energies

Article

Gas Station Recognition Method Based on Monitoring Data of Heavy-Duty Vehicles

Yan Ding [1], Zhe Ji [1,2,3,*], Peng Liu [4], Zhiqiang Wu [4], Gang Li [1], Dingsong Cui [4,*], Yizhong Wu [4] and Sha Xu [5]

1 State Environmental Protection Key Laboratory of Vehicle Emission Control and Simulation, Chinese Research Academy of Environmental Sciences, Beijing 100012, China; dingyan@vecc.org.cn (Y.D.); ligang@vecc.org.cn (G.L.)
2 University of Science and Technology of China, Hefei 230026, China
3 Key Laboratory of Environmental Optics and Technology, Anhui Institute of Optics and Fine Mechanics, Hefei Institute of Physical Science, Chinese Academy of Sciences, Hefei 230031, China
4 National Engineering Laboratory for Electric Vehicles, Beijing Institute of Technology, Beijing 100081, China; bieliupeng@bit.edu.cn (P.L.); 3220190341@bit.edu.cn (Z.W.); 3120200402@bit.edu.cn (Y.W.)
5 Beijing Bitnei Corp., Ltd., Beijing 100081, China; xusha@bitnei.cn
* Correspondence: jizhe@vecc.org.cn (Z.J.); dingsong_cui1994@163.com (D.C.)

Abstract: With the requirement of reduced carbon emissions and air pollution, it has become much more important to monitor the oil quality used in heavy-duty vehicles, which have more than 2/3 transportation emissions. Some gas stations may provide unqualified fuel, resulting in uncontrollable emissions, which is a big challenge for environmental protection. Based on this focus, a gas station recognition method is proposed in this paper. Combining the CART algorithm with the DBSCAN clustering algorithm, the locations of gas stations were detected and recognized. Then, the oil quality analysis of these gas stations could be effectively evaluated from oil stability and vehicle emissions. Massive real-world data operating in Tangshan, China, collected from the Heavy-duty Vehicle Remote Emission Service and Management Platform, were used to verify the accuracy and robustness of the proposed model. The results illustrated that the proposed model can not only accurately detect both the time and location of the refueling behavior but can also locate gas stations and evaluate the oil quality. It can effectively assist environmental protection departments to monitor and investigate abnormal gas stations based on oil quality analysis results. In addition, this method can be achieved with a relatively small calculation effort, which makes it implementable in many different application scenarios.

Keywords: gas stations recognition; oil quality evaluation; heavy-duty vehicles; DBSCAN clustering; CART algorithm; real-world data

Citation: Ding, Y.; Ji, Z.; Liu, P.; Wu, Z.; Li, G.; Cui, D.; Wu, Y.; Xu, S. Gas Station Recognition Method Based on Monitoring Data of Heavy-Duty Vehicles. *Energies* **2021**, *14*, 8011. https://doi.org/10.3390/en14238011

Academic Editor: Davide Astolfi

Received: 15 October 2021
Accepted: 23 November 2021
Published: 30 November 2021

1. Introduction

International agreements such as the Paris Agreement call for significant reductions in carbon emissions to mitigate global warming [1–3]. China aims to achieve a carbon peak by 2030 and is striving to achieve carbon neutrality by 2060. The transportation industry plays an important role in this process. Carbon emissions from heavy-duty vehicles (HDVs) are among the significant contributors to greenhouse gas emissions [4–6]. Although HDVs account for a small proportion of vehicle ownership, nitrogen oxides and particulate matter emissions are huge, accounting for 83.5% and 90.1% of total vehicle emissions, respectively [7,8]. To reduce these harmful effects on the atmospheric environment and human health, heavy-duty diesel vehicles (HDDVs) have mainly applied selective catalytic reduction technology by adding urea to reduce nitrogen oxide emissions such that vehicles could reach the level of the China V and China VI emission standards [9]. However, due to the material cost, some vehicle owners use various cheating methods to invalidate the nitrogen oxide control device or use inferior urea water solutions. Therefore, it is necessary to strengthen the supervision measures and law enforcement methods in this regard.

Some gas stations have not obtained the relevant permits in accordance with the law to engage in oil business activities without the qualification of dangerous goods business, which may result in serious safety hazards. The fuel sold by these gas stations cannot be supervised effectively, and uneven quality could lead to excessive emissions of vehicle nitrogen oxides and other pollutants. Additionally, lacking fuel and gas recovery devices may result in fuel leakage when refueling and unloading, which could cause direct pollution of the atmosphere, soil, and groundwater.

The current method for detecting fuel/urea refueling behavior mainly includes setting up vehicle identification and verification devices at gas stations and then using wireless communication to transmit vehicle fuel/urea refueling information to a remote server. The high cost, including the timely update rate and reading and writing storage data structure, has caused it to be unable to provide useful information to detect unregistered gas/urea stations and also their related fuel or urea refueling behaviors.

Gas station recognition is essentially outlier identification based on group behaviors. Few related studies have addressed this issue. However, some similar studies have been reviewed as follows: outlier identification is currently widely used in fault diagnosis, multimedia traffic detection, vehicle status detection, video and audio detection, and other fields. Muhammad et al. [10] applied an outlier detector based on Gosset's (Student's) *t*-test to distinguish a faulty condition and identify the nature of the fault for large-scale grid-tied photovoltaic power plants. Sahil et al. [11] proposed a hybrid deep-learning-based anomaly detection scheme for suspicious flow detection in the context of social multimedia. The anomaly detection module was based on leverage improved restricted Boltzmann machine and gradient descent-based support vector machine. Morteza et al. [12] developed an ARIMA-based anomaly detection framework to identify abnormal states of the vehicles based on the multiple-channel operating time-series data. The state anomaly is captured by the deviation of real-time values at different channels from the predictions. Rui-kai et al. [13] proposed an audio-based algorithm to detect faults of pumps of air-conditioning systems, which can monitor the abnormal sound of pumps, utilizing Fourier transform, a finite impulse response digital filter, and an autoregressive integrated moving average model.

In addition, with the rapid development of computer technology, researchers have proposed many new anomaly detection methods. Wen [14] presented the problem of detecting outliers in fixation gaze data through a novel mixed-integer optimization formulation and subsequently strengthened the formulation using two geometric arguments to provide enhanced bounds. Ji-hao [15] compared eight common anomaly detection methods based on the statistical distribution of data and features to detect anomalies in real-time body weight (BW) recorded by a precision feeding (PF) system. Ekin [16] proposed a new nonparametric outlier detection technique in the preprocessing stage of data analyses, which was based on the frequency-domain and Fourier transform definitions, called the frequency-domain based outlier detection (FOD). Hee [17] proposed a new two-stage procedure for detecting multiple outliers when the dimension of the data is much larger than the available sample size.

However, the existing methods lack research on road infrastructure identification, and the anomaly detection algorithms are computationally intensive and have few engineering applications. Research based on spatiotemporal GPS trajectory focuses on the driving behavior of vehicles and lacks the analysis of vehicle behavior during the stay. In response to the current problems of the high cost of gas station information collection, long-term update cycle and traditional oil analysis methods are very labor intensive. This study fully utilizes the vehicle trajectory data and provides the possibility of data-driven gas station location identification, fuel/urea refill behavior, and oil quality analysis. With the development of Internet of Vehicles technology, a large amount of vehicle monitoring data has been collected and integrated. China has successfully built a Heavy-duty Vehicle Emission Service and Management Platform (HVESMP) based on the China VI Emission Standard, and the sampling period of vehicle data is 1 s. The data are wirelessly transmitted to the cloud platform via T-BOX. The data include key parameter information such as vehicle speed, engine speed, engine fuel flow, etc. This paper implements a highly robust recognition method for the spatiotemporal characteristics of mobile vehicle fuel/urea refueling behavior in mobile source big data scenarios. This proposed method innovatively uses the time-series data change curve of the fuel tank or urea level detected by the vehicle-mounted T-BOX to detect the fuel/urea refueling behavior, which can ensure the robustness of the recognition of the fuel/urea refueling behavior. Additionally, the method can not only reduce the omission and lack of information but can also process large data scales. At the same time, in the process of identifying the behavior of fuel/urea refueling, the CART algorithm is introduced to reduce the interference of sensor noise, which can improve the recognition accuracy. In addition, this paper uses accumulative mileage, fuel/urea consumption/100 km, and the DBSCAN clustering algorithm to verify the accuracy of recognition results.

The remainder of this paper is structured as follows: Section 2 introduces the system framework, including gas stations recognition and oil quality analysis. Section 3 gives a brief introduction of the proposed method for identifying gas stations and evaluating oil quality. Section 4 provides the data collection method and shows a recognition result of Tangshan, China. Finally, the key conclusions are summarized in Section 5.

2. System Framework

2.1. Data Collection

To solve the problem of pollutant emission supervision of HDDVs, China has successfully built HVESMP, and the data sampling frequency is 1 s. The data collection process of the platform is provided as follows: (1) Collecting vehicle OBD information and engine emission data based on multiple sensors, such as fuel tank level, diesel particulate filter (DPF) pressure difference, selective catalytic reduction (SCR) inlet/outlet temperature, etc.; (2) Collecting the data to the onboard T-BOX; (3) All data of T-BOX will be transferred to the storage server of the platform via wireless network (4G or 5G), which obeys the transmission protocol named by the limits and measurement methods for emissions from HDDVs (China VI). Some parameters collected from the data resources are shown in Table 1.

Table 1. Parameters collected from the data resource.

Driving Information	Engine Emission Information
Timestamp	Fuel tank level (%)
Longitude	Rotational speed (rpm)
Latitude	Fuel flow (L/h)
Speed (km/h)	SCR outlet/inlet temperature (°C)
Mileage (km)	DPF Pressure difference (kPa)

A total of 183,804 vehicles, including 54,352 HDVs (China VI) [18,19] Ekin and 129,163 HDVs (China V), were connected to the HVESMP until April 2021. The average daily online duration of vehicles (China VI) is about 13 h, and the average daily operating duration is about 9 h. This massive real-world operating data could support the research effectively.

2.2. System Framework

The system framework of this paper is mainly shown in Figure 1:

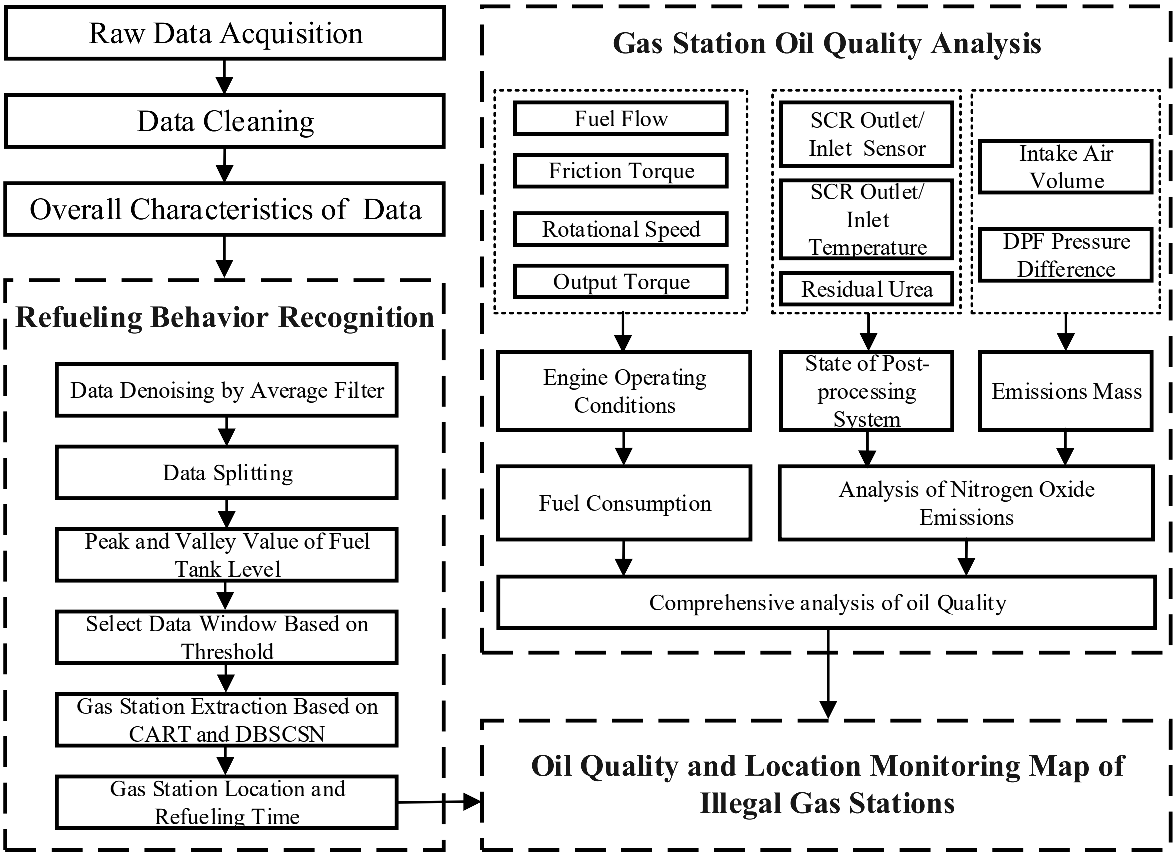


Figure 1. System framework of the gas station recognition and the oil quality analysis.

1. We set some rules for data cleaning and data preprocessing to eliminate abnormal and noisy data.
2. We establish some necessary characteristics based on extracting the data statistics such as location, fuel tank level, fuel flow, etc. These characteristics will be applied in the following models.
3. A refueling behavior recognition model and a gas station fuel analysis method are constructed to monitor gas station oil quality and spatiotemporal location information effectively.

The refueling behavior recognition model is based on the cleaned mobile terminal data and the refueling behavior data representation, combined with the judgment rules to extract and reconstruct the refueling feature detection window. The CART algorithm is used to effectively capture the spatiotemporal multidimensional information corresponding to a large number of refueling behaviors in the long-term dimension. Then, based on the DBSCAN clustering algorithm to realize the identification of gas station location, the gas amount and other information are accumulated and compared with the location data filed by legal gas stations to screen out suspicious gas stations.

In addition, to ensure the accuracy of the recognition of gas stations, based on the real-time information of onboard sensors, considering the two dimensions of vehicle fuel consumption and nitrogen oxide emissions, a quantitative evaluation of the oil quality of the gas station is carried out. First, the fuel consumption of the vehicle is calculated based on data such as fuel flow, speed, friction torque, and output torque, combined with the operating conditions of the vehicle engine. Secondly, based on data such as SCR temperature, residual urea, etc., the working status of the vehicle's postprocessing system is characterized. Intake air volume and DPF pressure difference are combined with other data to calculate the quality of vehicle emissions and achieve an effective evaluation of vehicle nitrogen oxide emissions.

3. Gas Stations Recognition Model

3.1. Data Preprocessing

In the process of platform data collection, there are many interference factors, such as onboard sensors, transmission lines, transmission network signals, and so on. It is very common to have invalid values and missing values in the original data. So, data preprocessing is an important step in the entire data analysis process [20]. For example, if the collection time of multiple frames is the same, only one frame of data can be selected, and also when the fuel tank level error data and missing data continuously exceeds 10 frames, the corresponding problem data frame would be directly deleted. In addition, in order to smooth the noise data existing in the sensor and data transmission process, this paper will adopt the moving average filter. The filter window size is set to 10 in this paper.

3.2. Refueling Feature Detection Window Selection

After data preprocessing and filtering, a full historical data set of the vehicle fuel tank level that can be analyzed is formed. The data volume is very large, and the hardware required for data processing at the same time is extremely high. Therefore, in order to realize the rapid screening of the characteristics of the refueling behavior data and improve the calculation efficiency of the algorithm, this paper divides the entire data into several basic data analysis windows according to the time series. The length of the window is 900 s (15 min). The full data of vehicle fuel tank level is represented as:

$$U_{all} = [U_1 \cdots U_i \cdots U_n]^T \tag{1}$$

where U_i is the i-th basic data analysis window, and $i = 1, 2, \ldots, n$. n represents the total number of basic data analysis windows. Using $u^i_{\max}$, $u^i_{\min}$ representing the maximum and minimum fuel tank level of the i-th basic data analysis window, respectively, the fuel tank level extremum matrix can be expressed as:

$$\boldsymbol{U}_{\max,\min} = \begin{bmatrix} u^1_{\max} & u^1_{\min} \\ \vdots & \vdots \\ u^i_{\max} & u^i_{\min} \\ \vdots & \vdots \\ u^n_{\max} & u^n_{\min} \end{bmatrix} \tag{2}$$

Furthermore, the matrix elements in (2) are sorted by time index to obtain the fuel tank level extremum series. Considering that the time sequence of the two extreme values in the same basic data analysis window is uncertain, and also in order to improve the efficiency of subsequent calculations, the basic windows that meet the rules will be merged. If the maximum/minimum values of two consecutive windows are adjacent in the extremum series, the corresponding two basic windows are merged, and the maximum/minimum value in the extreme value is updated as the maximum/minimum value of the new window. Finally, the number of updated basic data analysis windows is x, and the full data of vehicle fuel tank level can also be expressed as:

$$U_{all} = [U'_1 \cdots U'_j \cdots U'_x]^T \tag{3}$$

where U'_j is the j-th basic data analysis window after the windows are merged, and $i = 1, 2, \ldots, x$.

Then, the new fuel tank level extremum matrix after shape transformation and windows merging can be expressed as:

$$U_{\max,\min} = [u_1 \cdots u_j \cdots u_{2x}]^T \tag{4}$$

The difference between adjacent values in Matrix (4) is calculated, and the absolute value is taken. The result is as follows:

$$U_{diff} = \left[\left|u'_1\right| \cdots \left|u'_k\right| \cdots \left|u'_{2x-1}\right|\right]^T \tag{5}$$

where u'_k is the k-th tank level difference, and $k = 1, 2, \ldots, 2x - 1$.

Based on the feature that the fuel tank level suddenly increases in the fueling data windows, the judgment threshold values $\delta_1 = 30$ and $\delta_2 = 10$ are set to construct the filtering conditions of the fueling feature detection window. The specific conditions are as follows:

$$\begin{cases} k = 2m + 1, m = [1 \cdots x - 2] \\ u'_k > \delta_1 \\ u'_{k+1} < \delta_2 \\ u'_{k-1} < \delta_2 \end{cases} \tag{6}$$

If the above conditions are met, it indicates that the fluctuation range of the fuel tank level in the basic windows corresponding to the $k - 1$, $k + 1$ difference is relatively stable. The change in the level of the basic window corresponding to the k-th difference is likely to be caused by the user's fueling behavior rather than the sensor noise. Therefore, the value of k that meets the above conditions is recorded, considering that the input data of the CART algorithm needs to ensure enough data frames, so the basic data analysis window corresponding to the $k - 1$, k, $k + 1$ index is extracted as the refueling feature detection window. The result can be expressed as

$$U_{detect} = \begin{bmatrix} U'_{\frac{k-1}{2}} & U'_{\frac{k+1}{2}} & U'_{\frac{k+3}{2}} \end{bmatrix}^T = \begin{bmatrix} U'_m & U'_{m+1} & U'_{m+2} \end{bmatrix}^T \tag{7}$$

3.3. Gas Stations Recognition Method

Based on historical operating data, we constructed a CART classification model, which can identify whether it is a mutation threshold point through the sequence points in the input window. Each value in the level value sequence is used as an input to the CART model. The output value is determined by the Gini coefficient using u_y and Y representing the y-th value in the detection window and the length of the detection window matrix, respectively. The loss function of the CART algorithm can be expressed as:

$$\text{Loss} = \sum_{i=1}^{y} \left(\overline{x}_i - x_i\right)^2 + \sum_{j=y+1}^{Y} \left(\overline{x}_j - x_j\right)^2 \tag{8}$$

where $\overline{x}_i$ is the average of the first y values, $\overline{x}_j$ is the average of the rest values, and x_i and x_j are the ith and jth liquid level value in the detection window.

Finally, the y value is obtained when the loss function is the smallest. It is used as the refueling point where the tank level value changes suddenly, and the corresponding time stamp and latitude and longitude position are recorded. At the same time, the height t of the tank level value in the window is sorted to obtain the maximum and minimum values as the liquid level values after and before refueling, respectively.

In a relatively long time scale, the repeated refueling behavior of diesel vehicles at a fixed gas station shows spatial aggregation. Therefore, Grid-Search tuning and the DBSCAN clustering algorithm are used to cluster a large number of the identified refueling points within a certain period of time. The locations of the refueling behavior and the locations of the registered legal gas station are used as input to the DBSCAN model. If there is a possible station near the refueling location, it will be recognized as refueling at a legal station; otherwise, it will be recognized as an outlier, which means an illegal station. Then, the area and the total amount of refueling of the cluster rectangular box are calculated, which are determined by the four values of the maximum/minimum latitude and longitude of all points in the cluster. Based on the result, the optimal parameters as the gas station location are selected.

The clustering results can be matched with the existing registered gas station data based on the distance threshold. If the registered gas station cannot be matched, it may be an unregistered gas station, and this method is an auxiliary basis for the investigation of unregistered gas stations.

3.4. Oil Quality Analysis

The quality of oil sold at unregistered gas stations cannot be monitored, and substandard-quality fuel leads to increased vehicle fuel consumption and excessive vehicle emissions. Therefore, based on the collected data, vehicle fuel consumption and NOx emissions are calculated after each refueling act as a basis for judging the oil quality of the gas station and validating the gas station identification algorithm.

The first step is to preprocess the raw data to ensure the accuracy of the NOx emission calculation, extract the data suitable for the calculation, and exclude the emission exceedance caused by the abnormal operation of the SCR system. The rules for data screening are as follows:

(1) Urea level: >0;
(2) Atmospheric pressure: >76 kpa;
(3) SCR inlet temperature: >200 °C;
(4) Engine cooling temperature: >75 °C.

After the data has been processed, the NOx emission rate of the vehicle can be obtained from Equation (9) [9].

$$E_{\mathrm{NOx}} = \left(Q_{fV} \times \frac{\rho_f}{\rho_a} + \frac{Q_{am}}{\rho_a} \right) \times \frac{M \times P_{\mathrm{NOx(down)}}}{22.4 \times 10^6} \times 1000 \tag{9}$$

where E_{NOx} (g/h) is the NOx emission rate, Q_{fv} (L/h) is the engine fuel flow rate, Q_{am} (kg/h) is the mass air flow, $P_{\mathrm{NOx(down)}}$ (ppm) is the SCR downstream NOx sensor output, ρ_a is the air density, and M is the molecular weight of NOx. M = 30.4 according to NO: NO_2 = 95:5, and ρ_f is the diesel density.

In a realistic driving environment, drivers tend to refuel before they run out of fuel, and the refueling behavior is shown in Figure 2. When using vehicle emissions as an oil quality assessment indicator, it is necessary to adjust the weight distribution of emissions after one refueling act according to the percentage distribution of refueling volume at each refueling station, and the adjusted NOx emission rate of the refueling station is calculated, as shown in Equation (10).

$$\begin{aligned} m_i &= \frac{V_i}{Q_i + V_i} \\ m_k &= m_i \prod_{j=i}^{k} \frac{Q_{j+1}}{Q_{j+1} + V_{j+1}} (n \geq k > i) \\ T_i &= m_i E_i + \sum_{k=i+1}^{n} m_k E_k \end{aligned} \tag{10}$$

where T_i is the NOx emission rate of vehicle N_i in a single day at gas station S_i, Q_i is the amount of fuel remaining in the tank before the No. i refueling, V_i is the amount of fuel added for the No. i refueling, E_i is the NOx emission rate between the No. i refueling and the No. i + 1 refueling, and n is the number of refueling of the vehicle after the No. i refueling act.

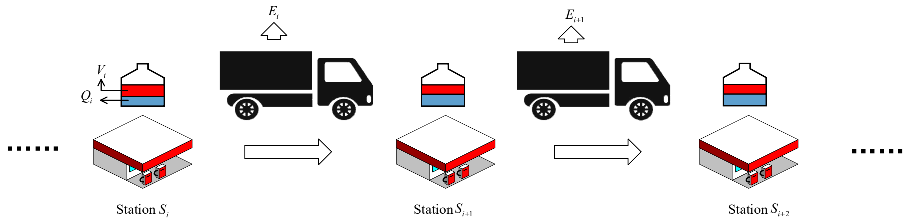


Figure 2. Refueling behavior diagram.

The NOx emission rate per vehicle at each gas station was calculated using the above algorithm, and the daily NOx emissions were further calculated using 10 h of driving per day. The results are shown in Figure 3. Each dot in the diagram represents a refueling act. The orange dots represent vehicles refueling at unregistered gas stations, the purple dots represent refueling at legal gas stations, and it can be concluded that there is a significant correlation between single-vehicle NOx emissions and the oil quality of gas stations. Vehicles refueling at unregistered gas stations for a long time emitted a daily average of 82 g of NOx much higher than 22 g after refueling at legal gas stations. There was a steep rise in NOx emissions after each unregistered refueling act, which caused serious pollution to the environment.

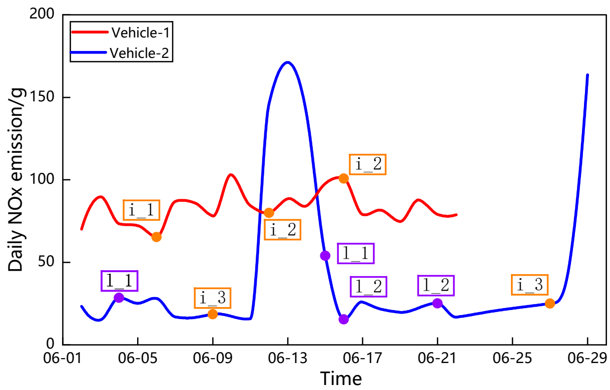


Figure 3. NOx emissions of two vehicles in June.

4. Case Study and Discussion

To verify the effectiveness of the method proposed in this paper, the historical operation data of all diesel vehicles in Tangshan area from November 2019 to March 2020 were retrieved from the platform. The model calculation was performed according to the above-mentioned gas station identification method. The overall distribution of detected refueling points in Tangshan area is shown in Figure 4.

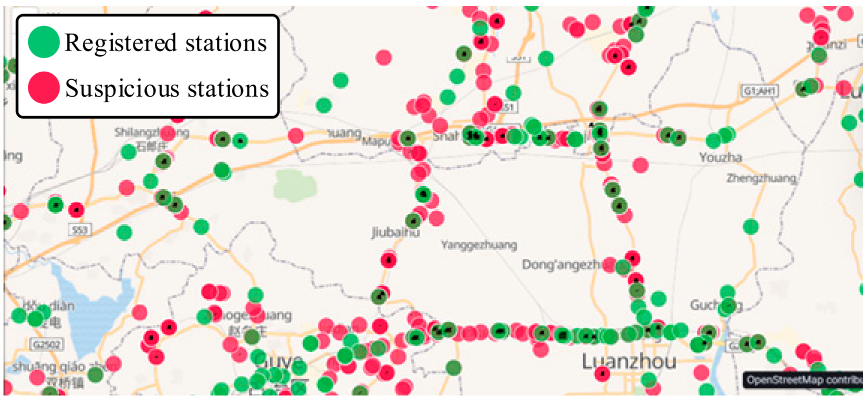
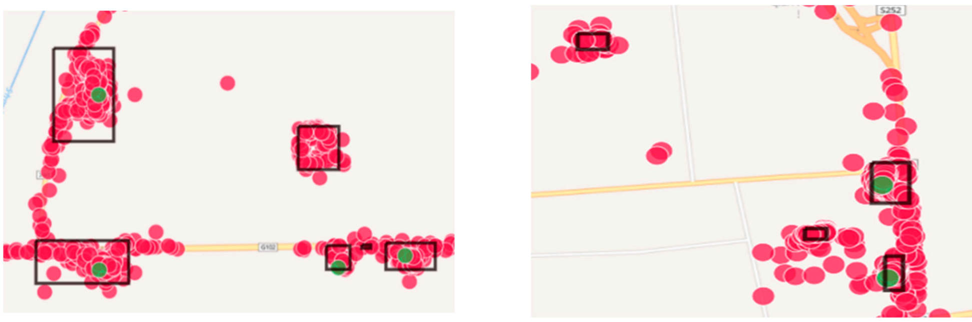


Figure 4. Distribution of detected refueling points in Tangshan.

In the identification results, there are a total of 5086 diesel vehicles that were online during this time period, and also there are a total of 71,087 refueling points in Tangshan. After using the DBSCAN clustering algorithm, 523 suspected gas station locations were obtained, in which a total of 303 gas stations successfully matched the registered legal gas station. It can be inferred that there are a total of 220 suspected unregistered gas stations. Therefore, the proposed method in this paper can effectively identify the refueling point time and location information based on the vehicle operation big data, thereby realizing effective detection and precise positioning of fueling stations.

In addition, we selected a partially enlarged view of a certain area for further analysis, as shown in Figure 5. It can be seen that the extracted refueling points show obvious aggregation characteristics, and the cluster spatial position formed by the aggregation matches well with the registered gas stations. Some categories cannot match known gas stations but have obvious aggregation characteristics. So, there is a certain probability that the clusters are unregistered gas stations. About 8% of the data points in the refueling points did not match the appropriate cluster centers. However, after a comparative analysis with the actual gas station situation, 96% of the suspected gas stations were found to be real illegal gas stations. According to the data quality analysis of a single refueling point, it may be related to the time positioning error caused by the lack of data and the spatial positioning error caused by GPS drift.



(**a**) (**b**)

Figure 5. Partial map of the refueling points in Tangshan. (**a**) Partial map No. 1; (**b**) Partial map No. 2.

In order to further analyze the oil quality of the gas stations in the identification results, based on the identification results in Tangshan area, five registered gas stations and four unregistered gas stations were selected. Based on the identified refueling points, the operation data of heavy diesel vehicles of the same model that had refueling behavior at these gas stations during June 2021 were extracted, which is shown in Figures 6 and 7. On this basis, daily emissions of all refueling vehicles in these gas stations can be calculated based on the NOx emission rate per minute of a single vehicle and vehicle running time. Figure 8 shows the NOx emissions of all refueling vehicles in June for the nine gas stations identified in Tangshan area, of which Nos. 1–5 are legal gas stations, and Nos. 6–9 are unregistered gas stations. It can be seen that there is a significant difference between unregistered and legal gas stations in terms of the amount of data for refueling vehicles and the NOx emissions from a single vehicle. Although the emissions from different vehicles at the same station vary greatly due to the different maintenance conditions and driving habits of drivers, the overall emissions from unregistered stations are much higher than those from legal stations. In addition, the daily emissions from legal gas stations are relatively stable, within the range of 0–40 g, while the daily emissions from unregistered gas stations fluctuate greatly, within the range of 0–300 g. This demonstrates the validity of single-vehicle NOx emissions as an assessment of oil quality at gas stations.

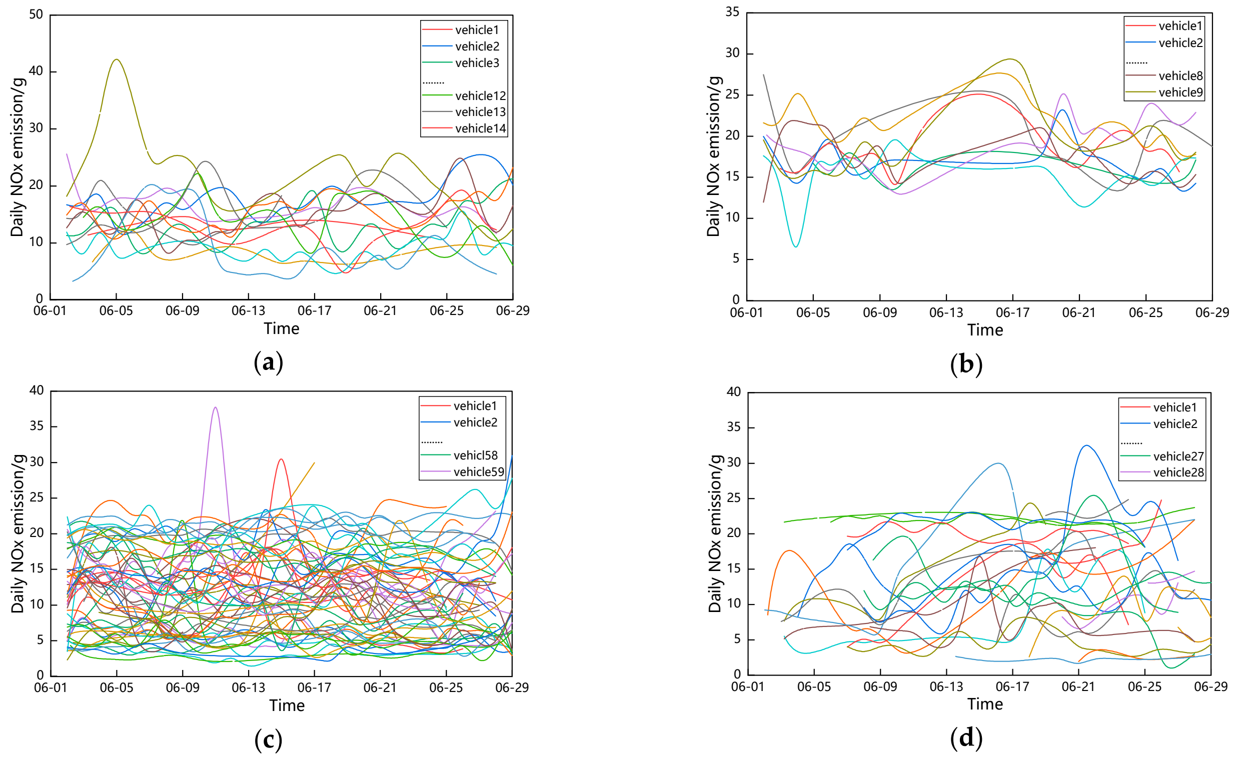


Figure 6. Refueling vehicles NOx emission at legal stations in June. (**a**) Legal gas station No. 1; (**b**) Legal gas station No. 2; (**c**) Legal gas station No. 3; (**d**) Legal gas station No. 4.

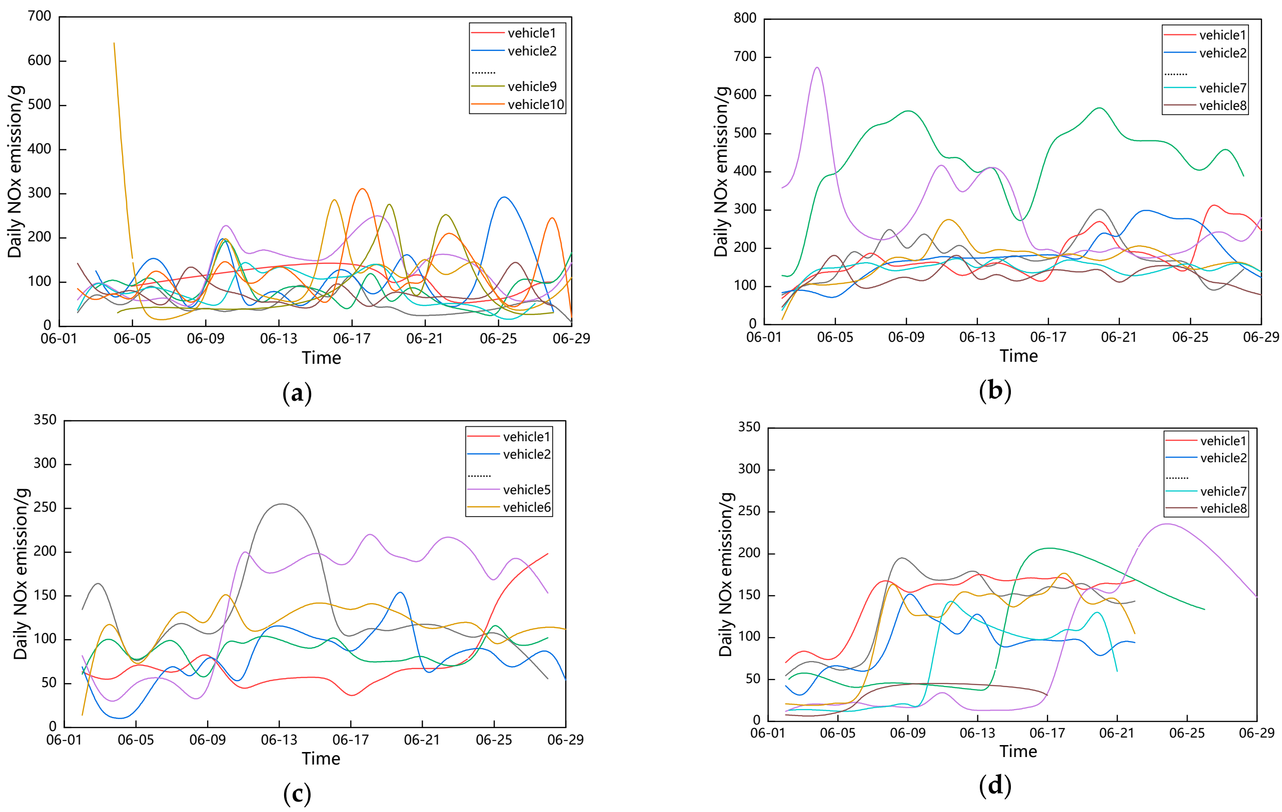


Figure 7. Refueling vehicles NOx emission at unregistered gas stations in June. (**a**) Unregistered gas station No. 6; (**b**) Unregistered gas station No. 7; (**c**) Unregistered gas station No. 8; (**d**) Unregistered gas station No. 9.

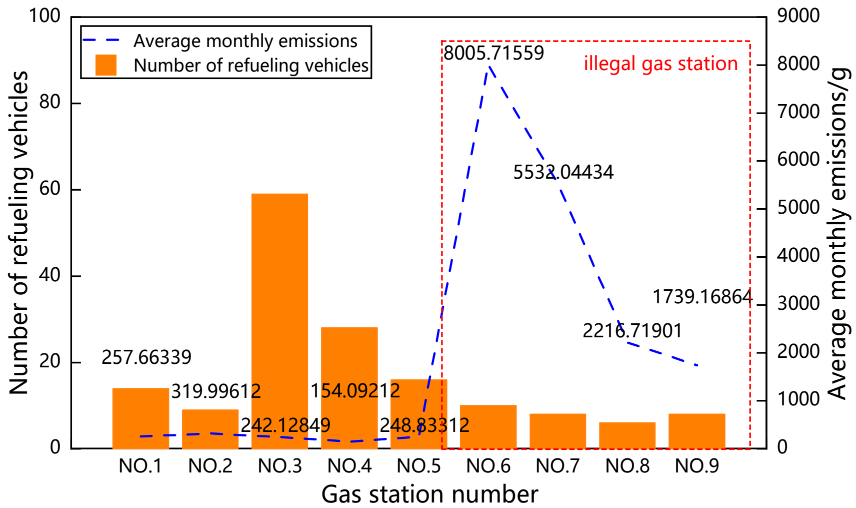


Figure 8. NOx emissions and number of refueling vehicles for 9 gas stations in June.

In addition, in order to analyze the overall situation of oil quality of the nine gas stations in June, the total NOx emissions and the total number of refueling vehicles for each of the nine gas stations in June were obtained. The result is shown in Figure 9. It is indicated that the total number of refueling vehicles in the registered gas stations is generally more than that of the unregistered gas station. However, the number of refueling vehicles at some unregistered gas stations is similar to that of the registered gas station. In terms of total NOx emissions, the maximum emissions of unregistered gas stations are 8005.72 g, and the minimum emissions are 1739.17 g. Overall, the average emissions of the four unregistered gas stations are 4373.41 g, and the average emissions of the five registered gas stations are 244.54 g, which is a 17.88 times difference. It can be seen that the oil quality of the unregistered gas stations is extremely poor, which can easily cause serious environmental pollution problems.

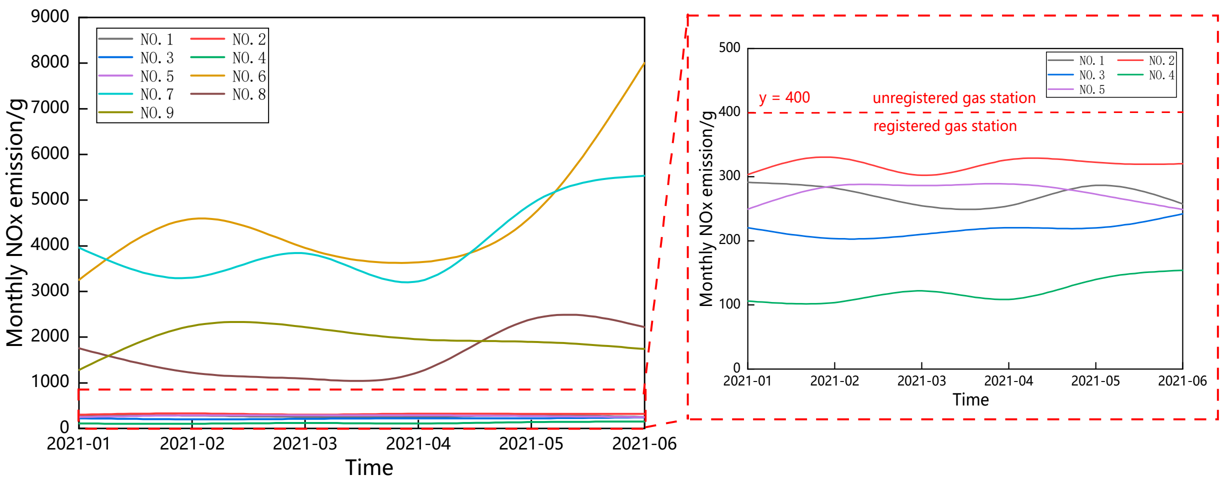


Figure 9. Emission trend change for 9 gas stations.

Further, in order to analyze the stability of oil quality in the nine gas stations, the monthly average total NOx emissions per vehicle for these stations from January to June was calculated, as shown in Figure 9. It can be found that there is a significant difference between the monthly NOx emissions from both unregistered and legal gas stations. The average monthly emissions from unregistered gas stations fluctuate in range and have extremely large emission values, and the quality of the oil is extremely unstable. On the contrary, legal gas stations are below 400 g and remain largely stable over time. Based on this conclusion, it is possible to distinguish between legal and unregistered gas stations by setting a threshold value for total monthly NOx emissions, which was set to 400/g in this batch of data, which can accurately distinguish between unregistered and legal gas stations. In addition, the above analysis further verifies the validity and reliability of the unregistered gas station identification method proposed in this paper.

In summary, the gas station identification method proposed in this paper was successfully applied to Tangshan area. Based on the oil quality analysis results of the identified gas stations, the accuracy of the identification of gas stations was further verified. Based on the full amount of diesel vehicle operation data, a regional gas station supervision map was constructed, which effectively detected and located gas stations in Tangshan area. In addition, the results provide a reliable and reasonable basis for the ecological protection department law enforcement and supervision.

5. Conclusions

This paper presents a gas station detection and location method based on the Heavy-duty Vehicle Remote Emission Service and Management Platform. A vast quantity of real-time diesel vehicle monitoring data in Tangshan area was collected from this big data platform to verify the effectiveness of the presented method. This paper formulates the selection rule of the detection window based on the refueling characteristics. The CART algorithm was applied to identify the refueling points. The DBSCAN cluster algorithm was applied to locate the gas stations. On this basis, based on the information of registered gas stations, the detection and location of gas stations are realized. This paper also analyzed the oil quality of the nine gas stations based on the data of refueling points, which verified the accuracy of the proposed method. The analysis results in Tangshan area showed that the proposed method could detect and locate gas stations. The detection results can effectively provide a detailed and reliable basis for supervision and investigation by law enforcement agencies and accurately and effectively regulate unregistered gas stations. The presented method can ensure the robustness and accuracy of the recognition of the refueling behavior. These were achieved with a relatively small calculation effort, which makes it implementable in a large number of application scenarios, thereby significantly ensuring fuel quality at gas stations, reducing carbon and NOx emissions from heavy-duty vehicles, and contributing to carbon neutrality and carbon peaking goals.

Author Contributions: Data curation, P.L.; Funding acquisition, Z.J.; Investigation, G.L.; Methodology, Z.W.; Project administration, Y.D.; Resources, S.X.; Writing—original draft, D.C. and Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by State Environmental Protection Key Laboratory of Vehicle Emission Control and Simulation (Grant Number: VECS2021S03), Research and Demonstration of RPTR(remote sensing-period testing insection-telematics-roadside inspection) Integrated Mobile Source Monitoring System in Tangshan (Grant Number: 19150259E) and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant Number: XDA23010202).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Greene, L.A. EHPnet: United Nations Framework Convention on Climate Change. *Environ. Health Perspect.* **2000**, *108*, 353. [CrossRef]
2. Cui, D.; Wang, Z.; Zhang, Z.; Liu, P.; Wang, S.; Dorrell, D.G. Driving Event Recognition of Battery Electric Taxi Based on Big Data Analysis. *IEEE Trans. Intell. Transp. Syst.* **2021**, 1–10. [CrossRef]
3. Liang, X.; Zhang, S.; Wu, Y.; Xing, J.; He, X.; Zhang, K.M.; Wang, S.; Hao, J. Air quality and health benefits from fleet electrification in China. *Nat. Sustain.* **2019**, *2*, 962–971. [CrossRef]

4. Grigoratos, T.; Fontaras, G.; Giechaskiel, B.; Zacharof, N. Real world emissions performance of heavy-duty Euro VI diesel vehicles. *Atmos. Environ.* **2019**, *201*, 348–359. [CrossRef]
5. Lv, L.; Ge, Y.; Ji, Z.; Tan, J.; Wang, X.; Hao, L.; Wang, Z.; Zhang, M.; Wang, C.; Liu, H. Regulated emission characteristics of in-use LNG and diesel semi-trailer towing vehicles under real driving conditions using PEMS. *J. Environ. Sci.* **2020**, *88*, 155–164. [CrossRef] [PubMed]
6. Giechaskiel, B.; Schwelberger, M.; Delacroix, C.; Marchetti, M.; Feijen, M.; Prieger, K.; Anderssong, S.; Karlssong, L.H. Experimental assessment of solid particle number Portable Emissions Measurement Systems (PEMS) for heavy-duty vehicles applications. *J. Aerosol Sci.* **2018**, *123*, 161–170. [CrossRef]
7. Quan-shun, Y.; Jian-wei, T.; Yun-shan, G.; Li-jun, H.; Zi-hang, P. Application of Diesel Particulate Filter on in-use On-road Vehicles. *Energy Procedia* **2017**, *105*, 1730–1736. [CrossRef]
8. Tan, L.; Guo, Y.; Liu, Z.; Feng, P.; Li, Z. An investigation on the catalytic characteristic of NO x reduction in SCR systems. *J. Taiwan Inst. Chem. Eng.* **2019**, *99*, 53–59. [CrossRef]
9. Ministry of Ecological Environment. *Limits and Measurement Methods for Emissions from Diesel Fueled Heavy-Duty Vehicles (CHINA VI) GB 17694-2018*; Chinese Academy of Environmental Sciences: Beijing, China, 2018.
10. Iqbal, M.S.; Niazi, Y.A.K.; Khan, U.A.; Lee, B.W. Real-time fault detection system for large scale grid integrated solar photovoltaic power plants. *Int. J. Electr. Power Energy Syst.* **2021**, *130*, 106902. [CrossRef]
11. Garg, S.; Kaur, K.; Kumar, N.; Rodrigues, J.J. Hybrid Deep-Learning-Based Anomaly Detection Scheme for Suspicious Flow Detection in SDN: A Social Multimedia Perspective. *IEEE Trans. Multimed.* **2019**, *21*, 566–578. [CrossRef]
12. Alizadeh, M.; Hamilton, M.; Jones, P.; Ma, J.; Jaradat, R. Vehicle Operating State Anomaly Detection and Results Virtual Reality Interpretation. *Expert Syst. Appl.* **2021**, *177*, 114928. [CrossRef]
13. He, R.; Xu, P.; Chen, Z.; Luo, W.; Su, Z.; Mao, J. A non-intrusive approach for fault detection and diagnosis of water distribution systems based on image sensors, audio sensors and an inspection robot. *Energy Build.* **2021**, *243*, 110967. [CrossRef]
14. Luo, W.; Liu, W.; Gao, S. Normal Graph: Spatial Temporal Graph Convolutional Networks based Prediction Network for Skeleton based Video Anomaly Detection. *Neurocomputing* **2020**, *444*, 332–337. [CrossRef]
15. You, J.; Lou, E.; Afrouziyeh, M.; Zukiwsky, N.M.; Zuidhof, M.J. A supervised machine learning method to detect anomalous real-time broiler breeder body weight data recorded by a precision feeding system. *Comput. Electron. Agric.* **2020**, *185*, 106171. [CrossRef]
16. Ece, A.; Vpa, B. Outlier detection and quasi-periodicity optimization algorithm: Frequency domain based outlier detection (FOD). *Eur. J. Oper. Res.* **2021**, *291*, 560–574.
17. Chung, H.C.; Ahn, J. Subspace rotations for high-dimensional outlier detection. *J. Multivar. Anal.* **2020**, *183*, 104713. [CrossRef]
18. Liu, Y.; Tan, J. Green traffic-oriented heavy-duty vehicle emission characteristics of China VI based on portable emission measurement systems. *IEEE Access* **2020**, *8*, 106639–106647. [CrossRef]
19. Wang, J.; Wu, Q.; Liu, J.; Yang, H.; Yin, M.; Chen, S.; Guo, P.; Ren, J.; Luo, X.; Linghu, W.; et al. Vehicle emission and atmospheric pollution in China: Problems, progress, and prospects. *PeerJ* **2019**, *7*, e6932. [CrossRef] [PubMed]
20. Liang, K.; Zhang, Z.; Liu, P.; Wang, Z.; Jiang, S. Data-driven ohmic resistance estimation of battery packs for electric vehicles. *Energies* **2019**, *12*, 4772. [CrossRef]