to significant emissions. In the future, engine designers may need to consider the uniformity of heating in SCR systems to further reduce vehicle emissions.

4. Conclusion

This study integrates previous research and establishes a data analysis framework that combines data processing, high-emission vehicle identification, and NOx emission analysis. Through bin-method analysis and machine learning analysis, this framework effectively reveals the causes of emissions for both normal and high-emission vehicles. The main findings are as follows:

- (1) The quality of OBD data from China VI vehicles on the data platform is generally acceptable, but malfunctioning OBD devices must be addressed. Among 32 vehicles, 6 had no analyzable data, but for the remaining 26, almost all had over 60% of analyzable data.
- (2) High-emission vehicles identified using the FWM closely correspond to those with fuel-based EFs exceeding the limits. This suggests that total EFs can be used for preliminary high-emission vehicle screening. Of the 26 vehicles, 5 were identified as high-emission (#9, #23–#26), with fuel-based EFs ranging from 1.13 to 13.6 times the PEMS standard.
- (3) Bin-method analysis can preliminarily diagnose the causes of NOx emissions. For vehicles #5, #9, and #23 selected in this study, the causes were as follows: prolonged idling and high-power operation (#5); reduced SCR conversion efficiency in the 210°C–250 °C Inlet Tem range (#9); and near failure of the SCR system below 430 °C (#23). The thermal decomposition of vanadium-based catalysts may be a major cause of SCR failure.
- (4) Machine learning captures the primary emission characteristics of vehicles, excluding the influence of other variables, with higher resolution, revealing the distinct impacts of multiple factors on emissions. For vehicles #5, #9, and #23, the five-fold cross-validation R² values were 0.58, 0.66, and 0.85, respectively. The machine learning models highlight the different impacts of SCR Inlet and Outlet Temp on instantaneous NOx emissions. For #5, the discrepancy between Inlet and Outlet Temp, particularly when the Outlet Temp exceeds the Inlet, was a major contributor to NOx emissions. For #9, it was the catalyst deterioration in the 200°C–250 °C range, and for #23, the near failure of the SCR system below 425 °C.

Nowadays, networked OBD devices are becoming increasingly prevalent. This paper reviews the current research on OBD and goes a step further by pioneering two innovative methods for investigating the causes of high vehicle emissions: bin-based analysis and machine

learning analysis. The proposed methods demonstrate excellent transferability and can be integrated into OBD big data platforms to provide policymakers with an overview of vehicle conditions and offer scientific support for law enforcement. In the next phase of research, we aim to explore further through the following steps: (Kowalska et al., 2020): Developing sampling algorithms to optimize computational efficiency (Wei et al., 2019). Expanding the dataset by downloading more diverse vehicle types and data spanning longer time periods, thereby overcoming the challenges of identifying vehicles across different emission standards (Salmon et al., 2018). Collaborating with law enforcement agencies to validate the accuracy of the proposed methods.

CRediT authorship contribution statement

Zeping Cao: Writing – original draft, Software, Methodology, Conceptualization. **Kai Shi:** Validation, Methodology. **Hao Qin:** Formal analysis. **Zhou Xu:** Formal analysis. **Xiaoyang Zhao:** Methodology. **Jiawei Yin:** Investigation. **Zhenyu Jia:** Software. **Yanjie Zhang:** Supervision. **Hailiang Liu:** Investigation. **Qijun Zhang:** Supervision, Conceptualization. **Hongjun Mao:** Funding acquisition.

Associated content

Supplementary Material. Detailed process to identify high-emission vehicles (Text S1). Detailed descriptions of the corresponding formulas and procedures for generating PDP and ICE plots (Text S2). Detailed descriptions of XGBoost (Text S3). Detailed vehicle information (Table S1). Upper and lower limits for each variable (Table S2). XGBoost Hyperparameter Specifications (Table S3). Additional features beyond those outlined in the text (Table S4). Additional features beyond those outlined in the text (Table S5). The data dispersion across different MOVES bins, represented by the coefficient of variation (Table S6). Examples of time series errors, exceeding boundaries errors, and constant value errors (Figure S1). The error proportions of different columns. V,RPM,F_{fuel},MAF,T_{SCR-in},T_{SCR-out},ECT,NO_x_up,NO_x_down refer to Vehicle speed, Engine speed, Fuel rate, Intake flow rate, SCR upstream temperature, SCR downstream temperature, Engine coolant temperature, SCR upstream NOx concentration and SCR downstream NOx concentration (Figure S2). Proportions of time series errors, exceeding boundaries errors, constant value errors, and analyzable data in the total dataset for each vehicle (Figure S3). Scatter plot based on fuel-based emission factors and compliant window ratio (Figure S4). Emission factors based on fuel (top) and distance (bottom) across different Moves bins (Figure S5). Scatter plot of fuel-based NOx emission factors versus R² values of trained machine learning models on the testing set (Figure S6). Feature importance of the three models. The color gradient of the bars represents Feature Weight (Figure S7). ICE plots of machine learning models for #5, #9, and #23 (Figure S8). Bivariate partial dependence plot for Outlet Temp and Torque, and SCR Inlet Tempe and Torque (Figure S9). Plot of Inlet and Outlet Temp Data (Samples #5, #9, and #23) (Figure S10).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the National Key Research and Development Program of China (2023YFC3705405), the National Natural Science Foundation of China (42107114, 42177084) and the Fundamental Research Funds for the Central Universities of China (63241318, 63241322, 63243126).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envpol.2025.125751>.

Abbreviations

HDDVs	heavy-duty diesel vehicles
PEMS	portable emissions measurement system
OBD	On-Board Diagnostic
MCUs	microcontroller units
OBM	on-board monitoring
FWM	fuel-consumption based window method
EFs	emission factors
PDP	partial dependence plots
ICE	individual conditional expectation
C-WTVC	China heavy-duty commercial vehicle test cycle
MSE	mean squared error
Torque	net output torque
RPM	engine revolutions per minute
SCR	selective catalytic reduction
R ²	coefficient of determination

Data availability

The authors are unable or have chosen not to specify which data has been used.

References

- Agarwal, A.K., Mustafi, N.N., 2021. Real-world automotive emissions: monitoring methodologies, and control measures. *Renew. Sustain. Energy Rev.* 137, 110624.
- Akimoto, H., Tanimoto, H., 2022. Rethinking of the adverse effects of NOx-control on the reduction of methane and tropospheric ozone—Challenges toward a denitrified society. *Atmos. Environ.* 277, 119033.
- Badshah, H., Posada, F., Muncrief, R., 2022. Current State of NOx Emissions from In-Use Heavy-Duty Diesel Vehicles in the United States. White Paper. The International Council on Clean Transportation. https://theicct.org/sites/default/files/publications/NOx_Emissions_In_Use_HDV_US_20191125.pdf.
- Barbier, A., Salavert, J.M., Palau, C.E., Guardiola, C., 2024. Analysis of the Euro 7 on-board emissions monitoring concept with real-driving data. *Transport. Res. Transport Environ.* 127, 104062.
- Bray, C.D., Battye, W.H., Aneja, V.P., Schlesinger, W.H., 2021. Global emissions of NH3, NOx, and N2O from biomass burning and the impact of climate change. *J. Air Waste Manag. Assoc.* 71 (1), 102–114.
- Xgboost: a scalable tree boosting system. In: Chen, T., Guestrin, C. (Eds.), 2016. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- Chen, Z., Liu, Q., Liu, H., Wang, T., 2024. Recent advances in SCR systems of heavy-duty diesel vehicles—low-temperature NOx reduction technology and combination of SCR with remote OBD. *Atmosphere* 15 (8), 997.
- China MoEaEoPsRo, 2018. Limits and measurement methods for emissions from diesel vehicles under free acceleration and lugdown cycle. Available from: https://www.mee.gov.cn/ywggz/fgbz/bz/bzwb/dqhjhb/dqydywrwpfbz/201811/t20181113_673588.shtml.
- Davis, S.J., Lewis, N.S., Shaner, M., Aggarwal, S., Arent, D., Azevedo, I.L., et al., 2018. Net-zero emissions energy systems. *Science* 360 (6396), eaas9793.
- Deng, F., Lv, Z., Qi, L., Wang, X., Shi, M., Liu, H., 2020. A big data approach to improving the vehicle emission inventory in China. *Nat. Commun.* 11 (1), 2801.
- Frazier, P.I., 2018. Bayesinan optimization. *Recent Advances in Optimization and Modeling of Contemporary Problems. Informs*, pp. 255–278.
- Grange, S.K., Farren, N.J., Vaughan, A.R., Davison, J., Carslaw, D.C., 2020. Post-dieselgate: evidence of NOx emission reductions using on-road remote sensing. *Environ. Sci. Technol. Lett.* 7 (6), 382–387.
- Gren, M., Brutemark, A., Jägerbrand, A., 2021. Air pollutants from shipping: costs of NOx emissions to the baltic sea. *J. Environ. Manag.* 300, 113824.
- Guan, B., Zhan, R., Lin, H., Huang, Z., 2014. Review of state of the art technologies of selective catalytic reduction of NOx from diesel engine exhaust. *Appl. Therm. Eng.* 66 (1), 395–414.
- Ho, C.S., Lv, Z., Peng, J., Zhang, J., Choe, T.-H., Zhang, Q., et al., 2023. Optical properties of vehicular brown carbon emissions: road tunnel and chassis dynamometer tests. *Environ. Pollut.* 320, 121037.
- Jia, Y., Hu, X., Kang, W., Dong, X., 2024a. Unveiling microbial nitrogen metabolism in rivers using a machine learning approach. *Environ. Sci. Technol.* 58 (15), 6605–6615.
- Jia, Z., Yin, J., Cao, Z., Wei, N., Jiang, Z., Zhang, Y., et al., 2024b. Large-scale deployment of intelligent transportation to help achieve low-carbon and clean sustainable transportation. *Sci. Total Environ.* 949, 174724.
- Jiang, Y., Tan, Y., Yang, J., Karavalakis, G., Johnson, K.C., Yoon, S., et al., 2022. Understanding elevated real-world NOx emissions: heavy-duty diesel engine certification testing versus in-use vehicle testing. *Fuel* 307, 121771.
- Ke, W., Zhang, S., Wu, Y., Zhao, B., Wang, S., Hao, J., 2017. Assessing the future vehicle fleet electrification: the impacts on regional and urban air quality. *Environ. Sci. Technol.* 51 (2), 1007–1016.
- Kowalska, M., Skrzypek, M., Kowalski, M., Cyrys, J., 2020. Effect of NOx and NO2 concentration increase in ambient air to daily bronchitis and asthma exacerbation, Silesian Voivodeship in Poland. *Int. J. Environ. Res. Publ. Health* 17 (3), 754.
- Kumbhar, V., Pandey, A., Sonawane, C.R., Panchal, H., Ağbulut, Ü., 2023. Numerical and experimental investigation of the influence of various metal-oxide-based nanoparticles on performance, combustion, and emissions of CI engine fuelled with tamarind seed oil methyl ester. *Energy* 265, 126258.
- Li, X., Ai, Y., Ge, Y., Qi, J., Feng, Q., Hu, J., et al., 2022. Integrated effects of SCR, velocity, and Air-fuel Ratio on gaseous pollutants and CO2 emissions from China V and VI heavy-duty diesel vehicles. *Sci. Total Environ.* 811, 152311.
- Lv, Z., Zhang, Y., Ji, Z., Deng, F., Shi, M., Li, Q., et al., 2023. A real-time NOx emission inventory from heavy-duty vehicles based on on-board diagnostics big data with acceptable quality in China. *J. Clean. Prod.* 422, 138592.
- McCaffery, C., Zhu, H., Tang, T., Li, C., Karavalakis, G., Cao, S., et al., 2021. Real-world NOx emissions from heavy-duty diesel, natural gas, and diesel hybrid electric vehicles of different vocations on California roadways. *Sci. Total Environ.* 784, 147224.
- Moradi, R., Berangi, R., Minaei, B., 2020. A survey of regularization strategies for deep models. *Artif. Intell. Rev.* 53 (6), 3947–3986.
- Mulholland, E., Miller, J., Bernard, Y., Lee, K., Rodríguez, F., 2022. The role of NOx emission reductions in Euro 7/VII vehicle emission standards to reduce adverse health impacts in the EU27 through 2050. *Transport Eng.* 9, 100133.
- Pandey, A.K., Nandgaonkar, M., Varghese, A., Kohil, R., Warke, A., 2023. Comparison and Evaluation of Engine Wear, Engine Performance, NOx Reduction and Nanoparticle Emission by Using Gasoline, JP-8, Karanja Oil Methyl Ester Biodiesel, and Diesel in a Military 720 kW, Heavy-Duty CIDI Engine Applying EGR with Turbo Charging. *SAE Tech. Paper* 0148-7191. <https://doi.org/10.4271/2023-01-0318>.
- Pandey, A., Suresh, S., Nandgaonkar, M., Sonawane, CR., 2020. Comparison and Evaluation of Performance, Combustion and Particle Emissions of Diesel and Gasoline in a Military Heavy Duty 720 kW CIDI Engine Applying EGR. *SAE Tech. Pap.* 01–2057. <https://doi.org/10.4271/2020-01-2057>.
- Pandey, A., Nandgaonkar, M., Varghese, A., Sambasivan, S., 2021. Evaluation and Comparison of Performance, Combustion Characteristics, NOx and Particle Emission of Gasoline, JP-8 and Diesel Fuel in a Military Heavy Duty 38 Liters CIDI Engine Applying EGR. *SAE Tech. Pap.* 01–1183. <https://doi.org/10.4271/2021-01-1183>.
- Pandey, A.K., Nandgaonkar, M., Sonawane, C., Kumbhar, V., 2022. Experimental Investigation of the Effect of Gasoline Fuel on Engine Performance, NOx reduction, and Engine Wear of a 38.8L Military Heavy Duty CIDI Diesel Engine Applying EGR. *SAE Tech. Pap.* 01–1027. <https://doi.org/10.4271/2022-01-1027>.
- Pavlovic, J., Fontaras, G., Broekaert, S., Ciuffo, B., Ktistakis, M., Grigoratos, T., 2021. How accurately can we measure vehicle fuel consumption in real world operation? *Transport. Res. Transport Environ.* 90, 102666.
- Pivetta, D., Tafone, A., Mazzoni, S., Romagnoli, A., Taccani, R., 2024. A multi-objective planning tool for the optimal supply of green hydrogen for an industrial port area decarbonization. *Renew. Energy* 232, 120979.
- Programme UNE, 2024. Used heavy-duty vehicles and the environment: a global overview of used heavy-duty vehicles: flow, scale and regulation. Available from: <https://www.unep.org/resources/report/used-heavy-duty-vehicles-and-environment-global-overview-used-heavy-duty-vehicles>.
- Salmón, P., Stroh, E., Herrera-Dueñas, A., von Post, M., Isaksson, C., 2018. Oxidative stress in birds along a NOx and urbanisation gradient: an interspecific approach. *Sci. Total Environ.* 622, 635–643.
- Tan, P-Q., Wang, S-Y., Hu, Z-Y., Lou, D-M., 2019a. Durability of V2O5-WO3/TiO2 selective catalytic reduction catalysts for heavy-duty diesel engines using B20 blend fuel. *Energy* 179, 383–391.
- Tan, Y., Henderick, P., Yoon, S., Herner, J., Montes, T., Boriboonsomsin, K., et al., 2019b. On-board sensor-based NOx emissions from heavy-duty diesel vehicles. *Environ. Sci. Technol.* 53 (9), 5504–5511.

- Tong, Z., Li, Y., Lin, Q., Wang, H., Zhang, S., Wu, Y., et al., 2022. Uncertainty investigation of plume-chasing method for measuring on-road NOx emission factors of heavy-duty diesel vehicles. *J. Hazard Mater.* 424, 127372.
- Wang, J., Wang, R., Yin, H., Wang, Y., Wang, H., He, C., et al., 2022. Assessing heavy-duty vehicles (HDVs) on-road NOx emission in China from on-board diagnostics (OBD) remote report data. *Sci. Total Environ.* 846, 157209.
- Wei, C.-C., Lin, H.-J., Lim, Y.-P., Chen, C.-S., Chang, C.-Y., Lin, C.-J., et al., 2019. PM2. 5 and NOx exposure promote myopia: clinical evidence and experimental proof. *Environ. Pollut.* 254, 113031.
- Wei, N., Zhang, Q., Zhang, Y., Jin, J., Chang, J., Yang, Z., et al., 2022a. Super-learner model realizes the transient prediction of CO₂ and NOx of diesel trucks: model development, evaluation and interpretation. *Environ. Int.* 158, 106977.
- Wei, N., Jia, Z., Men, Z., Ren, C., Zhang, Y., Peng, J., et al., 2022b. Machine learning predicts emissions of brake wear PM2. 5: model construction and interpretation. *Environ. Sci. Technol. Lett.* 9 (5), 352–358.
- Wei, N., Men, Z., Ren, C., Jia, Z., Zhang, Y., Jin, J., et al., 2022c. Applying machine learning to construct braking emission model for real-world road driving. *Environ. Int.* 166, 107386.
- Xu, Y., Xiao, H., Wu, D., 2019. Traffic-related dustfall and NOx, but not NH₃, seriously affect nitrogen isotopic compositions in soil and plant tissues near the roadside. *Environ. Pollut.* 249, 655–665.
- Yang, Z., Tate, J.E., Rushton, C.E., Morganti, E., Shepherd, S.P., 2022. Detecting candidate high NOx emitting light commercial vehicles using vehicle emission remote sensing. *Sci. Total Environ.* 823, 153699.
- Yang, Z., Han, K., Liao, L., Wu, J., 2024. Using multi-source data to identify high-emitting heavy-duty diesel vehicles. *arXiv preprint arXiv:240410243*.
- Zhang, S., Zhao, P., He, L., Yang, Y., Liu, B., He, W., et al., 2020. On-board monitoring (OBM) for heavy-duty vehicle emissions in China: regulations, early-stage evaluation and policy recommendations. *Sci. Total Environ.* 731, 139045.
- Zhang, X., Li, J., Liu, H., Li, Y., Li, T., Sun, K., et al., 2023. A fuel-consumption based window method for PEMS NOx emission calculation of heavy-duty diesel vehicles: method description and case demonstration. *J. Environ. Manag.* 325, 116446.
- Zhao, P., Wu, X., Zhang, S., He, L., Yang, Y., Hu, Q., et al., 2024. Regulatory insights for on-board monitoring of vehicular NOx emission compliance. *Environ. Sci. Technol.* 58 (18), 7968–7976.
- Ministry of Ecology and Environment of the People's Republic of China (MEEPRC), 2023. China Mobile Source Environmental Management Annual Report. Ministry of Ecology and Environment of the People's Republic of China (MEEPRC), Beijing, China (In Chinese).