Science of the Total Environment 928 (2024) 172427

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

On-road evaluation and regulatory recommendations for NOx and particle number emissions of China VI heavy-duty diesel trucks: A case study in Shenzhen

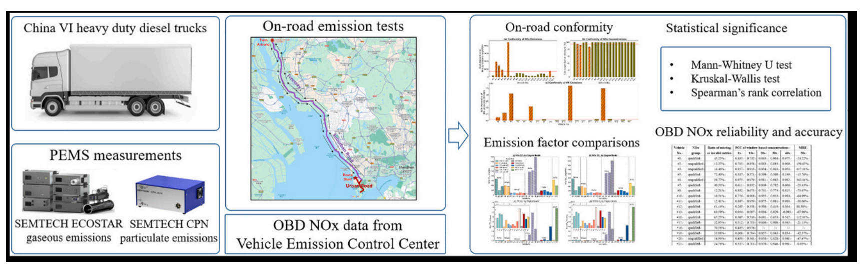
Weixia Li [a,*], Zhurong Dong [a], Ling Miao [a], Guoyuan Wu [b], Zhijun Deng [a], Jianfeng Zhao [a], Wenwei Huang [a]

[a] *School of Automotive and Transportation Engineering, Shenzhen Polytechnic University, 7098 Liuxian Avenue, Shenzhen, Guangdong, China*
[b] *Bourns College of Engineering - Center for Environmental Research & Technology (CE-CERT), University of California, Riverside 92507, CA, USA*

HIGHLIGHTS

- On-road NOx, PN emissions and OBD NOx data of 21 China VI HDDTs were analyzed.
- 38.1 % of the test vehicles emitted higher pollutants than certification standards.
- Vehicles emitted higher NOx (PN) at low (high) speeds or large (small) VSPs.
- Unqualified HDDTs emitted significantly higher pollutants at medium to high speeds.
- The integrity of OBD NOx was poor and time-averaged data show better reliability.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Pavlos Kassomenos

Keywords:
NOx and particle number emissions
China VI heavy duty diesel trucks
On-road conformity
Vehicle operating modes
Statistical significance
OBD NOx data reliability and accuracy

ABSTRACT

This research analyzed the real-world NOx and particle number (PN) emissions of 21 China VI heavy-duty diesel trucks (HDDTs). On-road emission conformity was first evaluated with portable emission measurement system (PEMS). Only 76.19 %, 71.43 % and 61.90 % of the vehicles passed the NOx test, PN test and both tests, respectively. The impacts of vehicle features including exhaust gas recirculation (EGR) equipment, mileage and tractive tonnage were then assessed. Results demonstrated that EGR helped reducing NOx emission factors (EFs) while increased PN EFs. Larger mileages and tractive tonnages corresponded to higher NOx and PN EFs, respectively. In-depth analyses regarding the influences of operating conditions on emissions were conducted with both numerical comparisons and statistical tests. Results proved that HDDTs generated higher NOx EFs under low speeds or large vehicle specific powers (VSPs), and higher PN EFs under high speeds or small VSPs in general. In addition, unqualified vehicles generated significantly higher NOx EFs than qualified vehicles on freeways or under speed≥40 km/h, while significant higher PN EFs were generated on suburban roads, freeways or under operating modes with positive VSPs by unqualified vehicles. The reliability and accuracy of on-board diagnostic (OBD) NOx data were finally investigated. Results revealed that 43 % of the test vehicles did not report reliable OBD data. Correlation analyses between OBD NOx and PEMS measurements further demonstrated that the consistency of instantaneous concentrations were generally low. However, sliding window averaged

* Corresponding author.
E-mail addresses: WeixiaLi@szpu.edu.cn (W. Li), arondong@szpu.edu.cn (Z. Dong), sdmiaoling@szpu.edu.cn (L. Miao), gywu@cert.ucr.edu (G. Wu), dengzhijun@szpu.edu.cn (Z. Deng), zjf@szpu.edu.cn (J. Zhao), ww6809@szpu.edu.cn (W. Huang).

https://doi.org/10.1016/j.scitotenv.2024.172427
Received 24 January 2024; Received in revised form 19 March 2024; Accepted 10 April 2024
Available online 16 April 2024

concentrations show better correlations, *e.g.*, the Pearson correlation coefficients on 20s-window averaged concentrations exceeded 0.85 for most vehicles. The research results provide valuable insights into emission regulation, *e.g.*, focusing more on medium- to high-speed operations to identify unqualified vehicles, setting higher standards to improve the quality of OBD data, and adopting window averaged OBD NOx concentrations in evaluating vehicle emission performance.

1. Introduction

On-road vehicles are the major contributors to air pollutants in the large and medium-sized cities in China. According to China Mobile Source Environmental Management Annual Report 2023 (MEE, 2023), the number of motor vehicles has increased to 417 million in 2022, resulting in 5.3 million tons of nitrogen oxides (NOx) emissions and 53 thousand tons of particulate matter (PM) emissions. Among all vehicle types, heavy duty diesel trucks (HDDTs) are the dominant pollutants emitters. More specifically, 74.6 % of NOx and 49.1 % of PM of the aforementioned emissions were contributed by HDDTs. Similar trends can be witnessed in highly-motorized countries around the world as well. For example, the United Kingdom reported that 32 % of NOx emissions and 14 % of PM2.5 (PM with an aerodynamic diameter smaller than 2.5 μm) emissions came from transportation in 2021 (Department for Transport U.K., 2023). And the emission share of heavy goods vehicles (HGVs) was up to 21 % although they made up only 6 % of the vehicle miles travelled. The report from European Union revealed that the contribution of road transportation to the 27 EU Member States (EU27) emissions in 2020 was 37 % for NOx and 9 % for PM2.5, and heavy-duty trucks were the second largest emission source following passenger buses (EMISIA, 2023). Similarly, data of California Air Resource Board (CARB) showed that on-road heavy-duty vehicles (HDVs) were responsible for >30 % of overall NOx emissions in California (CARB, 2020). NOx and PM can not only lead to environmental degradations such as acid rain, the greenhouse effect and smoggy weather, but also cause a variety of health issues such as respiratory impairment, cardiovascular, lung diseases, as well as cancer (Wolfe et al., 2019; EMISIA, 2023). Previous research estimated 339 annual premature deaths due to noncommunicable diseases plus lower respiratory infection were attributed to PM2.5 HDDT emissions in Beijing (Zhang et al., 2022). In EU27, 238,000 and 49,000 premature deaths were attributable to the exposure to PM2.5 and NO_2 concentrations above the World Health Organization (WHO) guideline level in 2020 (EEA, 2022).

To mitigate the negative impacts of road transportation activities, significant efforts have been dedicated to reduce vehicle emissions through technological innovations and legislative regulations. Regarding the former solutions, numerous after-treatment systems have been developed to bring down exhaust emissions with little or no impacts on engine performance (Xu et al., 2023; Ayodhya and Narayanappa, 2018), *e.g.*, advanced NOx emission control systems of selective catalytic reduction (SCR), PM emission control systems of diesel particulate filter (DPF), diesel oxidation catalyst (DOC) for oxidizing CO and HC while improving soot and NOx removal of downstream DPF and SCR, and ammonia slip catalyst (ASC) for eliminating the slipped NH_3 out of SCR. The latter solutions are implemented through enacting more and more stringent emission standards. In China, the vehicle emission control program started from the 1990s and has been developing rapidly since then (Wu et al., 2017; Zhang et al., 2023). The latest emission standard for heavy duty diesel vehicles (HDDVs), *i.e.*, "Limits and measurement methods for emissions from diesel fuelled heavy-duty vehicles (CHINA VI)" (GB17691-2018) was released in 2018 and implemented nationwide in 2021 (MEE and SAMR, 2018). The legislation set extremely low limits for NOx and PM emissions, and made on-board diagnostics (OBD) with remote emission management terminals mandatory for newly manufactured, sold or registered HDDVs from July 1 st, 2021 (which are regarded as China VI HDDVs). Despite the strict limits in regulations, vehicle emissions in the real-world operation might significantly differ from what were expected, as they can be impacted by various factors such as road conditions, operating statuses, vehicle usage features. For example, studies for the in-use NOx and black carbon emissions from diesel, natural gas, and diesel hybrid electric vehicles in California (Ma et al., 2024; McCaffery et al., 2021) and NOx emissions from medium-duty diesel trucks in Europe (Papadopoulos et al., 2020) all demonstrated higher emission levels than certification standards. For the impacting factors, research revealed that: 1) higher mileage might lead to higher NOx emissions of diesel trucks (Ma et al., 2024; Lyu et al., 2023; Li et al., 2022); 2) highway driving might result in lower black carbon (BC) emission factors than non-highway conditions (Shen et al., 2021); 3) both PM and gaseous emissions of NOx, CO, CO_2, THC (total hydrocarbons) showed descending orders by road types (urban, suburban and freeways); and 4) emission factors were highly dependent on vehicle speed (Li et al., 2022) and VSP bins (Dhital et al., 2021). Besides the aforementioned objective factors from the environment or vehicle dynamics, subjective factors such as intentionally tampering on-board sensors might improperly impact vehicle emissions as well (He et al., 2020).

Since China VI standard has been implemented for a few years, it is essential to investigate the real-world gaseous and particulate emission characteristics of China VI HDDTs to evaluate the effectiveness of the new standard. Meanwhile, with the mandatory deployment of OBD devices, the demands of in-use surveillance and analysis for HDVs *via* on-board monitoring have been increasing, *e.g.*, estimating potential benefits of the latest China VI standard (Wang et al., 2022), comparing emissions by operating environment (*i.e.*, city roads *versus* highway) or vehicle types (Ge et al., 2023a), identifying high emissions in the case of an in-use fleet (Ge et al., 2023b), and estimating real-time NOx emission inventory (Lv et al., 2023). Nevertheless, the reliability and accuracy of the OBD data, especially from China VI HDDTs, have not been fully investigated. One earlier study that compared OBD data with portable emission measurement system (PEMS) measurements only involved several China IV and China V HDDVs (Zhang et al., 2020). Other researchers focused on lab environment-based evaluation of OBD data from China VI vehicles using chassis dynamometers (Hao et al., 2023; Ge et al., 2023a, 2023b) rather than PEMS.

To make up the deficiencies of the previous literatures that lack 1) the accurate measurement of on-road emissions and 2) the quality evaluations of OBD data for China VI HDDTs, this research conducted on-road evaluations for NOx and particle number (PN) emissions of China VI HDDTs using portable emission measurement systems (PEMS) and statistical analyses, and compared OBD NOx data with PEMS measurements to evaluate the reliability and accuracy. The contributions can be summarized as follows.

- The research carried out large-scale road tests on vehicle emissions that involved 21 China VI HDDTs. To the authors' best knowledge, this is the largest on-road experiment to date that measured the real-world emissions of China VI HDDTs with PEMS. Since China VI emission standard is regarded as the stringiest standard ever enforced around the world, assessing the real-world emissions characteristics with on-road tests may not only provide regulators quantitative supports to learn about the effects of the new emission standard, but also provide potential references for other countries to consider and enact their next-step regulations.

- The research delivers significant insights derived from extensive real-world emission measurements. On-road conformity, comparative vehicle emissions with respect to vehicle features and operating conditions, and the statistical significance of the impact factors were investigated for the test vehicles. Both fundamental analyses of vehicle emissions regarding vehicle features (*e.g.*, engine model, mileage, *etc.*) and in-depth analyses of vehicle emissions regarding operating conditions (*e.g.*, road type, operating mode) were provided. Specific operational conditions and vehicle characteristics that influence emissions performance were identified through meticulous statistical analyses, which shed light on potential improvements for emission regulations and compliance mechanisms.
- The research collected the OBD-transmitted NOx data of the test fleet from Vehicle Emission Control Center (VECC) of MEE and evaluated OBD data quality by comparing them with PEMS measurements. The results of data integrity and accuracy revealed the major data issues encountered in the current regulatory stage, and pointed out the direction for further improvements. Preliminary analyses on the correlation and linearity between the two datasets were also carried out, which may provide valuable insights for further emission regulations with OBD NOx data.

2. Methodology

To fulfill the on-road evaluations for NOx and PN emissions of China VI HDDTs, large-scale road tests were first conducted to collect real-world vehicle operation and emission data. In the meanwhile, the OBD NOx data were also collected from VECC of MEE for evaluation. After that, vehicle operation and emission data were processed to generate operating modes and emission factors for on-road conformity analyses and emission comparisons. In the end, several statistical tests were applied to assess the significance of associations between vehicle emissions and impacting factors, and OBD NOx data accuracy. Details on experimental design, data processing and calculations, and statistical analyses were described below.

2.1. Experimental design

2.1.1. Test fleet

In this research, 21 HDDTs that comply with China VI emission standard were selected to form a test fleet, including 7 vehicle brands (*i.e.*, Chenglong, Dongfeng, Jiefang, Jianghuai, Haowo, Shandeka and Shanqi) and 6 engine brands (*i.e.*, Dachai, Dongfengkangmingsi, Weichai, Xichai, Yuchai and Zhongguozhongqi) that were widely used in the City of Shenzhen. The vehicles corresponded to 13 distinct vehicle models and 11 different engine models, respectively. One test vehicle was manufactured in 2019, while the rest ones were manufactured in either 2021 or 2022. The mileages ranged from 10,000 km to 160,000 km, and were categorized into four groups of '<20,000 km', '20,000–40,000 km', '40,000– 80,000 km' and '>80,000 km' from low to high in the following analyses. The vehicle curb weight was in the range of 5300 kg to 8800 kg, with tractive tonnage of 24,700 kg to 40,000 kg, which was approximated as three groups of 25 t, 34 t and 40 t from small to large for further analyses. All vehicles were installed with after-treatment systems of DOC, SCR, ASC and DPF, and 11 of them were also equipped with exhaust gas recirculation (EGR). The detailed vehicle specifications are summarized in Table S1. To fully understand vehicle emission characteristics and potential impacts of vehicle features, emission factors are calculated by and compared among engine models, EGR groups (equipped *versus* unequipped), mileage groups and tractive tonnage groups in Section 3.2, respectively.

2.1.2. Emission measurement systems

The gaseous and particulate emissions were measured using SEMTCH ECOSTAR (Sensors Inc., 2023a) and SEMTECH CPN (Sensors Inc., 2023b), respectively, which are widely-used PEMS manufactured by Sensors, Inc.. With SEMTECH ECOSTAR, second-by-second NO and NO_2 emissions were measured *via* the nondispersive ultraviolet (NDUV) module, CO and CO_2 emissions were measured by the nondispersive infrared (NDIR) analyzer, and the THC emissions were measured using the heated flame ionization detector (HFID). In this research, the NO and NO_2 measurements were used to analyze vehicle NOx emission characteristics. SEMTECH CPN was utilized to measure the instantaneous PN emissions. Technical specifications of the equipment are described in Tables S2 and S3. In addition to the gaseous and PN emissions, real-time engine operating information such as vehicle speed, engine speed, fuel flow rate, engine power, engine torque, *etc.*, were obtained using an OBD data logger. Then the emission rates and vehicle operating information were used to calculate emission factors. The trajectories of the obtained instantaneous speed, NOx emission rate and PN emission rate for a test vehicle, *i.e.*, vehicle #5 are provided in Fig. S1 in the Supplementary material. Besides instantaneous vehicle emission rates and operation status, the OBD NOx data of the test fleet during road tests were collected from VECC, and they are compared with PEMS measurements in Section 3.4 to evaluate the reliability and accuracy.

2.1.3. Test routes

Test routes were designed according to the PEMS measurement regulations in China (MEE, 2017), which specified the proportions of urban, suburban and freeway operations of the test routes for different types of vehicles. For HDDVs, there are three subcategories of N1, N2, and N3, corresponding to HDDVs with gross vehicle weight (GVW) $\leq$ 3500 kg, 3500 kg $<$ GVW $\leq$ 12,000 kg, and GVW $>$ 12,000 kg, respectively. For different subcategories of HDDVs, proportions of urban roads, suburban roads and freeways in the test routed are different. The test vehicles in our research were all N3 HDDTs, which were usually used for long-distance and heavy-duty logistics delivery. For these vehicles, the regulations require urban, suburban and freeway operations, with proportions in each test trip being around 20 %, 25 % and 55 %, respectively. In response to this requirement, on-road emission measurements were carried out on the test routes depicted in Fig. S2. More specifically, the test route was constituted by Xinghai Boulevard, Yueliangwan Boulevard and Guangshen Coastal Expressway in southwest of Shenzhen. Test vehicles departed from the intersection of Mawaner Road and Linhai Boulevard, and conducted urban operations on Xinghai Boulevard and Yueliangwan Boulevard with an average speed of 15 km/h to 30 km/h. And then vehicles entered Guangshen Coastal Expressway at Yueliangwan Boulevard entrance, and started suburban driving with an average speed of 45 km/h to 70 km/h that lasted for 45 min. After that, vehicles performed freeway operations with an average speed higher than 70 km/h on Guangshen Coastal Expressway. Road tests were conducted from October 20th, 2022 to November 16th, 2022, and the corresponding information including test date, duration, distance, maximum speed and mean speed for each vehicle is illustrated in Table S4. Vehicular emissions on different types of roads and under various operating modes were logged using the SEMTECH equipment during the tests, and the potential relationships between them are explored in Section 3.3.

2.1.4. Test procedures

The test procedures were in strict accordance with the requirements in the regulation (MEE, 2017). Each on-road test went through the procedures of 1) measurement systems preparation, 2) measurement parameter determination, 3) test vehicle preparation, 4) measurement systems installation, 5) experimental pretreatment, 6) on-road test, 7) measurement systems verification, and 8) test data validity check. The test procedures are illustrated in Fig. S3. More specifically, the major measurement system is the PEMS described in Section 2.1.2. For the gas analyzers of the PEMS, linearity check and calibration were regularly conducted with calibration gases (*e.g.*, zero gas, span gas, and gases of known concentrations). The parameters to be measured are summarized in Table S5, including the concentrations of the gaseous and particulate pollutants, the flow rate and temperature of the exhausts, environment parameters (temperature and barometric pressure), engine parameters (engine speed, engine torque, fuel consumption rate, and coolant temperature), vehicle speed, and locations (latitude, longitude, and altitude). The load of the test vehicle was required to be between 50 % and 100 % of the maximum load, and it is set to be 50 % of the maximum load for all test vehicles in our research. The installed devices included PEMS, exhaust flowmeter (EFM), global positioning system (GPS), and electronic control unit (ECU) data logger. In experimental pretreatment, the measurement devices were started and fixed, the sampling system of PEMS and EFM were cleaned, the zero point and the span point of the gas analyzers were calibrated. During on road tests, the drivers were instructed to drive the test vehicle with speeds of specified ranges and on the specified route. When on-road test was finished, the drifts on the zero point and the span point were checked to ensure the drifts were <2 % of the full scale. In the end, the validity of the test data was checked to ensure the regulation requirements were satisfied. For example, the proportions of the urban, suburban, and freeway operations should be around 20 %, 25 %, and 55 %, respectively. The average vehicle speed should be 15 km/h ~ 30 km/h, 45 km/h ~ 70 km/h, and >70 km/h for urban roads, suburban roads, and freeways, respectively. The proportion of effective work-based windows (see Section 2.2.3 for the specific definition) should be larger than or equal to 50 %.

It is noted that real driving emissions (RDE) tests were different with tests using chassis dynamometer. Firstly, repetitions were not required in on-road tests with PEMS, thus each vehicle was tested once. Also, different tests were not necessarily required to be conducted by the same driver. Furthermore, no specific speed trajectories were designed for the on-road tests, but only average speed ranges for different types of roads were specified (*e.g.*, the average speed should be 15 km/h ~ 30 km/h for urban roads). The regulations stipulated the proportions of road types in the test route, the altitudes (no higher than 1700 m for China VI phase 6a), the temperatures (no lower than −7 °C and no higher than the calculated value based on barometric pressure), *etc.* as well. All the requirements were guaranteed satisfied in our research.

2.2. Data processing and calculations

2.2.1. Vehicle operating modes

To explore the vehicle emission characteristics under different operating conditions, the operating mode binning method was utilized in our research. Operating mode bins were proposed and used in MOtor Vehicle Emission Simulator (MOVEs). They were initially conceptualized for light-duty vehicles in "MOVES2004 Energy and Emission Inputs draft report" (Koupal et al., 2005) and developed for heavy-duty vehicles in MOVES2009 (USEPA, 2009). Due to the availability of test data in each bin, specifically in the high-speed segments, Wu et al. (2012) made slight modifications on the bins according to real-world driving patterns of HDDVs in China, and the modified operating mode bins were then widely used in the research of HDDV emissions in China (*e.g.*, He et al., 2020; Zheng et al., 2015). The same operating mode binning method (as those in Wu et al., 2012; Zheng et al., 2015; He et al., 2020) was adopted in this research. Firstly, vehicle specific power (VSP) that denotes the instantaneous power demand per unit mass of the vehicle was first calculated using Eq. (1) (Jiménez-Palacios, 1999), in which, a and v stand for the instantaneous vehicle acceleration (m/s^2) and speed (m/s), respectively; m is the vehicle mass (t); A, B and C denote the coefficients of rolling resistance (kW·s/m), rotational resistance (kW·s^2/m^2), and aerodynamic drag (kW·s^3/m^3), respectively. The same values of $\frac{A}{m}$, $\frac{B}{m}$ and $\frac{C}{m}$ as those in the previous research (Wu et al., 2012, Zheng et al., 2015, He et al., 2020), *i.e.*, 0.0857, 0 and 0.000331, were used in the calculation in this research. θ is the road grade (radians). Since the test route in the research is quite flat, the road grade effect can be neglected in the calculations.

$$\mathrm{VSP} = av + \frac{A}{m}\cdot v + \frac{B}{m}\cdot v^2 + \frac{C}{m}\cdot v^3 + gv\sin\theta \tag{1}$$

With the calculated VSPs, vehicle speeds and accelerations, vehicle instantaneous operating modes were then computed based on the definitions in Table S6, which classified vehicle operations into 22 bins, including a braking or decelerating bin, an idling bin, eight low-speed bins, eight medium-speed bins and four high-speed bins.

2.2.2. Emission factors

Emission factors (denoted as EFs hereinafter) play an important role in a variety of real-world applications. Firstly, they are utilized in evaluating whether vehicles meet the regulatory requirements on pollutants emissions. Chassis dynamometer tests and on-road tests are two typical ways to fulfill the evaluation function. The two kinds of tests use EFs with consistent units but different limit values. In chassis dynamometer tests, the parameter of power is utilized to describe the engine states, thus work-based EFs (in units of g/kWh or #/kWh) are adopted in the evaluations. On-road evaluations follow the conventions and adopt work-based EFs as well. In our research, on-road PEMS tests were conducted to evaluate the on-road conformity, and thus the work-based EFs were calculated. In the meanwhile, EFs are widely used in regional emission inventory estimations, in which they are usually distance-based (in units of g/km or #/km), and combined with vehicle distance travelled instead of vehicle engine powers as it is difficult to obtain an aggregate vehicle engine power. Therefore, we provided analyses on distance-based EFs in the research as well to promote the applicable scenarios of the research results. Work-based EFs (denoted as EF_w hereinafter) were calculated using the instantaneous emission rates and engine powers, while distance-based EFs (denoted as EF_d hereinafter) were derived by the instantaneous emission rates and vehicle speeds. The equations for the calculations are presented in (2) and (3), respectively.

$$EF_{d,t} = \frac{ER_t}{v_t} \times 3600 \tag{2}$$

$$EF_{w,t} = \frac{ER_t}{P_t} \times 3600 \tag{3}$$

where $EF_{d,t}$ is the instantaneous distance-based EF at time t (g/km for gaseous pollutants and #/km for particulate pollutants), and $EF_{w,t}$ (g/kWh or #/kWh) is work-based EF. ER_t, v_t, and P_t represent instantaneous emission rate (g/s or #/s), speed (km/h) and power (kW), respectively. In work-based window analyses (see Section 2.2.3) for EF_w, the numerator and denominator in Eq. (3) are replaced with the summations of the emission rates and powers of the data points within the work-based window. The multiplier of 3600 was for unit conversion. And for illustrative purpose, EFs of NOx were further multiplied by 1000 to convert the unit to mg/km or mg/kWh in the following analyses.

2.2.3. Work-based windows and on-road emission conformity

Following the instructions of the measurement method and technical specification for PEMS test of exhaust pollutants from heavy-duty diesel and gas fueled vehicles in China, on-road emission conformity was determined according to the emissions of work-based windows. A work-based window is defined as a continuous interval in which the accumulated engine power equals to the total power of the transient test cycle (usually using World Harmonized Transient Cycle (WHTC)), and it can be determined by Eq. (4).

$$\mathrm{W}(t_{2,i} - \Delta t) - \mathrm{W}(t_{1,i}) < W_{ref} < \mathrm{W}(t_{2,i}) - \mathrm{W}(t_{1,i}) \tag{4}$$

In Eq. (4), $t_{1,i}$ and $t_{2,i}$ denote the start and end time of the i^{th} window $(t_{1,i}, t_{2,i})$, $\mathrm{W}(t)$ is the accumulated power by time t, Δt means the sampling period (usually less than or equal to 1 s), and W_{ref} (kWh) stands for the reference power, *i.e.*, the total power of the transient test cycle. After the i^{th} window was determined, the start time of the next window was set to $t_{1,i} + \Delta t$, and the window generation process was repeated until the window end time reached the end time of the entire trajectory.

Based on Eq. (4), a number of work-based windows were generated for each vehicle. And then the EF_w of each window was computed. In the emission regulation standard, it requires that the 90th percentiles of NOx and PN EF_w for all effective work-based windows do not exceed 690 mg/kWh and 1.2×10^{12}#/kWh, respectively. In addition, the NOx emission concentration is required to be <500 ppm for >95 % of the valid data points. Therefore, the 90th percentiles of NOx and PN EF_w and the share of NOx concentrations below 500 ppm were calculated in the research to conduct the analyses of on-road conformity. A vehicle would be considered as NOx emission qualified if the 90th percentile of NOx EF_w was lower than the threshold of 690 mg/kWh and the share of NOx concentrations below 500 ppm was higher than the threshold of 95 %. Similarly, a vehicle would be considered as PN emission qualified if the 90th percentile of PN EF_w was lower than the threshold of 1.2×10^{12}#/kWh. Otherwise, the vehicle would be considered as emission unqualified.

2.3. Statistical analyses

In this research, three types of statistical tests were selected to conduct analyses on significance of associations. The first type of methods was selected to test the mean (or median) value differences among groups. The second type of test was used to determine the potential monotonic relationships between dependent and independent variables. And the third type of method was adopted for correlation analysis. Prior to the selection of the specific tests, data independence and normality were analyzed. Since the on-road test for each individual vehicle was conducted separately, data independence can be guaranteed. Data normality was evaluated with the Kolmogorov-Smirnov (KS) tests (Berger and Zhou, 2014). Since the KS tests rejected the null hypothesis of normal distributions of EF data, the nonparametric tests were selected in the first two types of tests. The selected methods including the Mann-Whitney *U* test and the Kruskal-Wallis one-way analysis of variance for testing median value differences, Spearman's rank correlation test for monotonic relationship determination, and the Pearson correlation coefficient for correlation analysis are described below.

2.3.1. Mann-Whitney U test

In this research, the distributions of the EFs for the emission qualified and unqualified vehicles under different conditions were compared with the Mann-Whitney U test (Mann and Whitney, 1947), MWU test for short, which is a frequently used non-parametric test for the difference in location (*e.g.*, median) between distributions, and it can be applied when 1) all the observations of both groups are independent of each other; 2) the responses are ordinal or continuous; and 3) the responses are not normally distributed and sample size is small.

2.3.2. Kruskal-Wallis one-way analysis of variance

The second statistical analysis method used in this research is the 'extended' method of the MWU test, namely the Kruskal-Wallis one-way analysis of variance (Kruskal and Wallis, 1952), KW ANOVA for short, a non-parametric equivalent of the one-way ANOVA that tests whether or not there is a statistically significant difference among the medians of multiple groups of independent and non-normally distributed samples. As the MWU test can only be used for comparisons of two groups, *e.g.*, inter-group comparisons between qualified and unqualified vehicles in our research, KWANOVA is adopted in the within-group analysis, in which emissions under different conditions (*e.g.*, on different road types, under different operating modes) within each group of qualified and unqualified vehicles are compared.

2.3.3. Spearman's rank correlation test

Spearman's rank correlation test (Spearman, 1904) is the third non-parametric method that used in this research to measure the potential monotonic relationships between emissions and operating mode bins. In Section 3.3.2, the coefficients and corresponding *p* values between vehicle EFs and operating mode bins (including all bins and several conditional bins) are calculated to evaluate how well the relationship can be described using a monotonic function, which provide quantitative metrics for the statistical dependence of vehicle emissions on operating statuses.

2.3.4. Pearson correlation coefficient

To evaluate the accuracy of OBD NOx data, the Pearson correlation coefficient (Pearson, 1895), PCC for short, is adopted in this research to measure the correlation between OBD NOx data and PEMS measurements. PCC provides the ratio between the covariance of the two variables and the product of their standard deviations, thus measuring the linear correlation with a value between −1 and 1. The PCC value should be close to 1 if the OBD NOx data is highly consistent with PEMS measurements.

3. Results and discussions

3.1. On-road conformity of vehicle samples

This section presents the on-road emission conformity results of the test vehicles according to the work window-based method in Section 2.2.3. The calculated 90th percentiles of the NOx EF_w, PN EF_w and the share of NOx concentrations below 500 ppm are depicted with bar plots in Fig. 1, in which the limits of the certification standard are represented with dashed lines, and vehicles with emissions exceeding the limits are highlighted with orange-colored and hatched bars. Detailed numerical results corresponding to the figure can also be found in Table S7 in the Supplementary material. As shown in subfigure (a), five vehicles, *i.e.*, vehicle #2, #3, #4, #6 and #20, generated higher NOx EF_w than certificate standard, and three of them, *i.e.*, vehicle #2, #3 and #6 also generated higher NOx concentrations than limits as illustrated in subfigure (b). Subfigure (c) presents the comparison results on vehicle PN emissions, in which six vehicles including vehicle #1, #2, #3, #6, #12 and #17 were demonstrated unqualified as they emit notably higher PNs than the certification standard. In addition, it is noted that three vehicles, *i.e.*, vehicle #2, #3 and #6, failed both tests on NOx and PN emissions. Using qualification rates to summarize the aforementioned on-road emission conformity, only 76.19 %, 71.43 % and 61.90 % of the test vehicles passed the NOx test, PN test and both tests, respectively.

3.2. Comparative emission results by vehicle features

This section provides the analyses of vehicle emission characteristics with respect to vehicle features. Vehicle emissions were compared by engine model, EGR group, mileage group and tractive tonnage group, respectively. In the analyses, emission qualified vehicles and unqualified vehicles were distinguished and compared.

3.2.1. Emission analyses by engine model

The 21 test vehicles were divided into six engine brands, *i.e.*, Dachai, Dongfengkangmingsi (DFKMS for short), Weichai, Xichai, Yuchai and Zhongguozhongqi (ZGZQ for short). And they could be further subdivided into eleven categories based on engine models. The calculated EFs, including NOx EF_d, NOx EF_w, PN EF_d and PN EF_w, are presented with bar plots in Fig. 2, in which engine models are differentiated with different colors, engine brands are separated with dashed vertical lines, and emission unqualified vehicles (*i.e.*, vehicles that did not pass the emission tests) are also highlighted with hatched bars. The numerical EFs of each vehicle are also presented in Table S8 in the Supplementary material. According to Fig. 2, vehicles with Dachai or Xichai engines

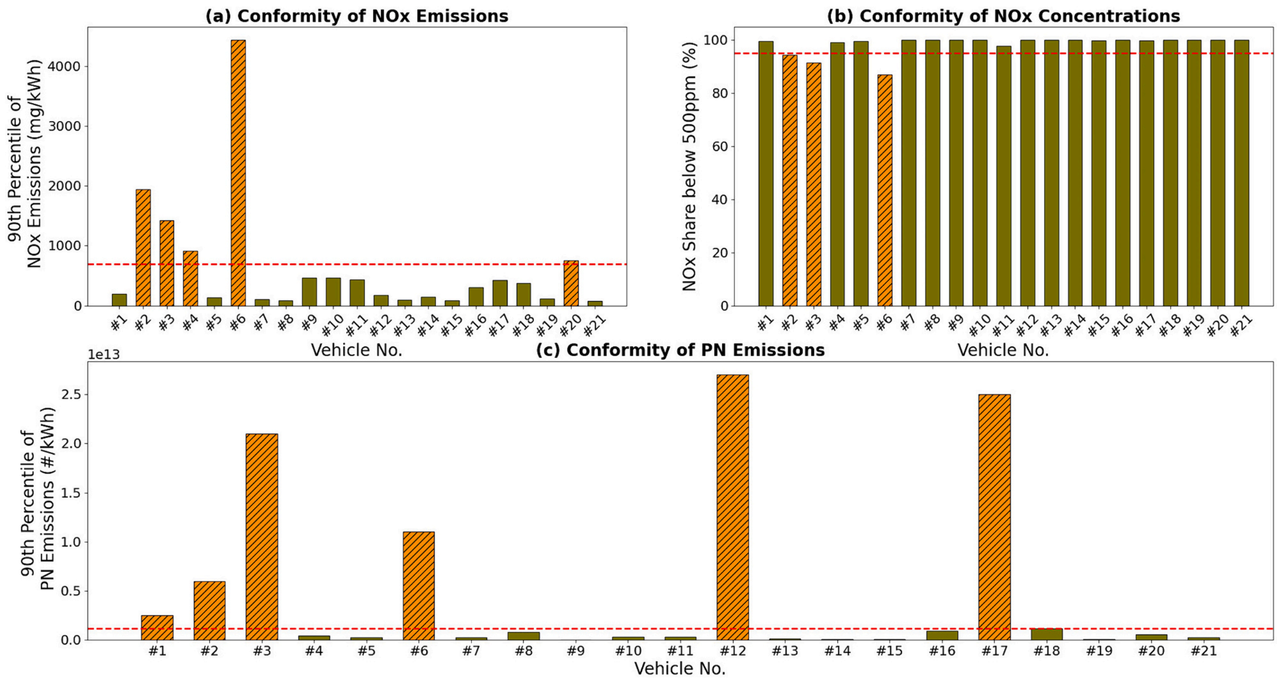


Fig. 1. On-road conformity results of the test vehicles.

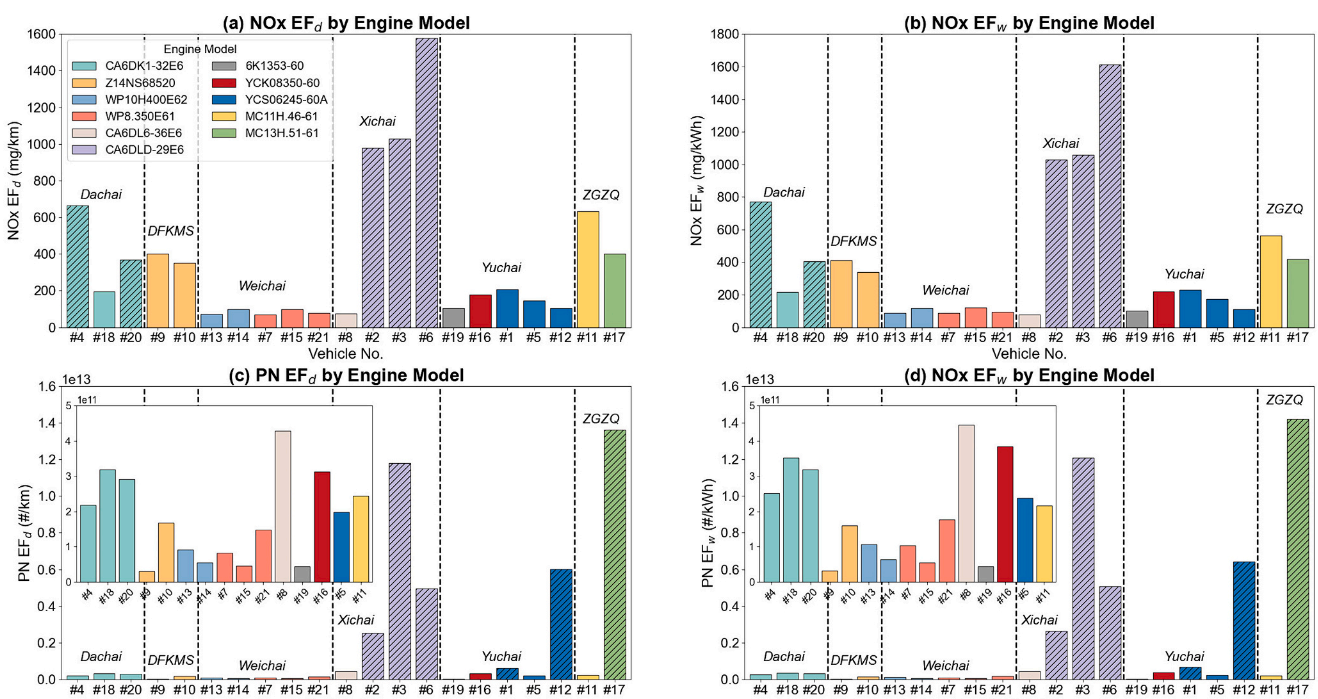


Fig. 2. Emission factors by engine model.

show poor performance in controlling NOx emissions. Two vehicles equipped with the 'CA6DK1-32E6' model of Dachai engine, *i.e.*, vehicle #4 and #20, and three vehicles equipped with the model 'CA6DLD-29E6' of Xichai engine, *i.e.*, vehicle #2, #3 and #6, failed to meet the NOx emission standards, and the latter three vehicles resulted in notably higher EFs than the rest vehicles, as depicted in subfigure (a) and (b). Subfigure (c) and (d) show that vehicles with Xichai, Yuchai or ZGZQ engines performed poorly in particulate emission control. Three vehicles equipped with the model 'CA6DLD-29E6' of Xichai engine, two vehicles equipped with the model 'YSC06245-60A' of Yuchai engine, and one vehicle equipped with the model 'MC13H.51-61' of ZGZQ engine did not meet the PN emission standards. Since the PN EFs of the unqualified vehicles were one or two orders of magnitudes larger than the qualified vehicles, the PN EFs of the qualified vehicles were illustrated in the zoomed subfigures for visual contrast.

3.2.2. Emission analyses by vehicle feature group

The EFs of qualified and unqualified vehicles by vehicle features of EGR group, mileage group and tractive tonnage group are compared in this section, and the results are presented in the three columns of subfigures in Fig. 3. It is noted that the EFs of unqualified vehicles were higher than emission qualified vehicles in all feature groups except for those with no samples of emission unqualified vehicles, *e.g.*, mileage group of '>80,000 km' and tractive tonnage group of '40 t' for NOx EFs, mileage groups of '<20,000 km' and '>80,000 km' for PN EFs. For emission qualified vehicles, the emissions show the following characteristics. EGR-equipped vehicles resulted in lower NOx EFs but higher PN EFs. NOx EFs increased with the increase of mileages, and PN EFs of the vehicles with mileages >40,000 km were higher than those with mileages <40,000 km. In terms of tractive tonnage, vehicles of '34 t' group generated lower NOx EFs than those of '25 t' and '40 t' groups, while vehicle PN EFs show a descending tendency with the increase of tractive tonnage. For emission unqualified vehicles, higher NOx and PN EFs were obtained by vehicle groups with EGR unequipped, larger mileage or larger tractive tonnage.

3.3. Comparative emission results by operating conditions

This section provides the in-depth analysis of vehicle emission characteristics with respect to operating conditions. Vehicle emissions were calculated and compared for qualified and unqualified vehicles by road type and operating mode, respectively.

3.3.1. Emission analyses by road type

To compare the road type based vehicle EFs, the instantaneous EFs of each vehicle were first divided into three categories that correspond to road types of urban roads (UR), suburban roads (SU) and freeways (FW). The operation duration of each vehicle on the three types of roads are summarized in Table S9. Then the mean values were calculated for each category. The average EFs and standard errors for qualified and unqualified vehicles are illustrated with the bar plots and error bars in Fig. 4, respectively.

As illustrated in Fig. 4, the NOx EFs of qualified vehicles were lower on UR and FW, but higher on SU than those of unqualified vehicles. And the PN EFs of qualified vehicles were lower than those of unqualified vehicles on all types of roads. It seems counterintuitive that qualified vehicles generated higher NOx EFs than unqualified vehicles on suburban roads. Thus, we looked into the potential reasons by checking on the suburban road-based NOx EFs vehicle by vehicle, and the corresponding EF values are presented in Table S10 in the Supplementary materials. It turns out that there were two vehicles, *i.e.*, vehicle #11 and vehicle #17, that generated obviously higher NOx EFs than the rest vehicles. With further exploration on their trajectories of the instantaneous NOx EFs and the comparison with other qualified vehicles, we found that there were several spikes of extremely high NOx EFs for vehicle #11 and vehicle #17, as shown in Fig. S4 in the Supplementary materials. The spikes increased the average suburban road-based NOx EFs of the two vehicles without impacting the determination results on emission qualification, as the determination was based on percentiles and thus was robust to outliers of spikes. As a result, the high NOx EFs of the two vehicles increased the average NOx EFs for qualified vehicles. However, we cannot conclude that qualified vehicles generated significantly higher NOx EFs than unqualified vehicles on suburban roads as the variances of NOx EFs from qualified vehicles are obviously larger than those from unqualified vehicles, which were presented in Table S11 in the Supplementary materials. In addition, the standard errors were large for most cases in Fig. 4, *e.g.*, NOx EFs of unqualified vehicles on UR and PN EFs of unqualified vehicles on SU. Therefore, statistical tests including MWU test and KWANOVA were conducted to further investigate the differences of the EF distributions and mean values, respectively. The sample sizes were between 5 and 16 in this research, and the p values are reported in Table 1, in which statistically significant results ($p < 0.05$) are highlighted in bold and notation * means $p < 0.0001$. The MWU tests demonstrate that the NOx EFs of qualified and unqualified vehicles were significantly different on FW. And the PN EFs of qualified and unqualified vehicles were significantly different on both SU and FW. The KWANOVA results reveal that both qualified and unqualified vehicles show significantly different NOx EFs on different types of roads. Road type significantly impacted both types of PN EFs of qualified vehicles and PN EF_w of unqualified vehicles as well. As for the numerical comparison among road type based EFs, qualified vehicles generated the highest NOx and PN EFs on UR and FW, respectively. The potential

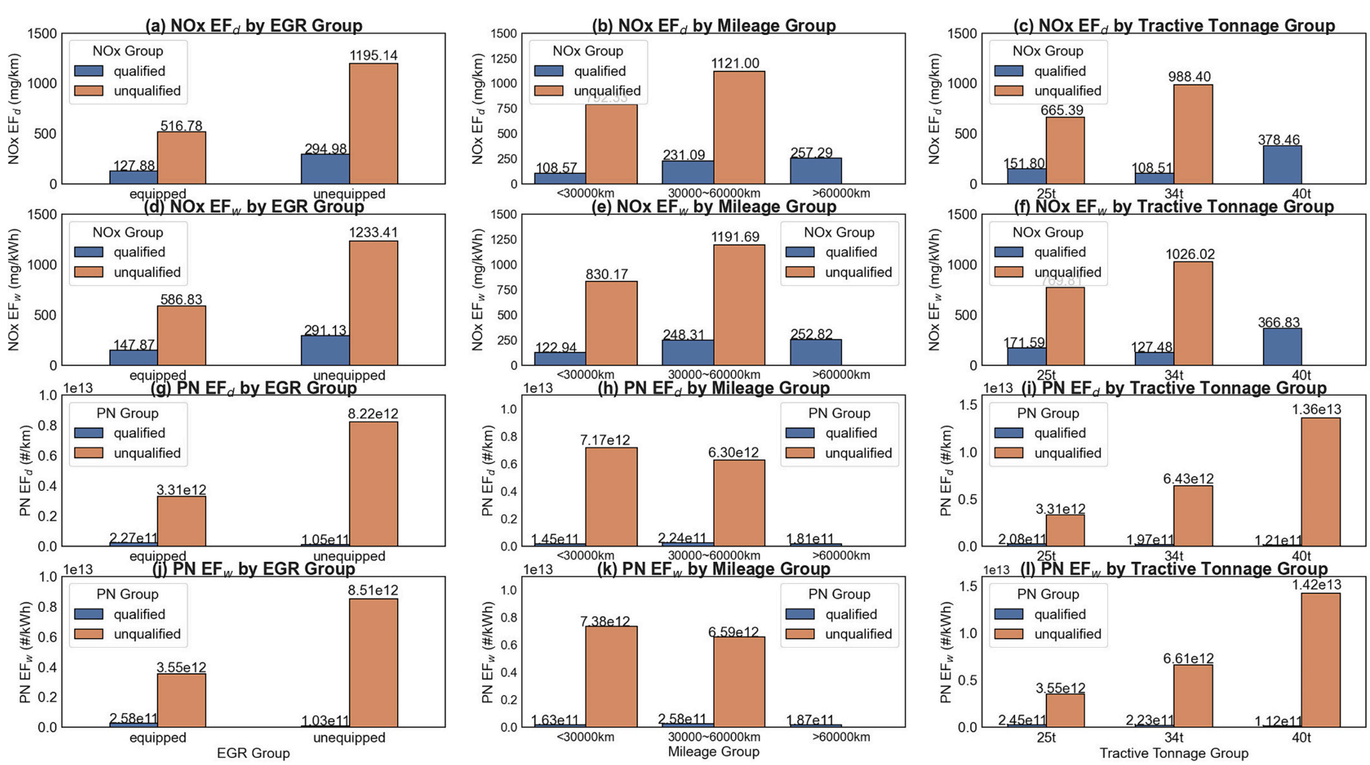


Fig. 3. Emission factors of qualified and unqualified vehicles by vehicle feature groups.

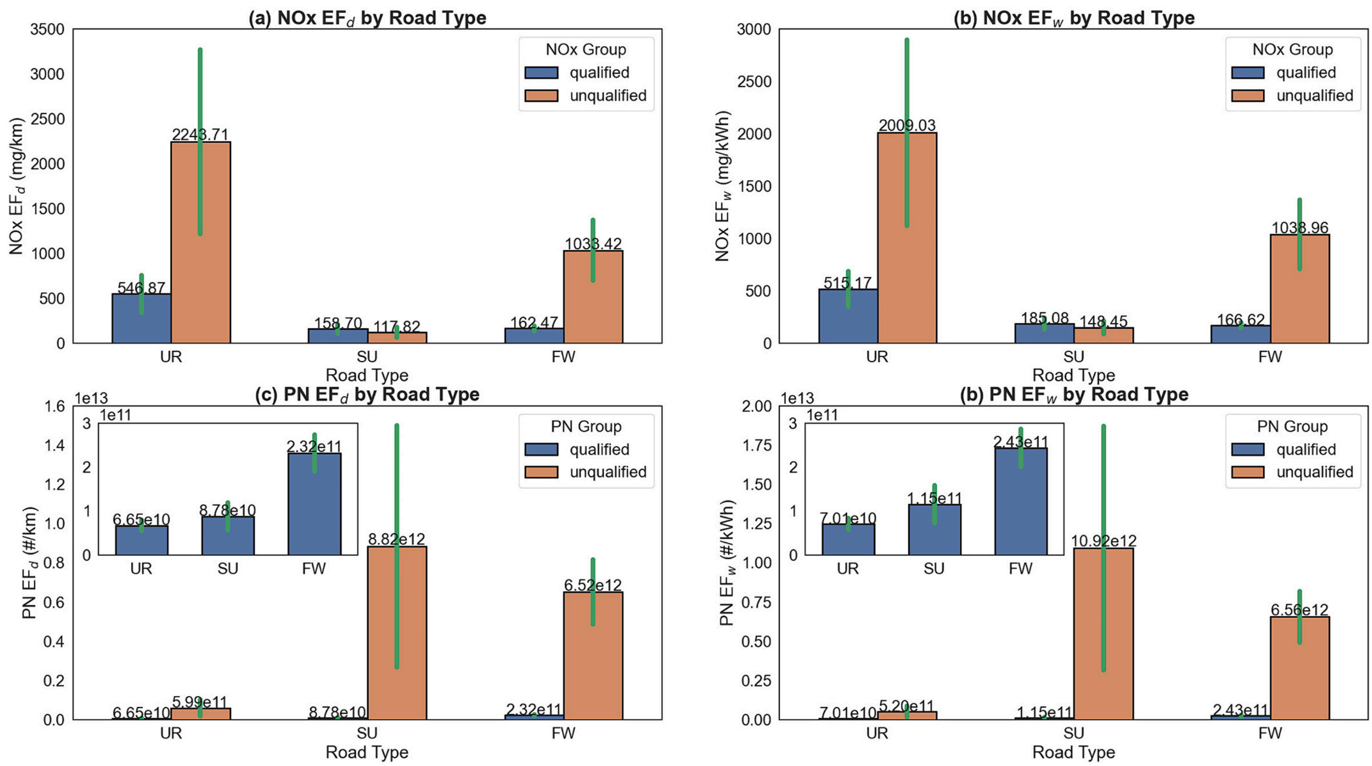


Fig. 4. Emission factors of qualified and unqualified vehicles by road type.

Table 1
Statistical test results for road type based EFs (bold numbers represent $p < 0.05$ and notation * means $p < 0.0001$).

Method	Group	p-Value NOx EF$_d$	NOx EF$_w$	PN EF$_d$	PN EF$_w$
MWU test for EFs of qualified and unqualified vehicles	UR	0.109	0.153	0.622	0.791
	SU	0.660	0.842	***0.018***	***0.014***
	FW	***9.8e^{-5*}***	***9.8e^{-5*}***	***3.7e^{-5*}***	***3.7e^{-5*}***
KWANOVA for road type based EFs	Qualified	***0.003***	***0.010***	***0.002***	***0.003***
	Unqualified	***0.019***	***0.019***	0.052	***0.034***

reasons may include: 1) when vehicles were running on urban roads or in low-speed operations, the exhaust temperature was usually low, which prevented the SCR system from working effectively and thus the NOx emissions were high; 2) when vehicles were running on freeways or in high-speed operations, both the cylinder temperature and the injected fuel increased, which could boost the fuel vaporization and result in the lack of oxygen, and thus the particle formation was promoted. While for unqualified vehicles, the highest NOx and PN EFs were generated on UR and SU, respectively. One hypothesis was that the EFs of unqualified vehicles varied greatly on suburban roads, and the average emission levels were increased due to the notably high value(s).

3.3.2. *Emission analyses by operating mode*

To analyze the emission characteristics of vehicles under different operating modes, the EFs of qualified and unqualified vehicles under 22 operating mode bins were also calculated, respectively. Vehicle operation durations under each operating mode bin are provided in Table S12. It is noted that the idling mode was omitted for EF$_d$ analysis due to the issue of zero divisor (*i.e.*, idling distance), and several bins (*i.e.*, bin 11, 17, 18 and 22) were neglected in the analyses due to the small sample sizes (<10 for most vehicles). The average EFs and standard errors are depicted with the bar plots and error bars in Fig. 5. Since the extremely high PN EFs of the unqualified vehicles in subfigure (c) and (d) make it difficult to see the PN EFs of the qualified vehicles, the PN EFs of the qualified vehicles were individually presented in subfigure (e) and (f) additionally. Statistical test results are summarized in Table 2, in which statistically significant results ($p < 0.05$) are highlighted in bold and notation * means $p < 0.0001$.

Subfigure (a)–(d) in Fig. 5 demonstrate that both the NOx EFs and PN EFs of unqualified vehicles were higher than those of qualified vehicles. The statistical significance tests of MWU in Table 2 further demonstrated that the differences of the NOx EFs were statistically significant in most of the medium and high speed operating modes, while the PN EFs were significantly different in all of the medium and high speed operating modes and most of the low speed modes. Therefore, unqualified vehicles generated significantly higher NOx EFs under medium to high speed modes, and higher PN EFs under all operating modes with positive VSPs than qualified vehicles.

Fig. 5 also demonstrates that low speed operating modes resulted in higher NOx EFs but lower PN EFs than medium and high speed operating modes for both qualified and unqualified vehicles. The statistical significances were further evaluated by applying KWANOVA and Spearman's rank tests on EFs and operating mode bins (denoted with all modes in Table 2). KWANOVA reveals that both the NOx EFs and PN EFs were significantly different among the investigated operating modes for qualified vehicles, and the KWANOVA results for unqualified vehicles were significant on NOx EF$_d$, PN EF$_d$ and PN EF$_w$ as well. For the significantly different EFs, Spearman's rank correlation tests show that there was a moderate negative correlation (with ρ value between −0.6 and − 0.4) between the NOx EF$_w$ and operating mode bins for qualified vehicles, a weak negative correlation (with ρ value between −0.4 and − 0.2) between the NOx EF$_w$ and operating mode bins for unqualified vehicles, a moderate positive correlation (with ρ value between 0.4 and 0.6) between PN EF$_d$ and operating modes for both qualified and unqualified vehicles, and a moderate positive correlation between the PN EF$_w$ and operating modes for unqualified vehicles. This demonstrated that qualified vehicles generated significantly higher NOx EF$_d$ and NOx

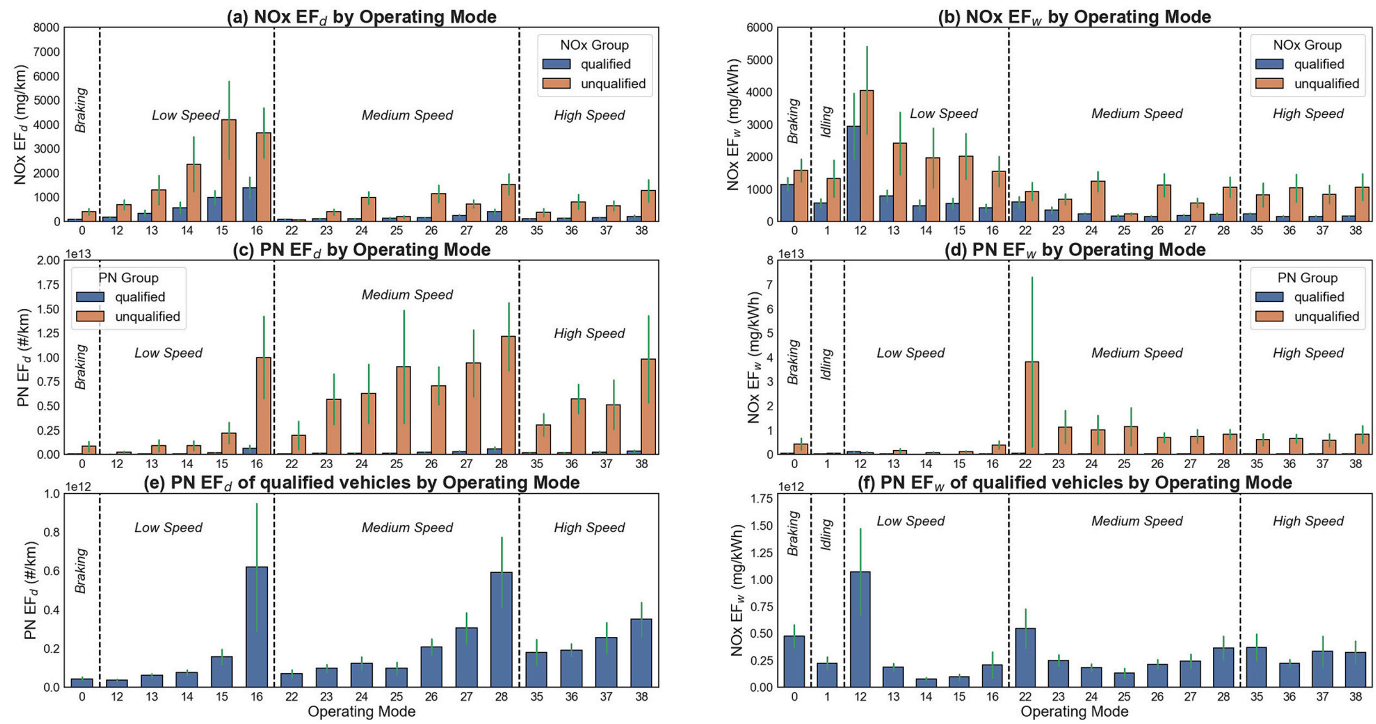


Fig. 5. Emission factors of qualified and unqualified vehicles by operating mode.

EF$_w$ and lower PN EF$_d$ in low speed modes than high speed modes, and unqualified vehicles generated significantly lower PN EF$_d$ and PN EF$_w$ in low speed modes than high speed modes.

To investigate the impact of VSP on vehicle emissions, KWANOVA and Spearman's rank coefficient tests were also conducted for conditional operating mode bins on each speed interval, *i.e.*, low speed modes, medium speed modes and high speed modes, respectively. KWANOVA tests show that both NOx EFs and PN EFs of qualified vehicles were significantly different among operating mode bins under low and medium speed separately. Unqualified vehicles did not generate consistent results with qualified vehicles as the out-of-standard emissions might be caused by various reasons, and they only generate significantly different NOx EF$_d$ and NOx EF$_w$ among medium speed mode bins, and significantly different PN EF$_d$ among low speed mode bins. When the vehicle speed was higher than 80 km/h, no significant differences on operating mode based EFs were detected with our data. The monotonicity of the identified significant differences was tested with Spearman's rank coefficients. For qualified vehicles, the NOx EF$_d$ increased with operating mode bins in low and medium speed modes, respectively, while the NOx EF$_w$ show a descending tendency in the two speed intervals. The PN EF$_d$ was positively associated with the operating mode bins in low and medium speed intervals as well, and the PN EF$_w$ was negatively associated with low speed operating mode bins. For unqualified vehicles, the NOx EF$_d$ and PN EF$_d$ show a significant increasing tendency with operating mode bins in medium and low speed intervals, respectively. The results demonstrated that both the NOx EF$_d$and PN EF$_d$ increased with VSP in low to medium speeds, while the EF$_w$ show the opposite tendency (for qualified vehicles only).

3.4. OBD NOx data reliability and accuracy

In this section, the reliability and accuracy of OBD NOx data were analyzed in term of data reliability and consistency with PEMS measurements. Among the 21 test vehicles, OBD NOx data of 4 vehicles were not available, *i.e.*, vehicle #4, #9, #14 and #15. The first two vehicles did not transmit remote OBD data to VECC, while the latter two vehicles were not connected to the network during the road tests. For the rest 17 vehicles that successfully reported remote OBD data, the data quality was first evaluated by analyzing their integrity and validity, and then the OBD NOx concentrations were compared with PEMS NOx data in terms of PCC and mean relative errors (MRE, the average of the relative errors between OBD NOx and PEMS measurements). Since cycle-averaged or trip-averaged NOx concentrations are usually used in real-world regulations, we adopt the same method of averaging NOx concentrations by time sliding windows as in the research of Zhang et al. (2020), and the window duration varied among 1 s (*i.e.*, the instantaneous concentration), 10s, 20s, 30s and 60s. The metrics of OBD NOx data quality including the ratio of missing or invalid entries, PCC and MRE for each vehicle are summarized in Table 3.

According to Table 3, the data consistencies on instantaneous concentrations varied greatly with PCCs ranging from 0.054 to 0.979. There are three, five, five and two vehicles with '$0.8 \leq$ PCC_1 s < 1', '$0.5 \leq$ PCC_1 s < 0.8', '$0.3 \leq$ PCC_1 s < 0.5' and '$0 <$ PCC_1 s < 0.3', corresponding to vehicle categories with high, moderate, weak and no correlation between OBD NOx and PEMS NOx, respectively. The scatter plots and correlation results of the OBD NOx and PEMS NOx data for four test vehicles that belong to the aforementioned four categories respectively were presented in Fig. S5 in the Supplementary material. For consistency analyses of window averaged concentrations, as window samples are impacted by the missing or invalid entry ratios, we first divided vehicles into three groups with the ratios of missing or invalid entries fell in the intervals of '>60 %', '20 % ~ 50 %' and '10 % ~ 20 %', respectively, and then analyze PCCs of window averaged concentrations for vehicles in each group. There are five vehicles, *i.e.*, vehicle #5, #12, #13, #16 and #18, that had >60 % of missing or invalid OBD NOx entries during road tests. For these vehicles, the PCCs between OBD NOx and PEMS NOx were low for all sliding windows except for the 10s window for vehicle #18, in which the high PCC of 0.976 might not represent the true correlation since it was generated by only five window-averaged samples. In addition, for this vehicle, the sample size reduced to 0 for sliding windows of 20s, 30s and 60s because of the large amount of missing or invalid entries. There are six vehicles, *i.e.*, vehicle #1, #6, #7, #17, #19 and #21, that had 20 % ~ 50 % missing or invalid data entries. Among these vehicles, the OBD NOx data of vehicle #6

Table 2
Statistical test results for operating mode based EFs (bold numbers represent $p < 0.05$ and notation * means $p < 0.0001$).

Method	Group		p-Value (ρ) NOx EF_d	NOx EF_w	PN EF_d	PN EF_w
MWU test for EFs between qualified and unqualified vehicles	Braking mode 0		0.050	0.208	***0.045***	0.235
	Idling mode 1		/	0.313	/	0.154
	Low speed (LS) modes	12	***0.006***	0.179	***0.029***	0.470
		13	0.091	0.130	0.055	0.161
		14	0.130	0.130	***0.008***	***0.014***
		15	0.109	0.075	***0.008***	***0.005***
		16	0.109	***0.032***	***0.005***	***0.002***
	Medium speed (MS) modes	21	0.062	0.179	***0.001***	***0.006***
		23	***0.004***	***0.032***	$\boldsymbol{1.5e^{-4}}$	$\boldsymbol{2.6e^{-4}}$
		24	$\boldsymbol{9.8e^{-5*}}$	***0.001***	***0.002***	$\boldsymbol{4.4e^{-4}}$
		25	0.153	0.109	$\boldsymbol{3.7e^{-5*}}$	$\boldsymbol{7.4e^{-5*}}$
		26	$\boldsymbol{9.8e^{-5*}}$	$\boldsymbol{9.8e^{-5*}}$	$\boldsymbol{2.6e^{-4}}$	$\boldsymbol{1.5e^{-4}}$
		27	***0.008***	***0.008***	$\boldsymbol{4.4e^{-4}}$	$\boldsymbol{4.4e^{-4}}$
		28	***0.008***	$\boldsymbol{3.9e^{-4}}$	***0.002***	$\boldsymbol{7.0e^{-4}}$
	High speed (HS) modes	35	***0.019***	***0.008***	$\boldsymbol{1.5e^{-4}}$	$\boldsymbol{1.5e^{-4}}$
		36	$\boldsymbol{9.8e^{-5*}}$	$\boldsymbol{9.8e^{-5*}}$	$\boldsymbol{3.7e^{-5*}}$	$\boldsymbol{3.7e^{-5*}}$
		37	***0.002***	$\boldsymbol{3.9e^{-4}}$	***0.006***	***0.002***
		38	***0.001***	$\boldsymbol{2.0e^{-4}}$	$\boldsymbol{1.5e^{-4}}$	$\boldsymbol{1.5e^{-4}}$
KWANOVA for operating mode based EFs	Qualified	All modes	$\boldsymbol{4.8e^{-15*}}$	$\boldsymbol{1.8e^{-16*}}$	$\boldsymbol{4.2e^{-12*}}$	$\boldsymbol{1.4e^{-7*}}$
		LS modes	$\boldsymbol{1.0e^{-4}}$	$\boldsymbol{8.0e^{-4}}$	$\boldsymbol{2.0e^{-4}}$	$\boldsymbol{1.8e^{-5*}}$
		MS modes	$\boldsymbol{5.3e^{-6*}}$	$\boldsymbol{7.1e^{-4}}$	$\boldsymbol{1.3e^{-5*}}$	***0.017***
		HS modes	0.321	0.085	0.088	0.786
	Unqualified	All modes	***0.001***	0.062	$\boldsymbol{6.0e^{-4*}}$	***0.001***
		LS modes	0.311	0.511	***0.033***	0.526
		MS modes	$\boldsymbol{3.1e^{-4}}$	***0.008***	0.270	0.951
		HS modes	0.080	0.733	0.513	0.857
Spearman's rank correlation test for operating mode based EFs	Qualified	All modes	0.434 (−0.05)	$\boldsymbol{5.4e^{-24*}}$ ***(−0.55)***	$\boldsymbol{1.6e^{-14*}}$ ***(0.46)***	0.678 (0.03)
		LS modes	$\boldsymbol{2.4e^{-7*}}$ ***(0.54)***	$\boldsymbol{5.8e^{-5*}}$ ***(−0.43)***	$\boldsymbol{6.3e^{-7*}}$ ***(0.54)***	$\boldsymbol{1.9e^{-6*}}$ ***(−0.52)***
		MS modes	$\boldsymbol{1.7e^{-9*}}$ ***(0.53)***	$\boldsymbol{7.5e^{-5*}}$ ***(−0.36)***	$\boldsymbol{8.2e^{-9*}}$ ***(0.53)***	0.12 (−0.15)
		HS modes	0.088 (0.21)	***0.025 (−0.28)***	***0.012 (0.32)***	0.54 (−0.08)
	Unqualified	all modes	0.71 (0.04)	***0.005 (−0.29)***	$\boldsymbol{4.0e^{-8}}$ ***(0.51)***	$\boldsymbol{3.3e^{-7*}}$ ***(0.47)***
		LS modes	***0.035 (0.42)***	0.10 (−0.33)	$\boldsymbol{4.7e^{-4}}$ ***(0.60)***	0.16 (0.26)
		MS modes	$\boldsymbol{3.7e^{-5*}}$ ***(0.64)***	0.41 (−0.14)	$\boldsymbol{7.9e^{-3}}$ ***(0.40)***	0.86 (−0.03)
		HS modes	***0.021 (0.51)***	0.56 (0.14)	0.30 (0.22)	0.98 (0.005)

show extremely high correlation with PEMS data even on instantaneous concentrations. For the other five vehicles, the PCCs on instantaneous concentrations were low, but high PCCs of around 0.85 could be obtained when using 20s-window averaged concentrations. The rest six vehicles, *i.e.*, vehicle #2, #3, #8, #10, #11 and #20, show better data reliability as the ratios of missing or invalid entries were between 10 % and 20 %. The OBD data of all these vehicles except for vehicle #8 and vehicle #20 show high correlations of 0.756– 0.897 with PEMS measurements on instantaneous NOx concentrations. And when using 10s-window averaged concentrations, the PCC could be increased to 0.858 ∼ 0.959. For vehicle #8 and vehicle #20, longer time windows were required to obtain high PCCs, *e.g.*, PCCs of 0.815 and 0.941 could be obtained with 60s sliding windows. Besides the correlation analyses, a preliminary exploration about the linear relationship between the OBD NOx data and PEMS measurements was also conducted in this research, and the R^2 of the linear regression models for each vehicle and on length-vary window-averaged NOx concentrations are summarized in Table S13 in the Supplementary materials. Similar to the results of correlations, the linear relationship varied greatly among vehicles, and the value of R^2 was lower than 0.80 for most cases. These results demonstrated that although OBD NOx shows good correlations with the PEMS measurements for most vehicles, they are not linearly correlated

Table 3
Statistical metrics of data quality.

Vehicle no.	NOx group	Ratio of missing or invalid entries	PCC of window-based concentrations 1 s	10 s	20 s	30 s	60 s	MRE 20 s
#1	Qualified	45.25 %	0.435	0.742	0.863	0.904	0.975	−24.22 %
#2	Unqualified	12.27 %	0.785	0.858	0.883	0.893	0.909	159.67 %
#3	Unqualified	16.46 %	0.857	0.915	0.934	0.943	0.958	417.11 %
#5	Qualified	72.40 %	0.387	0.571	0.599	0.509	0.199	−15.76 %
#6	Unqualified	39.77 %	0.975	0.979	0.981	0.982	0.982	362.11 %
#7	Qualified	40.54 %	0.611	0.832	0.849	0.782	0.606	−23.43 %
#8	Qualified	12.21 %	0.432	0.675	0.741	0.774	0.815	−73.67 %
#10	Qualified	10.51 %	0.756	0.908	0.935	0.953	0.980	−64.99 %
#11	Qualified	12.41 %	0.897	0.959	0.975	0.981	0.988	−38.06 %
#12	Qualified	61.16 %	0.265	0.358	0.380	0.419	0.304	88.58 %
#13	Qualified	63.59 %	0.054	0.087	0.036	0.029	−0.093	−67.96 %
#16	Qualified	65.23 %	0.307	0.510	0.601	0.653	0.542	112.11 %
#17	Qualified	32.95 %	0.512	0.703	0.860	0.908	0.943	−21.13 %
#18	Qualified	70.58 %	0.433	0.976	/	/	/	/
#19	Qualified	33.98 %	0.608	0.784	0.857	0.863	0.854	−42.37 %
#20	Unqualified	14.96 %	0.438	0.541	0.650	0.828	0.941	−47.47 %
#21	Qualified	24.78 %	0.527	0.701	0.870	0.946	0.981	−0.03 %

in most cases. Therefore, OBD NOx data cannot be directly used as the substitute of the real NOx emissions, and additional efforts are required to develop real NOx emissions prediction models based on the obtained OBD NOx data.

On the metric of MRE, the complete results with varying lengths of windows are provided in Table S14 in the Supplementary materials. Fluctuations among different windows were not large for most vehicles, except for those with dramatically changed samples of window-averaged NOx concentrations, *e.g.*, vehicle #5 and #21, for which the sample sizes on instantaneous concentrations were 1853 (the small samples were caused by high ratio of missing or invalid entries) and 6046, respectively, but were reduced to 106 and 107 on 60s-window averaged concentrations due to large amounts of inconsecutive entries. Since 20s-window averaged OBD NOx concentrations show good consistency with PEMS measurements, they were used in the analyses hereinafter. According to Table 3, NOx unqualified vehicles could report notably higher concentrations than PEMS measurements. For example, the MRE of vehicle #2, #3 and #6 were 159.67 %, 417.11 % and 362.11 %, respectively. Most qualified vehicles reported lower OBD NOx concentrations than PEMS results. More specifically, ten of the twelve qualified vehicles generated MREs between −73.67 % and −0.03 %. The rest two qualified vehicles (*i.e.*, vehicle #12 and #16) that resulted in positive MREs of 88.58 % and 112.11 % had >60 % of missing or invalid entries, which might introduce bias in the calculated MREs. These MRE results show that the accuracy of the OBD NOx data should be further improved as they could be far apart (with absolute values larger than 50 %) from 0 % for nearly half of the test vehicles. One potential method is to delve into the analysis of potential confidence intervals to enhance the accuracy and reliability of OBD data interpretation. For example, for a specific OBD NOx reading, there might be multiple PEMS measurements (real NOx emissions) conditioned on specific operating status (speed, acceleration, VSP, operating mode, *etc.*). Statistical analysis can be applied to the set of the real NOx emissions to construct a "confidence interval" for each conditional OBD reading. Thus, the confidence intervals can be constructed for all the OBD NOx readings and the operating statuses. By setting some criteria for all the confidence intervals (*e.g.*, to minimize the total difference between the confidence intervals and the truth emissions), the corresponding confidence intervals could be solved and obtained.

4. Conclusions and future work

In this research, the NOx and PN emissions of 21 China VI HDDTs were evaluated with on-road tests in Shenzhen, China. In addition, the NOx data that transmitted to the VECC by vehicle OBD were also evaluated with the PEMS measurements of NOx emissions. In the evaluations, fundamental analysis with respect to the on-road conformity of vehicle emissions was conducted firstly, which provides general knowledge of the real-world implementation effects of the recent China VI emission standard to regulators. After that, emission factors in terms of vehicle features (*i.e.*, engine model, EGR group, mileage group, and tractive tonnage group) and operating conditions (*i.e.*, road type, and operating mode) were calculated and compared. Specific vehicle characteristics and operational conditions that influence emissions performances were identified through meticulous statistical analyses, which are valuable for the potential improvements of emission regulations and compliance mechanisms. In the end, the reliability and accuracy of the OBD NOx data were analyzed in terms of their integrity and consistency with PEMS measurements, which holds great significance to the further development of using OBD NOx data to monitor and regulate vehicle emissions.

The research uncovers that a notable fraction, *i.e.*, 38.1 %, of the vehicles failed to comply with the regulatory emissions standards. Therefore, more efforts are demanded from the regulators to make sure the China VI emission standard work correctly and widely. There are several directions that regulators could put more focuses on. On the one hand, the regulators could pay close attention to certain diesel engine models that were demonstrated to be more likely to fail the emission tests. On the other hand, specific operating conditions where unqualified vehicles show significantly different emission features from the qualified ones could be the regulatory focuses as well. More specifically, the qualified vehicles show clear emission patterns in general. Their EFs disparities were significant on different types of road segments and under different operating modes, and monotonic patterns between EFs and operating modes were witnessed for most cases. However, unqualified vehicles show different emission features. There were significant NOx EFs differences on freeways and under medium to high-speed operating modes for qualified and unqualified vehicles. Significantly different PN EFs were identified on suburban roads, freeways and under all operating modes with positive VSPs as well. Therefore, emissions under freeway operations or medium to high-speed operating modes could be utilized to effectively differentiate both NOx and PN emission unqualified vehicles from the qualified ones, and thus regulators could put more emphasis on emission monitoring and regulations under these scenarios. For the evaluations of the OBD NOx data, results reveal that the data quality of the OBD NOx data was not satisfactory in terms of their reliability and accuracy, either. Around 43 % of the 21 test vehicles did not provide reliable OBD NOx data, including 4 vehicles that did not report remote data and 5 vehicles that transmitted data with high ratios (>60 %) of missing or invalid entries. The consistency between OBD NOx concentrations and PEMS measurements was not universally satisfactory on instantaneous emissions for all test vehicles, but when the ratio of missing or invalid entries was lower than 50 %, 20s-window averaged concentrations from OBD and PEMS show high correlations over 0.85 for most vehicles. Therefore, window averaged OBD NOx concentrations might be used as substitutes of PEMS measurements in real-world analyses and regulations. However, the accuracy of OBD NOx data should be further enhanced as the MREs between OBD NOx and PEMS NOx varied a lot and around half of the MREs had absolute values larger than 50 %. In addition, the OBD NOx concentrations were generally lower than PEMS measurements as most of the MREs were negative. Hence, regulators could pay closer attention to the vehicles with OBD NOx concentrations larger than the specified limit as the actual NOx concentrations might be even higher.

For future work, the mandatory implementation of OBD in China VI HDDTs made it easier to obtain the on-road NOx emissions. However, our research demonstrated that although OBD NOx data show good correlations with PEMS measurements, the linear relationship could be poor for most cases. Therefore, the OBD NOx data cannot be treated as the real NOx emissions. In the future, it is indispensable and valuable to develop real NOx prediction models based on the OBD NOx data, along with some emission-predictive explanatory variables such as vehicle specific factors and environmental parameters. In addition to the development of prediction models, enhancing the accuracy of OBD data interpretation is another promising research direction. Along this direction, researchers can investigate methods to construct effective confidence intervals to denote the uncertainty associated with OBD readings, and thus provide regulators a clearer understanding of the data's precision and guide improvements in emissions monitoring methodologies.

CRediT authorship contribution statement

Weixia Li: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Zhurong Dong:** Writing – review & editing, Supervision, Resources. **Ling Miao:** Validation, Investigation, Data curation. **Guoyuan Wu:** Writing – review & editing, Methodology. **Zhijun Deng:** Supervision, Funding acquisition. **Jianfeng Zhao:** Writing – review & editing, Funding acquisition. **Wenwei Huang:** Supervision, Resources, Project administration.

Declaration of competing interest

The authors declare that they have no financial or personal relationships that could inappropriately influence or bias the content of the paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This research was funded by Shenzhen Polytechnic University Research Fund (grant number: 6024310009K) and Research Projects of Department of Education of Guangdong Province (grant numbers: 2023ZDZX3078, 2021ZDZX1140).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2024.172427.

References

Ayodhya, A.S., Narayanappa, K.G., 2018. An overview of after-treatment systems for diesel engines[J]. Environ. Sci. Pollut. Res. 25, 35034–35047.
Berger, V.W., Zhou, Y.Y., 2014. Kolmogorov–smirnov test: Overview[J]. Wiley statsref: Statistics reference online. https://doi.org/10.1002/9781118445112.stat06558.
California Air Resources Board, 2020. Staff report: initial statement of reasons, for public hearing to consider the proposed heavy-duty engine and vehicle omnibus regulation and associated amendments, June. https://ww3.arb.ca.gov/regact/2020/hdomnibuslownox/isor.pdf (accessed on December 10, 2023).
Department for Transport U.K., 2023. Transport and environment statistics: 2023. https://www.gov.uk/government/statistics/transport-and-environment-statistics-2023/transport-and-environment-statistics-2023 (accessed on December 10, 2023).
Dhital, N.B., Wang, S.X., Lee, C.H., et al., 2021. Effects of driving behavior on real-world emissions of particulate matter, gaseous pollutants and particle-bound PAHs for diesel trucks[J]. Environ. Pollut. 286, 117292.
EMISIA, 2023. European Road Transport & Emissions (ERTE) Trends Report 2023. https://erte.report/ertefiles/ERTE2023_Complete_Downloable.pdf (accessed on December 10, 2023).
European Environment Agency (EEA), 2022. Health impacts of air pollution in Europe, 2022. https://www.eea.europa.eu/publications/air-quality-in-europe-2022/health-impacts-of-air-pollution (accessed on December 10, 2023).
Ge, Y., Hou, P., Lyu, T., et al., 2023b. Machine learning-aided remote monitoring of NOx emissions from heavy-duty diesel vehicles based on OBD data streams[J]. Atmosphere 14 (4), 651.
Ge, Z., Li, W., Wang, J., et al., 2023a. Real-road NOx and CO2 emissions of city and highway China-6 heavy-duty diesel vehicles[J]. J. Environ. Sci. 2023 https://doi.org/10.1016/j.jes.2023.11.018.
Hao, L., Ren, Y., Lu, W., et al., 2023. Assessment of heavy-duty diesel vehicle NOx and CO2 emissions based on OBD data[J]. Atmosphere 14 (9), 1417.
He, L., Zhang, S., Hu, J., et al., 2020. On-road emission measurements of reactive nitrogen compounds from heavy-duty diesel trucks in China[J]. Environ. Pollut. 262, 114280.
Jiménez-Palacios, J.L., 1999. Understanding and Quantifying Motor Vehicle Emissions with Vehicle Specific Power and TILDAS Remote Sensing[J]. Massachusetts Institute of Technology. http://hdl.handle.net/1721.1/44505.
Koupal, J., Landman, L., Nam, E., et al., 2005. MOVES2004 energy and emission inputs draft report[J]. US Environmental Protection Agency Report No. EPA420-P-05-003.
Kruskal, W.H., Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis[J]. J. Am. Stat. Assoc. 47 (260), 583–621.
Li, X., Ai, Y., Ge, Y., 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[J]. Sci. Total Environ. 811, 152311.
Lv Z, Zhang Y, Ji Z, et al. A real-time NOx emission inventory from heavy-duty vehicles based on on-board diagnostics big data with acceptable quality in China[J]. J. Clean. Prod., 2023, 422: 138592. https://doi.org/https://doi.org/10.1016/j.jclepro.2023.138592.
Lyu, L., Ji, Z., Yin, H., et al., 2023. NOx emission deterioration in modern heavy-duty diesel vehicles based on long-term real driving measurements[J]. Environ. Res. 116396.
Ma, T., Li, C., Luo, J., et al., 2024. In-use NOx and black carbon emissions from heavy-duty freight diesel vehicles and near-zero emissions natural gas vehicles in California's San Joaquin Air Basin[J]. Sci. Total Environ. 907, 168188.
Mann, H.B., Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other[J]. Ann. Math. Stat. 50–60.
McCaffery, C., Zhu, H., Tang, T., et al., 2021. Real-world NOx emissions from heavy-duty diesel, natural gas, and diesel hybrid electric vehicles of different vocations on California roadways[J]. Sci. Total Environ. 784, 147224.
MEE (Ministry of Ecology and Environment of the People''s Republic of China) and SAMR (State Administration for Market Regulation of the People''s Republic of China), 2018. Available at: https://www.mee.gov.cn/ywgz/fgbz/bz/bzwb/dqhjbh/dqydywrwpfbz/201807/W020180703397013304274.pdf (accessed on December 10, 2023) (in Chinese).
Ministry of Ecology and Environment of the People's Republic of China (MEE), 2017. Measurement method and technical specification for PEMS test of exhaust pollutants from heavy-duty diesel and gas fueled vehicles in China. Available at: https://www.mee.gov.cn/ywgz/fgbz/bz/bzwb/dqhjbh/dqydywrwpfbz/201709/W020170921587147401788.pdf (accessed on December 10, 2023) (in Chinese).
Ministry of Ecology and Environment of the People''s Republic of China (MEE), China Mobile Source Environmental Management Annual Report, 2023. Available at: https://www.mee.gov.cn/hjzl/sthjzk/ydyhjgl/202312/W020231211531753967096.pdf (accessed on January 15, 2024) (in Chinese).
Papadopoulos, G., Ntziachristos, L., Tziourtzioumis, C., et al., 2020. Real-world gaseous and particulate emissions from euro IV to VI medium duty diesel trucks[J]. Sci. Total Environ. 731, 139137.
Pearson, K., 1895. Notes on regression and inheritance in the case of two parents. Proc. R. Soc. Lond. 58 (347–352), 240–242. https://doi.org/10.1098/rspl.1895.0041.
Sensors Inc.. Product of Semtech-Model ecostar plus. https://www.environmental-expert.com/products/semtech-model-ecostar-plus-gaseous-and-flow-measurement-81851 (Accessed on January 16, 2024).
Sensors Inc.. Particle number measurement SEMTECH CPN. https://sensors-inc.com/Products/SEMTECH/CPN (Accessed on January 16, 2024).
Shen, X., Lv, T., Zhang, X., et al., 2021. Real-world emission characteristics of black carbon emitted by on-road China IV and China V diesel trucks[J]. Sci. Total Environ. 799, 149435.
Spearman, C., 1904. The proof and measurement of association between two things[J]. Am. J. Psychol. 15 (1), 72–101.
US EPA (US Environmental Protection Agency): Development of emission rates for heavy-duty vehicles in the Motor Vehicle Emissions Simulator (Draft MOVES2009). EPA-420-P-09-005, Washington DC, USA, 2009.
Wang, J., Wang, R., Yin, H., et al., 2022. Assessing heavy-duty vehicles (HDVs) on-road NOx emission in China from on-board diagnostics (OBD) remote report data[J]. Sci. Total Environ. 846, 157209.
Wolfe, P., Davidson, K., Fulcher, C., et al., 2019. Monetized health benefits attributable to mobile source emission reductions across the United States in 2025[J]. Sci. Total Environ. 650, 2490–2498.
Wu, Y., Zhang, S.J., Li, M.L., et al., 2012. The challenge to NO x emission control for heavy-duty diesel vehicles in China[J]. Atmos. Chem. Phys. 12 (19), 9365–9379.
Wu, Y., Zhang, S., Hao, J., et al., 2017. On-road vehicle emissions and their control in China: a review and outlook[J]. Sci. Total Environ. 574, 332–349.
Xu, G., Shan, W., Yu, Y., et al., 2023. Advances in emission control of diesel vehicles in China[J]. J. Environ. Sci. 123, 15–29.
Zhang, B., Cheng, S., Lu, F., et al., 2022. Estimation of exposure and premature mortality from near-roadway fine particulate matter concentrations emitted by heavy-duty diesel trucks in Beijing[J]. Environ. Pollut. 311, 119990.
Zhang, S., Zhao, P., He, L., et al., 2020. On-board monitoring (OBM) for heavy-duty vehicle emissions in China: regulations, early-stage evaluation and policy recommendations[J]. Sci. Total Environ. 731, 139045.
Zhang, S., Fu, M., Zhang, H., et al., 2023. Emission control status, and future perspectives of diesel trucks in China[J]. J. Environ. Sci. https://doi.org/10.1016/j.jes.2023.06.010.
Zheng, X., Wu, Y., Jiang, J., et al., 2015. Characteristics of on-road diesel vehicles: black carbon emissions in Chinese cities based on portable emissions measurement[J]. Environ. Sci. Technol. 49 (22), 13492–13500.