LETTER • OPEN ACCESS

Using snapshot measurements to identify high-emitting vehicles

View the article online for updates and enhancements.

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

Using snapshot measurements to identify high-emitting vehicles

Minghao Qiu[1,3,*] and Jens Borken-Kleefeld[2,*]

1   Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, United States of America
2   International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
3   Current address: Department of Earth System Science, Stanford University, Stanford, USA.
*   Authors to whom any correspondence should be addressed.

E-mail: mhqiu@mit.edu and borken@iiasa.ac.at

Abstract

Policy makers have long been interested in detecting 'high-emitters', a supposedly smallfraction of vehicles that make disproportionally large contributions to total fleet emissions. However, existing identification schemes often exclusively rely on snapshot measurements (i.e. emissions within less than a second), and thus simply identify vehicles with high instantaneous emissions, instead of vehicles with high average emissions over a driving period as regulated by emission standards. We design a comprehensive scheme to address this challenge by combining fleetwide remote sensing measurements with detailed second-by-second emission measurements from individual vehicles. We first determine the trip-average NO_x emission rates of individual vehicles in a Euro-5 diesel fleet measured across European locations; this allows, second, to calculate the fraction and emission contributions of high-emitters based on trip-average emission. We demonstrate that the identification of high-emitters is quite uncertain as long as it is based on single snapshots only; but 80% of the high-emitters can be identified with over 75% precision with five or more repeated measurements of the same vehicle. Compared to the conventional detection schemes, our scheme can increase the identified high-emitters and associated emission reductions by over 140%. Our method is validated and shown to be superior to the conventional interpretation of snapshot measurements.

1. Introduction

Motor vehicles are a key contributor to air pollutant emissions, most notably nitrogen oxides (NO_x), which poses substantial threat to air quality, public health, and climate change worldwide [1–6]. It has been previously suggested that a small fraction of vehicles (i.e. 'high-emitters') dis-proportionally contribute to a large fraction of total fleet emissions, due to lack of maintenance, tampering, and/or usage of defeat devices [7–11]. Therefore, identification and then repairing or removing these high-emitters is often viewed as a cost-effective strategy to reduce on-road vehicle emissions [12–14]. However, the existing inspection programs in many countries often fail to detect high-emitters in the real world [15–21]. Increasing effort has been invested in measuring on-road emissions from in-use vehicles (e.g. the in-use surveillance tests by US Environmental Protection

Agency [22]). However, these tests only cover a limited number of vehicles, as testing every in-use vehicle constitutes a significant logistical and financial burden to both vehicle owners and authorities.

Remote sensing (RS) offers great potential for identification of high-emitters, as RS devices can remotely measure the emission rates of a large number of vehicles during pass-by without interrupting traffic [23–25]. However, only a few countries/regions have used or proposed to use on-road RS measurements to assist the identification of high-emitters (e.g. mainland China [26], Hong Kong [27], Scotland [28], and certain states in the US [25]). One key issue is that RS devices only take snapshot measurements as vehicles pass by the sensor (equivalent to emissions within less than a second), and therefore cannot characterize the average emission level of individual vehicles over a representative driving period [29–31]. Existing identification programs have

used excessively high identification thresholds (e.g. 1500 ppm for NO emission for heavy-duty trucks in China, ∼9× the China 5 standard [32, 33]); then high instantaneous emission rates do indeed suggest the identified vehicles have high average emissions, but such high identification thresholds may result in a large number of undetected potential high-emitters.

Most previous studies define high-emitters and estimate their contributions to total fleet emissions purely based on snapshot measurements derived from RS. These studies find highly-skewed distributions of the instantaneous emissions and calculate that the dirtiest 10% vehicles contributed to over 50% of the total pollutant emissions, mainly focusing on gasoline fleets [8, 11, 30, 34–37]. However, the distributions of instantaneous emissions measured by RS, have shown to be very different from the distributions of average emission factors calculated with high-frequency emission measurements from chassis dynamometer or portable emissions measurement system (PEMS) [31]. Contrary to RS measurements, measurements from chassis dynamometer or PEMS consist of continuous records of vehicle emissions with resolutions of one measurement per second for an extended period of time (e.g. half an hour), but are only available for a very limited number of vehicles due to the measurement cost [38].

Only a few studies define high-emitters based on average vehicle emissions and combine high-frequency measurements and snapshot measurements to identify them. Huang *et al* define high-emitters as vehicles whose average emission factors are above two times the emission limits, and then use the 99th percentile of the instantaneous emissions from chassis dynamometer tests of the vehicles that passed tests as threshold for high-emitter identification [39]. Researchers have also used statistical models such as neural network methods to predict high-emitters using snapshot measurements, speed, vehicle age, and meteorology information [40, 41]. However, these approaches do not establish a relationship between snapshot emission and average emission, and often just focus on the binary label (whether a vehicle is a high-emitter or not). They also fail to explore the advantage of using repeated instantaneous measurements, and lack rigorous evaluation of the uncertainty and effectiveness of the identifications. To date, very little is known about the fraction of high-emitters (defined based on their average emission factors) in real-world fleets, their contributions to total emissions, and the associated precision and effectiveness of identifying these high-emitters with snapshot measurements.

Here, we develop a comprehensive scheme which combines second-by-second PEMS and chassis dynamometer data with a large RS dataset, to identify vehicle with high average emission factors. We focus on NO_x emissions from diesel passenger cars certified to the Euro 5 emission standards (EU-5D), due

to the importance of NO_x from diesel vehicles for air quality and relatively low measurement error of the RS devices [42]. Applying our method to over 130 000 RS measurements of EU-5D passenger cars from nine European cities, we estimate the distribution of average emission factors and calculate the fraction of high-emitting and clean vehicles for each fleet. We then quantify the accuracy and uncertainty of identifying individual high-emitting vehicles with a few snapshot RS measurements, as a function of the number of repeated RS measurements and the classification rules. Finally, we present a versatile method to validate our algorithm for estimating the average emission factors, demonstrating that our algorithm is significantly superior to the conventional interpretations of snapshot measurements. We conclude by discussing the implications to policy makers and practitioners on using RS snapshot measurements to identify high-emitters.

2. Method

Throughout our analysis, a high-emitting vehicle is defined as having an average emission factor higher than an absolute threshold value (here, usually two times or more the emission limit value for EU-5D). To identify them with RS measurements, we first design an iterative algorithm that can estimate the distribution of vehicle-level average emission factors in the measured fleet. To do this, we use an extensive set of second-by-second PEMS/chassis measurements to establish the relationship between instantaneous emissions and the average emission of the test cycle, i.e. the variability of instantaneous emissions around the mean. With these variability relationships, our algorithm then iteratively estimates the distribution of the average emission factors that can reproduce the set of observed instantaneous RS measurements. We then use the derived distribution of average emission factors to simulate a set of instantaneous emissions with known average vehicle emissions. Using this simulation dataset, we validate our algorithm and quantify the precision and effectiveness of identifying individual high-emitting and clean vehicles, under different numbers of repeated RS measurements and classification rules. We further perform several sensitivity analyses to explore how varying algorithm assumptions and modifications of the second-by-second measurements influence our results (see supplementary methods available online at stacks.iop.org/ERL/17/044045/mmedia).

2.1. On-road remote sensing (RS) measurements
We use a collection of 131 284 RS records of the EU-5D passenger cars measured in 9 cities (Antwerp, Basel, Bruges, Ghent, Gothenburg, London, undisclosed cities in Spain, Stockholm, and Zurich) during 2011–2019 [24]. Due to the strong association between ambient temperature and vehicle emissions

[43], we differentiate RS records into three categories based on ambient temperature: Below 10°C, 10°C–20°C, and above 20°C. To ensure the comparability between RS measurements and emission limit values established in laboratory test environments, we only focus on vehicles whose instantaneous vehicle specific power (VSP, calculated following [44]) are in the range of 3–22 kW per metric ton. We calculate fuel-based NO_x emission factors (unit: $g\,NO_x\,kg^{-1}$ fuel) as the product of NO_x/CO_2 ratio (unit: ppm/ppm), the molecular weights of CO_2 and NO_X, and CO_2 intensity of diesel ($3.13\,kg\,CO_2\,kg^{-1}$ diesel following [45]). We omit the highest and lowest records (0.1% each) to remove outliers of instantaneous measurements. Our final sample consists of 79 576 RS records. A summary of the testing conditions and characteristics of measured fleets can be found in table S1.

2.2. Second-by-second laboratory test data

,We use a collection of 163 second-by-second test cycles performed on 48 EU-5D passenger cars from the European Research for Mobile Emission Sources (ERMES) database [46]. This collection includes 18 ERMES cycle tests, 39 New European Driving Cycle (NEDC) tests, 60 Common Artemis Driving Cycle (CADC) tests, 18 Worldwide harmonized Light vehicles Test Cycles (WLTC) tests, and 28 Portable Emissions Measurement System (PEMS) tests. Their test lengths vary from 900 s in CADC cycles to over 6000 s in PEMS trips. To be comparable with the driving conditions measured at the RS sites, we only use the test cycles performed under urban and rural conditions (dropping cycles under motorway conditions), and only use data with VSP of $3–22\,kW\,ton^{-1}$.

For each test cycle, we calculate the fuel-based NO_x emission factors in a way similar to the RS data. Average NO_x emission factors of test vehicles (over the test cycle) range from $2.2\,g\,NO_x\,kg^{-1}$ fuel to $25.2\,g\,NO_x\,kg^{-1}$ fuel, covering both clean and high-emitting vehicles. Following recommendations from researchers directly working with these data, we calculate the three second moving averages of NO_x and CO_2 emissions to address the potential issue of misaligned NO_x and CO_2 measurements. The highest and lowest (0.5% each) emission factors are omitted for PEMS/chassis measurements to remove the outliers of the measurements.

2.3. Decomposition of RS emission records

For a vehicle i, an observed RS record of instantaneous emission can be decomposed into the sum of its (unknown) average emission factor and the variability around its average emission:

$$RS_i = AE_i + f_i$$

where RS_i denotes the instantaneous emission of vehicle i measured by RS. AE_i denotes the unknown average emission factor of vehicle i over some average

driving conditions (characterized by the test cycles in our database). f_i is the underlying instantaneous variability of this measurement. f_i can be viewed as a draw from an underlying distribution of the instantaneous variability of vehicle i ($F_i, f_i \sim F_i$), conditioned on the VSP value. The mean value of F_i is always zero, but the shape of F_i describes the variability of instantaneous emissions around the average emission factor of vehicle i. F_i is difficult to derive from RS data, as most vehicles were only measured a few times and the small number of repeated measurements (usually below 5) is insufficient to constrain the variability distribution. Therefore, we use PEMS/chassis measurements to derive information of the variability distributions.

2.4. Iterative algorithm to estimate the distribution of average emission factors

We design an iterative algorithm to estimate the distribution of average emission factors using a set of RS measurements in which every vehicle is measured only once. Given a set of variability distributions derived from PEMS/chassis measurements, our algorithm estimates the distribution of average emission factors which produces a set of instantaneous emission records whose distribution is indistinguishable from the original distribution of the RS records. In this section, we only provide a high-level summary of the algorithm, with more details reported in the SI.

The main idea of the algorithm is to match each RS record (and therefore each vehicle in the RS data) with a second-by-second measurement profile (i.e. a test cycle). The matched second-by-second profile characterizes the potential relationship between instantaneous emissions and the average emission of the matched vehicle measured in RS data. The average emission factor of the associated second-by-second profile is used to estimate the average emission factor of the matched vehicle in the RS data. The algorithm performs the matching process in an iterative manner. The algorithm starts with an initial estimate of the average vehicle emission factor for each vehicle in the RS data. At each iteration, the algorithm matches a vehicle in the RS data with one second-by-second measurement of similar average emission factor. After the matching step at each iteration, the algorithm simulates a set of instantaneous emission measurements by sampling one instantaneous emissions record from the matched second-by-second data. Only the instantaneous emission record with a similar VSP value as the original RS record is sampled from the second-by-second measurement profile. The algorithm compares the distribution of the simulated instantaneous emission with the distribution of the RS data, and then updates the estimated average emission factor of each vehicle in the RS data, to reduce the distance between the distribution of simulated instantaneous records and the distribution of RS records. The algorithm terminates if no further improvement could be

made to reduce the distance between these two distributions.

2.5. Validation of the algorithm

We design the following experiment to validate our algorithm. We create a set of instantaneous emission records by sampling from the PEMS/chassis measurement data, as proxies for RS measurements. By doing this, the average emission factors associated with these instantaneous emissions are known to us. We then apply the iterative algorithm to the simulated test set to estimate the distribution of average emission factors and compare the results with distributions of true average emission factors of the test set. We randomly select half of the PEMS/chassis measurements (that cover both clean and high-emitting vehicles) to generate the test dataset and use the other half as inputs of our algorithm.

2.6. Identification of individual high-emitting and clean vehicle

To identify individual high-emitting and clean vehicles, we first simulate a hypothetical fleet using the estimated distribution of average emission factors and the PEMS/chassis measurements. We illustrate the idea with the fleet measured in Zurich with ambient temperature above 20°C due to the largest number of records available. The average emission factor of each vehicle in the hypothetical fleet is randomly drawn from the distribution of the estimated average emission factors for Zurich (derived from our iterative algorithm). Each vehicle is then matched with a PEMS/chassis cycle with a similar average emission factor. For each vehicle, we then simulate ten independent instantaneous emission records with ten independent random draws from the matched PEMS/chassis cycle. In practice, the repeated instantaneous emission measurements for one vehicle can be obtained if a vehicle either drives through one device multiple times (which occasionally occurred in real-world measurement campaigns), or is measured by multiple deployed RS devices (in a future campaign).

We focus on the identification of 'super high-emitters', defined here as vehicles with average emission factors more than five times the type approval limit ($5 \times 3.5 = 17.5$ g NO_x kg^{-1} fuel), since more than 90% of EU-5D passenger cars already have average on-road NO_x emissions above the type approval limit. We convert the distance-based emission limit value over the test cycle (here the NEDC for the Euro 5 cars) to a fuel-based emission limit using the cycle average fuel economy, which has been measured well for classes of vehicles. For each potential cut-off threshold of RS measurement, we calculate the resulting identification precision (i.e. correctly identified high-emitters/identified high-emitters) and the identified fraction (i.e. correctly identified high-emitters/actual high-emitters, or 'recall'). We also apply the similar idea to the identification of 'clean

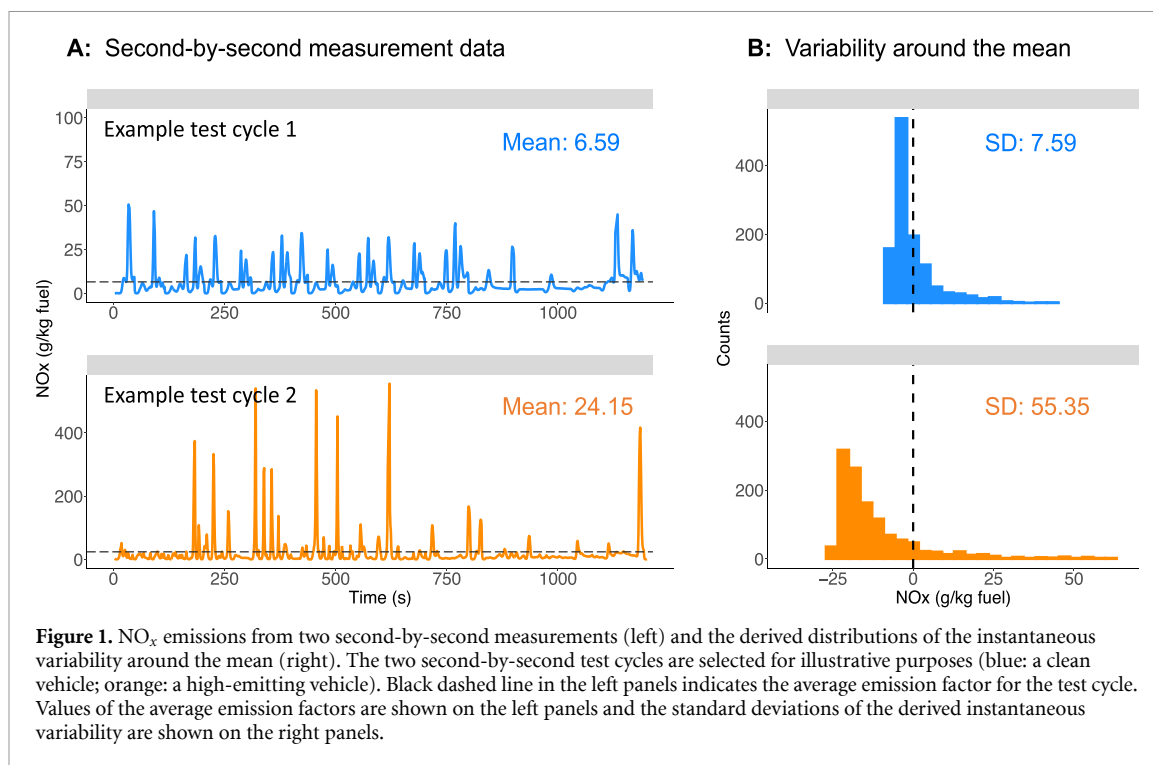vehicles' with average on-road emissions below the type approval limit (3.5 g NO_x kg^{-1} fuel).

With repeated instantaneous measurements of one vehicle, we evaluate two simple classification rules. The first rule classifies a vehicle as a high-emitter/clean vehicle if the average of repeated instantaneous emissions are above/below a threshold ('average rule'). The second rule counts all measurements and classifies a vehicle as a high-emitter/clean vehicle only if all repeated measurements are above/below some threshold ('count all rule'). The two rules are selected based on their simplicity to operate. The 'count all rule' can be overly stringent under a large number of repeated measurements (e.g. very few vehicles may have all five repeated measurements above certain threshold), but the high precision levels might be appealing to policy makers as it puts a light burden on drivers of normal vehicles.

3. Results

3.1. Determining the distribution of average emission factors

Figure 1 shows the second-by-second measurements and the derived distributions of instantaneous variability for two vehicles with different average emission factors, as illustrative examples. Over an extensive driving period, we find that the instantaneous emissions are highly skewed towards lower values. On average for the 163 test cycles, 65% of the instantaneous records are below the average emission rate of the respective trip. Therefore, it is more likely than not that a random snapshot RS measurement has an emission rate less than the average value of the same vehicle. At the same time, the few extremely high instantaneous emission records are also not representative for the average behavior of the vehicle; the 99th percentile of the instantaneous emissions is 1.6–9.8 times the average emission factor. We also find large differences in the shape of variability distributions between clean and high-emitting vehicles. The variability distributions of high-emitting vehicles are flatter with a more negative mode and a longer tail, despite all distributions having zero mean. This highlights the importance of having an extensive set of second-by-second measurement data which can differentially characterize the variability of high-emitting and clean vehicles.

Differences between instantaneous and average emission shown in the second-by-second measurements are reflected in the analysis performed on EU-5D fleets measured across Europe. Figure 2 shows the distributions of instantaneous RS measurements and the derived average emission factors for each city in our RS dataset. Consistent with previous literature on EU-5D vehicles, we observe excessively high instantaneous NO_x emissions with mean RS emission rates of 9.2–20.1 g NO_x kg^{-1} fuel (2.6–5.7 times the type approval limit for EU-5D)

Figure 1. NO_x emissions from two second-by-second measurements (left) and the derived distributions of the instantaneous variability around the mean (right). The two second-by-second test cycles are selected for illustrative purposes (blue: a clean vehicle; orange: a high-emitting vehicle). Black dashed line in the left panels indicates the average emission factor for the test cycle. Values of the average emission factors are shown on the left panels and the standard deviations of the derived instantaneous variability are shown on the right panels.
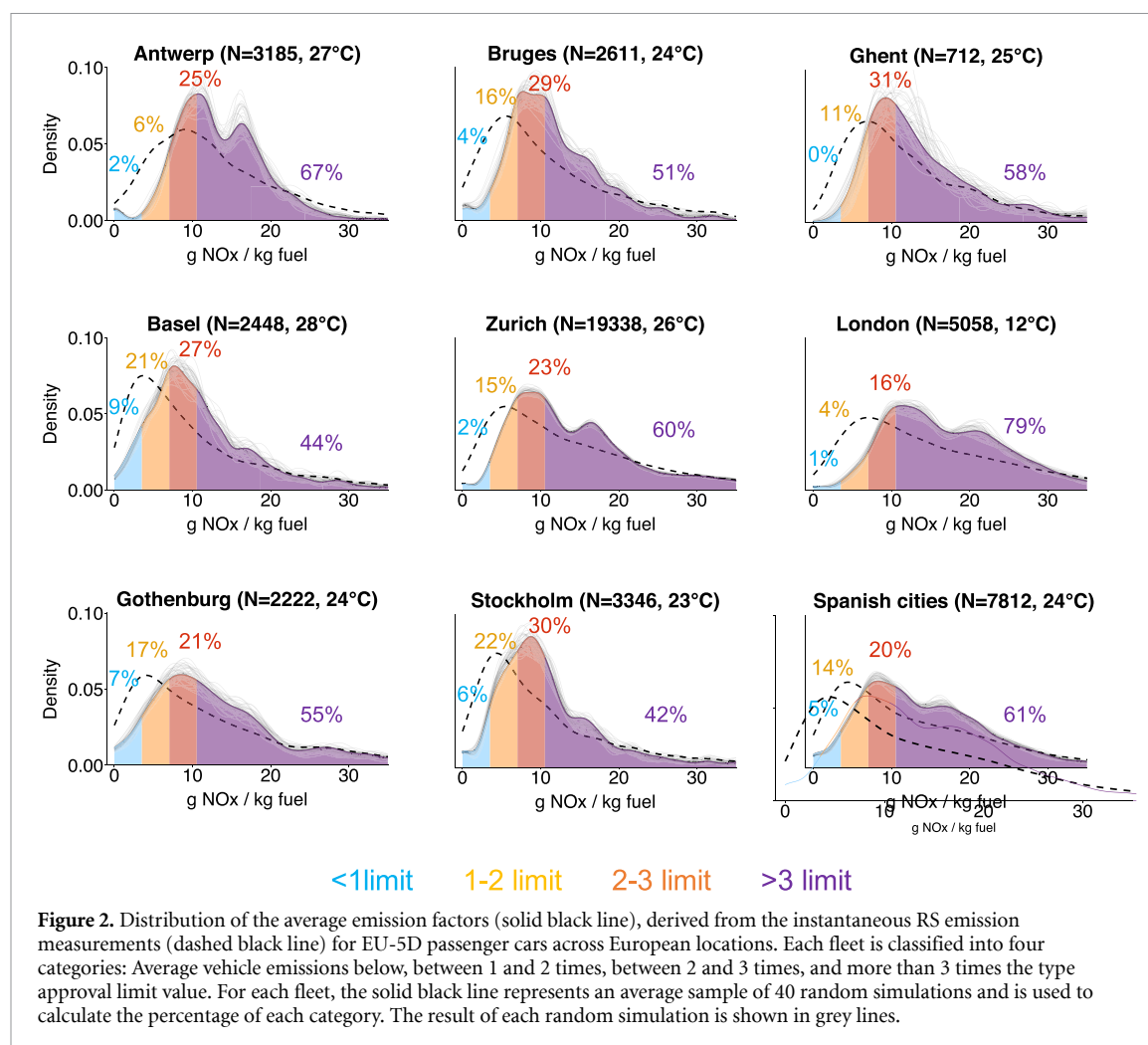
across different cities and temperature conditions. These high instantaneous emissions are associated with vehicles with excessively high average emission factors. However, distributions of the estimated average emission factors (solid lines in figure 2) differ substantially from distributions of the RS emission rates (dashed lines in figure 2). The distribution of average emission factors peaks at larger values (approximately $10 \, g \, NO_x \, kg^{-1}$ fuel) compared to the distribution of RS records (approximately $5 \, g \, NO_x \, kg^{-1}$ fuel); in addition, there are significantly less extreme values (both low and high). This suggests that in fact there are substantially less vehicles with extremely low or high average on-road emissions than previous estimates purely based on single instantaneous RS records. For example, we estimate that only 2.2% of the measured fleet in Zurich have an average emission factor below the type approval limit, while 11% of the instantaneous RS records are below the same limit value (an overestimation of 413%). Across nine cities, only 0.1%–8.7% measured vehicles have an average emission factor below the type approval limit, while 8%–20% of the RS records are below the same limit value. We observe similar differences between the estimated average emission factors and instantaneous measurements across all temperature conditions and ages of the measured vehicles (see figure S1). These results are largely independent of the algorithm assumptions and treatment of PEMS/chassis measurements (see supplementary methods).

There is substantial variability of the estimated average emission factors of individual vehicles within each city and between cities. The average emission factors decrease as the ambient temperature increases

(see figure S1); this reflects the 'thermal window' mechanism reported earlier that the emission control devices are optimized for the temperature conditions of the type approval testing procedures (around 24°C) [43]. There is no evident relationship between average emission factors and ages of the vehicles, extending earlier analysis [47]: Newer EU-5D vehicles do not have lower emission rates than older vehicles (see figure S1). We also find distinctly different emission rates of different vehicle brands (figure S2). For example in Zurich, the cleanest brand (BMW, $n = 2373$) has a median average emission factor of $11 \, g \, NO_x \, kg^{-1}$ fuel, while the dirtiest brand (Renault, $n = 952$) has a median average emission factor of $31 \, g \, NO_x \, kg^{-1}$ fuel (or 2.9 times the cleanest brand). However, there is large variability of vehicles' average emission factors within each brand as well, with most vehicles emitting excessively high NO_x over a representative test cycle.

3.2. Identification of individual high-emitting and clean vehicles

Next, we evaluate the uncertainty and effectiveness of using RS snapshot measurements to identify individual high-emitting vehicles with inferred average emission factors. As an illustration, we focus on the Zurich fleet measured with ambient temperature >20°C. Our algorithm estimates that 28% of the measured vehicles in Zurich are super high-emitters with average emission factors more than 5 times the type approval limit, which accounts for 76% of the total emissions of the fleet (calculated with their average emission factors). The identification of high-emitting vehicles is highly uncertain if based on one
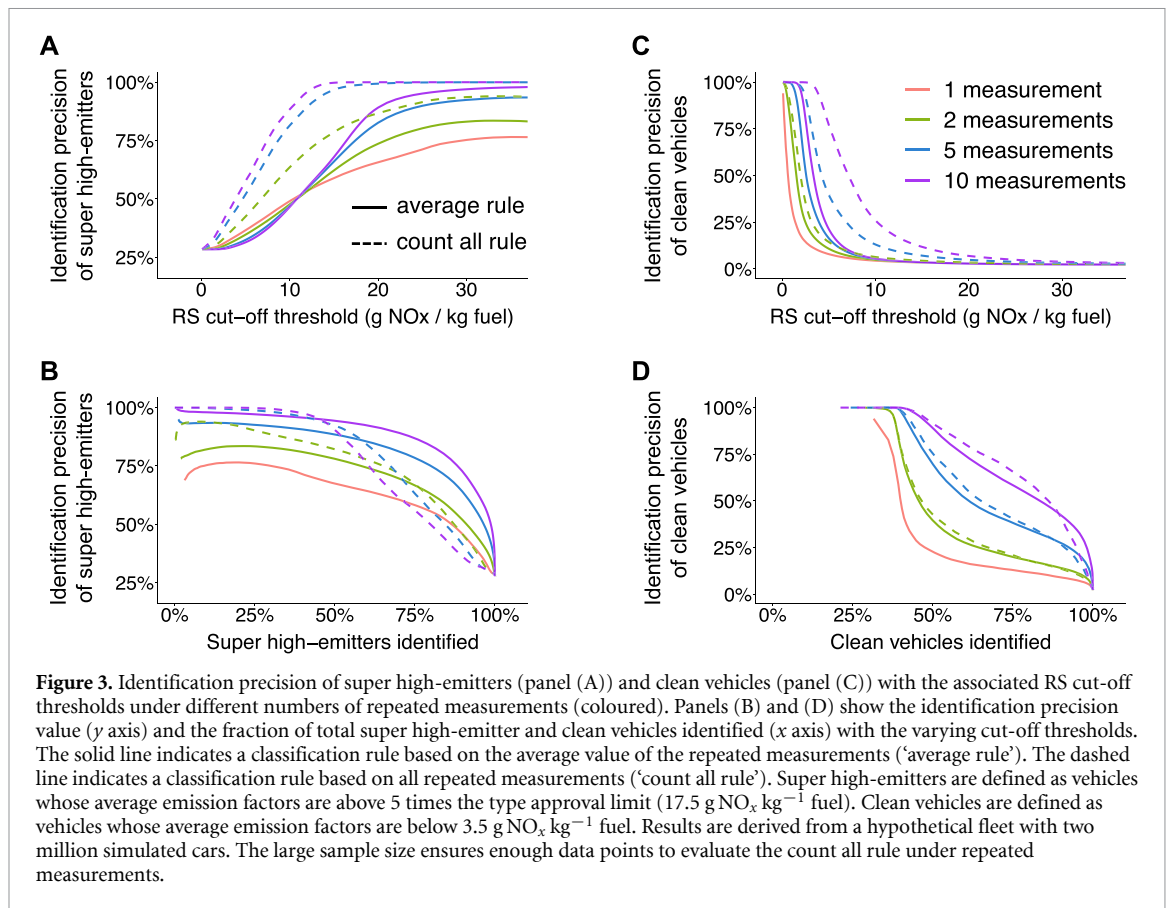
Figure 2. Distribution of the average emission factors (solid black line), derived from the instantaneous RS emission measurements (dashed black line) for EU-5D passenger cars across European locations. Each fleet is classified into four categories: Average vehicle emissions below, between 1 and 2 times, between 2 and 3 times, and more than 3 times the type approval limit value. For each fleet, the solid black line represents an average sample of 40 random simulations and is used to calculate the percentage of each category. The result of each random simulation is shown in grey lines.

RS record alone. For instance, if the cut-off threshold is set at 17.5 g NO_x kg^{-1} fuel (i.e. the car will be naively classified as a super high-emitter based on this one record), there is only a chance of 63% that this car is actually a super high-emitter (figure 3(A)). Increasing the cut-off threshold will increase precision to detect actual super high-emitters. But surprisingly, the identification precision of a super high-emitter levels off at about 77%, even if the cut-off threshold is increased further. This is because vehicles with relatively low average emissions can also have quite high instantaneous emissions, making it impossible to distinguish between super high-emitters and vehicles with less high levels of average emissions based on only one instantaneous measurement.

The precision increases substantially, however, if the same vehicle is measured several times; this is shown by simulations of multiple independent instantaneous emissions based on RS and PEMS/chassis data, not from the rather limited repeated measurements in the RS data. Here, we evaluate two classification rules—one based on the average of repeated instantaneous emissions ('average rule', solid lines in figure 3) and the other that counts all measurements ('count all rule', dashed lines in figure 3).

Already with two repeated RS records that are both five times above the type approval limit value, the precision of identifying super high-emitters increases to 83%; with five RS records the identification is 99% accurate ('count all rule'). An increasing number of repeat RS records also allows lowering the RS threshold while maintaining the precision level. For instance, if 75% precision is required for detection of super high-emitters, the cut-off threshold for a single RS record would need to be 30 g NO_x kg^{-1} fuel (or 8.6 times the type approval limit value). The threshold could substantially decrease to 14 g NO_x kg^{-1} fuel or 8.7 g NO_x kg^{-1} fuel respectively with two or five repeated records (using the 'count all rule'). Under the same RS thresholds and same number of repeated measurements, the 'average rule' is less stringent than the 'count all rule', and hence results in a lower precision level for the same threshold (the difference between solid and dashed lines in figure 3(A)).

Similar relations hold for the identification of clean vehicles (figure 3(C)): Only 11% of vehicles with an instantaneous RS record below the emission limit are actually clean vehicles with average emissions below the type approval limit. To ensure a 75% precision with a single record, the RS threshold would
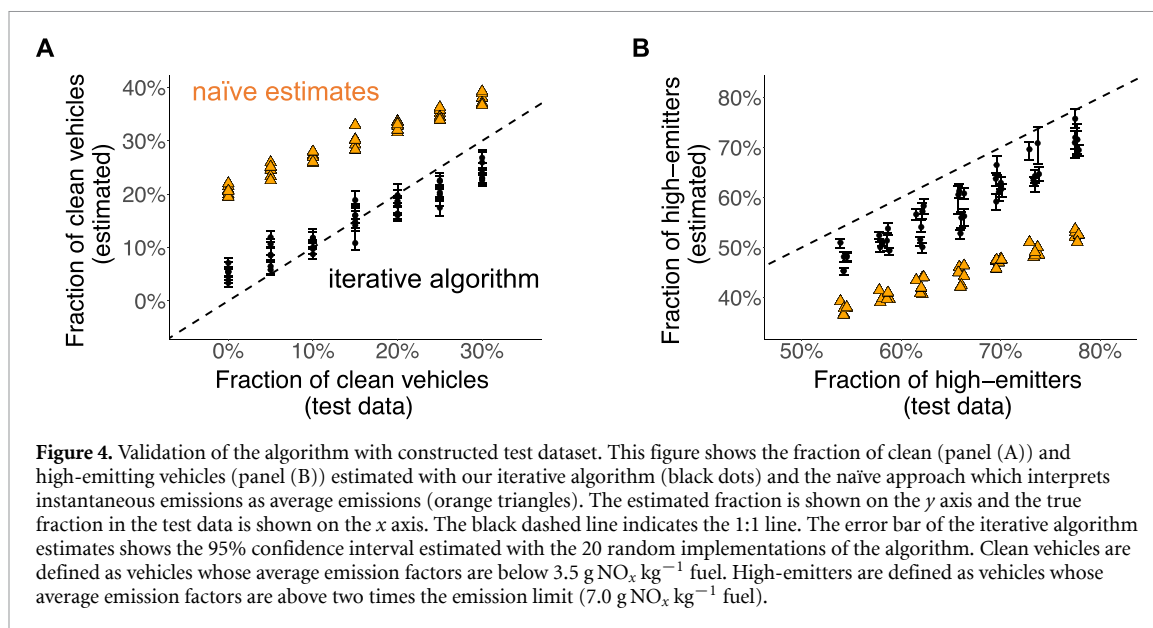
Figure 3. Identification precision of super high-emitters (panel (A)) and clean vehicles (panel (C)) with the associated RS cut-off thresholds under different numbers of repeated measurements (coloured). Panels (B) and (D) show the identification precision value (*y* axis) and the fraction of total super high-emitter and clean vehicles identified (*x* axis) with the varying cut-off thresholds. The solid line indicates a classification rule based on the average value of the repeated measurements ('average rule'). The dashed line indicates a classification rule based on all repeated measurements ('count all rule'). Super high-emitters are defined as vehicles whose average emission factors are above 5 times the type approval limit (17.5 g NO_x kg^{-1} fuel). Clean vehicles are defined as vehicles whose average emission factors are below 3.5 g NO_x kg^{-1} fuel. Results are derived from a hypothetical fleet with two million simulated cars. The large sample size ensures enough data points to evaluate the count all rule under repeated measurements.

need to be set as low as 0.2 g NO_x kg^{-1} fuel; this low value is however not operational as it is at the detection limit of common RS devices. Using higher cut-off thresholds (e.g. 1 g NO_x kg^{-1} fuel) yet reduces precision to 34%. However, the precision value increases to 85% or 99%, if two or five RS measurements for the same vehicle are both below 1 g NO_x kg^{-1} fuel respectively.

Figures 3(B) and (D) show the trade-offs between the identification precision and identified fraction with varying cut-off thresholds, for different numbers of repeated measurements. With one measurement, one can only identify 29% of the total super high-emitters and 36% of the total clean vehicles at a 75% precision level. With repeated measurements, less stringent RS thresholds can be adopted for the same precision level and more super high-emitters or clean vehicles could be identified. At 75% precision level (using 'average rule'), one can identify 40% of the total clean vehicles with two repeated measurements and 49% of the total clean vehicles with five repeated measurements. Improvements for super high-emitter identification are even more substantial—one can identify 58% of all super high-emitters with two repeated measurements and 80% of super high-emitters with five repeated measurements. A higher fraction of identified super high-emitters leads to substantial increases in NO_x emission reductions. If all identified super high-emitters at the 75% precision level were replaced with vehicles

with average emissions at the type approval limit, an identification program based on single measurement can reduce total NO_x emissions (of the Zurich fleet) by 15%, while a program based on five repeated measurements could reduce total NO_x emissions by 36% (with the 'average rule') or 31% (with the 'count all rule'). We observe very similar performances offered by either classification rule in clean vehicles identification, but meaningful differences between the two classification rules in high-emitter identifications. Using 'count all rule' can achieve almost perfect identification precision at the cost of a smaller fraction of super high-emitters identified, while using 'average rule' can identify almost all super high-emitters if a lower precision level is allowed (see figure 3(B)).

3.3. Algorithm validation
Our algorithm estimates the fraction of clean and high-emitting vehicles of measured fleets based on instantaneous RS emissions. The ideal way to validate a high-emitter identification scheme would be to do confirmatory testing of these vehicles with PEMS/ chassis tests, and calculate their average emission factors and share of high-emitters in the fleet, preferably on the spot. In practice, this is so expensive and time-consuming that it is completely impractical on a large number of vehicles. Here, we propose a simulation method validating any such algorithm without a life experiment.

Figure 4. Validation of the algorithm with constructed test dataset. This figure shows the fraction of clean (panel (A)) and high-emitting vehicles (panel (B)) estimated with our iterative algorithm (black dots) and the naïve approach which interprets instantaneous emissions as average emissions (orange triangles). The estimated fraction is shown on the *y* axis and the true fraction in the test data is shown on the *x* axis. The black dashed line indicates the 1:1 line. The error bar of the iterative algorithm estimates shows the 95% confidence interval estimated with the 20 random implementations of the algorithm. Clean vehicles are defined as vehicles whose average emission factors are below 3.5 g NO_x kg^{-1} fuel. High-emitters are defined as vehicles whose average emission factors are above two times the emission limit (7.0 g NO_x kg^{-1} fuel).

Using the PEMS/chassis measurements, we simulate a test dataset of instantaneous emissions as proxy for RS measurements; the average emissions of the simulated vehicles are exactly known to us. Figure 4 compares the fraction of high-emitting and clean vehicles estimated by our algorithm with the true fraction in the test dataset. As shown in figure 4, our algorithm performs well in estimating the fraction of high-emitting and clean vehicles in the test datasets, and significantly outperforms the naive estimates that treat instantaneous emissions as the average emission factors. Across fleets with different fractions of high-emitting and clean vehicles, the naive approach estimates a higher fraction of clean vehicles and a lower fraction of high-emitting vehicles by 10%–30%, while the biases of our algorithm estimates are only 5%–10%. Our algorithm offers the biggest improvement when applied to the dirtiest fleet (which is closest to the real-world conditions of EU-5D). The improvement offered by our algorithm slightly decreases but remains significant as the test simulated fleet becomes cleaner. Figure S3 shows the estimated and true distribution of average emission factors. Across all simulated fleets, we observe that the distributions of estimated average emission factors are substantially more similar to the distributions of true average emission factors, compared with the distribution of the instantaneous measurements; the Kolmogorov–Smirnov distances between the two distributions (a standard metric that measures the distance between two distributions) reduce by 40%–80%.

4. Discussion

Here we develop a scheme to combine instantaneous RS records with second-by-second measurements from PEMS or chassis dynamometer tests to identify vehicles with high average emission factors. With only one snapshot measurement for each vehicle, our algorithm can successfully determine the distribution of average emission factors of fleets measured by RS and subsequently calculate the fraction of vehicles as below/above a chosen 'clean' or 'dirty' threshold, respectively. Our algorithm aims to identify vehicles with high average emission factors of varying thresholds, whether the fraction of high-emitters is over 50% as for EU-5D, or just a few percentages as for gasoline cars or EU-6D. In its application to EU-5D passenger cars, we find the instantaneous measurements significantly overestimate the fraction of vehicles in compliance. This is likely the case for all other vehicle classes and pollutants as well, since instantaneous emissions are highly skewed towards zero. Furthermore, the contributions of the dirtiest 10%–20% vehicles are overestimated if one directly uses instantaneous measurements (see figure S4). This suggests that targeting only the dirtiest few percentages of EU-5D fleets is misleading but that rather the top 50% should be targeted as their average emission factors are not much lower than the top 10%, at least in the case of NO_x emissions from diesel cars.

Our analysis is highly relevant to the policy discussions on the RS-based identification programs globally. The trade-off between precision and identified fraction of such programs is quantified for the first time. It provides a rigorous way to design these programs by choosing the RS thresholds, number of repeated measurements, and classification rules depending on desired precision, identified fractions, and program budgets. Local decision makers can choose any point on the trade-off curves based on local priorities. Our analysis suggests the popular high-emitter detection based on single measurement is highly uncertain and ineffective for the current

EU-5D fleets, while identification programs with repeated measurements of the same vehicle can yield substantially higher precision and would detect more high-emitters (or clean vehicles).

Joining RS data and high-frequency measurement data from PEMS or chassis dynamometers has great potential. Linking the RS data to detailed high-frequency measurements can help researchers analyze RS data down to individual vehicles, which can provide a much more comprehensive understanding of the vehicle emitting behaviors of the measured fleet beyond instantaneous emissions. In fact, one can also try to understand, for example, extreme emissions or emissions under certain driving conditions with a similar framework. Vice versa, RS data and the estimated distribution of average emission factors provide context for the detailed measurements of individual vehicles with PEMS or chassis dynamometers, for example, to understand whether the measured vehicle is representative of the local fleet.

Our analysis has several limitations. The iterative algorithm introduced here estimates a distribution of the average vehicle emission factors, which can recover the distribution of the measured RS records when combined with the variability distributions derived from second-by-second PEMS/chassis measurements. However, multiple solutions may exist and our algorithm is only able to find one of them. The impacts seem limited in our case, as our results are robust to many variants of the approach, including different ways of processing the PEMS/chassis measurement data, number of candidate profiles, and different ways of algorithm initialization (see supplementary method). More research is needed to better understand how different assumptions can influence the algorithm performances under different contexts. Like most previous studies analyzing the RS data, we only calculate the fuel-based emission factor and do not consider the potential heterogeneity in fuel economy. This issue is likely small in our case for EU-5D, as the biggest deterministic factor of average vehicle emission is not fuel economy but whether a vehicle has installed a defeat device or not. For future analysis, our algorithm could be extended to incorporate the fuel economy heterogeneity among vehicles by adopting identification threshold values as a function of vehicle sub-class or fuel economy.

Our framework can be further applied to gasoline cars and Euro-6 diesel (EU-6D) cars. With a reduction in overall NO_x emissions and a lower fraction of super high-emitters, identification of high-emitters among gasoline and EU-6D cars may become more rewarding despite greater challenges. One thing that needs to be checked is whether the current RS and PEMS/chassis measurements are accurate enough at lower emission values to detect high-emitters with low absolute emission levels. We may expect that in such situations a few more repeated measurements may be needed to achieve the same precision compared with the EU-5D fleet studied here. While second-by-second test data have been limited in the past, this is no obstacle any more as manufacturers are required to provide such data publicly as part of the Euro 6 legislation in Europe. This framework can also be easily adopted to study different pollutants with available RS and second-by-second measurement data, e.g. hydrocarbons. Extending the algorithm to heavy-duty vehicles and emissions of particulate matter would be highly desirable for air quality management in the transportation sector.

5. Conclusion

Our research establishes a framework of using RS and high-frequency measurement data for developing the link between instantaneous and average emissions and identifying high-emitters, at the example of NO_x emissions of EU-5D passenger cars. Our method shows that interpreting instantaneous emissions as the average emitting behavior can be highly misleading, while a combination of RS and second-by-second test data can help address this gap. Compared to the currently-used conventional method of high-emitter detection based on a single measurement, we demonstrate that programs based on 2–5 repeated measurements capture 98%–174% more high-emitting vehicles with the same precision, and lead to 88%–140% increase in NO_x emission reductions. Combining RS and high-frequency measurement data, we present a simulation approach that can allow testing and verification of algorithms without (costly or complicated) life experiments.

Data availability statement

The code and sample data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.5281/zenodo.6341957. The original data of remote sensing and PEMS/chassis measurements are available upon request.

Acknowledgments

and innovation programme, as part of the CARES project under Grant Agreement No. 814966.

Author contributions

M Q developed the approach and analyzed the data; J B K designed the research; both wrote the paper together.

Conflict of interest

The authors declare no competing interests.

References

[1] Shindell D, Faluvegi G, Walsh M, Anenberg S C, Van Dingenen R, Muller N Z, Austin J, Koch D and Milly G 2011 Climate, health, agricultural and economic impacts of tighter vehicle-emission standards *Nat. Clim. Change* **1** 59–66

[2] Anenberg S C *et al* 2017 Impacts and mitigation of excess diesel-related NO_x emissions in 11 major vehicle markets *Nature* **545** 467–71

[3] Chossière G P, Malina R, Ashok A, Dedoussi I C, Eastham S D, Speth R L and Barrett S R 2017 Public health impacts of excess NO_x emissions from Volkswagen diesel passenger vehicles in Germany *Environ. Res. Lett.* **12** 034014

[4] Jonson J E, Borken-Kleefeld J, Simpson D, Nyíri A, Posch M and Heyes C 2017 Impact of excess NO_x emissions from diesel cars on air quality, public health and eutrophication in Europe *Environ. Res. Lett.* **12** 094017

[5] Tanaka K, Lund M T, Aamaas B and Berntsen T 2018 Climate effects of non-compliant Volkswagen diesel cars *Environ. Res. Lett.* **13** 044020

[6] Chossière G P, Malina R, Allroggen F, Eastham S D, Speth R L and Barrett S R 2018 Country-and manufacturer-level attribution of air quality impacts due to excess NO_x emissions from diesel passenger vehicles in Europe *Atmos. Environ.* **189** 89–97

[7] Zhang Y, Stedman D H, Bishop G A, Guenther P L and Beaton S P 1995 Worldwide on-road vehicle exhaust emissions study by remote sensing *Environ. Sci. Technol.* **29** 2286–94

[8] Beaton S P, Bishop G A, Zhang Y, Ashbaugh L L, Lawson D R and Stedmam D H 1995 On-road vehicle emissions: regulations, costs and benefits *Science* **268** 991–3

[9] Ahn K, Rakha H and Trani A 2004 Microframework for modeling of high-emitting vehicles *Transp. Res. Rec.* **1880** 39–49

[10] Park S S, Kozawa K, Fruin S, Mara S, Hsu Y K, Jakober C, Winer A and Herner J 2011 Emission factors for high-emitting vehicles based on on-road measurements of individual vehicle exhaust with a mobile measurement platform *J. Air Waste Manage. Assoc.* **61** 1046–56

[11] Bishop G A, Schuchmann B G, Stedman D H and Lawson D R 2012 Multispecies remote sensing measurements of vehicle emissions on Sherman Way in Van Nuys, California *J. Air Waste Manage. Assoc.* **62** 1127–33

[12] U.S. Environmental Protection Agency 1996 *User Guide and Description for Interim Remote Sensing Program Credit Utility* (available at: https://esp-global.com/downloads/USEPA GrossEmitterGuidance.pdf) (Accessed 2021)

[13] Böhm M, Nanni M and Pappalardo L 2021 Improving vehicles' emissions reduction policies by targeting gross polluters (arXiv:2107.03282)

[14] Ministry of Ecology and Environment 2017 *Technical Guidelines for Emissions Standards for National Mobile Air Pollution Sources* (available at: www.mee.gov.cn/ ywgz/fgbz/bz/bzwb/dqhjbh/dqydywrwpfbz/201708/ t20170802_419063.shtml) (Accessed 2021)

[15] Smit R and Bluett J 2011 A new method to compare vehicle emissions measured by remote sensing and laboratory testing: high-emitters and potential implications for emission inventories *Sci. Total Environ.* **409** 2626–34

[16] Carslaw D C, Beevers S D, Tate J E, Westmoreland E J and Williams M L 2011 Recent evidence concerning higher NO_x emissions from passenger cars and light duty vehicles *Atmos. Environ.* **45** 7053–63

[17] Carslaw D C and Rhys-Tyler G 2013 New insights from comprehensive on-road measurements of NO_x, NO_2 and NH_3 from vehicle emission remote sensing in London, UK *Atmos. Environ.* **81** 339–47

[18] Chen Y and Borken-Kleefeld J 2014 Real-driving emissions from cars and light commercial vehicles–results from 13 years remote sensing at Zurich/CH *Atmos. Environ.* **88** 157–64

[19] Fontaras G, Franco V, Dilara P, Martini G and Manfredi U 2014 Development and review of Euro 5 passenger car emission factors based on experimental results over various driving cycles *Sci. Total Environ.* **468–469** 1034–42

[20] Franco V, Posada Sánchez F, German J and Mock P 2014 Real-world exhaust emissions from modern diesel cars. Part 1: aggregated results (The International Council On Clean Transportation) p 59 (available at: www.theicct.org/sites/de fault/files/publications/ICCT_PEMS-study_diesel-cars_ 20141010.pdf)

[21] Mock P and German J 2015 The future of vehicle emissions testing and compliance. How to align regulatory requirements, customer expectations, and environmental performance in the European Union (The International Council On Clean Transportation) (available at: www.theicct.org/future-of-vehicle-testing)

[22] US Environmental Protection Agency 2019 2014–2017 progress report vehicle and engine compliance activities (available at: https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey =P100WKFC.pdf)

[23] Bishop G A, Stedman D H, Hutton R B, Bohren L and Lacey N 2000 Drive-by motor vehicle emissions: immediate feedback in reducing air pollution *Environ. Sci. Technol.* **34** 1110–6

[24] Borken-kleefeld J, Bernard Y, Carslaw D and Sjödin Å 2018 Contribution of vehicle remote sensing to in-service/real driving emissions monitoring *CONOX Task 3 Report* (available at: www.bafu.admin.ch/dam/bafu/de/dokumen te/luft/externe-studien-berichte/contribution-of-vehicle- remote-sensing-to-inservice-real-driving-emissions- monitoring-conox-task-3-report.pdf.download.pdf/cont ribution-of-vehicle-remote-sensing-to-in-service-r)

[25] Bernard Y, German J and Muncrief R 2019 Worldwide use of remote sensing to measure motor vehicle emissions (International Council on Clean Transportation) (available at: https://theicct.org/publications/worldwide-use-remote- sensing-measure-motor-vehicle-emissions)

[26] Yang Z 2017 Remote-sensing regulation for measuring exhaust pollutants from in-use diesel vehicles in China (International Council on Clean Transportation) (available at: https://theicct.org/publications/remote-sensing-regulati on-measuring-exhaust-pollutants-use-diesel-vehicles-china)

[27] Environmental Protection Department, the Government of the Hong Kong Special Administrative Region 2020 Strengthened emissions control for petrol and LPG vehicles (available at: www.epd.gov.hk/epd/english/environmentinhk /air/guide_ref/remote_sensing_Petrol_n_LPG.ht)

[28] The Scottish Government 2020 Cleaner air for Scotland 2—draft air quality strategy consultation (available at: www.gov.scot/binaries/content/documents/govscot/publicat ions/consultation-paper/2020/10/cleaner-air-scotland-2- draft-air-quality-strategy-consultation/documents/cleaner- air-scotland-2draft-air-quality-strategy-consultation/clea ner-air-scotland-2dr)

[29] Bishop G A, Stedman D H and Ashbaugh L 1996 Motor vehicle emissions variability *J. Air Waste Manage. Assoc.* **46** 667–75

[30] Huang Y, Organ B, Zhou J L, Surawski N C, Hong G, Chan E F and Yam Y S 2018 Remote sensing of on-road vehicle emissions: mechanism, applications and a case study from Hong Kong *Atmos. Environ.* **182** 58–74

[31] Huang Y, Surawski N C, Yam Y-S, Lee C K, Zhou J L, Organ B and Chan E F 2020 Re-evaluating effectiveness of vehicle emission control programmes targeting high-emitters *Nat. Sustain.* **3** 904–7

[32] Ministry of Environmental Protection 2017 Measurement methods and specifications for exhaust pollutants from in-use diesel vehicles (remote sensing method) p 12 (available at: www.vecc.org.cn/xgbzzyc/2699.jhtml)

[33] Yang L, Bernard Y and Dallmann T 2019 Technical considerations for choosing a metric for vehicle remote-sensing regulations (International Council on Clean Transportation) (available at: https://theicct.org/publication/technical-considerations-for-choosing-a-metric-for-vehicle-remote-sensing-regulations/)

[34] Popp P J, Bishop G A and Stedman D H 1999 Development of a high-speed ultraviolet spectrometer for remote sensing of mobile source nitric oxide emissions *J. Air Waste Manage. Assoc.* **49** 1463–8

[35] Mazzoleni C, Kuhns H D, Moosmüller H, Keislar R E, Barber P W, Robinson N F, Watson J G and Nikolic D 2004 On-Road vehicle particulate matter and gaseous emission distributions in Las Vegas, Nevada, compared with other areas *J. Air Waste Manage. Assoc.* **54** 711–26

[36] McClintock P 2011 *Enhanced Remote Sensing Performance Based Pilot Program* (Tiburon, CA: Environmental Systems Products Inc.)

[37] Lau J, Hung W and Cheung C S 2012 Observation of increases in emission from modern vehicles over time in hong kong using remote sensing *Environ. Pollut.* **163** 14–23

[38] Borken-Kleefeld J *et al* 2018 Comparing emission rates derived from remote sensing with PEMS and chassis dynamometer tests *CONOX Task 1 Report* (available at: www.ivl.se)

[39] Huang Y, Organ B, Zhou J L, Surawski N C, Shing Yam Y and Chan E F 2019 Characterisation of diesel vehicle emissions and determination of remote sensing cutpoints for diesel high-emitters *Environ. Pollut.* **252** 31–38

[40] Jun Z, Hua-Fang G U O and Yue-Ming H U 2007 Artificial neural network model for identifying taxi gross emitter from remote sensing data of vehicle emission *J. Environ. Sci.* **19** 427–31

[41] Kang Y, Li Z, Lv W, Xu Z, Zheng W X and Chang J 2021 High-emitting vehicle identification by on-road emission remote sensing with scarce positive labels *Atmos. Environ.* **244** 117877

[42] Rushton C E *et al* 2018 Interinstrument comparison of remote-sensing devices and a new method for calculating on-road nitrogen oxides emissions and validation of vehicle-specific power *J. Air Waste Manage. Assoc.* **68** 111–22

[43] Grange S K, Farren N J, Vaughan A R, Rose R A and Carslaw D C 2019 Strong temperature dependence for light-duty diesel vehicle NO_x emissions *Environ. Sci. Technol.* **53** 6587–96

[44] Hausberger S 2020 *Horizon 2020 CAREs: simulation of fuel flow from vehicle speed and acceleration data* TU Graz

[45] Huss A and Weingerl P 2020 JEC tank-to-wheel report v5: passenger cars (available at: https://ec.europa.eu/jrc/en/publication/jec-tank-wheel-report-v5-passenger-cars)

[46] Matzer C, Weller K, Dippold M, Lipp S, Röck M, Rexeis M and Hausberge S 2019 *Update of Emission Factors for HBEFA Version 4.1* pp 1–138

[47] Chen Y and Borken-Kleefeld J 2016 NO_x emissions from diesel passenger cars worsen with age *Environ. Sci. Technol.* **50** 3327–32