Research on Remote Monitoring Data's Accuracy Verification for Heavy-duty Diesel Vehicles' Emission and Energy Consumption

Yu Liu, Hao Zhang, Hanzhengnan Yu, Jingyuan Li, Yanyan Yang, Xionghui Zou, Kunqi Ma, Yongkai Liang, Hang Xu, Xiaopan An, Xi Hu

***Abstract*—The application and development of remote monitoring big data for heavy-duty diesel vehicles (HDDVs) is a critical technology and a key focus for the effective supervision of vehicle emissions and energy consumption in the transportation sector. However, the lack of rigorous accuracy verification for these remote monitoring data significantly impedes their practical application and wider development. To address this challenge, this research investigates accuracy verification methods for remote monitoring data concerning HDDVs emissions and energy consumption. By integrating correlation analysis, linear regression modeling, the Bland-Altman method, and factor accounting theory, we propose a comprehensive framework for verifying HDDVs remote monitoring data accuracy. This framework evaluates the data from three perspectives: correlation strength, consistency, and suitability for application (specifically, emission factor calculation). Results demonstrate that key parameters within the remote monitoring dataset—vehicle speed, fuel flow, intake flow, and NOx emission rate—exhibit a strong positive correlation and excellent consistency with corresponding reference measurements. The coefficient of determination (R^2) exceeded 0.94 for all relevant comparisons. Furthermore, over 95.01% of data samples fell within the 95% consistency limits defined by the Bland-Altman method. Crucially, emission factors calculated using the remote monitoring data deviated by less than ±10% from actual measurements, while fuel consumption factors showed deviations within ±5%. These findings establish a robust scientific basis for the regulation of in-use vehicles.**

***Index Terms*—Heavy-duty diesel vehicle, Remote monitoring, Exhaust emissions, Fuel consumption, Data accuracy verification**

Manuscript received February 27, 2025; revised June 20, 2025.

Yu Liu is a senior engineer of China Automotive Technology and Research Center Co., Ltd., Tianjin 300300, China (e-mail: liuyu2016@catarc.ac.cn).

Hao Zhang is an engineer of China Automotive Technology and Research Center Co., Ltd., Tianjin 300300, China (corresponding author, phone: 86-15222647267; e-mail: zhanghao2023@catarc.ac.cn).

Hanzhengnan Yu is a senior engineer of China Automotive Technology and Research Center Co., Ltd., Tianjin 300300, China (e-mail: yuhanzhengnan@catarc.ac.cn).

Jingyuan Li is a senior engineer of China Automotive Technology and Research Center Co., Ltd., Tianjin 300300, China (e-mail: lijingyuan@catarc.ac.cn).

Yanyan Yang is a senior engineer of Beijing Municipal Ecological and Environmental Supervision Center, Beijing 100048, China (e-mail: 2294256364@qq.com).

Xionghui Zou is a senior engineer of CATARC Automotive Test Center (Tianjin) Co., Ltd., Tianjin 300300, China (e-mail: zouxionghui@catarc.ac.cn).

Kunqi Ma is an engineer of CATARC Automotive Test Center (Tianjin) Co., Ltd., Tianjin 300300, China (e-mail: makunqi@catarc.ac.cn).

Yongkai Liang is an engineer of CATARC Automotive Test Center (Tianjin) Co., Ltd., Tianjin 300300, China (e-mail: liangyongkai@catarc.ac.cn).

Hang Xu is an engineer of CATARC Automotive Test Center (Tianjin) Co., Ltd., Tianjin 300300, China (e-mail: xuhang@catarc.ac.cn).

Xiaopan An is a senior engineer of CATARC Automotive Test Center (Tianjin) Co., Ltd., Tianjin 300300, China (e-mail: anxiaopan@catarc.ac.cn).

Xi Hu is a senior engineer of CATARC Automotive Test Center (Tianjin) Co., Ltd., Tianjin 300300, China (e-mail: huxi@catarc.ac.cn).

I. Introduction

Heavy-duty diesel vehicles (HDDVs) constitute a major source of road transport emissions and energy consumption in China. This prominence stems largely from rapid economic development and expanding road freight demand over the past three decades [1-4]. Reports indicate HDDVs accounted for over 80% of total vehicular NOx emissions and exceeded 59% of the transportation sector's carbon emissions in 2023 [5]. Such substantial pollutant loads and fuel consumption pose severe threats to environmental integrity, public health, and national energy security [6-7]. To address these challenges, China promulgated the Three-Year Action Plan for Winning the Blue Sky Defense Battle (2018), reinforcing its commitment to air quality improvement [8]. This plan explicitly mandated enhanced emission oversight of HDDVs. Concurrently, the Ministry of Ecology and Environment enacted GB17691-2018 "Limits and measurement methods for emissions from diesel fuelled heavy-duty vehicles (China VI)", which requires the installation of onboard terminals from the China VIb phase onward for real-time monitoring of emissions and fuel use throughout the vehicle lifecycle [9]. Further amplifying these efforts, the Ministry jointly issued the "Action Plan for the Battle Against Diesel Truck Pollution" in late 2018, specifically advocating for "promoting the construction of remote monitoring systems for HDDVs" [10]. Consequently, China's HDDV remote monitoring system was formally introduced

Beijing stands among the earliest adopters of the HDDV remote monitoring system in China [11]. As of 2023, the system encompasses over 190,000 HDDVs in the Beijing region, involving more than 590 enterprises. Vehicle connectivity exceeds 60%, with an average daily online duration of approximately 7 hours. This infrastructure enables continuous data collection and oversight of connected vehicles, establishing a robust foundation for environmental regulators and manufacturers to monitor emissions, energy consumption, and operational status in real-time [12-14]. Furthermore, the system facilitates the acquisition of substantial volumes of HDDV operational data, termed remote monitoring data. These data effectively support regional emissions inventories, energy consumption accounting, and the iterative refinement of automotive products [15-17]. Consequently, advancing the application and development of HDDV remote monitoring data has emerged as a significant research focus, particularly in vehicle energy efficiency, pollution control, carbon mitigation, and environmental protection.

However, data acquisition and transmission are susceptible to multiple confounding factors, including sensor inaccuracies, signal conversion errors, data parsing inconsistencies, and potential data tampering [18-19]. These issues frequently introduce discrepancies between the monitoring data and reference measurements. Such discrepancies have generated significant dissatisfaction among data users and vehicle owners, undermining confidence in emission compliance policies and regulatory enforcement mechanisms reliant on remote monitoring data. Therefore, rigorous accuracy verification of remote monitoring data is a prerequisite for its reliable application and further development.

Addressing this need, Sun et al. [20] employed a full-flow emission analyzer to collect reference NOx measurements for validation. Their comparative analysis demonstrated that remote monitoring system NOx data complied with China VI regulatory requirements. Similarly, Xiong et al. [21] utilized cross-correlation analysis to assess data accuracy, identifying inherent latency within the remote monitoring system. They successfully corrected this latency by introducing noise signals (-10 dB signal-to-noise ratio), achieving a coefficient of determination (R^2) exceeding 0.9 between corrected remote data and reference vehicle measurements. Ren et al. [22] applied linear regression to evaluate correlations between remote monitoring data and Portable Emission Measurement System (PEMS) measurements. Their research confirmed strong correlations for parameters like vehicle speed, fuel flow rate, and intake flow, while noting discernible deviations in NOx emission data relative to PEMS measurements. Collectively, these studies underscore the critical importance and ongoing challenges of remote monitoring data validation.

Current research predominantly focuses on correlational analyses between remote monitoring data and reference measurements, while largely overlooking critical accuracy verification regarding data consistency and practical application. This gap highlights a significant limitation in existing methodologies for validating HDDV remote monitoring data. While strong correlation may exist between measurement methods, such statistical association does not inherently demonstrate methodological equivalence or interchangeability [23]. Consequently, rigorous assessment of data consistency is essential. Furthermore, the core purpose of HDDV remote monitoring—effective supervision of emissions and energy consumption—depends fundamentally on the reliable application of these data. Robust data consistency serves as a prerequisite for trustworthy application outcomes. Therefore, advancing accuracy verification research specifically addressing data consistency and application validity holds substantial scientific and regulatory significance.

To bridge this gap, this study integrates correlation analysis, linear regression modeling, Bland-Altman agreement assessment, and factor accounting theory to establish a comprehensive framework for HDDV remote monitoring data verification. We systematically evaluate data accuracy across three dimensions: correlation strength, statistical consistency, and applicability to emission/fuel consumption factor calculation. This approach delivers robust validation for this emerging big data source by directly addressing the critical limitations posed by the absence of consistency and application-layer verification in existing studies.

II. Materials & Methods

A. Real-road Data Acquisition

To verify the remote monitoring data's accuracy for HDDVs, the real road data acquisition test is conducted in this paper. The PEMS [24-25] is utilized to acquire HDDVs' operational and emission data. This provides a robust data foundation for the accuracy verification of remote monitoring data. Fig. 1 presents the details of the real road data acquisition test conducted in this paper. A total of 100 HDDVs are selected for the real road data acquisition test. These vehicles encompassed a diverse range of vehicle models, including tractors, trucks, buses, dumpers, and special vehicles, as illustrated in Fig. 1(a). These test vehicles are sourced from 18 different domestic and international manufacturers of HDDVs, covering all types of heavy commercial vehicles, as classified in Fig. 1(b). All test vehicles comply with China VI regulations, and each is equipped with an on-board terminal that met regulatory requirements. This terminal enables real-time retrieval of a vehicle's OBD information and data streams, which are transmitted to remote monitoring system in accordance with the communication protocol. Thus facilitating the generation of remote monitoring data for HDDVs. The primary objective of this data is to enable real-time monitoring of fuel consumption and emissions for HDDVs, supporting the identification and tracing of vehicles exhibiting high fuel consumption and emissions. The remote monitoring data included critical parameters such as vehicle speed, fuel flow, intake flow, and NOx emission rates, as shown on the right side of Fig. 1(d).

Furthermore, during real road data acquisition using PEMS, the driving conditions and loading parameters adhered to the stipulations of China VI regulations [25]. The driving conditions for the test vehicles consisted of 20% urban roads, 25% suburban roads, and 55% highways. Fig. 1(e) presents the driving conditions of HDDVs during the test. The data acquired by the PEMS included vehicle speed, CO_2 emission rate, exhaust flow, and NOx emission rate, as depicted on the left side of Fig. 1(d). At the test's conclusion, remote monitoring data for the same time intervals are concurrently collected, thereby preparing the dataset for subsequent accuracy verification of the remote monitoring data for HDDVs.

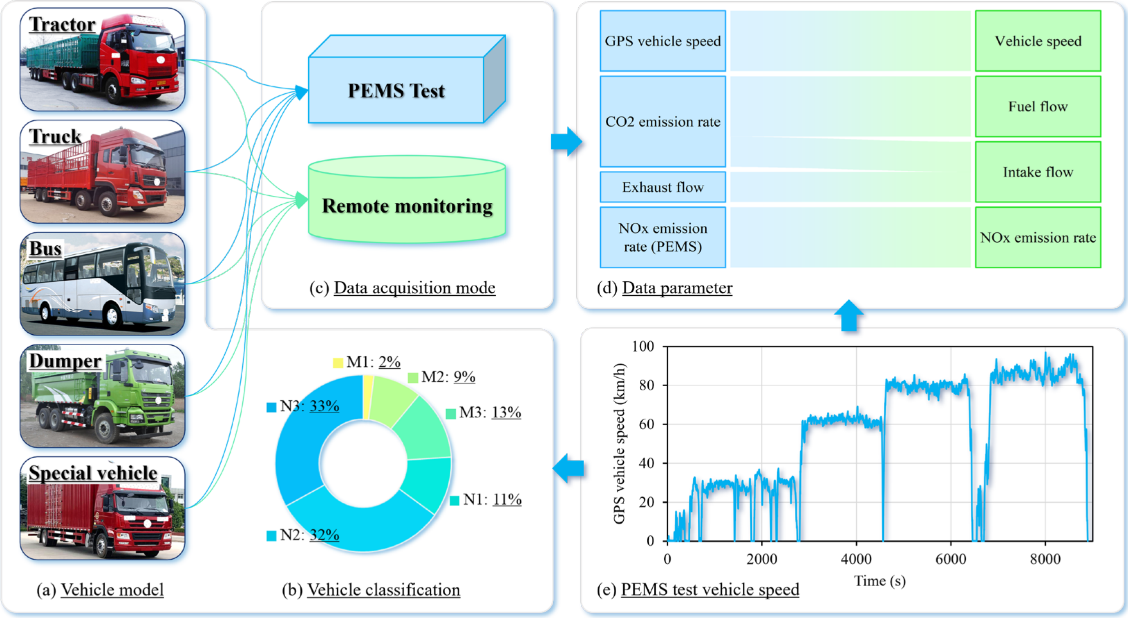


Fig. 1. Details of real road data acquisition test.

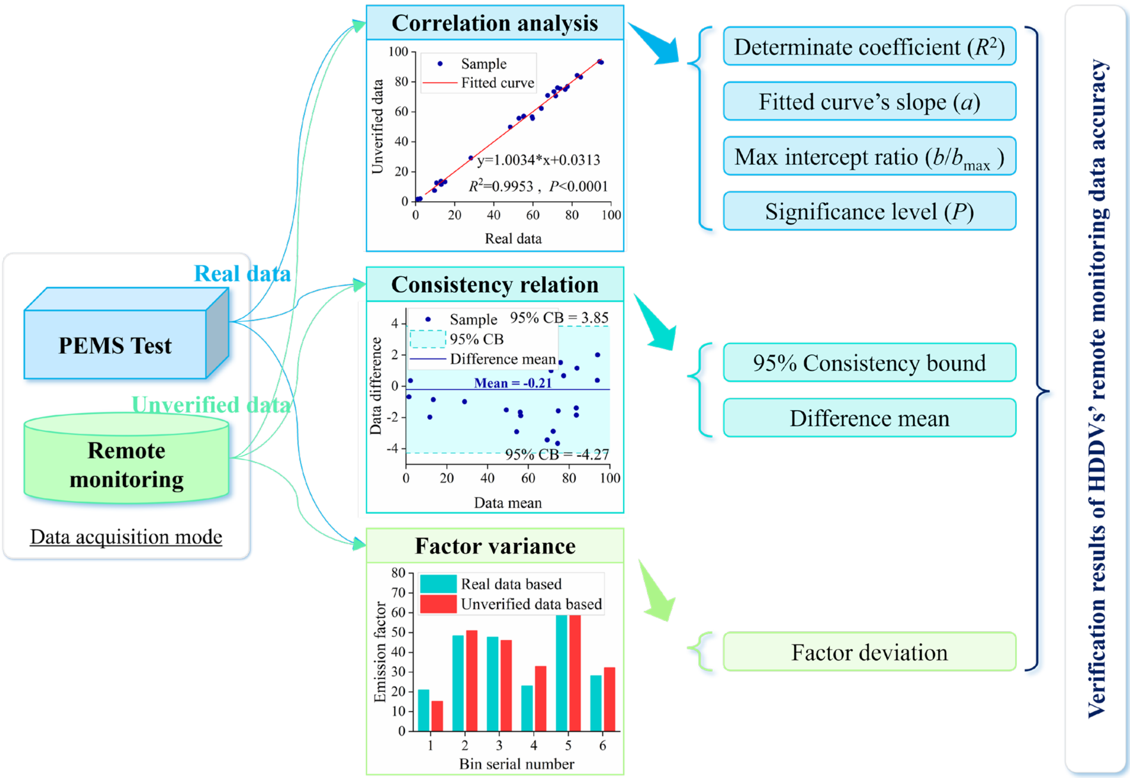


Fig. 2. Flow chart of remote monitoring data's accuracy verification method for HDDVs.

B. Remote Monitoring Data's Accuracy Verification Method

Building on HDDVs' data obtained from PEMS and the remote monitoring, research on remote monitoring data's accuracy verification is carried out. Combining correlation analysis, linear regression fitting, the Bland-Altman method [26] and factor accounting theory, the accuracy verification method of HDDVs' remote monitoring data is proposed, as illustrated in Fig. 2. Using PEMS data as a benchmark, the accuracy of the critical parameters within the remote monitoring data will be assessed.

Initially, partial correlation analysis [27] and linear regression fitting methods are utilized to determine the correlation between remote monitoring data and PEMS data. The determination coefficients (R^2), the fitted curve's slope (a), the max intercept ratio (b/b_{max}), and the significance level (P) are calculated. Among the max intercept ratio (b/b_{max}) is the ratio of the fitted curve's intercept to the remote monitoring data's maximum value. These results can provide insights into the correlation and its statistical significance between the remote monitoring data and PEMS data. According to the requirements of China VI regulations [28], accurate remote monitoring data should meet the following criteria: R^2 should approach 1, a should approach 1, and b/b_{max} should be as close to 0 as possible. Additionally, the P value less than 0.05 indicates a more significant correlation.

Subsequently, the Bland-Altman method is utilized to perform a consistency check between the remote monitoring data and PEMS data. The means, differences, and 95% consistency bound (95% CB) are calculated. The results from the Bland-Altman consistency check elucidate the accuracy of the remote monitoring data. Relevant studies have demonstrated that a higher sample proportion within the 95% CB signifies better consistency between the two datasets [29].

Finally, emission factors and fuel consumption factors are calculated using both remote monitoring data and PEMS data as independent data sources. The accuracy of the remote monitoring data is verified by comparing the differences in factors calculated from these two data sources. A smaller factor difference indicates greater accuracy of the remote monitoring data. This factor accounting methodology facilitates the verification of data accuracy in the application and development of remote monitoring big data.

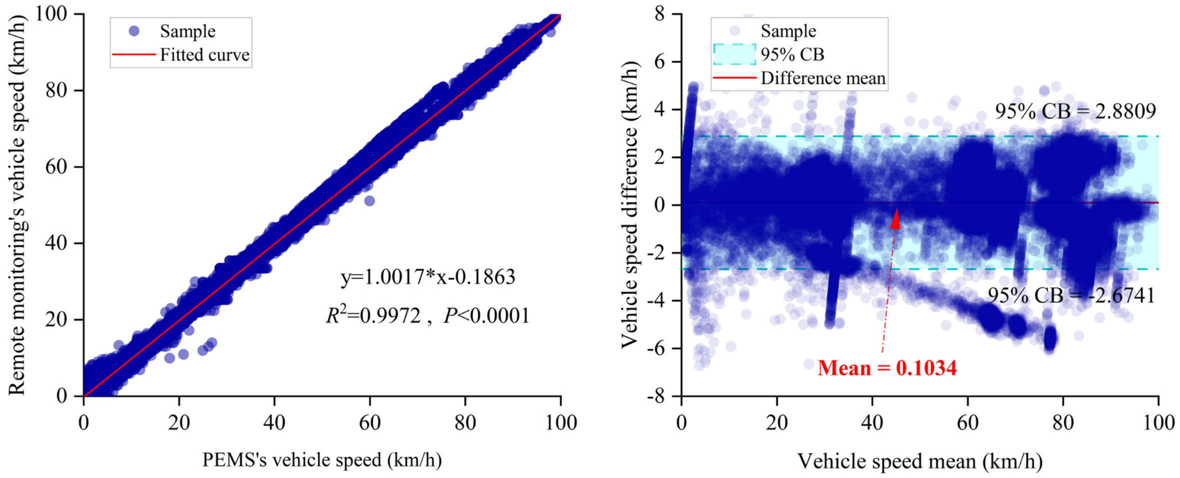
In addition, three methods, namely Kappa test [30], Intraclass correlation coefficient (ICC) [31] and Kendall cofficient of concordance (Kendall W) [32], were introduced again for the final test of remote monitoring data. The calculation results of the three methods are above 0.6, indicating good data consistency. The closer the calculation results of the three methods are to 1, the more accurate the remote monitoring data is. This can also provide validation for the calculation results of this method. In summary, through the correlation, consistency, and data application, a comprehensive accuracy verification study of the HDDVs' remote monitoring data is achieved. This research provides a crucial foundation for the application and development of remote monitoring data.

III. Data Verification Results Analysis

Based on the methodology proposed, research on remote monitoring data's accuracy verification for HDDVs is carried out. Correlation analysis and consistency relation are performed for critical parameters such as vehicle speed, fuel flow, intake flow, and NOx emission rate. Subsequently, the accuracy of the emission factor and fuel consumption factor calculated from the remote monitoring data source is evaluated from data application perspective.

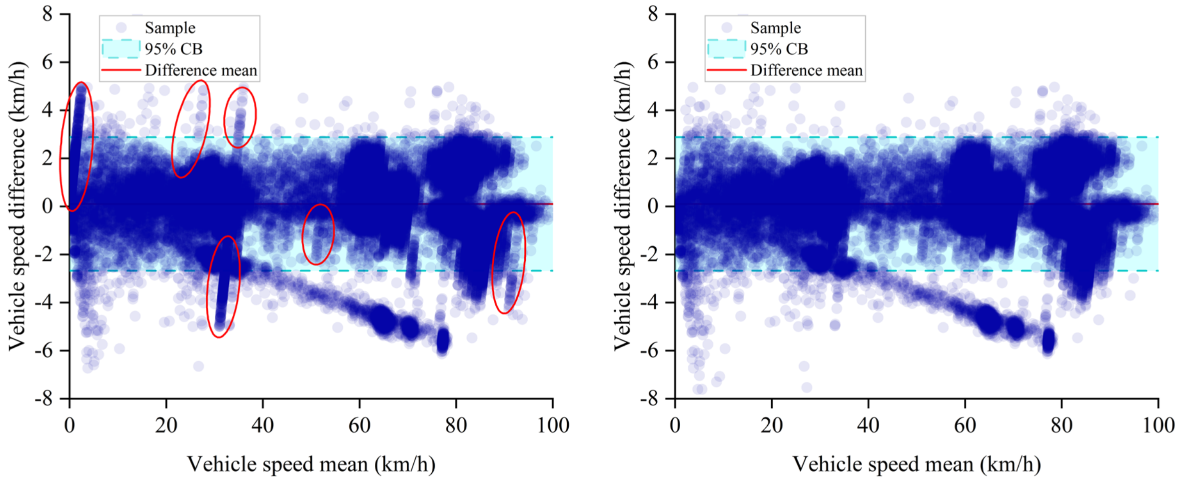
A. Vehicle Speed

Vehicle speed is one of the primary parameters in the HDDVs' remote monitoring data. The vehicle speed is measured using on-board speed sensors. In this study, the accuracy of the remote monitoring's vehicle speed is verified against the GPS vehicle speed from the PEMS dataset. Initially, all HDDVs' vehicle speed data are compiled. Subsequently, the data are segmented into several vehicle speed samples with a time interval of 1 second. Finally, correlation analysis and consistency relation for the remote monitoring' vehicle speed are calculated using the proposed method, and the results are shown in Fig. 3.



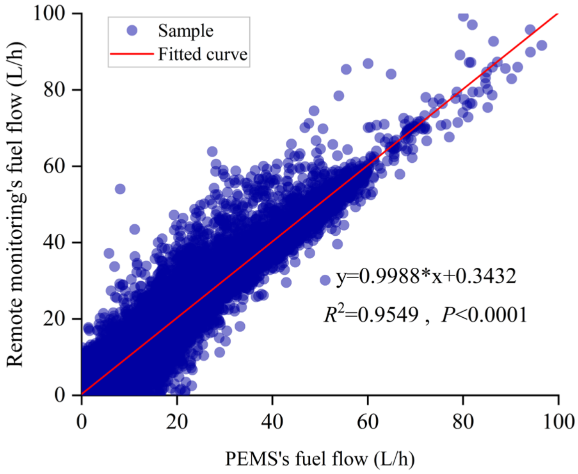
(a) Correlation analysis (b) Consistency relation

Fig. 3. Accuracy verification results of remote monitoring's vehicle speed.

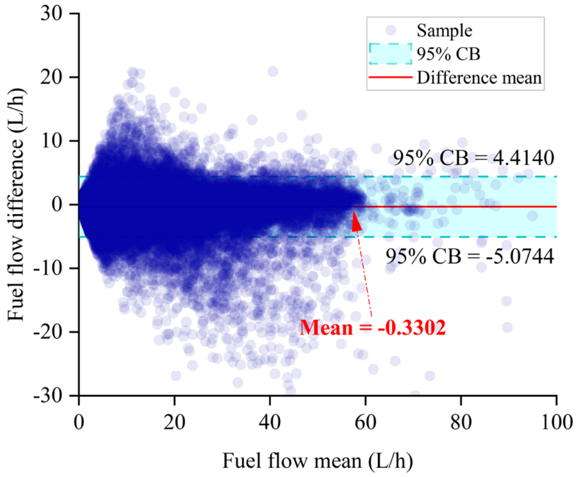


(a) Original consistency relation (b) Processed consistency relation

Fig. 4. Comparison of remote monitoring's vehicle speed consistency relation of before and after processing.



(a) Correlation analysis



(b) Consistency relation

Fig. 5. Accuracy verification results of remote monitoring's fuel flow.

The accuracy verification results of remote monitoring's vehicle speed are illustrated in Fig. 3. The deep blue scatter points represent the vehicle speed samples. In Fig. 3(a), the red solid line represents the linear regression fit between the PEMS's vehicle speed and the remote monitoring's vehicle speed. In Fig. 3(b), the shaded area bounded by the light blue dashed lines represents the 95% CB of the vehicle speed differences. The red solid line in Fig. 3(b) represents the mean value of the vehicle speed differences. According to the correlation analysis results presented in Fig. 3(a), it was observed that, except for a few vehicle speed samples, there is a strong fit between the remote monitoring's vehicle speed and the PEMS's vehicle speed. Specifically, the R^2 is found to be 0.9972; the a is 1.0017; the b/b_{max} is 0.0019; and the P value is less than 0.0001. The consistency results presented in Fig. 3(b) indicate that the majority of the vehicle speed samples fell within the 95% CB. Specifically, the 95% CB is determined to be [-2.6741, 2.8809] km/h, with 95.17% of the vehicle speed samples residing within these limits; the mean value of the vehicle speed differences is calculated to be 0.1034 km/h. Additionally, Fig. 3(b) reveals a small number of vehicle speed samples (problematic points) that exhibited a trend of clustering along multiple straight lines. These problematic points have been highlighted with red circles, as shown in Fig. 4(a). This phenomenon is primarily attributed to the effects of the vehicle speed sensor itself and the electronic signal transmission process, which resulted in constant or missing values in the remote monitoring's vehicle speed. Fig. 4(b) presents the consistency relation after the removal of these problematic points. Following the exclusion of problematic points, the processed sample count constituted 99.12% of the original vehicle speed sample size, indicating that the proportion of problematic points is less than 1%. Overall, the impact of these problematic points on the remote monitoring's vehicle speed is deemed minimal, and this analysis also provides guidance for subsequent preprocessing of remote monitoring data. In summary, there is a strong positive correlation and consistency between the remote monitoring's vehicle speed and the PEMS's vehicle speed. This finding confirms the accuracy of the remote monitoring's vehicle speed, thereby providing solid support for monitoring the operational status of heavy-duty vehicle in the designated areas through the remote monitoring system.

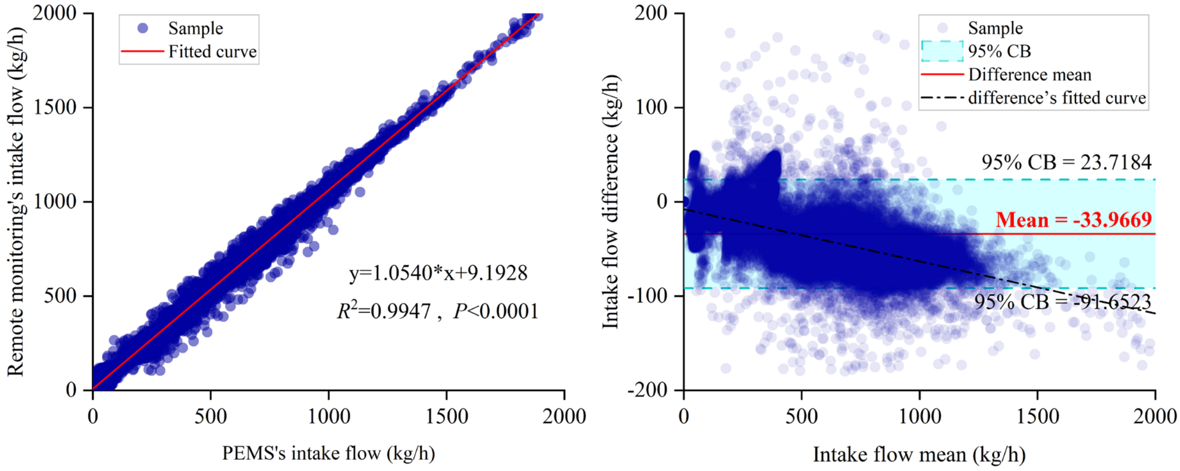
B. Fuel Flow

Fuel flow is one of the primary parameters in the HDDVs' remote monitoring data. Accurate remote monitoring's fuel flow provides a valuable information for optimizing energy consumption and controlling transportation costs for fleet operators. In accordance with national standards, the PEMS's fuel flow is calculated using the PEMS's CO_2 emission rate, as expressed in equation (1). This calculation results are then utilized to verify the accuracy of the remote monitoring's fuel flow. Initially, the remote monitoring's fuel flow and PEMS's CO_2 emission rate for all HDDVs are compiled, and the PEMS's fuel flow is calculated using equation (1). Subsequently, the fuel flow is segmented into several samples at 1-second intervals. Finally, correlation analysis and consistency relation for the remote monitoring's fuel flow are calculated using the proposed method, and the results are shown in Fig. 5.

$$Q_{\text{fuel}} = Q_{\text{CO}_2} \div K_{\text{CO}_2} \times 3600 \tag{1}$$

where Q_{fuel} represents the fuel flow of heavy-duty vehicles, which is regarded as the PEMS's fuel flow. Q_{CO2} represents the CO_2 emission rate from PEMS. K_{CO2} is the conversion factor, set to 2600 for diesel fuel vehicles.

The accuracy verification results of remote monitoring's fuel flow are illustrated in Fig. 5. The deep blue scatter points represent the fuel flow samples. In Fig. 5(a), the red solid line represents the linear regression fit between the PEMS's fuel flow and the remote monitoring's fuel flow. In Fig. 5(b), the shaded area bounded by the light blue dashed lines represents the 95% CB of the fuel flow differences. The red solid line in Fig. 5(b) represents the mean value of the fuel flow differences. According to the correlation analysis results presented in Fig. 5(a), it was observed that a significant positive correlation trend is observed in the distribution of the fuel flow samples. However, in comparison to the vehicle speed distribution depicted in Fig. 3(a), the distribution of the fuel flow samples is slightly more dispersed, with a small number of samples deviating from the fitted curve. Excluding these outliers, a strong fit is noted between the remote monitoring's fuel flow and the PEMS's fuel flow. Specifically, the R^2 is reported as 0.9549; the a is 0.9988; the b/b_{max} is 0.0035; and the P value is less than 0.0001. As indicated in Fig. 5(b), the consistency analysis reveals that



(a) Correlation analysis (b) Consistency relation

Fig. 6. Accuracy verification results of remote monitoring's intake flow.

the majority of fuel flow samples fall within the 95% CB. The 95% CB is determined to be [-5.0744, 4.4140] L/h, with 95.31% of the fuel flow samples residing within these limits. The mean value of the fuel flow differences is calculated as -0.3302 L/h. Meanwhile, the presence of problematic points observed in the vehicle speed issues of the Fig. 3(b) is not evident in the fuel flow samples. In summary, there is a strong positive correlation and a high degree of consistency between the remote monitoring's fuel flow and the PEMS's fuel flow. This finding supports the assertion that the fuel flow data obtained from remote monitoring is accurate.

C. Intake Flow

Intake flow is one of the primary parameters in the HDDVs' remote monitoring data. Accurate intake flow measurements provide a reliable basis for assessing the combustion status within the engine cylinders. In this paper, the PEMS's intake flow of HDDVs is calculated using the exhaust flow and fuel flow obtained from PEMS, as described by equation (2). Subsequently, the calculated PEMS's intake flow is utilized to verify the accuracy of the remote monitoring's intake flow. Initially, the remote monitoring's intake flow, PEMS's exhaust flow, and fuel flow from all HDDVs are compiled. The PEMS's intake flow is calculated using equation (2). The intake flow data are then segmented into several samples at 1-second intervals. Finally, the proposed method is utilized to calculate the correlation analysis and consistency relation of the remote monitoring's intake flow, and the results are shown in Fig. 6.

$$Q_{\text{in}} = Q_{\text{out}} - Q_{\text{fuel}} \cdot \rho_{\text{diesel}} \tag{2}$$

where Q_{in} represents the PEMS's intake flow. Q_{out} represents the PEMS's exhaust flow. Q_{fuel} represents the PEMS's fuel consumption. And ρ_{diesel} is the density of the diesel fuel used by the heavy-duty vehicles.

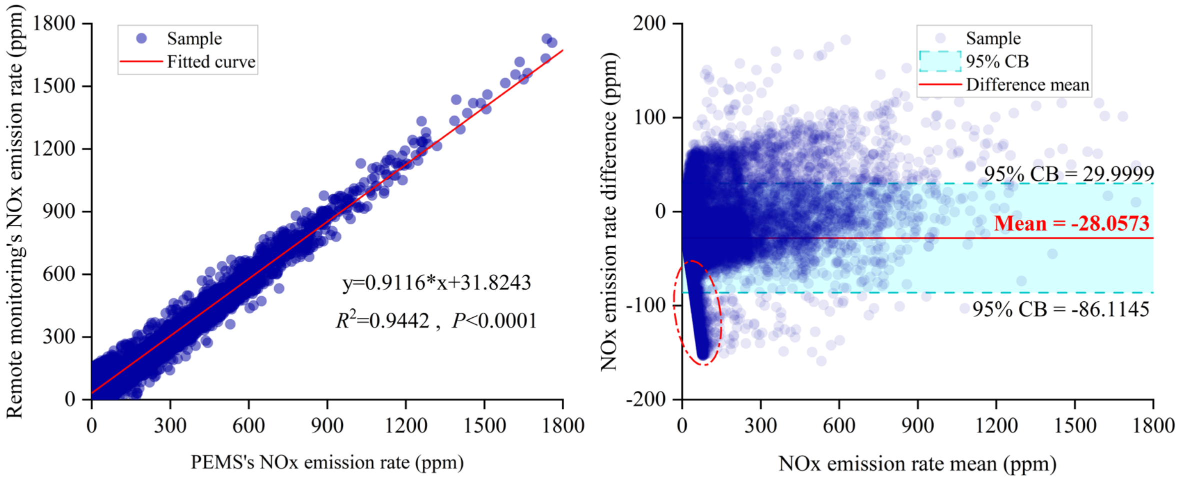
The accuracy verification results of remote monitoring's intake flow are illustrated in Fig. 6. The deep blue scatter points represent the intake flow samples. In Fig. 6(a), the red solid line represents the linear regression fit between the PEMS's intake flow and the remote monitoring's intake flow. In Fig. 6(b), the shaded area bounded by the light blue dashed lines represents the 95% CB of the intake flow differences. The red solid line in Fig. 6(b) represents the mean value of the intake flow differences. According to the correlation analysis results presented in Fig. 6(a), it was observed that, except for a few intake flow samples, there is a strong fit between the remote monitoring's intake flow and the PEMS's intake flow. Specifically, the R^2 is found to be 0.9947; the a is 1.0540; the b/b_{max} is 0.0045; and the P value is less than 0.0001. The consistency relation illustrated in Fig. 6(b) indicates that the majority of the intake flow samples fall within the 95% CB. Specifically, the 95% CB is determined to be [-91.6523, 23.7184] kg/h, with 95.01% of the intake flow samples residing within these limits, and the mean intake flow difference is calculated as -33.9669 kg/h. Furthermore, Fig. 6(b) reveals that the intake flow differences tend to be concentrated in the negative range. To quantify this trend, the linear regression fitting method is utilized to calculate the fitting curve for the intake flow difference sample data, represented by the black dotted line in Fig. 6(b). It is found that the intake flow differences decrease as the intake flow's mean increases. Since the remote monitoring data is used as the minuend during the calculation of the intake flow differences, it can be concluded that, for the majority of samples, the remote monitoring's intake flow is greater than that of PEMS. Moreover, with the increase of the intake flow's mean, this degree of greater than gradually expanded. In summary, there is a strong positive correlation and good consistency degree between remote monitoring's intake flow and the PEMS's intake flow. This demonstrates the accuracy of the intake flow data obtained from the remote monitoring system, providing a valid basis for analyzing the combustion status of HDDVs' engines through remote monitoring.*NOx Emission Rate*

NOx emission rate is recognized as the most critical parameter in the HDDVs' remote monitoring data. In the remote monitoring system, the NOx emission rate is measured using an upstream NOx sensor installed before the vehicle's diesel oxidation catalyst and a downstream nitrogen oxide sensor located at the end of the exhaust pipe. The parameters are categorized into upstream and downstream NOx emission rates. From the perspectives of environmental protection and carbon reduction, the downstream NOx emission rate is of primary concern to the public. This study employs PEMS data to verify the accuracy of the downstream NOx emission rate. First, all NOx emission rate data for heavy-duty vehicles are compiled. Subsequently, the NOx emission rate data are segmented into several samples with a 1-second interval. Finally, the proposed method is utilized to calculate the correlation and consistency relation for the remote monitoring's NOx emission rate, and the results are shown in Fig. 7.

The accuracy verification results of remote monitoring's NOx emission rate are illustrated in Fig. 7. The deep blue scatter points represent the NOx emission rate samples. In Fig. 7(a), the red solid line represents the linear regression fit between the PEMS's NOx emission rate and the remote monitoring's NOx emission rate. In Fig. 7(b), the shaded area bounded by the light blue dashed lines represents the 95% CB of the NOx emission rate differences. The red solid line in Fig. 7(b) represents the mean value of the NOx emission rate differences. According to the correlation analysis results presented in Fig. 7(a), it was observed that, the vast majority of the NOx emission rate sample points are distributed near the fitted curve, primarily concentrated within the [0, 1200] ppm range. A good fit is observed between the remote monitoring's NOx emission rate and the PEMS's NOx emission rate. Specifically, the R^2 is found to be 0.9442; the a is 0.9116; the b/b_{max} is 0.0184; and the P value is less than 0.0001. The consistency relation shown in Fig. 7(b) indicates that most NOx emission rate sample points fall within the 95% CB. Specifically, the 95% CB is determined to be [-86.1145, 29.9999] ppm, with 95.01% of the NOx emission rate samples residing within these limits.

The mean value of the NOx emission rate differences is calculated as -28.0573 ppm. Furthermore, Fig. 7(b) reveals that certain NOx emission rate samples (problematic points) exhibit a trend of aggregation along a linear pattern. This problematic region is highlighted with a dashed red circle in Fig. 7(b). To elucidate the underlying reasons, the original NOx emission rate data for a heavy-duty vehicle test is provided in Fig. 8.

In Fig. 8, the cyan line represents vehicle speed, the solid red line represents the PEMS's NOx emission rate, and the deep blue dashed line represents the remote monitoring's NOx emission rate. It is observed that the remote monitoring data accurately reflects the variations in the actual measurements obtained from PEMS. During the latter half of the test (as indicated by the red dot line), the vehicle's SCR system becomes operational, resulting in a reduction in NOx emission rates. Notably, during rapid acceleration, an increase in fuel injection leads to a sharp rise in NOx emissions, which depletes the ammonia stored in the SCR system, resulting in several peaks in the NOx emission rate. Following these peaks, the actual PEMS measurements approach zero, while the remote monitoring's NOx emission



(a) Correlation analysis (b) Consistency relation

Fig. 7. Accuracy verification results of remote monitoring's NOx emission rate.

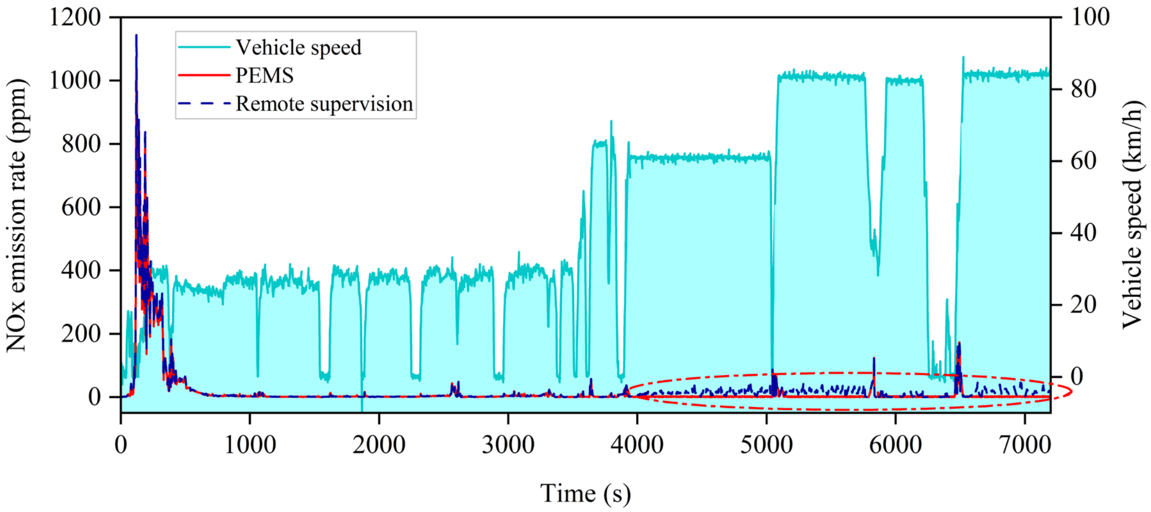


Fig. 8. Original NOx emission rate data for a heavy-duty vehicle test.

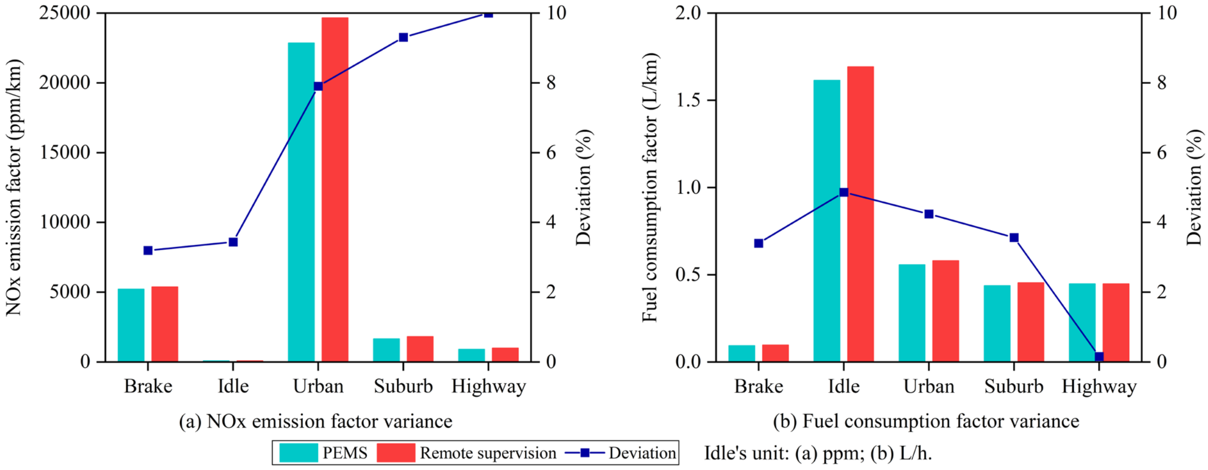


Fig. 9. Emission and fuel consumption factor's deviations.

rate to fluctuate within the [0, 150] ppm range. This phenomenon accounts for the problematic points identified in Fig. 7(b). These observations are consistent with the findings reported in literature [22]. In summary, there is a strong positive correlation and consistency between remote monitoring's NOx emission rate and PEMS's NOx emission rate. This indicates that the NOx emission rates derived from remote monitoring data are accurate, supporting the application of critical parameters in the supervision of HDDVs' remote monitoring systems.

E. Factor Variance Analysis

The application development of HDDVs' remote monitoring data is the most important task. Among these applications, the most critical are the calculations of emission factors and fuel consumption factors. By calculating emission and fuel consumption factors under various vehicle operating conditions, this study verifies the accuracy of remote monitoring data from the aspects of data application. Initially, by referring to the interval division method in reference [33-34], the vehicle operating states are divided into 5 intervals, as shown in Table 1. Subsequently, emission factors and fuel consumption factors are calculated for both data sources within each interval. Finally, deviations are utilized to quantify the variance between the factors derived from the two data sources, as expressed in Equation (3). The results of this accuracy verification of remote monitoring data applications are illustrated in Fig. 9.

$$De_{\mathrm{F}} = | F_{\mathrm{PEMS}} - F_{\mathrm{Remote\ supervision}} | \div F_{\mathrm{PEMS}} \times 100\% \quad (3)$$

where De_{F} represents the factor deviation based on the two data sources. F_{PEMS} represents the factor calculated from PEMS data. $F_{\mathrm{Remote\ supervision}}$ represents the factor derived from remote monitoring data.

TABLE 1.
VEHICLE OPERATING STATES' INTERVAL.

Number	Interval	Vehicle driving state
1	Brake	acceleration < -0.15 m/s2
2	Idle	vehicle speed < 0.5 km/h and -0.15 m/s2 ≤ acceleration < 0.15 m/s2
3	Urban	0.5 km/h ≤ vehicle speed < 45 km/h and acceleration ≥ 0.15 m/s2
4	Suburb	45 km/h ≤ vehicle speed < 70 km/h and acceleration ≥ 0.15 m/s2
5	Highway	vehicle speed ≥ 70 km/h and acceleration ≥ 0.15 m/s2

The results of accuracy verification for emission and fuel consumption factors calculated based on remote monitoring data are illustrated in Fig. 9. The cyan bars represent factors derived from PEMS data, while the red bars represent factors derived from remote monitoring data. The dark blue line represents the deviations between the factors derived from the two data sources. Due to the limited mileage in the idle interval, "ppm" and "L/h" are employed as units for emission and fuel consumption factors. In Fig. 9(a), it is demonstrated that the overall deviation of emission factors based on remote monitoring data is below 10%. The minimum deviation occurs in the brake interval, recorded at 3.20%. The maximum deviations in the suburb and highway intervals are observed to range between 9% and 10%. This observation aligns with the trends presented in the original data shown in the red dot line region of Fig. 8. In both suburb and highway areas, the actual measurements from PEMS approach zero, while remote monitoring data fluctuate within the range of [0, 150] ppm. This discrepancy accounts for the larger emission factors' deviations observed in the suburb and highway intervals. Fig. 9(b) indicates that the overall deviation of fuel consumption factors based on remote monitoring data is below 5%. The minimum deviation is identified in the highway interval at 0.16%, whereas the maximum deviation occurs in the idle interval at 4.86%. In comparison to emission factors, the accuracy of fuel consumption factors is relatively higher, particularly in the urban, suburb, and highway intervals where data accuracy is notably improved. In summary, the deviations of emission factors are maintained below 10%, while those of fuel consumption factors are kept below 5%. The results verify remote monitoring data's accuracy from the aspects of data application.

F. Result Verification

In this paper, three methods of Kappa test, ICC and Kendall W were used again for calculation. The calculation results are given in Table 2.

TABLE 2.
CALCULATION RESULT.

	Kappa	ICC	Kendall W
Vehicle speed	0.61	0.68	0.83
Vehicle speed (processed)	0.72	0.85	0.96
Fuel flow	0.75	0.77	0.95
Intake flow	0.68	0.69	0.94
NOx emission rate	0.72	0.76	0.92
Emission / fuel consumption factor	0.71	0.74	0.91

According to the calculation results in Table 2, the results of Kappa, ICC and Kendall W of each parameter in the remote monitoring data are all higher than 0.6. The calculated results of some parameters are higher than 0.9. In addition, the calculated result of the speed after processing is obviously better than that before processing. These results are basically consistent with the calculation results of the method presented in this paper. According to the calculation results of the three methods in Table 2, it can be seen that the remote monitoring data has a good consistency with the real road PEMS data, and the data has a certain accuracy, which can provide a solid foundation for engineering application development.

IV. Conclusion

This study proposes a comprehensive framework for validating HDDVs remote monitoring data, assessing accuracy across three critical dimensions: statistical correlation, measurement consistency, and application reliability. Our methodology establishes a robust foundation for advancing telematics-based emission supervision, with key findings detailed below:

(1) The parameters of vehicle speed, fuel flow, intake flow, and NOx emission rates from remote monitoring data exhibit a strong positive correlation and good consistency with actual PEMS data. Specifically, R^2 exceeding 0.94; and the sample proportion within the 95% CB exceeds 95.01% for all parameters. Among these, the vehicle speed shows the highest R^2 value, while the fuel flow shows the greatest sample proportion within the 95% CB. The R^2 value and the sample proportion within the 95% CB for the NOx emission rate are the lowest among all parameters.

(2) The accuracy of remote monitoring data is further verified from data application perspective by calculating emission and fuel consumption factors. The results indicate that the deviation of emission factors based on remote monitoring data from actual measurements remains within 10%, with the largest deviations occurring in suburb and highway intervals. The deviation of fuel consumption factors based on remote monitoring data from actual measurements is within 5%, with the greatest deviation observed in the idle interval.

(3) The remote monitoring data exhibits the following characteristics: (i) A small proportion of vehicle speed data presents constant or missing values, amounting to less than 1%. (ii) Most intake flow values are greater than actual measurements, and this discrepancy increases as the intake flow's mean increases. (iii) Certain NOx emission rate samples show a trend of aggregation along a linear pattern. The reason is that the NOx emission rates collected by remote monitoring in suburb and highway interval (where SCR systems begin operation) are still fluctuating in a small range. The above characteristics can provide guidance for the pre-processing of remote monitoring data's application and development.

(4) In addition, three methods, including Kappa test, ICC and Kendall W, were introduced in this paper to verify the calculation results of the proposed method. The calculation results of all parameters of remote monitoring data are above 0.6, which indicates that the data has good consistency.

References

[1] S. Zhang, P. Zhao, L. He, Y. Yang, B. Liu, W. He, and Y. Wu, "On-board monitoring (OBM) for heavy-duty vehicle emissions in China: Regulations, early-stage evaluation and policy recommendations," *Science of The Total Environment*, vol.731, pp.139045, 2020.

[2] H. Liu, H. Tian, Y. Hao, S. Liu, X. Liu, C. Zhu, and B. Wu, "Atmospheric emission inventory of multiple pollutants from civil aviation in China: Temporal trend, spatial distribution characteristics and emission features analysis," *Science of The Total Environment*, vol.648, pp.871-879, 2019.

[3] C. Keramydas, L. Ntziachristos, C. Tziourtzioumis, G. Papadopoulos, T. S. Lo, K. L. Ng, and C. K. L Wong, "Characterization of real-world pollutant emissions and fuel consumption of heavy-duty diesel trucks with latest emissions control," *Atmosphere*, vol.10, no.9, pp.535, 2019.

[4] B. Lu, and J. Li, "NOx Emission Prediction of Diesel Engine Based on GWO-LSTM," *International Journal of Automotive Technology*, pp.1-10, 2024.

[5] C. Tong, S. Ren, T. Li, D. Liu, L. Liu, P. Zhang, Z. Liu, X. Jing, and Z. Yang, "Testing and statistics of the heavy-duty vehicle chassis dynamometer fuel consumption measurement," *Journal of Automotive Safety and Energy*, vol.15, no.04, pp.553-560, 2024.

[6] X. Liang, S. Zhang, Y. Wu, J. Xing, X. He, K. M. Zhang, and J. Hao, "Air quality and health benefits from fleet electrification in China," *Nature Sustainability*, vol.2, no.10, pp.962-971, 2019.

[7] J. L. Breuer, R. C. Samsun, D. Stolten, and R. Peters, "How to reduce the greenhouse gas emissions and air pollution caused by light and heavy duty vehicles with battery-electric, fuel cell-electric and catenary trucks," *Environment international*, vol.152, pp.106474, 2021.

[8] Q. Zhang, Y. Zheng, D. Tong, M. Shao, S. Wang, Y. Zhang, and J. Hao, "Drivers of improved PM2. 5 air quality in China from 2013 to 2017," *Proceedings of the National Academy of Sciences*, vol.116, no.49, pp.24463-24469, 2019.

[9] G. Xu, W. Shan, Y. Yu, Y. Shan, X. Wu, Y. Wu, and H. He, "Advances in emission control of diesel vehicles in China," *Journal of Environmental Sciences*, vol.123, pp.15-29, 2023.

[10] G. Li, H. Yin, Z. Ji, "Formulation ideas and implementation suggestions on technical specification for emission remote supervision system of heavy-duty vehicles," *Journal of Environmental Engineering Technology*, vol.13, no.2, pp.867-872, 2023.

[11] K. Zhen, G. Li, Z. Ji, B. Liu, Y. Yang, K. Qin, and K. Guo, "Study on applicability of data collection frequency for heavy-duty vehicles based on remote monitoring," *In E3S Web of Conferences*, vol.360, pp.01005, 2022.

[12] X. Zhao, Y. Fang, H. Min, X. Wu, W. Wang, and R. Teixeira, "Potential sources of sensor data anomalies for autonomous vehicles: An overview from road vehicle safety perspective," *Expert Systems with Applications*, vol.236, pp.121358, 2024.

[13] W. Fang, H. Chen, and F Zhou, "Fault diagnosis for cell voltage inconsistency of a battery pack in electric vehicles based on real-world driving data,". *Computers and Electrical Engineering*, vol.102, pp.108095, 2022.

[14] J. Wang, R. Wang, H. Yin, Y. Wang, H. Wang, C. He, and K. He, "Assessing heavy-duty vehicles (HDVs) on-road NOx emission in China from on-board diagnostics (OBD) remote report data," *Science of The Total Environment*, vol.846, pp.157209, 2022.

[15] A. Wang, J. Xu, M. Zhang, Z. Zhai, G. Song, and M. Hatzopoulou, "Emissions and fuel consumption of a hybrid electric vehicle in real-world metropolitan traffic conditions," *Applied Energy*, vol.306, pp.118077, 2022.

[16] Y. Wen, R, Wu, Z. Zhou, "A data-driven method of traffic emissions mapping with land use random forest models," *Applied Energy*, vol.305, pp.117916, 2022.

[17] Z. Lv, and W. Shang, "Impacts of intelligent transportation systems on energy conservation and emission reduction of transport systems: A comprehensive review," *Green Technologies and Sustainability*, vol.1, no.1, pp.100002, 2023.

[18] N. Dimokas, D. Margaritis, M. Gaetani, K. Koprubasi, and E Bekiaris, "A Big Data application for low emission heavy duty vehicles," *Transport and Telecommunication Journal*, vol.21, no.4, pp.265-274, 2020.

[19] Z. Lv, Y. Zhang, Z. Ji, F. Deng, M. Shi, Q. Li, and K. He, "A real-time NOx emission inventory from heavy-duty vehicles based on on-board diagnostics big data with acceptable quality in China," *Journal of Cleaner Production*, vol.422, pp.138592, 2023.

[20] Y. Sun, Y. Guo, and C. Wang, "Research on data consistency of remote emission management vehicle terminals for heavy-duty vehicles,"

Small Internal Combustion Engine and Vehicle Technique, vol.48, no.2, pp.1-6, 2019.
[21] X. Xiong, Y. Yang, K. Zhen, and B. Liu, "Correction of Data Delay between Heavy-Duty Vehicle Monitoring Platform and OBD System," *Vehicle Engine*, no.02, pp.82-87, 2021.
[22] S. Ren, C. Tong, G. Li, C. Zhang, T. Li, D. Liu, P. Zhang, Zhi. Liu, and M. Guan, "A comparison study between heavy-duty remote monitoring and on-road experimental data," *Journal of Automotive Safety and Energy*, vol.14, no.02, pp.239-248, 2023.
[23] P. Sedgwick, "Limits of agreement (Bland-Altman method)," *Bmj*, pp.346, 2013.
[24] X. Zhang, J. Li, H. Liu, Y. Li, T. Li, K. Sun, and T. Wang, "A fuel-consumption based window method for PEMS NOx emission calculation of heavy-duty diesel vehicles: Method description and case demonstration," *Journal of Environmental Management*, vol.325, pp.116446, 2023.
[25] C. Liu, Y. Pei, C. Wu, F. Zhang, and J. Qin, "Novel insights into the NOx emissions characteristics in PEMS tests of a heavy-duty vehicle under different payloads," *Journal of Environmental Management*, vol.348, pp.119400, 2023.
[26] M. A. Mansournia, R. Waters, M. Nazemipour, M. Bland, and D. G. Altman, "Bland-Altman methods for comparing methods of measurement and response to criticisms," *Global Epidemiology*, vol.3, pp.100045, 2021.
[27] L. Zhang, L. Yang, C. M. Zohner, T. W. Crowther, M. Li, F. Shen, and C. Zhou, "Direct and indirect impacts of urbanization on vegetation growth across the world's cities," *Science advances*, vol.8, no.27, pp.eabo0095, 2022.
[28] Ministry of Ecology and Environment releases China Mobile Source Environmental Management Annual Report (2023). https://www.mee.gov.cn/ywgz/dqhjbh/ydyhjgl/202312/t20231207_1058461.shtml
[29] D. Tsikas, "Mass Spectrometry-Based Evaluation of the Bland–Altman Approach: Review, Discussion, and Proposal," *Molecules*, vol.28, no.13, pp.4905, 2023.
[30] M. A. Bujang and N. Baharum, "Guidelines of the minimum sample size requirements for Kappa agreement test," *Epidemiology, biostatistics, and public health*, vol.14, no.2, 2017.
[31] T. K. Koo, and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *Journal of chiropractic medicine*, vol.15, no.2, pp.155-163, 2016.
[32] P. Legendre, "Species associations: the Kendall coefficient of concordance revisited," *Journal of agricultural, biological, and environmental statistics*, vol.10, pp.226-245, 2005.
[33] X. Zhang, J. Li, Z. Yang, Z. Xie, and T. Li, "Accuracy Analysis of Carbon Emissions Measurement of Heavy Heavy-Duty Diesel Vehicles Based on Remote Data," *China Environmental Science*, vol.42, pp.4565-4570, 2022.
[34] Q. Feng, B. X. Liu, Y. Y. Yang, S. Zhang, X. Shen, Y. Lu, and Y. Wu, "Characteristics of actual road NOX and CO2 emissions from diesel vehicles in Beijing based on remote monitoring," *China Environmental Science*, vol.43, no.8, pp.4418-4426, 2023.

Yu Liu was born in 1987 in Huanghua City, Hebei Province, China. In 2010, he obtained a bachelor's degree in Thermal Energy and Power Engineering from Harbin Engineering University. In 2012, he received a master's degree in Power Machinery and Engineering from Tianjin University. In 2016, he obtained a doctor's degree in Power Machinery and Engineering from Tianjin University. His main research directions include big data mining and analysis of motor vehicle energy consumption and emissions, as well as the research on the development of vehicle driving cycles.