Science of the Total Environment 832 (2022) 154943

Contents lists available at ScienceDirect

Science of the Total Environment

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

Developing an optimal sampling design to monitor the vehicle fuel consumption gap

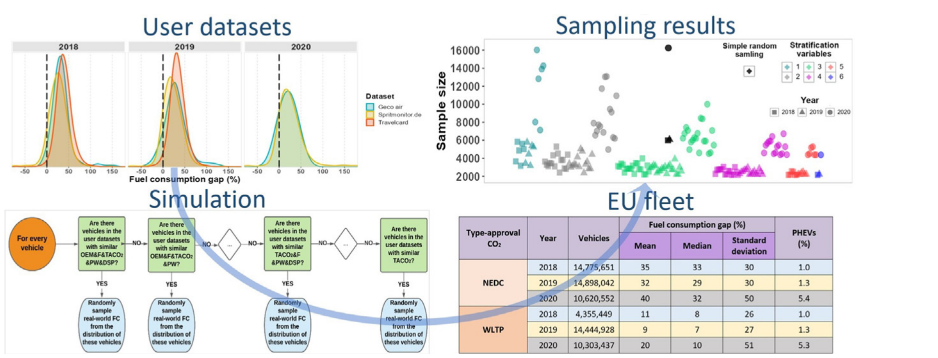
Markos A. Ktistakis, Jelica Pavlovic, Georgios Fontaras *

European Commission, Joint Research Centre (JRC), Ispra 21027, Italy

HIGHLIGHTS

- Sampling schemes to monitor the gap between official and real-world fuel consumption (CO_2 emissions) are examined.
- The average fuel consumption gap can be accurately estimated with a sample of less than 0.05% of the fleet.
- To estimate the gap per manufacturer more than three times as many vehicles are required.
- The sample size and the gap were higher in 2020, due to the increase of PHEVs sales.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Pavlos Kassomenos

Keywords:
CO_2 gap
Real-world fuel consumption
Real-world CO_2 emissions
Sampling methods
RDE
PHEV

ABSTRACT

Monitoring the fuel consumption gap between official and real-world measurements is of great interest to policy makers and researchers. This study explores how sampling methods (simple random, stratified and quota sampling) can be used to supplement and validate the monitoring. Three user datasets were utilised to simulate the fuel consumption gap of the 11.6–15.5 million vehicles registered annually in the European Union (2018–2020). Results suggest that a simple random sample of 16,240 vehicles is sufficient to estimate accurately the fleets' average fuel consumption gap. Stratified sampling can reduce the sample size to less than 4,500 vehicles. To estimate accurately the fuel consumption gap of each manufacturer, the sample size increases to approximately 17,200 vehicles. The increase in sales of plug-in hybrid vehicles in 2020 led to an increase of the average fuel consumption gap by 8% and its standard deviation (variability) by 20%. This higher variability resulted in a more than double sample size, compared to previous years. It was also found that the introduction of the Worldwide Harmonized Light vehicles Test Procedure (WLTP) reduced the average gap by 20–24%. This study highlights the viability of a sampling scheme to estimate the fuel consumption gap by monitoring less than 0.05% of the fleet. Moreover the study draws attention to the need for further analysis and understanding of the real-world use and fuel consumption of plug-in hybrid vehicles.

1. Introduction

The European Union (EU) has set 2030 and 2050 climate and energy targets with the goal to make Europe the first climate neutral continent in the world, in order to contribute to global climate change mitigation. To achieve this it is crucial to reduce the Carbon Dioxide (CO_2) emissions of transport (Krause et al., 2020; Logan et al., 2020; Rottoli et al., 2021; Siskos et al., 2018). Light-Duty Vehicles (LDVs) are the greatest contributor to CO_2 emissions in the EU transport sector, amounting to 15% of the EU's

Abbreviations: EC, European Commission; LDV, Light-Duty Vehicle; NEDC, New European Driving Cycle; WLTP, Worldwide Harmonized Light vehicles Test Procedure; OBFCM, On-Board Fuel and/or energy Consumption Monitoring; PTI, Periodic Technical Inspection; OTA, Over-The-Air; PHEV, Plug-in Hybrid Electric Vehicle; EEA, European Environmental Agency; SM, Spritmonitor.de; TC, Travelcard; GA, Geco air; ME, Margin of Error; CL, Confidence Level; CV, Coefficient of Variation.

* Corresponding author.
E-mail address: georgios.fontaras@ec.europa.eu (G. Fontaras).

http://dx.doi.org/10.1016/j.scitotenv.2022.154943
Received 29 January 2022; Received in revised form 20 March 2022; Accepted 27 March 2022
Available online 5 April 2022

total emissions in 2018 (Buysse and Miller, 2021; EEA, 2020). Since 2009 the European Commission (EC) has put in place binding targets to reduce the CO_2 emissions of LDVs (EC, 2009; EU, 2019; Frondel et al., 2011). These are based on measurements from controlled laboratory tests for the type-approval procedure and officially have been a success, reaching the targets and reducing the average CO_2 emissions of new passenger cars from 170 g/km in 2001 to 118 g/km in 2016, with slight rebounding to 122 g/km in 2019 (EEA, 2021a). Nonetheless, this decrease did not generally translate into an equivalent reduction of the CO_2 emission in real-world conditions. The divergence between official and real-world CO_2 emissions is generally known as the fuel consumption (CO_2 emissions) gap. Voluntarily self-reported consumer and fleet operator fuel consumption data and experimental tests report that the gap increased substantially in the EU, from about 6% in 2001 to almost 40% in 2017 (Fontaras et al., 2017; Jiménez et al., 2019; Ntziachristos et al., 2014; Tietge et al., 2019; Zacharof et al., 2016). The gap has also been found to increase in other leading markets around the world (Ben Dror et al., 2019; Huang et al., 2019; Tietge et al., 2017; Tu et al., 2022; Yu et al., 2021).

Up to 2017, the official type-approval fuel consumption figures were based on the New European Driving Cycle (NEDC), which was created in the seventies and did not reflect the 21st century's testing and vehicle technologies and therefore has been identified as a major reason for this gap (Cames and Helmers, 2013; Fontaras and Dilara, 2012; Triantafyllopoulos et al., 2019). The EC, acknowledging the issue, introduced in September 2017 the new Worldwide Harmonized Light vehicles Test Procedure (WLTP) in the type-approval procedure (EU, 2017). While the emissions of pollutants (NOx, CO, PN, THC) were tested from the beginning under the new regulation and procedure, for CO_2 emissions there was a phase-in period from 2017 to the end of 2020 when each vehicle had both (WLTP and NEDC) emissions recorded in its Certificate of Conformity (Fontaras et al., 2018). During this period, some aspects of the European Legislation were still based on the NEDC procedure (CO_2 emission targets, vehicle labelling, etc.).

As of 2021, WLTP is the only cycle used for LDVs CO_2 certification in the EU. The average real-world to NEDC fuel consumption gap was reported to be 39% in 2018 (Dornoff et al., 2020) and 37% for petrol and 38% for diesel passenger cars in 2019 (Van Gijlswijk et al., 2020). The WLTP is expected to reduce the CO_2 gap by more than half, to about 20% (Pavlovic et al., 2018; Tsiakmakis et al., 2016). Dornoff et al. (2020) confirmed these estimates, suggesting that WLTP bridges the gap by 22 to 24% compared to the NEDC. Van Gijlswijk et al. (2020) found the average WTLP fuel consumption gap being 16% for petrol passenger cars and 10% for diesel passenger cars.

To monitor the evolution of the CO_2 emissions and the real-world fuel or energy consumption, avoid a gradual recurrence of the gap's increase and inform the public, all new passenger cars registered in the EU from 2021 onwards have to be equipped with an On-Board Fuel and/or energy Consumption Monitoring (OBFCM) device that records the fuel or energy consumed (EU, 2019). According to Regulation 2021/392 (EU, 2021), the manufacturers shall collect real-world data for new vehicles registered from 1st January 2021, either via direct transfer from vehicle or when vehicles are brought in for service or repairs. From May 2023, the real-world fuel and energy consumption data shall also be collected by the Member States when vehicles undergo the roadworthiness tests (Periodic Technical Inspections - PTIs) and shall be reported to the EC each calendar year. The Regulation foresees the review of some aspects of the provisions on the monitoring and reporting the real-world fuel and energy consumption data, in particular Over-The-Air (OTA) direct data transfers from vehicles. A large sample from the data collected from the authorised dealers and repairers or PTIs is not expected to be available for new vehicle models for several years (Dornoff, 2019). For most vehicles, PTIs take place for the first time, on a compulsory basis, 2–4 years after registration, depending on the member state.

Although the most accurate results and analysis can be expected from accessing the real-world fuel consumption of the whole vehicle fleet equipped with OBFCMs and having the data transferred to the EC, this may not be possible for the reasons mentioned above, particularly before 2025. An alternative way to monitor the real-world fuel consumption is to sample a small set of vehicles drawn from the fleet and statistically process them. Sampling-based approaches can also be utilised to supplement and validate the other data transfer methods that will be employed in the future. This study aims to explore different sampling methods (and their accuracy) that can be used to estimate the average fuel consumption gap of the fleet and of specific sub-fleets (e.g. per manufacturer, fuel type, powertrain type, etc.). The main question this study attempts to answer is: What is the optimal sampling design and the minimum sample size required to precisely estimate the average fuel consumption gap of the entire annual EU fleet? The research used 2018, 2019 and 2020 real-world fuel consumption data from three user-based datasets (Geco Air, Spritmonitor.de and Travelcard) to calculate and better understand the most recent fuel consumption gap and its variability and simulate a reasonable real-world fuel consumption of the EU fleet. This in turn is used to examine the statistical methods and calculate the required sample size to precisely estimate the fuel consumption gap. Special attention has been given to Plug-in Hybrid Electric Vehicles (PHEVs), because their market share is increasing (Blanck et al., 2020; EEA, 2021c), while they exhibit a much larger fuel consumption gap (Plötz et al., 2021; Sánchez et al., 2019).

2. Data and methods

This section presents the datasets used, the data pre-processing performed, the simulation of the real-world fuel consumption and the sampling methods utilised.

2.1. Datasets

Three user-based datasets (Spritomonitor.de, Travelcard and Geco air) were used to quantify the real-world fuel consumption and subsequently the fuel consumption gap. The European Environmental Agency (EEA) CO_2 monitoring datasets provide data for all the vehicles registered in the EU every year. They do not include the real-world fuel consumption, which was needed to run the sampling methods and hence it has been simulated using the available user-based datasets.

2.1.1. Spritmonitor.de

Spritmonitor.de (SM) is a free web service where vehicle owners report real-world fuel consumption (Spritmonitor.de, 2021). It was launched as a website in Germany in 2001, to give users a simple way to monitor their fuel consumption. The owners register a vehicle by providing basic specifications (i.e. manufacturer, model, powertrain type, fuel type, registration year, transmission type and engine power). Optionally, SM users can report the type-approval fuel consumption value. To start making use of the service, users need to fill the fuel tank, and the first event provides the reference for calculating the fuel consumption. After every fuelling entry, the user reports the mileage and the fuel or energy inserted.

The SM dataset consisted of approximately 58,500 vehicles built in 2018, 2019 and 2020. Battery electric vehicles and vehicles using alternative fuels (i.e. Hydrogen, CNG and LPG), representing about 6% of the total vehicles, have been excluded from the analysis. Vehicles with unknown or erroneous type-approval or real-world fuel consumption were also discarded (53%), as were the vehicles with recorded mileage less than 1000 km (3%). This left approximately 22,200 vehicles suitable for analysis. The full description of all exclusion criteria can be found in the Supplementary material.

2.1.2. Travelcard

Travelcard (TC) Nederland BV is a fuel card provider based in the Netherlands (Travelcard, 2021). Fuel cards are utilised by companies to trace their fleets' fuel expenses. TC users are drivers who use the fuel cards as payment cards at gas stations, usually drive new company cars and change vehicles every couple of years (Van Gijlswijk and Ligterink, 2018). Usually, the expenses of TC users are covered by the employers.

Real-world CO_2 emissions were estimated based on pairs of consecutive fuelling events, utilising the fuel recorded on the TC system and odometer readings recorded by the users. The dataset included information about the manufacturer, fuel type, powertrain type, NEDC type-approval CO_2 emissions, registration year, engine power, engine displacement and the number of fuelling events.

Data were provided for vehicles registered in 2018 and 2019, they were already processed and the distance was not given. After data cleaning, using the same criteria as explained above with SM, except the distance criterion, which was replaced by removing vehicles with less than 3 fuelling events, the initial dataset covering about 37,700 vehicles was brought down to approximately 36,800. Details about the vehicles removed can be found in the Supplementary material.

2.1.3. Geco air

Geco air (GA) is a mobile application designed by the Institut Français du Pétrole Energies Nouvelles (IFPEN) to assist drivers to become aware of their NOx, CO and CO_2 emissions and to evaluate their driving behaviour (Geco Air, 2021; Thibault et al., 2017). The real-world CO_2 emissions are simulated using a physical model of the vehicle (Michel et al., 2021) which uses the speed and altitude real-world data profiles calculated based on the GPS measurements recorded every second by the users' mobile phones, and the vehicle technical specifications that the users register when enlisting the vehicles. These include the manufacturer, fuel type, powertrain type, type-approval CO_2 emissions, build year, engine displacement and engine power.

IFPEN provided data of approximately 500 vehicles registered from 2018 to 2020. Using the same criteria as with SM resulted in decreasing the dataset to about 300 vehicles, which were analysed in this research.

2.1.4. European Environment Agency

The EEA documents annually the new LDVs registered in the EU and publishes anonymized datasets (EEA, 2021b). The annual datasets include the registration country, type-approval CO_2 emission, manufacturer, fuel type, engine power and engine displacement. They do not contain real-world fuel consumption and CO_2 emissions data.

In this study, the final 2018 and 2019 EEA datasets covering approximately 15.27 and 15.50 million passenger cars, respectively, were used. For 2020, the provisional (preliminary) EEA dataset containing about 11.60 million passenger cars was utilised. After removing BEVs, vehicles using alternative fuels, and vehicles with unknown or erroneous NEDC type-approval CO_2 emissions, approximately 14.78, 14.90 and 10.62 million vehicles remained from 2018, 2019 and 2020, respectively. The reason that NEDC (and not WLTP) type-approval CO_2 emissions were used is because of their wider availability. In 2018, from the approximately 14.81 million non-BEVs vehicles using conventional fuels, the WLTP type-approval CO_2 emissions were reported for only around 4.38 million vehicles. In 2019 and 2020 the WLTP type-approval CO_2 emissions were reported for the majority of vehicles, but still more NEDC than WLTP values are available (Table SM1 Supplementary material).

Simulated real-world fuel consumption was assigned to every vehicle registered in EEA in these three years. Together with the type-approval fuel consumption, the fuel consumption gap was calculated and used to test the sampling methods. The user datasets' real-world fuel consumption could not be directly used and a simulation procedure was deemed necessary. This is because the user datasets compositions are not the same as EEA's. The simulated fuel consumption was computed separately for the PHEVs and the rest of the vehicles (Internal Combustion Engine and Not-off-Vehicle Charging Hybrid Electric Vehicles) and separately for each year. Fig. 1 summarizes the steps taken to assign a real-world fuel consumption to each vehicle of the EEA datasets. All vehicles' type-approval CO_2 emissions fall in one of the bins (strata) when grouping only according to the type-approval CO_2 emissions. Hence, at the last step of the procedure "Are there vehicles in the user datasets with similar $TACO_2$?" the only possible outcome is "YES". It should be noted that the procedure was designed in such a way that the most important variable in this simulation scheme is the manufacturer, followed by the type-approval CO_2 emissions.

2.2. Sampling methods

Three sampling methods were deemed the most appropriate and were investigated in this paper: simple random sampling, stratifying sampling and quota sampling. Simple random sampling and stratifying sampling are probability-sampling methods with known formulas for determining the accuracy of the fuel consumption gap, the equations used can be found in (Cochran, 1977; Ktistakis et al., 2021). Utilising these formulas allows calculating in advance the required sample size to achieve the prespecified accuracy. Quota sampling is a non-probability method; therefore, the requirements for choosing a sample are less restrictive. However making inferences for the whole population is risky, because the sampling error cannot be calculated.

2.2.1. Simple random sampling

Simple random sampling is a common probability sampling method. The principle is that each vehicle from the fleet has the same probability of being selected. Simple random sampling provides a benchmark for the maximum sample size. Advanced knowledge of the population is not needed. It only requires the whole population size and an estimate of the standard deviation of the population. Using those, the sample size to achieve the required precision can be calculated. The degree of precision and accuracy of the estimate is determined by the Margin of Error (ME) and Confidence Level (CL). In this paper, the terms precision and accuracy

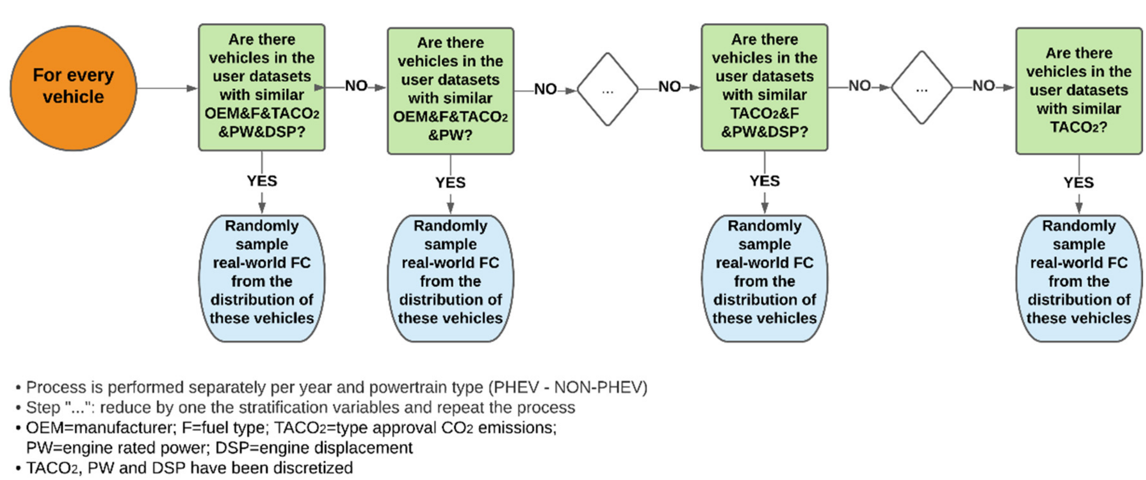


Fig. 1. Schematic representation of the simulation procedure of EEA vehicles' real-world fuel consumption.

will be used interchangeably to quantify the uncertainty of the estimate. If not stated otherwise, for population wise estimates, a ME of 1% and CL of 99% were used.

2.2.2. Stratifying sampling

If the population is heterogeneous (the fuel consumption gap of distinct subpopulations differs substantially), or when certain homogeneous subpopulations can be isolated, stratified sampling can be used to optimize the sampling procedure and reduce the sampling size. Stratification schemes are practical and have the advantage of minimizing the impact of the fleet composition (e.g. fuel type, region, season etc.) when averaging the fuel consumption rates of a mixed fleet. In stratified sampling, the population is broken down to subpopulations (strata). The more homogenous the strata, in respect to the fuel consumption gap, the more accurate the results (or equivalently the smaller the sample size to achieve a required precision). Sample allocation methods are dependent on the number of vehicles in each stratum, the variability of the fuel consumption gap within each stratum and the sampling cost in each stratum. In this research, cost considerations were not taken into account; the sampling cost was considered the same for all strata. JRC Science for Policy Report (Ktistakis et al., 2021) has examined equal, proportional and Neyman allocations. Neyman allocation was found to produce the best results in almost all cases but requires a-priori knowledge of the variability of the fuel consumption gap and the number of vehicles, per stratum. It allocates larger sample sizes to the larger and more variable strata. In this paper, if not mentioned otherwise, the Neyman allocation was used. Proportional allocation, which allocates the sample size proportionally to the strata sizes, without taking into consideration the variability of each stratum, was also examined. Similarly to simple random sampling, the required sample size can be calculated after choosing the ME and CL. As already stated, when using stratified sampling, the precision of a sample size increases and the required sample size decreases (if the class priors are correct). To quantify this decrease in sample size and to have comparable results for the three years/datasets examined, the sample efficiency of a stratification scheme has been defined as the percentage of decrease in sample size, compared to using simple random sampling.

Methods that use one stratification variable are called univariate stratification methods. If more than one stratification variable is utilised, they are called multivariate stratification methods. For this study, a cross-classification design was used for multivariate stratification. In this scheme, if there are L variables, with the j having H_j strata, then they form $H = \prod_{j=1}^{L} H_j$ strata. Strata consisting of a single vehicle were annexed into a superstratum. If there was only one stratum with one vehicle, the vehicle was reallocated to a random stratum.

When precise estimators (low ME and high CL) are not only needed for the whole population but also for all/some strata, a larger number of vehicles need to be sampled. This way a sufficient number of vehicles would be sampled from each stratum of interest to ensure that comparisons and a separate analysis of these strata can be performed. The higher the required precision of these strata specific estimators, the higher the sample size per stratum and consequently the total sample size. In this research, if not mentioned otherwise, a ME of 5% with a CL of 90% were used for the strata specific estimators.

2.2.3. Quota sampling

Quota sampling is a non-probability adaption of stratified sampling. Non-random criteria are used to select vehicles, and not all vehicles have a possibility of being sampled. This kind of sampling costs less and is easier to implement, however it does not provide valid statistical inferences about the whole population. In quota sampling, the population is separated into relevant subgroups depending on variables of interest (quota controls). Quotas can be either independent or dependent. Independent quotas are based on a single characteristic, whereas dependent are a combination of more than one. The number of vehicles sampled from each subgroup is proportional to the subgroup's size. Thus, concerning the quota controls, the sample is representative of the population. In the case of independent quotas, the sample has the same proportion of vehicles, as the population, for each quota control. In the case of dependent variables, the sample is representative of their combination. For example, if two quota controls are used, the fuel type and the manufacturer, when using independent quotas the method ensures that the sample has the same percentage of diesel and petrol vehicles as the population and the same percentage of each manufacturer. With dependent quotas, it also ensures that it has the same percentage of each combination of fuel type and manufacturer (e.g. diesel fuelled Fiat vehicles). A good understanding of the fuel consumption gap and the contributors to its variability (fuel type, manufacturer, engine power etc.) is crucial for designing an accurate quota sampling scheme and avoiding large biases. The higher cost of having more detailed quota controls and the increased accuracy and representativeness have to be balanced. In this study results from the probability sampling methods examined are used to propose an optimal quota sampling scheme.

3. Results and discussion

The analysis is separated in four parts. In Section 3.1, the results of a statistical analysis of the fuel consumption gap, for the three user-based datasets and the simulated gap for the annual EU fleets, are presented. In Section 3.2 the required sample size to measure accurately the fuel consumption gap across the whole EU, for different probability-sampling schemes, is analysed. In addition, in Subsection 3.2.3 the case where strata specific accurate estimators are needed is analysed for different sampling stratification schemes. In Section 3.3 quota sampling and its practical implementation is discussed. Finally, in Section 3.4 the impact of non-sampling errors on the sample size of the fuel consumption gap is examined.

3.1. Real-world fuel consumption gap

3.1.1. User-based data

In this section, the fuel consumption gap between type-approval and real-world driving is analysed using the three user datasets. The evolution of the gap is studied from 2018 to 2020 and the impact of different individual factors, such as the fuel type, is assessed.

Table 1 summarizes the real-world fuel consumption from three user-based datasets. It should be noted that the fuel consumption gap shown in Table 1 for the SM dataset is calculated based on user inputs for what regards the official type-approval fuel consumption values. For most of these vehicles, users had both NEDC and WLTP values due to the phase-in of the WLTP CO_2/fuel consumption Regulations in that period. Therefore, the fuel consumption gap in Table 1 is somehow the mixture of NEDC and WLTP gap, with most likely a much higher tendency towards the NEDC side, especially in 2018.

The mean gap in all three datasets decreased from 2018 to 2019 by 3–4%. The gap was found to be 36–40% for owners of company cars (TC). For owners of private vehicles, it was 24–27% (SM) and 29–33% (GA). In 2020, the average fuel consumption gap for SM rebounded to 33%, while 2020 data from TC were not available and from GA not enough vehicles (7) were available to make reliable inferences. This increased fuel

Table 1
Real world fuel consumption gap statistical results for three user-based datasets.

Dataset	Year	Vehicles	Fuel consumption gap (%)			PHEVs (%)
			Mean	Median	Standard deviation	
Geco air	2018	172	33	32	23	0.6
	2019	84	29	27	25	0.0
	2020	7	20	24	16	0.0
Spritmonitor.de	2018	9,920	27	26	29	2.9
	2019	8,294	24	21	30	2.5
	2020	3,970	33	22	55	8.4
Travelcard	2018	27,231	40	38	26	0.6
	2019	8,548	36	33	29	1.0

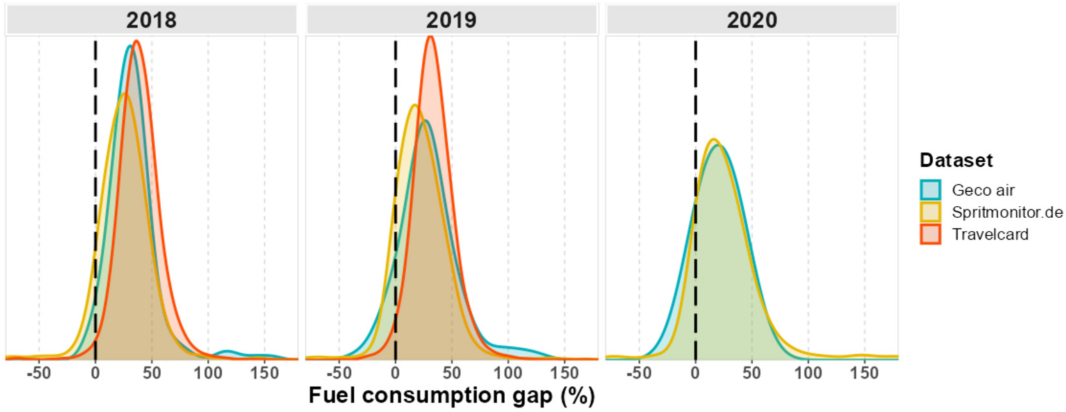


Fig. 2. Distribution of the fuel consumption gap. The black vertical dashed lines at 0% correspond to equal real-world and type-approval fuel consumption.

consumption gap in 2020 (SM) is mainly attributed to the fact that the proportion of PHEVs more than tripled in 2020 (from 2.5% in 2019 to 8.4% in 2020). Plötz et al. (2021) found that the fuel consumption of PHEVs, in real-world driving, on average is about two to four times higher compared to type-approval values, translating to a gap of 100–300%. The median fuel consumption gap, which is not influenced by extreme values, only increased by 1% (21–22%) from 2019 to 2020.

The variability of the fuel consumption gap, (quantified mostly by the Standard Deviation in this research), showed an increase ranging from 1 to 3%, between 2018 and 2019, for the three datasets. GA dataset has the lowest variability, mainly because of the lower number of vehicles and because of its low percentage of PHEVs. SM's 2020 registered vehicles exhibit the highest variability (55%). Again, this can be mainly attributed to the steep increase of the proportion of PHEVs. This difference in variability can also be seen in Fig. 2, where the 2020 distributions of the fuel consumption are more spread out.

The average fuel consumption gap of SM was significantly lower than that of TC (by 12–13%). Reasons for this are that TC consists of company cars and as their fuel bill is usually paid by the company, these drivers have a lower incentive to drive in a fuel-efficient way and in the case of PHEVs to use the electric mode. Which leads to a small utility factor and hence a large real-world fuel consumption (Plötz et al., 2018). On the other hand, SM's users are ones who care to monitor their fuel consumption and CO_2 emissions and consequently the dataset suffers from a self-selection bias. In addition, as mentioned before, some SM users may have reported the WLTP type-approval fuel consumption, which translates to a lower gap. The average gap of GA vehicles was 7% lower than TC's in 2018 and 2019.

Concerning the fuel type, the average fuel consumption gap did not vary a lot between petrol and diesel vehicles (Fig. 3), except for 2020 GA vehicles, which are only seven and hence not enough to make solid inferences.

The biggest difference between SM and TC was found in SM 2020, with an average divergence of 9% (petrol vehicles 36% and diesel 27%). For GA the average differences in 2018 and 2019 were larger, at 5% and 7%, respectively. In all cases, except SM 2020, the mean fuel consumption gap of diesel vehicles is higher. The reason SM for 2020 is an exception is that the proportion of PHEVs is by far the highest (8.6%) and the vast majority of PHEVs are fuelled by petrol. This also explains why the mean divergence is bigger in 2020, while the median divergence is constantly low. The interquartile ranges are similar in all cases; however, the standard deviations are not. For TC, the standard deviation is 11–13% higher for petrol vehicles, compared to diesel vehicles, for SM 2020 it is 24% higher. The main reason why TC petrol vehicles have a much higher variability is that most PHEVs are fuelled by petrol, and these PHEVs have a much higher average fuel consumption gap. The even higher divergence for SM 2020 is again because of the PHEVs and the large proportion of them in this sub-fleet.

On the other hand, as shown in Fig. 4A, depending on the type-approval CO_2 emissions significant variations are noted. In average, vehicles with low type-approval CO_2 emissions (<75 g/km) have a much larger fuel consumption gap, reaching up to 253% for 2019 TC vehicles with type-approval CO_2 emissions lower than 50 g/km. This can be explained by the fact that most of these vehicles are PHEVs. It can also be seen that for SM and TC vehicles, as the type-approval CO_2 emissions increase the average fuel consumption gap decreases. The only exception are 2019 TC's vehicles with very high (>200 g/km) type-approval CO_2 emissions. Vehicles with low type-approval CO_2 emissions exhibit a substantially higher variability of fuel consumption gap. The highest was found for vehicles of the SM dataset registered in 2020, with type-approval CO_2 emissions lower than 50 g/km, these vehicles' fuel consumption gap had a standard deviation of 123%. The variability of the fuel consumption gap exhibits a decreasing trend as the type-approval CO_2 emissions increase and a small

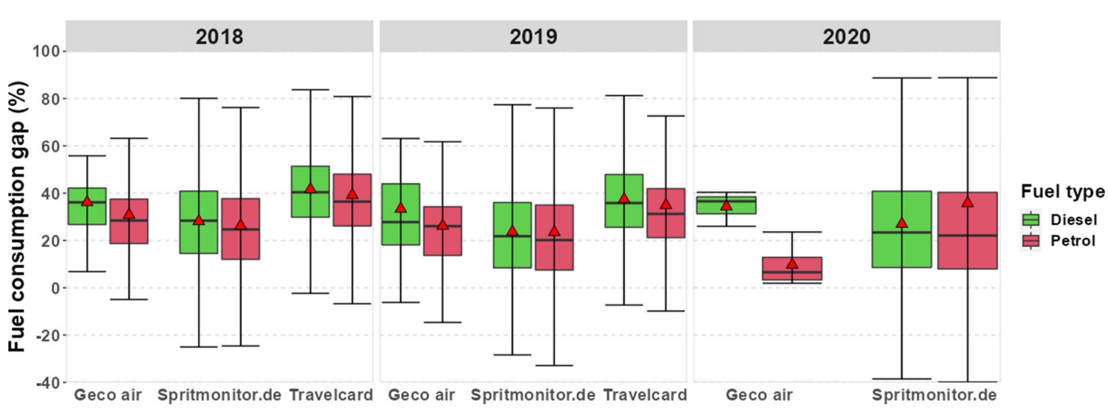


Fig. 3. Fuel consumption gap by fuel type. Red triangles depict the mean values.

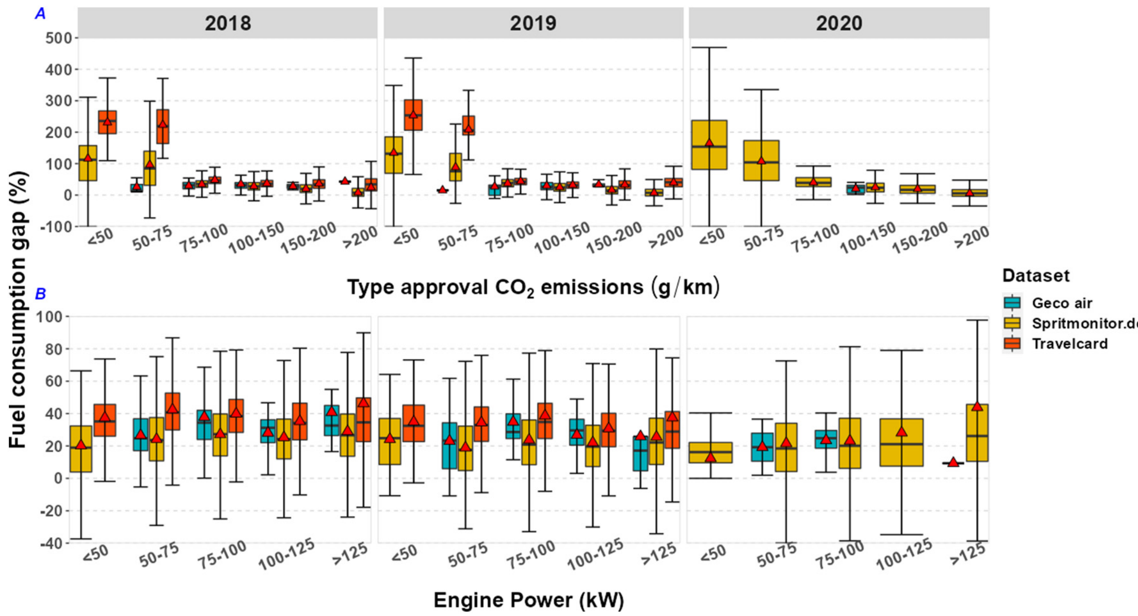


Fig. 4. Fuel consumption gap by: A) type-approval CO_2 emissions and B) engine power. Red triangles depict the mean values.

rebound to a more variable fuel consumption gap for vehicles with high type-approval CO_2 emissions (>200 g/km).

The relationship between the fuel consumption gap and the engine power is depicted in Fig. 4B. It can be seen that the average fuel consumption gap and its variability do not differ significantly between the subgroups, as when stratifying by type-approval CO_2 emissions. In most cases vehicles with high engine power (>125 kW) have the largest average and the most variable fuel consumption gap. This can be partially attributed to the fact that PHEVs are in average larger vehicles with larger engine power. For example, the bin with the largest standard deviation (72%) is composed of SM 2020 vehicles with high engine power (>125 kW) and 14% of these vehicles are PHEVs. The engine power and the engine displacement are highly correlated and the impact of the engine displacement to the fuel consumption gap is similar to the engine's power. It should be noted that the two discretizations adopted in Fig. 4 are not the same as the ones used when stratifying. In the stratification schemes, the type-approval CO_2 emissions and the engine power have been grouped in 19 and 13 bins, respectively. While in Fig. 4, they have been grouped in 6 and 5 bins, respectively.

3.1.2. EEA simulated data

Statistical results of the EEA simulated fuel consumption gap for 2018, 2019, and 2020 are presented in Table 2. The dataset that influenced the simulation process the most is TC, because it is larger and more representative of the European fleet in respect to the technical characteristics utilised (Ktistakis et al., 2021). This is especially true in respect to the manufacturer and type-approval CO_2 emissions, which were the two most important variables in simulating the real-world fuel consumption (Fig. 1). The SM dataset contains in average larger vehicles than the European fleet and is more representative of the German sub-fleet (Ktistakis et al., 2021b) and GA is much smaller than the other two user datasets. Therefore, for 2018 and 2019 their impact on the simulation procedure is lower. In 2020, there were no TC data available; hence, the simulation is mainly based on SM data. However, this does not affect the methodology and results are expected to be not significantly different even if TC data were available. This is because, firstly, the simulation method was made in such a way that only the technical characteristics of the vehicles matter. Secondly, because the percentage of PHEVs in TC is smaller than in SM and these vehicles are the ones that mainly contribute to the increase of the average 2020 fuel consumption gap and its variability.

Table 2
Simulated fuel consumption gap of EEA datasets.

Type-approval CO_2	Year	Vehicles	Fuel consumption gap (%)			PHEVs (%)
			Mean	Median	Standard deviation	
NEDC	2018	14,775,651	35	33	30	1.0
	2019	14,898,042	32	29	30	1.3
	2020	10,620,552	40	32	50	5.4
WLTP	2018	4,355,449	11	8	26	1.0
	2019	14,444,928	9	7	27	1.3
	2020	10,303,437	20	10	51	5.3

In this research, unless mentioned otherwise, NEDC type-approval CO_2 emissions are examined and from these values the NEDC fuel consumption gap is calculated (Table 2). The average fuel consumption gap, when the WLTP type-approval CO_2 emissions are used, is also presented in Table 2. The WLTP fuel consumption gap was available for less than one-third (4.36 out of 14.76 million) of the vehicles in 2018. In 2019 and 2020, the vast majority of vehicles analysed had both NEDC and WLTP type-approval CO_2 emissions available, 14.15 M out of 14.90 M and 10.30 M out of 10.62 M, respectively.

The distribution of the 2018 and 2019 simulated EEA fuel consumption gap is closer to that of GA. However, its variability, because it takes into account data from more than one country, is higher. The average fuel consumption gap of the fleet decreased from 35% in 2018 to 32% in 2019, while its variability remained the same at about 30% (Table 2). From 2019 to 2020 an increase of the average fuel consumption gap of the fleet from 32% to 40% and a steep increase of the variability from 30% to 50% was observed. The main reason for this increase is that the share of PHEVs more than quadrupled (from 1.3% in 2019 to 5.4% in 2020). This means that for the mean value, the influence of PHEVs was more than 4 times higher in 2020, compared to 2019.

For all three years, the average WLTP gap is 20–24% lower compared to the NEDC one. These values perfectly agree with findings of Dornoff et al. (2020) who reported that WLTP reduces the gap by 22 to 24% compared to NEDC. Concerning the variability of the gap, in 2018 and 2019 the WTLP standard deviation was 3–4% lower, compared to the NEDC gap variability, while in 2020 it was 1% higher. This supports the claim that the following sampling calculations and results would hold for vehicles introduced after 2021, using only WLTP type-approval CO_2 emissions to compare the gap.

3.2. Sample size for different sampling methods

In this section, the required sample size to achieve specified levels of accuracy for probability sampling is examined.

3.2.1. Simple random sampling

For simple random sampling, the sample size to reach a ME of 1% with a CL of 99% (high accuracy) and a ME of 2% with a CL of 95% (lower accuracy), is presented in Fig. 5. The sample size was calculated for a standard deviation starting from 20% up to 120% (numbers chosen based on the strata with the highest and the lowest variability when stratifying by type-approval CO_2 emissions). The population sizes examined were 10, 15, 30 and 45 million vehicles, which correspond approximately to the number of vehicles registered in the EU in the previous years and to a possible number of registered vehicles the next 1, 2 and 3 years. The exact sample sizes are presented in Fig. 5B for all the cases of the simulated fuel consumption gap. The large increase in population size does not necessarily translate in a substantial increase of the sample size. For 10, 15, 30 and 45 million vehicles the sample sizes, for the same ME, CL and standard deviation, are almost the same. The biggest increase is found for high accuracy and a standard deviation of 120%, it is only 702 vehicles (94,639 and 95,341) for a population of 10 and 45 million vehicles, respectively. This means that if instead of sampling vehicles annually, the sampling procedure was repeated every three years, the required sample size remains almost the same. This is oversimplified, as additional variabilities would be included as the models and vehicles change every year. Also for a stratification sampling scheme, further considerations are needed; however, the conclusion still holds.

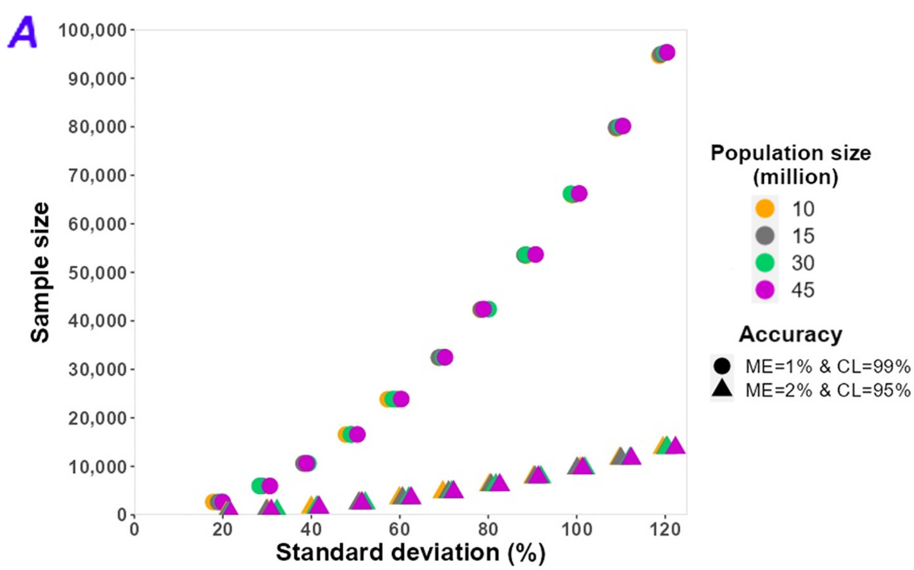
For the fleet of the EEA 2018 vehicles, which are about 14.78 million with a standard deviation of about 30%, a sample size of 6,025 and 873 is required to reach a high and lower accuracy, respectively. In 2019, there are approximately 14.90 million vehicles with a standard deviation of approximately 30% and a sample size of 5,851 and 848 is required to reach the high and the lower accuracy, respectively. The reason that in 2019 less vehicles are required to reach both accuracies is because the standard deviation in 2019 was lower compared to 2018 (29.7 and 30.1%, respectively) and this 0.4% difference weights more than that the 2019 population size being higher by 0.12 M vehicles. The 2020 fleet has a much smaller population size (10.62 million vehicles), but a substantially larger standard deviation (50%), which leads to a steep increase of required sample size. For a high accuracy, 16,240 vehicles have to be randomly sampled, while for the lower accuracy, 2,354 are sufficient.

The difference in required population size for high and lower accuracy is also shown in Fig. 5. As the population standard deviation increases, so does the divergence between the required sample sizes. A population of 15 million vehicles, for a standard deviation of 20% requires 385 and 2,654 vehicles for a lower and high accuracy, respectively. For a standard deviation of 50% 2,401 and 16,569 vehicles and for 120% 13,817 and 94,938 vehicles are needed. For the rest of this paper, concerning the accuracy of the average fuel consumption gap of the fleet, a ME equal to 1% with a CL equal to 99% (high accuracy) is used.

3.2.2. Stratified sampling

In stratified sampling, firstly the stratification variable(s) have to be chosen. In this section, various stratification variables are investigated and comparisons and conclusions are presented. The main purpose was to determine the stratification variables that need the smallest sample size to accurately predict the population's fuel consumption gap. This is equivalent to identifying the stratification schemes that would result in the highest precision with the same sample size.

For stratifying, six variables were used: powertrain type, fuel type, manufacturer, type-approval CO_2 emissions, engine power and engine displacement. The 6 stratification variables can be combined in 63 different stratification schemes. There are 6 univariate stratifications, 15 stratifications using 2 variables, 20 with 3, 15 with 4, 6 with 5 and 1 using all 6 variables. In Fig. 6 the required sample sizes of all combinations of variables



B

Year	Vehicles	SD	Accuracy High	Accuracy Lower
2018	14,775,651	30	6,025	873
2019	14,898,042	30	5,851	848
2020	10,620,552	50	16,240	2,354

Fig. 5. Sample size for simple random sampling. High Accuracy: ME = 1% & CL = 99%; Lower Accuracy: ME = 2% & CL = 95%; SD = Standard deviation.

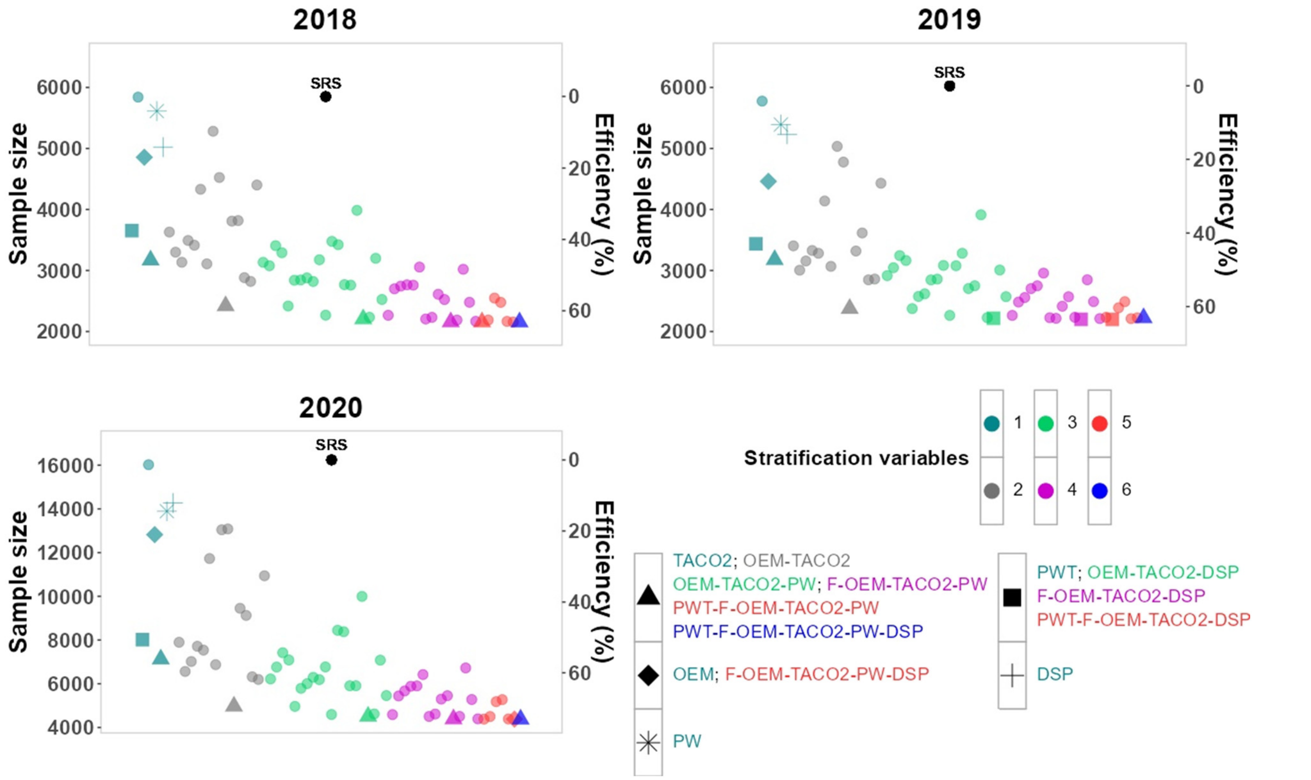


Fig. 6. Sample size and efficiency using univariate stratification and best (lowest sample size) stratification schemes. SRS = sample size for simple random sampling; F = fuel type; PWT = powertrain type; OEM = manufacturer; TACO2 = type-approval CO_2 emissions; PW = engine power; DSP = engine displacement.

are presented for the populations of all vehicles registered in 2018–2020 (Table 2). The results refer to Neyman allocation, as explained in Section 2.2.2. In addition, for comparison reasons the sample size of simple random sampling and the efficiency of each stratification scheme (reduction in the sample size compared to the sample size of simple random sampling) are also shown on the same figure. The 6 univariate stratifications, the stratification using all 6 variables and the best (requiring the smallest sample size) stratifications with 2, 3, 4 and 5 variables are labelled. In Table SM2 (Supplementary material), the sample size and efficiency of these and other combinations of interest are presented.

The most efficient univariate stratification for all three years was achieved when stratifying by the type-approval CO_2 emissions. In this case, the sample efficiency ranged from 46% to 56%, for the three years and the sample size from 3,170 vehicles in 2018 to 7,129 in 2020. The second most efficient univariate stratification was when stratifying by powertrain type, which resulted in an efficiency of 39–51% followed by manufacturer stratification (19–24%). The reason the type-approval CO_2 emissions and the powertrain type are the most efficient sampling variables is because, as shown in Fig. 4, the means and standard deviations of the fuel consumption gap between the strata differ a lot when stratifying using these variables. These differences can be directly linked to the sample efficiency (Cochran, 1977; Ktistakis et al., 2021). In Fig. 3 it was evident that the fuel consumption gap between diesel and petrol vehicles did not differ a lot and hence stratifying by fuel type leads to the smallest sample efficiency (1–3%) and results in the biggest sample size (5,841, 5,776 and 16,021 for 2018, 2019 and 2020, respectively).

The most efficient stratification with two variables is achieved, for all three years, when stratifying by manufacturer and type-approval CO_2 emissions (59–69% efficiency) followed by the type-approval CO_2 emissions and engine displacement (51–62% efficiency). Stratifying by type-approval CO_2 emissions and engine power has a similar efficiency of 51–61%. Stratifying by powertrain type and type-approval CO_2 emissions has an efficiency of 46–57%. There is almost no benefit when stratifying by these two variables compared to when stratifying only by type-approval CO_2 emissions (46–56% efficiency). This is because the type-approval CO_2 emissions capture almost all the variability of the powertrain type and almost all vehicles with very low type-approval CO_2 emissions are PHEVs (Fig. 4). The least efficient combinations of two variables are stratifying by fuel type and engine power (12–20% efficiency) and fuel type and engine displacement (18–25% efficiency). In both cases, adding the fuel type provides a marginal increase of 6–8% in the efficiency, compared to stratifying only by engine displacement or engine power.

The best stratification schemes with three variables, for all three years, are by the manufacturer, type-approval CO_2 emissions and engine power (62–72% efficiency) and by manufacturer, type-approval CO_2 emissions and engine displacement (61–72% efficiency). The worst is by fuel type, engine power and engine displacement (33–38% efficiency) which is again worse than stratifying just by type-approval CO_2 emissions (46–58% efficiency). A combination which could be of future interest is to stratify by the powertrain, fuel type and manufacturer, this has an efficiency of 48–62% and sample size ranging from 2,916 to 6,220 vehicles, depending on the year. This stratification scheme could be used in practice because it does not require checking any vehicle documentation (certificate of conformity, registration document) and can be easily retrieved by the drivers.

Adding more stratification variables produces only a marginal increase in efficiency and decrease in sample size. When stratifying with all six variables the efficiency is 62–73%. Therefore, one can conclude that by adding more stratification variables the analysis becomes more complex in practical implementation, but the gains are marginal and do not offset technical difficulties. Consequently, the combinations of three variables: manufacturer, type-approval CO_2 emissions and engine power and manufacturer, type-approval CO_2 emissions and engine power are optimal. For the rest of this study, the first will be called the optimal stratification scheme.

Until the OBFCM data are available for the whole fleet, one cannot fully trust the strata standard deviations, used in the Neyman stratification. A way to avoid this is to use proportional allocation instead of Neyman allocation. In this approach, the sample sizes per stratum are proportional to the strata sizes, without taking into consideration the variability of each stratum. The consequence of feeding less information to the model is that the sampling efficiency decreases (sample size increases). When stratifying by manufacturer, the efficiency is 3, 7 and 5% respectively for 2018, 2019 and 2020 (Table SM2 Supplementary material) which is a substantial decrease from 19, 24 and 21% calculated for Neyman allocation. For stratifying by type-approval CO_2 emissions there is an 8, 9 and 18% decrease, respectively per year. For the optimal stratification of the manufacturer, type-approval CO_2 emissions and engine power the efficiency decreases from 62 to 72% (Neyman allocation) to 52–54% (proportional allocation).

The decrease in sample size, when using stratification sampling with Neyman allocation compared to simple random sampling (efficiency), can be partitioned in two: 1.The decrease in sample size between simple random sampling and stratification sampling with proportional allocation and 2. The decrease between proportional and Neyman allocation. The first depends on the variability of the strata means and the second on how variable are the strata standard deviations. To quantify these variabilities, the Coefficient of Variation (CV) was used. As shown in Fig. SM1 (Supplementary material) there is a strong correlation between the sample efficiency and the CV of the standard deviation (CV_{SD}). The CV_{SD} explains 56% of the efficiency's variability, as found by performing a simple linear regression analysis ($R^2 = 56\%$). It was observed in Fig. SM1 that there is a correlation between the number of stratification variables and the efficiency. Taking this into account, a multiple linear regression model was utilised, using as explanatory variables the CV_{SD} and the number of stratification variables. This way the R^2 increased to 74%. Adding as explanatory variables the year and the CV of the strata mean, the R^2 only marginally increased to 75%.

3.2.3. Strata specific estimators

When precise estimators (5% ME and 90% CL) are required for all strata and not only for the mean value of the whole fleet, the required sample size increases. For univariate stratification, a larger number of strata and a higher variability of the fuel consumption gap per stratum, result in a larger increase in the sample size, compared to the normal stratified sampling (Section 3.2.2). In multivariate stratification, there are two options, either to demand precise estimators for all strata or only for some of the stratification variables. If the stratification variables are the manufacturer and type-approval CO_2 emissions, there are 1,463 strata (77 manufacturers □19 type-approval CO_2 bins) for the fleet of vehicles registered in 2018. Out of these 1,463 strata, 557 have at least one vehicle. If precise estimators for all 557 strata are needed, the required total sample size would be 32,158 vehicles (Table SM2 Supplementary material). However, if only precise estimators per manufacturer are required (77 strata), then the sample size would be 8,403 vehicles. The number of vehicles in the case of stratifying by manufacturer and type-approval CO_2 emissions and requiring precise estimators only per manufacturer (8,403 vehicles) is close to the number of vehicles when stratifying by manufacturer and requiring precise estimators for each manufacturer (8,491) (Table SM2 Supplementary material). This applies for all cases, hence only sample sizes for when estimators for every stratum are required will be discussed.

Fig. 7 shows the sample size required to get precise estimators of the mean fuel consumption gap, for the fleet of newly registered vehicles in the EU, in 2018, 2019 and 2020, for different stratification schemes, when precise estimators are also required for all strata. The required sample sizes of all combinations of variables are presented. In addition, for comparison reasons the sample size for simple random sampling and the sample size increase (%) of each stratification scheme (compared to the sample size for simple random sampling) are also shown on the same figures. The stratification schemes requiring the smallest sample size per each number of stratification variables (1, 2, . . , 6) and the case of stratifying per OEM (because of its importance) are labelled. In Table SM2 (Supplementary material) the sample size and sample size increase of combinations of interest are presented.

When only the mean fuel consumption gap of the whole population is required to be precise, increasing the stratification variables leads to a lower sample size. On the other hand, when the mean fuel consumption gap should be precise for all strata, increasing the stratification variables leads to a steep increase in the sample size. The stratification variables for which the largest sample size is needed to estimate accurately the fuel consumption gap are the manufacturer and the type-approval CO_2 emissions. This is because they have a larger number of strata and the variabilities within the strata they form are not small. When precise estimators are required for each manufacturer, compared to normal stratified sampling, an additional 3,635 (4,856 ⇒ 8,491), 3,395 (4,461 ⇒ 7,856) and 4,380

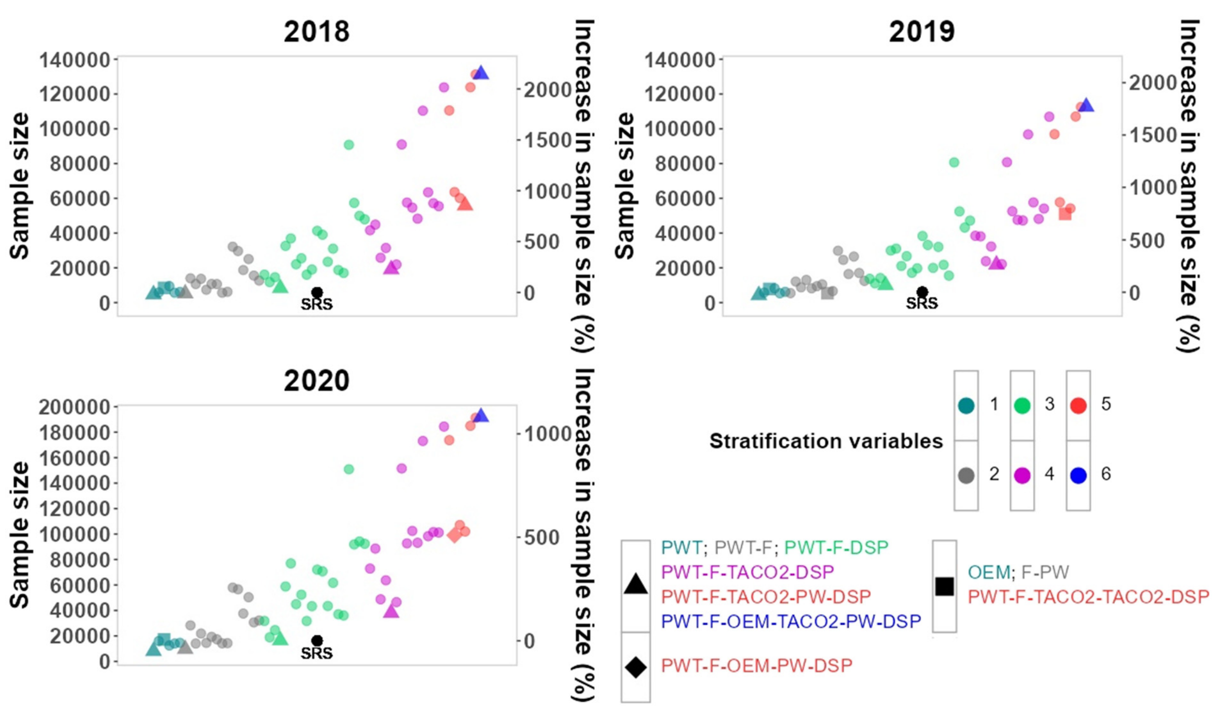


Fig. 7. Sample size when in-strata estimators are required for stratification schemes with lowest sample size and by OEM. SRS = sample size for simple random sampling; F = fuel type; PWT = powertrain type; OEM = manufacturer; TACO2 = type-approval CO_2 emissions; PW = engine power; DSP = engine displacement

(12,823 ⇒ 17,203) vehicles are needed, respectively in 2018, 2019 and 2020. To get precise estimators per type-approval CO_2 emissions 9,488, 8,218 and 12,486 vehicles are required. It should be noted that in 2020, stratifying by type-approval CO_2 emissions and requiring precise estimators for all strata, increases the sample size, but it is still lower than the sample size required with simple random sampling (16,240). For the optimal sampling scheme of manufacturer, type-approval CO_2 emissions and engine power 90,782, 80,661 and 158,879 vehicles are required respectively for 2018, 2019 and 2020. These numbers translate to a 1,407, 1,279 and 859% increase in sample size compared to simple random sampling. When stratifying by all 6 variables, if precise estimators are needed for all strata, the required sample size is 131,330, 112,658 and 191,894 vehicles for 2018, 2019 and 2020, respectively (2,080, 1,825 and 1,082% increase in sample size compared to simple random sample).

If a higher accuracy for every stratum is required and instead of a CL of 90% and ME of 5%, a CL of 95% and ME of 2% is used, the total sample size steeply increases (Table SM2 Supplementary material). For stratifying by manufacturer, type-approval CO_2 emissions and engine power the required sample sizes are 491,587, 444,048 and 693,540 vehicles, respectively, for 2018, 2019 and 2020 (which corresponds to 3.3, 3.0, and 6.5% of the fleets' population). When strata specific estimators are required for each OEM, instead of 8,491, 7,856 and 17,203 vehicles that were needed with the lower precision requirements, 41,192, 39,789 and 71,664 vehicles should be sampled.

3.3. Practical implementations of quota sampling

In quota sampling, selecting as quota controls (stratification variables) those that affect the fuel consumption gap and contribute to its variability can give some degree of control over the selection biases. The relationship between the fuel consumption gap and other variables were examined in Section 3.1 and the efficiencies of using these variables in univariate and multivariate stratified sampling designs were compared in Section 3.2. The vehicle manufacturer, the type-approval CO_2 emissions and the powertrain type are the three variables that required the least sample size in a univariate stratification scheme. However, the combination of type-approval CO_2 emissions and powertrain type is not optimal, because the type-approval CO_2 emissions include in a way the information of the powertrain. Instead of the powertrain type, the engine power or the engine displacement (two highly correlated variables) should be used as a third quota control. This leads to the optimal sampling scheme: using as dependent quota controls the manufacturer, type-approval CO_2 emissions and engine power (or engine displacement). When using this sampling scheme, for stratification sampling, the sample size required to reach a pre-specified accuracy was small (2,206–45,10 vehicles). Equivalently with a sufficiently big pre-specified sample size and composition, the accuracy is expected to be high. For quota sampling a sample of at least 4,500 vehicles (0.04% of the 2020 fleet) would be befitting. Adding more quota controls can increase the representativeness and accuracy of the sample, however, the difficulty and the sampling cost escalate.

3.4. Impact of non-sampling errors

Using probability-sampling methods (simple random sampling and stratifying sampling), this study attempted to minimize the sample size, while keeping under control the sampling error (low ME and high CL). In the case of non-probability sampling (quota sampling) by using appropriate quota controls and choosing a representative sample of sufficient sample size, the sampling error can be controlled. However, in practice, there are also non-sampling errors. By designing carefully all steps of the sampling procedure most types of non-sampling errors can be minimized or even eliminated (Ktistakis et al., 2021). One example of non-sampling error, of interest in this research, is the instrument error (laboratory and real-world OBFCM error), which could even result in a systematic error.

It should be noted that the user-based datasets that were used in this study to simulate the fuel consumption gap of the EU fleet do not come from the OBFCMs. However, in the future data will be extracted from the OBFCM devices whose accuracy is set to ±5% compared to the fuel consumption measured in the laboratory during WLTP type-approval test (Pavlovic et al., 2020). OBFCM accuracy is expected to improve as the new technologies are introduced and the requirements and tests become stricter. Hence, it is of importance to quantify the measurement error of these devices and to understand the impact that it has on the sampling scheme. According to previously published research (Pavlovic et al., 2021), in real-world conditions OBFCM devices were found to have a mean error/inaccuracy of 2.7% with a standard deviation equal to 5.6%, compared to the values measured by Portable Emissions Measurement Systems (PEMS). During laboratory (WLTP) tests the OBFCM devices had a mean error/inaccuracy of 1.5% (standard deviation of 4.5%).

To get an indication of the impact of the OBFCM inaccuracies on the sampling procedure and the sample size, a "noise" amounting to the respective on-road and type-approval OBFCM error was added to the real-world and to the type-approval fuel consumption of all vehicles. The noise is not a cumulative addition to the fuel consumptions, but a percentage dispersion. By doing this, the standard deviation increased by 2–5% (Table 3). Consequently, when using simple random sampling, the sample size increased by 10–39%. In 2020, the average fuel consumption and its variability are higher and hence the impact of the noise is smaller. In addition, the sample size is larger so an increase in standard deviation affects it less. The increase in standard deviation for 2018 data is 5% and for 2019 3%, this difference is deemed to be a matter of chance. The noise happened to influence the simulated 2018 data in such a way that the fuel consumption gap's standard deviation increased more in 2018. For stratified sampling, the increases are expected to be of the same order. To get results that are more precise an estimation/understanding of the OBFCM accuracy for different sub-groups (strata) is necessary.

4. Conclusions

The fuel consumption gap issue is a subject extensively studied and legislation has been established to monitor it across the EU with a view to prevent its increase. This study focused on sampling schemes that would allow estimating accurately the gap by monitoring the smallest possible number of vehicles. Data from three user datasets were used to simulate the fuel consumption gap of the fleet of newly registered vehicles in EU countries in 2018, 2019 and 2020.

Results of simple random sampling analysis suggest that by examining 16,240 vehicles (0.14% of the total fleet) the fuel consumption gap can be estimated accurately (ME of 1% and CL of 99%). The optimal sample scheme was achieved when stratifying by manufacturer, type-approval CO_2 emissions and engine power (or engine displacement), in which case the sampling size reduced by 72%, to approximately 4,500 vehicles. When precise estimators are required per manufacturer, a sample size of around 17,200 vehicles is required. Quota sampling is an alternative non-probability sampling technique that can be utilised in case the whole fleet is not accessible. Using as dependent quota controls the manufacturer, type-approval CO_2 emissions and engine power (or engine displacement), as well as a sample size of at least 4,500 vehicles (0.04% of the 2020 fleet) should be considered.

The results presented in this paper provide a basis for setting up a sampling procedure to monitor and quantify the fuel consumption gap of the future EU fleet and the manufacturers' sub-fleets. The study confirmed that by monitoring less than 0.05% of the total fleet, the fuel consumption gap can

Table 3
Impact of non-sampling error (OBFCM error) in simple random sampling.

Dataset	Fuel consumption gap standard deviation increase (absolute)	Sample size before	Sample size after	Sample size increase (%)
EEA 2018	5%	6,025	8,382	39
EEA 2019	3%	5,851	6,938	19
EEA 2020	2%	16,240	17,847	10

be estimated with high accuracy (ME of 1% and CL of 99%). A periodical sample scheme can be established that will utilise previous years' data to improve the future estimates. The results of this study also highlight that special attention should be given to PHEVs due to their higher and more variable fuel consumption gap and their expected increase in sales in the years to come.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Markos A. Ktistakis: Conceptualization, Methodology, Formal analysis, Data curation, Software, Visualisation, Writing - original draft. **Jelica Pavlovic:** Conceptualization, Methodology, Visualisation, Writing - Review & Editing. **Georgios Fontaras:** Conceptualization, Methodology, Writing - Review & Editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The scientific output expressed does not imply a policy position of the European Commission.

Acknowledgements

The authors gratefully acknowledge Filip Francois and Carlos Serra from the European Commission's Directorate-General for Climate Action (DG CLIMA) for their constructive comments and revising the manuscript. The authors would like to express their appreciation to Pierre Michel, Alexandre Chasse, Sol Selene Rodriguez, Ilias Cheimariotis and Gilles Corde from the Institut Français du Pétrole Energies Nouvelles (IFPEN) for supplying the French (Geco air) real-world dataset. The authors would like to acknowledge the support of Travelcard B.V., and Norbert Ligterink from the Netherlands Organisation for Applied Scientific Research (TNO) for providing the Dutch (Travelcard) dataset.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2022.154943.

References

Ben Dror, M., Qin, L., An, F., 2019. The gap between certified and real-world passenger vehicle fuel consumption in China measured using a mobile phone application data. Energy Policy 128, 8–16. https://doi.org/10.1016/j.enpol.2018.12.039.

Blanck, R., Kasten, P., Jöhrens, J., Kräck, J., Räder, D., Kräck, J., Mathieu, L., 2020. Plug-in hybrid electric cars: market development, technical analysis and CO_2 emission scenarios for Germany. [WWW Document]. URL https://www.oeko.de/uploads/oeko/oekodoc/PHEV-Report-Market-Technology-CO2.pdf. (Accessed 25 December 2021).

Buysse, C., Miller, J., 2021. Transport could burn up the EU's entire carbon budget. [WWW Document]. URLInternational Council on Clean Transportation (ICCT). https://theicct.org/blog/staff/eu-carbon-budget-apr2021. (Accessed 25 December 2021).

Cames, M., Helmers, E., 2013. Critical evaluation of the European diesel car boom - global comparison, environmental effects and various national strategies. Environ. Sci. Eur. 25, 15. https://doi.org/10.1186/2190-4715-25-15.

Cochran, W.G., 1977. Sampling Techniques. 3rd edition. John Wiley & Sons, New York.

Dornoff, J., 2019. One goal, multiple pathways: a review of approaches for transferring on-board fuel consumption meter data to the European Commission. White paper [WWW Document]. URLInternational Council on Clean Transportation. https://theicct.org/publications/transferring_obfcm_fuel_data_ec. (Accessed 25 December 2021).

Dornoff, J., Tietge, U., Mock, P., 2020. On the way to "real-world" CO2 values: the European passenger car market in its first year after introducing the WLTP. White paper [WWW Document]. URLInternational Council on Clean Transportation. https://theicct.org/publications/way-real-world-co2-values-european-passenger-car-market-its-first-year-after. (Accessed 25 December 2021).

EC, 2009. Regulation (EC) No 443/2009 of the European Parliament and of the Council of 23 April 2009 setting emission performance standards for new passenger cars as part of the Community's integrated approach to reduce CO 2 emissions from light-duty vehicles. Off. J. Eur. Union OJ L 140, 1–15 5.6.2009.

EEA, 2020. Greenhouse gas emissions from transport in Europe — European Environment Agency [WWW Document]. URL https://www.eea.europa.eu/data-and-maps/indicators/transport-emissions-of-greenhouse-gases-7/assessment. (Accessed 25 October 2021).

EEA, 2021. CO2 performance of new passenger cars in Europe — European Environment Agency [WWW Document]. URL https://www.eea.europa.eu/data-and-maps/indicators/average-co2-emissions-from-motor-vehicles-1/assessment. (Accessed 25 December 2021).

EEA, 2021. Monitoring of CO2 emissions from passenger cars – Regulation (EU) 2019/631 — European Environment Agency. URL https://www.eea.europa.eu/data-and-maps/data/co2-cars-emission-20. (Accessed 25 October 2021).

EEA, 2021. New registrations of electric vehicles in Europe — European Environment Agency [WWW Document]. URL https://www.eea.europa.eu/data-and-maps/indicators/proportion-of-vehicle-fleet-meeting-5/assessment. (Accessed 25 December 2021).

EU, 2017. Regulation (EU) 2017/1151 of 1 June 2017 supplementing Regulation (EC) No 715/2007 of the European Parliament and of the Council on type-approval of motor vehicles with respect to emissions from light passenger and commercial vehicles (Euro 5 and Euro 6) and on access to vehicle repair and maintenance information, amending Directive 2007/46/EC of the European Parliament and of the Council, Commission Regulation (EC) No 692/2008 and Commission Regulation (EU) No 1230/2012 and repealing Commission Regulation (EC) No 692/2008. Off. J. Eur. Union OJ L 175 7.7.2017.

EU, 2019. Regulation (EU) 2019/631 of the European Parliament and of the Council of 17 April 2019 setting CO2 emission performance standards for new passenger cars and for new light commercial vehicles, and repealing Regulations (EC) No 443/2009 and (EU) No 510/2011. Off. J. Eur. Union OJ L 111 25.4.2019.

EU, 2021. Regulation (EU) 2021/392 of 4 March 2021 on the Monitoring and Reporting of Data Relating to CO2 Emissions From Passenger Cars and Light Commercial Vehicles Pursuant to Regulation (EU) 2019/631 of the European Parliament and of the Council and Repealing Commission Implementing Regulations (EU) No 1014/2010, (EU) No 293/2012, (EU) 2017/1152 and (EU) 2017/1153. OJ L 77 5.3.2021.

Fontaras, G., Dilara, P., 2012. The evolution of European passenger car characteristics 2000–2010 and its effects on real-world CO2 emissions and CO2 reduction policy. Energy Policy 49, 719–730. https://doi.org/10.1016/j.enpol.2012.07.021.

Fontaras, G., Zacharof, N.G., Ciuffo, B., 2017. Fuel consumption and CO2 emissions from passenger cars in Europe – laboratory versus real-world emissions. Prog. Energy Combust. Sci. 60, 97–131. https://doi.org/10.1016/j.pecs.2016.12.004.

Fontaras, G., Valverde, V., Arcidiacono, V., Tsiakmakis, S., Anagnostopoulos, K., Komnos, D., Pavlovic, J., Ciuffo, B., 2018. The development and validation of a vehicle simulator for the introduction of worldwide harmonized test protocol in the European light duty vehicle CO2 certification process. Appl. Energy 226, 784–796. https://doi.org/10.1016/j.apenergy.2018.06.009.

Frondel, M., Schmidt, C.M., Vance, C., 2011. A regression on climate policy: the European Commission's legislation to reduce CO2 emissions from automobiles. Transp. Res. A Policy Pract. 45, 1043–1051. https://doi.org/10.1016/j.tra.2009.12.001.

Geco Air, 2021. Mobility and Air Quality. URLGeco Air. https://www.gecoair.fr/en/. (Accessed 25 December 2021).

Huang, Y., Surawski, N.C., Organ, B., Zhou, J.L., Tang, O.H.H., Chan, E.F.C., 2019. Fuel consumption and emissions performance under real driving: comparison between hybrid and conventional vehicles. Sci. Total Environ. 659, 275–282. https://doi.org/10.1016/j.scitotenv.2018.12.349.

Jiménez, J.L., Valido, J., Molden, N., 2019. The drivers behind differences between official and actual vehicle efficiency and CO2 emissions. Transp. Res. Part D: Transp. Environ. 67, 628–641. https://doi.org/10.1016/j.trd.2019.01.016.

Krause, J., Thiel, C., Tsokolis, D., Samaras, Z., Rota, C., Ward, A., Prenninger, P., Coosemans, T., Neugebauer, S., Verhoeve, W., 2020. EU road vehicle energy consumption and CO2 emissions by 2050 – expert-based scenarios. Energy Policy 138, 111224. https://doi.org/10.1016/j.enpol.2019.111224.

Ktistakis, M.A., Pavlovic, J., Fontaras, G., 2021a. Sampling Approaches for Road Vehicle Fuel Consumption Monitoring, EUR 30420 EN. 2021. Publications Office of the European Union, Luxembourg 978-92-76-23986-4. https://doi.org/10.2760/39369 JRC122063.

Ktistakis, M.A., Pavlovic, J., Fontaras, G., Michel, P., Rodriguez, S.S., Chasse, A., 2021b. Understanding the Fuel Consumption Gap Using Self-Reported User and Smartphone Application Datasets. Presented at the International Transport and Air Pollution Conference (TAP), Graz, Austria. Proceedings of the 24th International Transport and Air Pollution (TAP) Conference. 2022. Publications Office of the European Union, Luxembourg 978-92-76-43803-8. https://doi.org/10.2760/019404 JRC126978.

Logan, K.G., Nelson, J.D., Lu, X., Hastings, A., 2020. UK and China: will electric vehicle integration meet Paris agreement targets? Transp. Res. Interdiscip. Perspect. 8, 100245. https://doi.org/10.1016/j.trip.2020.100245.

Michel, P., Rodriguez, S.S., Chasse, A., Fontaras, G., Pavlovic, J., Ktistakis, M.A., Komnos, D., 2021. Conventional Engine Vehicles Model Parametrization for Real-World Fuel Consumption Estimation. Presented at the International Transport and Air Pollution Conference TAP, Graz, Austria. Proceedings of the 24th International Transport and Air Pollution (TAP) Conference. 2022. Publications Office of the European Union, Luxembourg 978-92-76-43803-8. https://doi.org/10.2760/019404 JRC126978.

Ntziachristos, L., Mellios, G., Tsokolis, D., Keller, M., Hausberger, S., Ligterink, N.E., Dilara, P., 2014. In-use vs. type-approval fuel consumption of current passenger cars in Europe. Energy Policy 67, 403–411. https://doi.org/10.1016/j.enpol.2013.12.013.

Pavlovic, J., Ciuffo, B., Fontaras, G., Valverde, V., Marotta, A., 2018. How much difference in type-approval CO2 emissions from passenger cars in Europe can be expected from changing to the new test procedure (NEDC vs. WLTP)? Transp. Res. A Policy Pract. 111, 136–147. https://doi.org/10.1016/j.tra.2018.02.002.

Pavlovic, J., Fontaras, G., Ktistakis, M.A., Anagnostopoulos, K., Komnos, D., Ciuffo, B., Clairotte, M., Valverde, V., 2020. Understanding the origins and variability of the fuel consumption gap: lessons learned from laboratory tests and a real-driving campaign. Environ. Sci. Eur. 32, 53. https://doi.org/10.1186/s12302-020-00338-1.

Pavlovic, J., Fontaras, G., Broekaert, S., Ciuffo, B., Ktistakis, M.A., Grigoratos, T., 2021. How accurately can we measure vehicle fuel consumption in real world operation? Transp. Res. Part D: Transp. Environ. 90, 102666. https://doi.org/10.1016/j.trd.2020.102666.

Plötz, P., Funke, S.Á., Jochem, P., 2018. The impact of daily and annual driving on fuel economy and CO2 emissions of plug-in hybrid electric vehicles. Transp. Res. A Policy Pract. 118, 331–340. https://doi.org/10.1016/j.tra.2018.09.018.

Plötz, P., Moll, C., Bieker, G., Mock, P., 2021. From lab-to-road: real-world fuel consumption and CO 2 emissions of plug-in hybrid electric vehicles. Environ. Res. Lett. 16, 054078. https://doi.org/10.1088/1748-9326/abef8c.

Rottoli, M., Dirnaichner, A., Pietzcker, R., Schreyer, F., Luderer, G., 2021. Alternative electrification pathways for light-duty vehicles in the european transport sector. Transp. Res. Part D: Transp. Environ. 99, 103005. https://doi.org/10.1016/j.trd.2021.103005.

Sánchez, P.P.R., Ndiaye, A.B., Martín-Cejas, R., 2019. Plug-in hybrid electric vehicles (PHEVS): a possible perverse effect generated by environmental policies. Int. J. Transp. Dev. Integr. 3, 259–270. https://doi.org/10.2495/TDI-V3-N3-259-270.

Siskos, P., Zazias, G., Petropoulos, A., Evangelopoulou, S., Capros, P., 2018. Implications of delaying transport decarbonisation in the EU: a systems analysis using the PRIMES model. Energy Policy 121, 48–60. https://doi.org/10.1016/j.enpol.2018.06.016.

Spritmonitor.de, 2021. MPG and Cost Calculator and Tracker. URL. https://www.spritmonitor.de/en/. (Accessed 25 December 2021).

Thibault, L., Dégeilh, P., Sabiron, G., Voise, L., Thanabalasingam, K., Corde, G., 2017. GNSS data-based approach to monitor real-world exhaust emissions at microscopic resolution. Presented at the International Transport and Air Pollution Conference (TAP), Zürich.

Tietge, U., Díaz, S., Yang, Z., Mock, P., 2017. From laboratory to road international: a comparison of official and real-world fuel consumption and CO2 values for passenger cars in Europe, the United States, China, and Japan | International Council on Clean Transportation. White paper [WWW Document]. URL https://theicct.org/publications/laboratory-road-intl. (Accessed 25 December 2021).

Tietge, U., Sonsoles, D., Mock, P., Bandivadekar, A., Dornoff, J., Ligterink, N., 2019. From laboratory to road. A 2018 update of official and 'real-world' fuel consumption and CO2 values for passengers cars in Europe. International Council on Clean Transportation, Netherlands Organisation for Applied Scientific Research. White paper [WWW Document]. URL https://theicct.org/publications/laboratory-road-2018-update. (Accessed 25 December 2021).

Travelcard, 2021. Tankpas of Laadpas van Travelcard. Overal tanken en laden. Travelcard. https://travelcard.nl. (Accessed 25 December 2021).

Triantafyllopoulos, G., Dimaratos, A., Ntziachristos, L., Bernard, Y., Dornoff, J., Samaras, Z., 2019. A study on the CO2 and NOx emissions performance of Euro 6 diesel vehicles under various chassis dynamometer and on-road conditions including latest regulatory provisions. Sci. Total Environ. 666, 337–346. https://doi.org/10.1016/j.scitotenv.2019.02.144.

Tsiakmakis, S., Ciuffo, B., Fontaras, G., Anagnostopoulos, K., Arcidiacono, V., Praksova, R., Marotta, A., 2016. Introducing a new emissions certification procedure for European light-duty vehicles: Monte Carlo simulation of the potential effect on fleet carbon dioxide emissions. Transp. Res. Rec. 2572, 66–77. https://doi.org/10.3141/2572-08.

Tu, R., Xu, J., Wang, A., Zhang, M., Zhai, Z., Hatzopoulou, M., 2022. Real-world emissions and fuel consumption of gasoline and hybrid light duty vehicles under local and regulatory drive cycles. Sci. Total Environ. 805, 150407. https://doi.org/10.1016/j.scitotenv.2021.150407.

Van Gijlswijk, R., Ligterink, N., 2018. Real-world Fuel Consumption of Passenger Cars Based on Monitoring of Dutch Fuel Pass Data 2017. TNO Report 2018 R10371. https://doi.org/10.13140/RG.2.2.23378.61123.

Van Gijlswijk, R., Paalvast, M., Ligterink, N., Smokers, R., 2020. Real-world Fuel Consumption of Passenger Cars and Light Commercial Vehicles. TNO report 2020 R11664. https://doi.org/10.13140/RG.2.2.24856.42249.

Yu, R., Ren, H., Liu, Y., Yu, B., 2021. Gap between on-road and official fuel efficiency of passenger vehicles in China. Energy Policy 152, 112236. https://doi.org/10.1016/j.enpol.2021.112236.

Zacharof, N.G., Tietge, U., Franco, V., Mock, P., 2016. Type approval and real-world CO2 and NOx emissions from EU light commercial vehicles. Energy Policy 97, 540–548. https://doi.org/10.1016/j.enpol.2016.08.002.