Manuscript – Final and authoritative paper published in *Environmental Science & Technology*, 18 Oct 2019. https://pubs.acs.org/doi/abs/10.1021/acs.est.9b04123

When is enough? Minimum sample sizes for on-road measurements of car emissions

Yuche Chen[‡] *, *Yunteng Zhang*[‡], *Jens Borken-Kleefeld*[†] *

[‡]Department of Civil and Environmental Engineering, University of South Carolina, USA

[†]International Institute for Applied System Analysis, Laxenburg, Austria

* Corresponding author: Yuche Chen, chenyuc@cec.sc.edu; Jens Borken-Kleefeld, borken@iiasa.ac.at

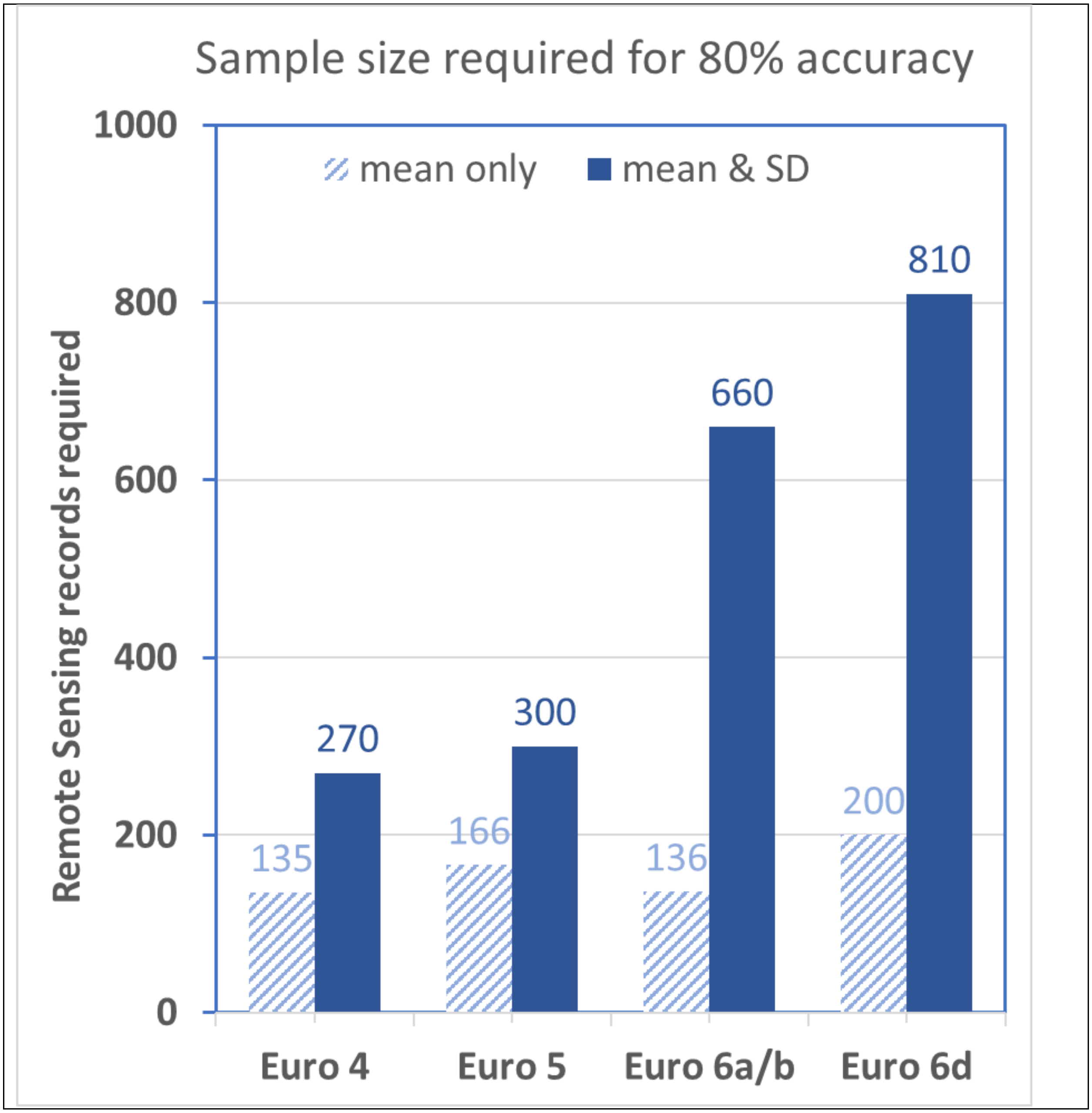
Key words: Monte Carlo simulation, bootstrap, in-use surveillance, diesel car, Europe, Euro 6, mean, variance, remote emission sensing.

Manuscript – Final and authoritative paper published in *Environmental Science & Technology*, 18 Oct 2019. https://pubs.acs.org/doi/abs/10.1021/acs.est.9b04123

ABSTRACT

The power of remote vehicle emission sensing stems from the big sample size obtained and its related statistical representativeness for the measured emission rates. But how many records are needed for a representative measurement and when does the information gain per record become insignificant? We use Monte Carlo simulations to determine the relationship between the sample size and the accuracy of the sample mean and variance. We take the example of NO emissions from diesel cars measured by remote emission monitors between 2011 and 2018 at various locations in Europe. We find that no more than 200 remote sensing records are sufficient to approximate the mean emission rate for Euro 4, 5 and 6a,b diesel cars with 80% certainty within a ±1 g NO per kg fuel tolerance margin (~±50 mg NO per km). Between 300 and 800 remote sensing records are needed to approximate also the variance of the mean NO emission rates for those diesel car technologies. This translates to only 2 and up to 9 measurement days respectively to characterize the means and their variance for a car fleet typical in Europe.

TOC Graphic


Sample size required for 80% accuracy
mean only
mean & SD
Remote Sensing records required
1000
800
600
400
200
0
135
270
166
300
136
660
200
810
Euro 4
Euro 5
Euro 6a/b
Euro 6d

Manuscript – Final and authoritative paper published in *Environmental Science & Technology*, 18 Oct 2019. https://pubs.acs.org/doi/abs/10.1021/acs.est.9b04123

INTRODUCTION

Vehicle emission remote sensing has been routinely applied in numerous US states for three decades now, followed by applications in Europe, in Hong Kong and more recently in mainland China.[1-5] Typical campaigns last between a couple of days to many weeks, with up to a million of emission records collected. An obvious assumption in these campaigns is that the larger the measurement sample the more representative it is for the fleet and driving conditions investigated. However, every further day of measurement adds to the costs. Too few records on the other hand mean that the validity of the whole sample can be compromised and hence efforts were wasted. Despite its fundamental importance we have not found sound guidance in the literature about the minimum number of emission records needed for a representative measurement nor on the statistical power of the data. Statistical theories provide tools to determine the relationship between sample size and confidence of population mean estimation. However, these formulations usually assume certain distribution characteristics of the data as a priori, such as normality, independence etc., which are usually not valid for vehicle exhaust emission rates. In addition, although population variance is an important statistic in assessing distribution and variability of vehicle emissions[17,27,28], there is no existing method that constitute relationship between sample size and confidence of population variance estimation.

Here we work on real-world data and explore their inherent relationships. We propose a bootstrap-sampling based Monte Carlo simulation to determine the relationship between size of the emission measurement sample and the statistical performance of sample mean and variance. We carry out the simulations on a set of 130,000 remote sensing emission records of Diesel cars measured between 2011 and 2018 at 23 locations across Switzerland, Sweden and the United Kingdom.[25] This unique dataset covers vehicles up to 25 years old, measurement ambient temperature from 0 to 43 Celsius degree and instantaneous vehicle specific power up to 54 kW per ton. These records comprehensively cover a wide range of real-world driving conditions and a broad spectrum of Europe's passenger car fleet. We consider this the best available sample of real emission rates to conduct our analysis on. This allows exploring of methods in deciding minimum sample size of vehicles with different emission standards under the control of vehicle age, power and ambient temperature conditions.

LITERATURE REVIEW

Sample size determination is an important component in empirical studies. A minimum number of measurements is needed to detect statistically significant effects. Traditional methods in determining sample sizes are dependent on the underlying population distribution. For example, equation (1) is used to determine a sample size whose sample mean is within E units from the population mean with (1-α) × 100% confidence.[7] σ is the standard deviation of sample mean and $z_{\alpha/2}$ is a critical value calculated based on normality assumption of population.

$$n = \frac{z_{\alpha/2}^2 \times \sigma^2}{E^2} \tag{1}$$

According to Central Limit Theory (CLT), the distribution of the sample means will be approximately normally distributed and sample mean is an unbiased estimator of population mean. Therefore, the closed-form solution in Eq. (1) is applicable in determining sample size whenever mean emission statistics are the focus. However, there is no closed-form equation that determines sample size based on accuracy of variance statistics. Variance of emission is an important statistic which determines the spread of on-road vehicle emission and has been used in various of studies focusing on estimating confidence interval of emission, distribution and variability of vehicle emissions.[17,27,28] However, limited attention is given to estimate population variance of emission based on sample variance. The variance of sample with size n is commonly assumed to follow chi-squared (χ^2) distribution with n-1 degree of freedom. But this requires sample to be drawn from normal distribution. Instantaneous vehicle emission rates however are known to be skewed[8] and are not normally distributed. Thus, there is a lack of knowledge in determining sample size to achieve accuracy in variance estimation statistics.

Monte Carlo simulation approach has recently been utilized for sample size determination to achieve accurate mean estimation. Monte Carlo is a numerical experiment that generates T-time sampling simulations each with n draws with or without replacement from a random sample with a prescribed probability distribution.[9] Each sample generates one sample mean estimator and one sample variance estimator. Given a sufficient large simulation time T, e.g. 1000 times, it is possible to examine the statistical robustness of using mean and variance of sample size n to approximate population mean and variance.

Muthén and Muthén is one of the early literatures to use Monte Carlo simulation in determining sample size.[10] Parameter estimate bias, standard error bias and coverage were reported using different sample sizes. It was found that non-normality and missing data are major factors of sample size. Shi and Lee utilized Monte Carlo simulations to calculate sample size needed for group randomized trials with unequal group sizes in cancer prevention and health promotion research.[11] They found that the widely used formula for sample size in group randomized trials is not applicable when group sizes vary, which is commonly observed in empirical research setup. Qumsiyeh utilized a bootstrap sampling technique in Monte Carlo simulations to find required sample sizes to achieve various confidence levels in health care related statistical experiments.[12] The required sample size was proven to be smaller than the one computed based on exact method as shown in equation (1). This has practical consequence because small sample size without sacrificing predicting power means less labor and cost in conducting research.

In on-road vehicle emission measurement studies, researchers install equipment on roadside and measure vehicle emissions for a certain period in one year or in multiple years to collect enough data for emissions analyses. Huang et al. compared vehicle emission measurement techniques under real-world driving conditions and concluded that on-road remote sensing is an effective and economic tool to monitor and control vehicle emissions.[4] However, they also pointed out major challenges in applying remote sensing technology, which include robustness of sampling process. In review of existing real-world vehicle emission studies, it shows that measurement sample sizes are either determined by researchers' experience or constrained by research budgets, but have not been derived systematically.[13-15] There is a knowledge gap in systematically determining necessary sample size to achieve statistical robustness.

INPUT DATA AND DATA HANDLING

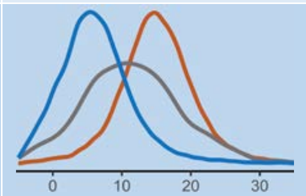
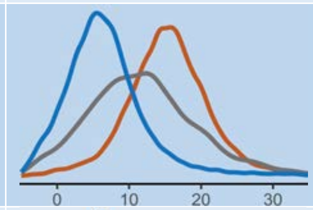
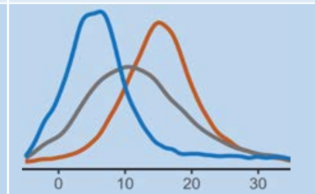
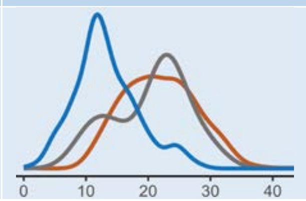
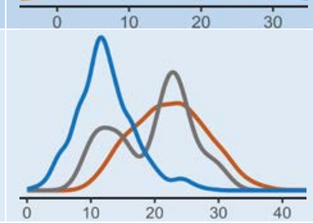
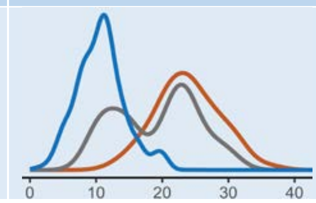
Three spectroscopic remote sensing (RS) instruments were used to conduct vehicle emission measurement in this study, including the FEAT instrument developed by the University of Denver and the Opus AccuScan RSD 4600 and RSD 5000. These instruments have been used and discussed extensively in previous studies.[1-2,16-18] RS instruments are placed at a roadside; the concentration of certain pollutants (CO_2, CO, HC, NO) in the plume of the vehicles passing is proportional to the attenuation of the light transmitted through the plume. The increment in the concentration relative to the background measured immediately before is then attributed to the vehicle. The incremental pollutant concentration is then divided by the incremental concentration of CO2, which in turn is proportional to the fuel burnt in the engine. This ratio presents the instantaneous fuel specific emission rate of the vehicle. Instantaneous speed and acceleration of passing vehicles are recorded as well. A possible instrument drift is corrected by regular calibration with an external reference gas. The information on the vehicle technology is retrieved via the recorded license plate from the vehicle registry. This provides information on the model year and emission certification standard, the make and model, the engine size and power, the fuel type and curb weight.

We use a collection of 130,000 RS records that were measured during eight year across Europe to serve as population. Table 1 summarizes the different measurement campaigns, testing conditions and passenger car fleet characteristics. The table groups the data by emission standards to facilitate comparison within and across country-specific measurements. In this paper, we focus on NO emissions of Euro 4, Euro 5 and Euro 6a,b diesel cars and demonstrate our approach for searching a minimum sample size. Euro 4 to Euro 6a,b cars account for 78 percent of all diesel car records. The choose of the NO emission is due to its importance for air quality, its low measurement error of ±15% (unlike NO_2) and because its emission rate is arguably the least variable among the emissions measured. Therefore, it presents an ideal case for the analysis and we consider our results as lower bounds for minimum sample sizes for other pollutants or emission concepts. The Monte Carlo simulation approach proposed here can however easily serve as a template for sample size determination of other pollutants, vehicle types and control stages.

Vehicle emissions increase with age or mileage, respectively. However, Chen and Borken-Kleefeld[15] did not find a relevant deterioration on NO and NOx emissions for Euro 4 diesel cars. More recently by Carslaw et al. [19] did likewise not find a relevant change in the NO emission rate with vehicle mileage for diesel Euro 5 and Euro 6a,b cars. Therefore, we need not discriminate records by vehicle age or mileage here.

Manuscript – Final and authoritative paper published in *Environmental Science & Technology*, 18 Oct 2019. https://pubs.acs.org/doi/abs/10.1021/acs.est.9b04123

Table 1. Summary of remote sensing testing conditions and passenger car fleet characteristics in UK (blue), Sweden (gray) and Switzerland (orange)

		Euro 4 Diesel Car	Euro 5 Diesel Car	Euro 6a,b Diesel Car
# of measurements	UK	23,825	32,071	12,136
	SE	617	5,106	3,426
	CH	17,257	29,725	7,072
Measurement Year and instrument	UK	FEAT: 2012, 2013, 2017, 2018; RSD 4600: 2013, 2015; RSD 5000: 2017, 2018		
	SE	RSD 5000: 2016		
	CH	RSD 4600: 2011-2015; RSD 5000: 2016, 2017		
Average age (years)	UK	7.4	3.8	2.0
	SE	8.3	5.0	1.7
	CH	8.3	4.3	2.5
Average NO emission rates +1 SD (g / kg fuel)	UK	11.0 (8.0)	12.6 (9.4)	6.0 (9.2)
	SE	10.9 (10.9)	11.4 (10.6)	5.9 (7.7)
	CH	11.0 (9.3)	12.6 (10.3)	6.0 (8.6)
VSP (kW/ton)				
Ambient Temperature (C)				

METHODOLOGY

Here, we aim to find the smallest sample size of emission records whose mean and standard deviation are reasonably close to the ‘true’ mean and standard deviation of the full population. The population is the collection of 130,000 RS records that were measured during eight year across Europe as shown in Table 1. We define terminologies as in Table 2.

Table 2. Terminology Definition

Terminology	Definition
Population	The set of all measured emission rates stratified by vehicle emission control technology, denoted as X

Population size	Total number of measurements in population, N
Population mean	Mean of all elements in the population, $\mu = \frac{\sum_{i=1}^{N} X_i}{N}$
Population Standard Deviation	Mean of all elements in the population, $\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - X)^\wedge 2}{N}}$
Sample mean	The mean of a n size sample from the population $\bar{x}_j{}^n = \frac{\sum_{i=1}^{n} x_i}{n}$
Sample Standard Deviation	The standard deviation of a n size sample from the population $s_j = \sqrt{\frac{\sum_{i=1}^{n}(x_i - x)^2}{n}}$
Error of sample mean	The absolute deviation of sample means from population mean, $v_j = \lvert \bar{x}_j{}^n - \mu \rvert$
Error of sample standard deviation (SD)	The absolute deviation of sample standard deviation from population standard deviation, $w_j = \lvert s_j - \sigma \rvert$
Tolerated error of estimation	The maximum error of sample means/standard deviation that counts as an accurate sample, V/W
Certainty ratio	Ratio of accuracy, $A_n = \frac{\sum_{j=1}^{T} I(v_j \leq V \,\&\, w_j \leq W)}{T}$, in T simulations with size n each time

We utilize a bootstrap-based Monte Carlo simulation approach that consists of T-fold random sampling simulations and evaluating the statistical performance of the sample mean to determine minimum sample size. T is a large number, i.e. 1000, as suggested in Monte Carlo simulation literature.[9,12] Specifically, in each iteration of the sampling simulation, we conduct a bootstrap experiment, i.e. drawing n samples with replacement from a stratified population X that has N emission records. With replacement means we replace an item once it is drawn from population. The purpose of bootstrap is to construct an empirical distribution based on observed data which can be used to asymptotically infer statistics of true stratified population. Literature have shown that bootstrap sampling can guarantee asymptotic feature of sample mean and variance distribution to that of population mean.[20]

The *T*-time simulations will generate a series of sample means $\bar{x}_1^n \ldots \bar{x}_j^n \ldots \bar{x}_T^n$.and sample variances, $s_1^2, \ldots s_j^2 \ldots s_T^2$.. The errors of sample mean and standard deviation for each simulation, v_j, w_j, defines closeness of j^{th} sample mean and standard deviation to the population mean and standard deviation (SD). We classify an accurate estimation as if $v_j \leq V$ and $w_j \leq W$ where V and W are our tolerated errors for mean and SD. The certainty ratio of T-time simulations, A_n, represents the consistency of sample mean and SD estimations when emission measurements are repeated. Higher A_n corresponds to greater confidence of observing accurate sample mean and SD estimations. Specifically, if *n* sample are drawn out of population dataset, we can be A_n confident that the sample mean and SD are within $\pm V$ and $\pm W$ of population mean and SD, respectively. Fixing V and W, we expect A_n to increase as sample size *n* enlarges. The advantage of Monte Carlo simulation is to explicitly explore the relationship between sample size, accuracy and confidence performance of the sample mean, here based on a large empirical dataset.

INFLUENCE OF POPULATION SIZE ON MINIMUM SAMPLE SIZE

First, we need to make sure that our populations are large enough not to constrain or bias the subsequent analysis. This is also referred to as finite population correction. It is recommended to use finite population correction factor to adjust variance/standard deviation estimate when sample size is greater than 5 percent of a population. Here, we aim to empirically find the population size that does not need to apply correction factor. The tolerated error of mean estimation is set to 1 g NO/kg fuel; this corresponds to about 50 mg NO per km, i.e. about 5% t o 10% of the average emission rate for Euro 5 diesel cars. The tolerated error of population standard deviation is set to be 0.5 g NO/kg fuel. In normal distribution, mean plus/minus 2 times standard deviation covers 95% of data in distribution. We borrow this idea and set the tolerated

Manuscript – Final and authoritative paper published in *Environmental Science & Technology*, 18 Oct 2019. https://pubs.acs.org/doi/abs/10.1021/acs.est.9b04123

error for standard deviation at 0.5 g NO/kg fuel so that two times of it equal to 1 g NO/kg fuel, which is comparable to tolerated error of mean estimation. As certainty rate of estimation, i.e. the confidence metric, we require 80%, which is based on our engineering knowledge. In the remainder of this paper, we vary the certainty rate to test sensitivity of our results to certainty rate.

In this analysis, we choose different size of sub-population by randomly draw from the actual population for Euro 4, 5 and 6a,b diesel car NO emission measurement. We use the Monte Carlo simulation approach to find minimum sample size for each size of sub-population. Figure 1 shows that the minimum sample sizes are not affected by statistical fluctuations if sub-population size is greater than 2500 for Euro 4/Euro 5 and 4500 for Euro 6a,b. Obviously, the exact numbers depend on the required tolerance and certainty, with more stringent requirements leading to bigger population size thresholds and larger minimum sample sizes. The important observation here is that we always have records larger than 2500 for Euro 4/Euro 5 or 4500 for Euro 6a,b, even in the stratified analysis below, so that our minimum sample size results are robust and will not be influenced by finite population.

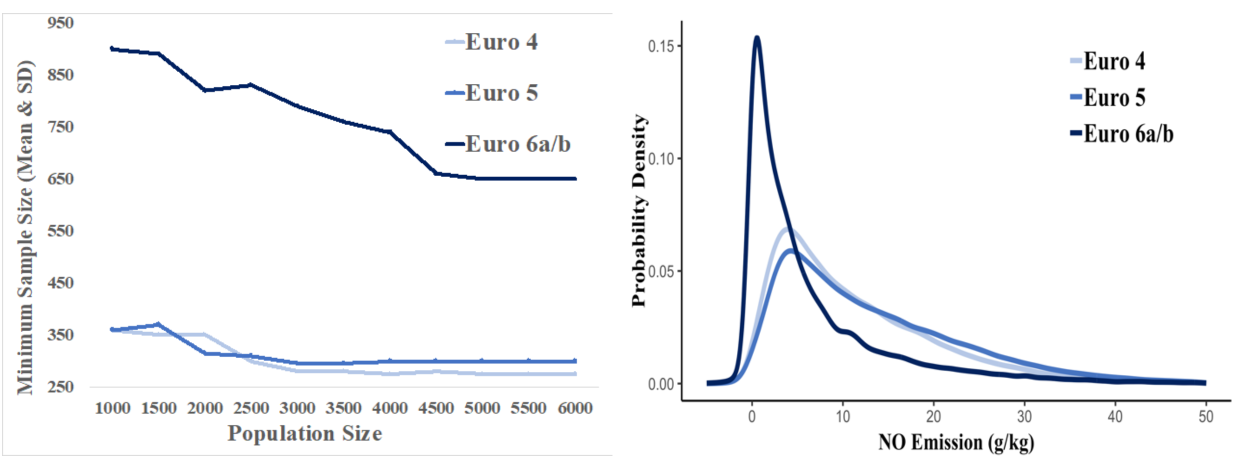


Figure 1. Minimum sample size (mean and SD) vs population size for NO emission estimation of Euro 4-6 diesel cars, all data from three countries. Default tolerated error of mean is 1 g NO / kg fuel, tolerated error of standard deviation at 0.5 g NO/kg fuel, certainty rate of estimation 80%.

SAMPLE SIZE AS FUNCTION OF TOLERATED ERROR AND ESTIMATION CERTAINTY RATE

The minimum sample size is clearly a function of the tolerated error for the mean and standard deviation V, W and the certainty ratio of estimation A_n. Figure 2a shows the minimum sample size for the NO emission rate of diesel cars first as a function of the certainty ratio. A larger sample size leads to higher confidence in using the sample mean and sample's standard deviation to estimate population mean and standard deviation, as expected. However, the increase in certainty diminishes as the sample size becomes bigger. To achieve 70%, 80% or even 90% certainty in both, mean and standard deviation estimate, a sample size of about 200, 300 or 500 records is needed for Euro 4 and Euro 5 diesel cars, and significantly large size of 440, 660 and 1010 records for Euro 6a,b cars, respectively. If only the mean is of interest then about half that number would be sufficient, with the notable exception of Euro 6a,b cars: The mean estimate requires only about 90, 140 and 230 records respectively for the chosen certainties. This is a consequence of the very different shape of distribution of the Euro 6a,b records compared to the earlier diesel generations. The stark difference in sample size between the mean and the standard deviation estimate is a result of the Monte Carlo simulation further discussed below. Next we explore the relationship between the tolerated error and the minimum sample size (Figure 2b). For illustration, we choose to vary the tolerated error for the mean estimation from 0.5 to 2.5 g NO / kg fuel and keep the certainty ratio fixed at 80%. As expected, the minimum sample size increases when the tolerated error is reduced, with the exact form established here from observations. We find that a tolerated error for mean of 1 g NO / kg fuel is actually a tipping point for the reduction rate for all three Euro classes investigated here: Below 1 g NO / kg fuel, a further decrease of the tolerated error results in a strong increase in the minimum sample size. For instance the minimum sample size increases from 50 to 100 to then about 600, when the tolerated error is reduced from 2 to 1 to finally 0.5 g NO / kg.

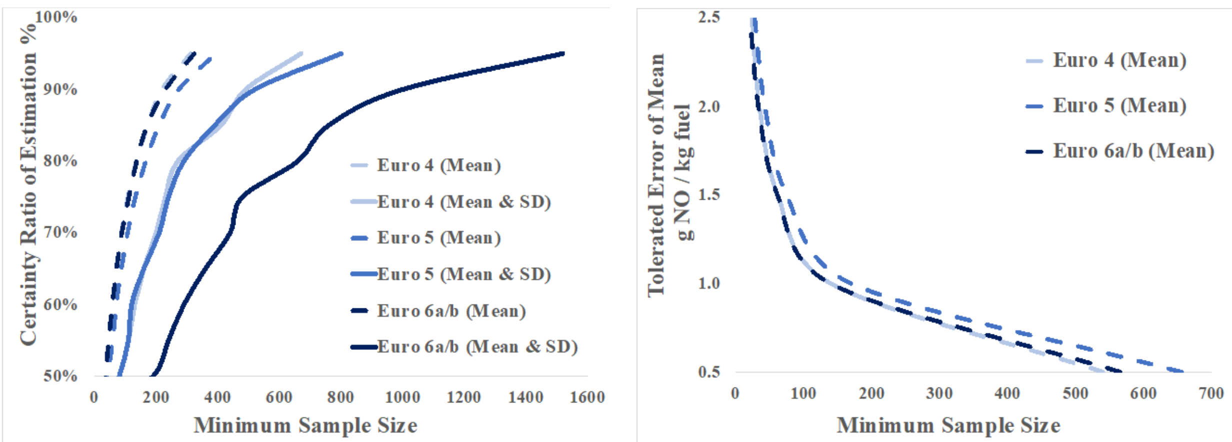


Figure 2. Minimum sample size vs certainty rate of estimation based on tolerated error of mean at 1 g NO/kg fuel and tolerated error of standard deviation at 0.5 g NO/kg fuel (left, 2a); minimum sample size vs tolerance error of mean based on 80% certainty ratio of mean estimation (right, 2b).

Table 3 compares the minimum sample sizes derived from the traditional closed form solution of equation (1) with the results from our Monte Carlo simulation. We confirm with empirical data the sample size results when the interest is only in the mean values of the population. However, we show at the same time that more than two times that number of records is needed to estimate the standard deviation of the distribution with the same accuracy. The distributions for Euro 4 and Euro 6a,b cars are quite different e.g. in terms of skewness, peak, possession of symmetry (Figure 1b). The Monte Carlo simulation approach utilizes both population variance and shape of distribution to determine sample sizes that can guarantee robustness estimation of both

Manuscript – Final and authoritative paper published in *Environmental Science & Technology*, 18 Oct 2019. https://pubs.acs.org/doi/abs/10.1021/acs.est.9b04123

population mean and variance. These sample sizes are consistently larger than sample sizes obtained based on closed form solution as shown in Equation (1).

Table 3. Minimum sample size as a function of required certainty in standard deviation estimation for diesel cars Euro 4, 5, 6a,b, numbers, in parenthesis are sample size calculated using Equation (1).

Certainty Ratio	70%		80%		90%	
	Mean	Mean & SD	Mean	Mean & SD	Mean	Mean & SD
Euro 4	87 (87)	200	135 (133)	270	218 (219)	490
Euro 5	107 (110)	210	166 (169)	300	273 (278)	520
Euro 6a,b	88 (85)	440	136 (131)	660	229 (215)	1010

INFLUENCE OF LOCATION ON MINIMUM SAMPLE SIZE

Previous studies of remote sensing have demonstrated heterogeneity of emissions behavior at different measurement locations.[21] As shown in Table 1, our data were collected from three countries and contain heterogeneous vehicle specific power and ambient temperature distributions. Thus, we differentiate the data by country and explore the relationship between minimum sample size and certainty ratio of estimation for each location specifically (Figure 3a). We observe that to achieve the same level of confidence in estimation, i.e. certainty ratio, Swiss data require the smallest sample size, followed by UK and Sweden. For example, to achieve 80% certainty ratio in mean and standard deviation estimation for Euro 5 cars, a sample of around 300 records are required for any location (exactly 277, 302, 326 in Switzerland, UK and Sweden respectively). This is a remarkably consistent result despite different fleets, different instruments, driving conditions and ambient temperatures in the different locations. This means that records from different sites, i.e. from different instruments, vehicle fleets and driving conditions can be collectively analysed together, at least for NO emissions from diesel cars Euro 4, Euro 5 and Euro 6a,b.

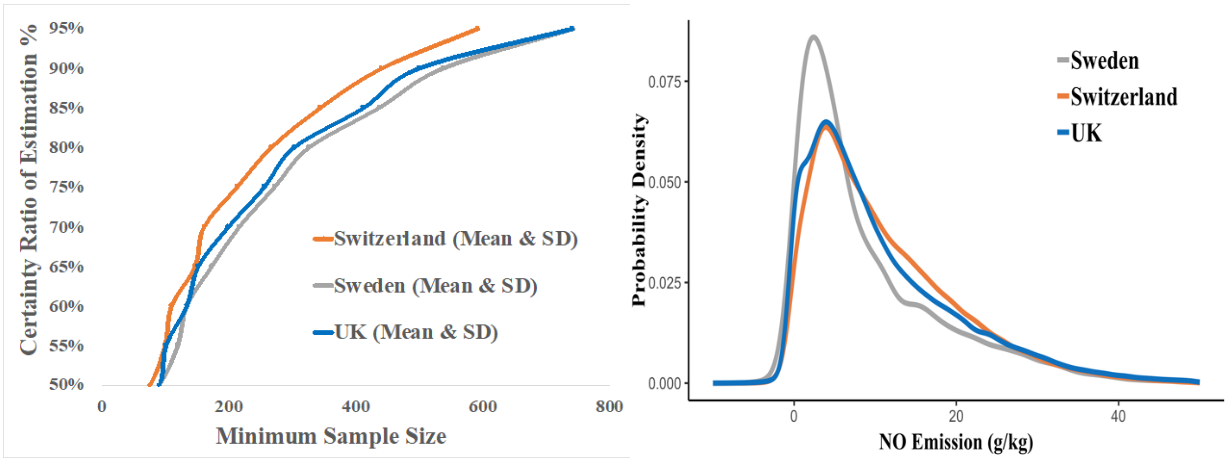


Figure 3. Minimum sample size (mean and SD) vs certainty ratio of estimation based on Euro 5 diesel cars location-specific NO emission data, default tolerated error is 1 g NO/kg fuel and standard deviation tolerated error is 0.5 g NO/kg fuel.

INFLUENCE OF POWER HOMOGENIZATION ON MINIMUM SAMPLE SIZE

Vehicle specific power (VSP) is a metric for estimating engine power demand of a vehicle and has been extensively used in emission models and remote sensing analysis.[17,22-23] Particularly, Carslaw et al. found a clear increase of the NO_x emission rate with increasing VSP.[13] In the NEDC type-approval driving cycle in Europe VSP ranges from 3 to 22 kW/ton, which is consist of normal urban driving and extra-urban driving cycles. To assess impacts of vehicle power on sample size determination, in Figure 4a, we present minimum sample size under various VSP bins and various accuracy performance metrics, i.e. 70%, 80% and 90% certainty ratio of estimation, A_n. The tolerated error of mean estimation V is fixed at 1 g NO / kg fuel and the tolerated error of standard deviation W is fixed at 0.5 g NO / kg fuel. We restrict the data to Euro 5 diesel cars measured in United Kingdom, i.e. the most abundant sample, to identify the influence of engine load as clearly as possible. Given the certainty ratio of 80%, the minimum sample size is relatively stable at 300 records up to a VSP of 18 kW/ton. When the certainty is reduced from 80% to 70%, meaning that roughly one third of records is allowed to be outside the tolerance margin, only 100 records are required. Vice versa, to increase the certainty to 90%, meaning that only 10% of the sample is allowed outside the tolerance error, then at least 450 records are needed to approximate the population mean.

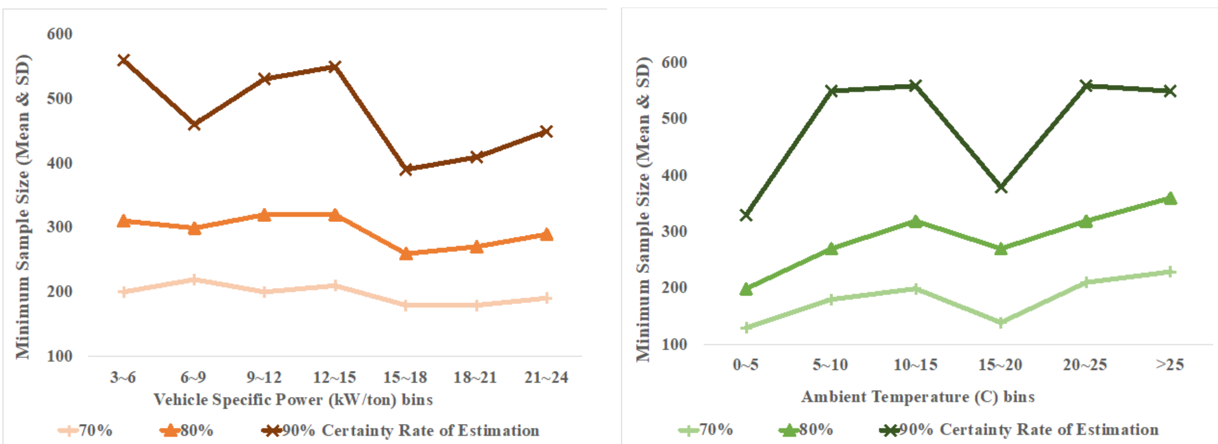


Figure 4. Minimum sample size (mean and SD) based on vehicle specific power bins (left, 4a), temperature bins (middle, 4b) with different certainty ratio of estimation, default tolerated error of mean 1 g NO/kg fuel and tolerated error of standard deviation 0.5 g NO/kg fuel, UK Euro 5 diesel car NO emission.

INFLUENCE OF TEMPERATURE ON MINIMUM SAMPLE SIZE

It has been shown before, that the NO emission rate increases significantly when the ambient temperature decreases from 20°C to 5°C.[24,26] While this affects the mean rate, it does actually not affect the sample size to achieve accurate mean and/or SD estimations, as we find empirically. The minimum sample size required is stable at about 300 records across temperatures from 5°C to 25°C (Figure 4b). The same stable behavior is observed when a higher or lower certainty is requested: Then the necessary sample size is either about 500 or 200 records to approximate the population mean. These stable relations are good news for experimentalists and analysts alike: The former can be assured that they do not need to spend time on finely controlling driving conditions and ambient temperatures; the latter can justify combining data from different external conditions in their analysis.

OUTLOOK: REQUIRED SAMPLE SIZE WHEN THE EMISSION RATE IS MUCH LOWER

The RS measurement data contain emission records from diesel cars certified up to Euro 6a,b, which are vehicles manufactured before September 2018. Their NO emission rate is approximately 400-500 mg NO per km, and thus much higher than for gasoline cars that emit less than 2 g NO per kg fuel or less than 80 mg NO per km on the road. Diesel cars certified to the current Euro 6d-temp emission standard are also measured in this range. Therefore, the question is whether many more RS records are needed for a reliable sampling at much lower average emission rates? For lack of data we cannot answer this from the existing set of RS records.

As a proxy we use modal PEMS data (courtesy S. Hausberger, TU Graz) from six Euro 6d-temp diesel cars all having an emission rate of no more than 40 mg NOx per km over 20 RDE compliant trips. This constitutes some 116,000 second-by-second emission records. We convert them into a format compatible with RSD as follows: To alleviate possible issues with time alignment notably between the NO and the CO2 sensor we take the running average over consecutive three second intervals. Next, we calculate the ratio of the three second NO to CO2 and fuel consumption respectively. Finally, we filter out all records for negative acceleration or VSP above 22 kW per ton. This way we generate a set of 39,000 instantaneous emission ratios comparable to RS measurement conditions. The average NO emission rate of these Euro 6d-temp cars is 0.8 g / kg fuel, about a factor eight lower than the earlier Euro 6a,b cars. We perform the sample size analysis on this set.

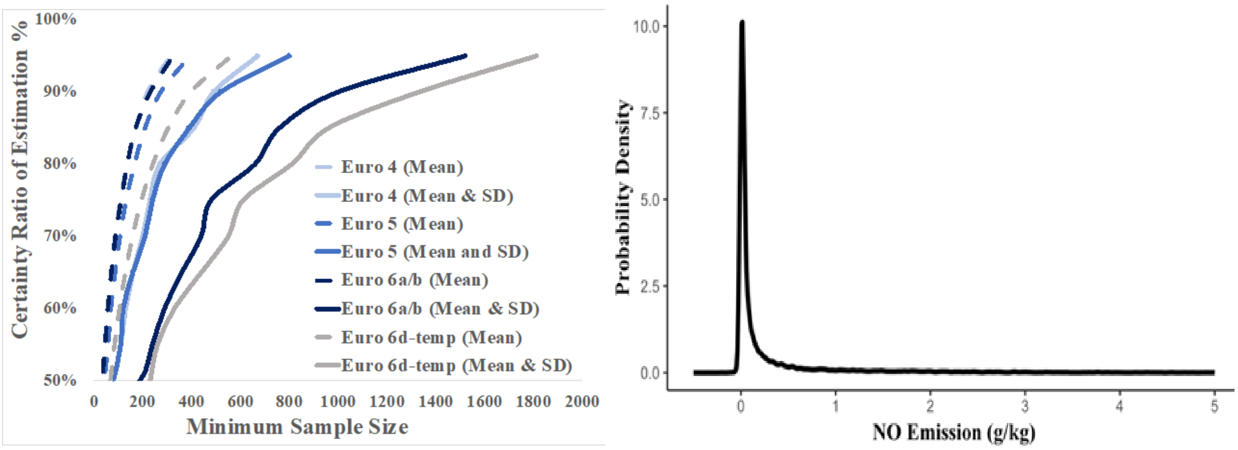


Figure 5. Minimum sample size vs average NO emission of Euro 4-6 diesel cars in RS data and Euro 6d in PEMS data (left, 5a), default tolerated error of mean at 1 g NO / kg fuel, tolerated error of standard deviation at 0.5 g NO/kg fuel, certainty rate of estimation 80%; Probability density function of NO emission rates of Euro 6d cars based PEMS experiment (right, 5b).

Figure 5a presents minimum sample size versus certainty ratio for Euro 4, 5 and 6a,b using remote sensing data and Euro 6d-temp using PEMS data. As before the tolerated error of mean estimation V is fixed at 1 g NO / kg fuel and the tolerated error of standard deviation W is fixed at 0.5 g NO / kg fuel. The certainty rate of estimation is set at 80%. As suspected, hat the minimum sample size for Euro 6d is larger than those of Euro 4, 5 and 6a,b. For example, at 80% certainty ratio, the minimum sample size for Euro 6d-temp diesel cars is 810 records which is larger than that of Euro 6a,b (660), and about 2.5 times of those of Euro 4 and Euro 5. However, that number is within the range found before, meaning that also for vehicles with very clean exhaust emissions our results indicate the range. One could assume that this then also holds true for gasoline cars as well.

As we see that increase in the required sample size is determined by the shape and in particular the variance of the underlying emission distribution: A small portion of higher emission records in Euro 6d-temp diesel cars leads to a high variance in the emission measurements and thus higher minimum sample size for estimating the average emission level. Figure 5b presents the probability density function of NO emission from Euro 6d, which shows majority of NO emission of Euro 6d cars are small and the average emission of Euro 6d-temp is well controlled.

DISCUSSION

In summary, we propose a bootstrap-based Monte Carlo simulation approach to determine the minimum sample size in remote sensing measurements of vehicle emissions. The sample size is given explicitly here as a function of required accuracy and robustness for both the population mean and its standard deviation. The minimum number depends on vehicle technology, fuel type and pollutant. Here we explore the empirical relationship for the NO emission rate of European diesel cars certified to Euro 4, 5 or 6 emission standards. We believe this pollutant presents a good opportunity to develop the method that is suitable for other vehicle concepts and pollutants. Because of their bigger variance we expect that the minimum sample will be higher for the other pollutants. Our results are important for planning measurement campaigns, for appropriately budgeting resources and for assessing the robustness of the records obtained. This is illustrated by a simplified example in Table 4: Suppose, RS measurements are conducted at a road with an average 2000 passenger cars passing during daytime, which is typical of many sites used so far in Europe. Assume for simplicity an even share of diesel and gasoline cars, meaning that there are about 1000 diesel cars passing, of which typically 90% have valid records. This fleet might be composed of around 40% Euro 5 cars (mandatory between 2009 and 2014), 40% newer Euro 6a,b and 10% Euro 6d cars, the rest being older. Then between 90 and 360 diesel cars of the respective certification standards could be measured in a single day (and about the same distribution for gasoline cars). Within half a day the mean values of Euro 5 and Euro 6a,b cars could be determined with more than 80% certainty; in less than two measurement days also their variance could be measured representatively. The same campaign would lend similar data for the other technology layers, and the more records come in the more accurate the sample mean and variance become for those and any other vehicle category and technology layer. These averages per technology layer are crucial input e.g. to traffic emission and air quality models.

If the objective of the RS campaign is market surveillance e.g. of individual engine families, then more measurement time is needed, depending on the frequency of occurrence of the respective engines. Assume the ten top selling engine families have at least 2% share in the Euro 6a,b cars. To determine the mean NO emission rates for these top ten engine families (i.e. needing at least 136 records each) about 19 measurement days would be needed, so roughly two days per month. Most days would be needed for Euro 6d engine families because they are (so far) less abundant and need most records for an accurate determination: For a Euro 6 engine family with only 1% share more than 200 measurement days would be needed to determine its mean emission rate. This would represent nearly continuous measurements or calls for a change in the measurement strategy: Either more RS units could be deployed to multiply the data capture, or they should be deployed to road with higher traffic volume, or to sites where a higher occurrence of the target engine families is known. The numbers can be easily adopted to a different local situation and different vehicle categories. Whatever the target, the campaign will always capture very useful data for the whole fleet and all vehicle categories and technologies occurring at once. How robust and accurate the values are is essentially 'only' a mater of the statistical sample. This can be boosted by cooperation between different RS campaigns, as illustrated by the CONOx project[24] that provided the initial sample for this analysis.

A campaign of 20 days would yield about 36,000 valid car records, which would very accurately provide mean NO emission rates of the top ten engine families for all light duty vehicles but the latest technology layer (Euro 6d). At indicative costs of 0.5 to 2 € per record this translates to about 18,000 to 72,000 Euros for the whole campaign for top engine families for light duty vehicles, diesel and gasoline alike. This is significantly cheaper than a series of PEMS measurement of dozens of individual vehicles and illustrates the important cost saving potential when RS is used for market surveillance and pre-screening before detailed emission testing. Flexible, small scale measurement campaigns allow capture of different driving conditions and fleets, longer term, stationary campaigns allow more detailed analysis down to individual vehicles when they are measured repeatedly.

Table 4. Example for planning the duration of a RS campaign for either fleet or family emission rate, mean only (with ±1 g NO/kg accuracy) or including standard deviation (as ±0.5 g NO/kg).

	Assumptions on traffic		Required sample size @80% certainty		Measurement days	
		Records per day [6-18hrs]	…for mean only	…for mean & SD	…for mean only	…for mean & SD
Volume of passenger cars	2000					
Valid records	90%	1800				
Share diesel cars	50%	900				
Share: Euro 4 and older	10%	90	135	270	1.5	3
Share Euro 5	40%	360	166	300	0.5	<1
Share Euro 6a,b	40%	360	136	660	0.4	<2
Share Euro 6d	10%	90	200	810	>2	9
a Euro 5 engine family	2%	7.2	136	660	19	92
a Euro 6a,b engine family	1%	<1	200	810	222	900

AUTHOR INFORMATION

Corresponding Author: Yuche Chen
Phone: +1-803 777 9105; Fax: +1-8037770670; E-mail: chenyuc@cec.sc.edu

Jens Borken-Kleefeld
Phone: +43-2236 807 570; Fax: +43-2236 807 530; E-mail: borken@iiasa.ac.at

ACKNOWLEDGEMENTS

The authors would like to thank the CONOx group for providing remote sensing data[24] and Prof. S. Hausberger of Graz University of Technology for providing detailed PEMS data.

REFERENCES

(1) Carslaw, D.C.; Rhys-Tyler, G. New Insights from Comprehensive On-Road Measurements of NOx, NO2 and NH3 from Vehicle Emission Remote Sensing in London, UK. Atmospheric Environment, 2013, 81 (December): 339–47. https://doi.org/10.1016/j.atmosenv.2013.09.026.

(2) Chen, Y.;Borken-Kleefeld, J. Real-driving emissions from cars and light commercial vehicles–Results from 13 years remote sensing at Zurich/CH. Atmospheric Environment, 2014, 88, pp.157-164. https://doi.org/10.1016/j.atmosenv.2014.01.040

(3) Bishop, G.A.; Haugen, M.J. The story of ever diminishing vehicle tailpipe emissions as observed in the Chicago, Illinois area. Environmental science & technology, 2018, 52(13), pp.7587-7593. doi: 10.1021/acs.est.8b00926.

(4) Huang, Y.; Organ, B.; Zhou, J.L.; Surawski, N.C.; Hong, G.; Chan, E.F.; Yam, Y.S. Remote sensing of on-road vehicle emissions: Mechanism, applications and a case study from Hong Kong. Atmospheric Environment, 2018, 182, pp.58-74. Doi: 10.1016/j.atmosenv.2018.03.035

(5) Guo, H., Zhang, Q., Shi, Y. and Wang, D. On-road remote sensing measurements and fuel-based motor vehicle emission inventory in Hangzhou, China. Atmospheric Environment, 2007, 41(14), pp.3095-3107. https://doi.org/10.1016/j.atmosenv.2006.11.045

(6) Borken-Kleefeld, J., and Dallmann, T.R. “Remote Sensing of Motor Vehicle Exhaust Emissions.” White Paper. Washington DC, USA: The International Council on Clean Transportation (ICCT), 2018.

(7) Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W. Applied Linear Statistical Models, 5th ed. McGraw-Hill, New York, 2005.

(8) Zhang, Y., Bishop, G.A., Stedman, D.H. Automobile Emissions Are Statistically Gamma Distributed. Environmental Science & Technology, 1994, 28 (7): 1370–1374. https://doi.org/10.1021/es00056a029

(9) Beasley, W.H. and Rodgers, J.L. Bootstrapping and Monte Carlo methods. APA handbook of research methods in psychology, 2012, 2, pp.407-425.

(10) Muthén, L.K. and Muthén, B.O. How to use a Monte Carlo study to decide on sample size and determine power. Structural equation modeling, 2002, 9(4), pp.599-620. https://doi.org/10.1207/S15328007SEM0904_8

(11) Shi, Y. and Lee, J.H. Sample size calculations for group randomized trials with unequal group sizes through Monte Carlo simulations. Statistical Methods in Medical Research, 2018, 27(9): 2569-2580, doi: 10.1177/0962280216682775.

(12) Qumsiyeh, M., 2013. Using the bootstrap for estimating the sample size in statistical experiments. Journal of Modern Applied Statistical Methods, 12(1), p.9.

(13) Carslaw, D.C., Williams, M.L., Tate, J.E. and Beevers, S.D. The importance of high vehicle power for passenger car emissions. Atmospheric environment, 2013, 68, pp.8-16. https://doi.org/10.1016/j.atmosenv.2012.11.033

(14) Pujadas, M., Domínguez-Sáez, A., De la Fuente, J. Real-driving emissions of circulating Spanish car fleet in 2015 using RSD Technology. Science of The Total Environment, 2017, 576, pp.193-209. https://doi.org/10.1016/j.scitotenv.2016.10.049

(15) Chen, Y., Borken-Kleefeld, J. NOx emissions from diesel passenger cars worsen with age. Environmental science & technology, 2016, 50(7), pp.3327-3332. https://doi.org/10.1021/acs.est.5b04704

(16) Borken-Kleefeld, J. and Chen, Y. New emission deterioration rates for gasoline cars–Results from long-term measurements. Atmospheric Environment, 2015, 101, pp.58-64. https://doi.org/10.1016/j.atmosenv.2014.11.013

(17) Bishop, G.A., Stedman, D.H., Ashbaugh, L., 1996. Motor Vehicle Emissions Variability. Journal of the Air & Waste Management Association 46 (7): 667–75. https://doi.org/10.1080/10473289.1996.10467501.

(18) Rushton, C. E., Tate, J. E., Shepherd, S. P., Carslaw, D. C. Inter-instrument 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 Management, 2018, 68:2, 111-122. https://doi.org/10.1080/10962247.2017.1296504

(19) Carslaw, D., Farren, N., Borken-Kleefeld, J., Sjödin, Å. "Study on the Durability of European Passenger Car Emission Control Systems Utilizing Remote Sensing Data." C 387. Gothenburg, Sweden: IVL Swedish Environmental Research Institute, 2019.

(20) Kuhn, M., Johnson, K. Applied predictive modeling (Vol. 26). New York: Springer, 2013.

(21) Dallmann, T., Bernard, Y., Tietge, U., Muncrief, R. Remote sensing of motor vehicle emissions in London. The International Council on Clean Transportation, white paper, 2018. https://www.theicct.org/publications/true-london-dec2018

(22) Jiménez, J.L. "Vehicle Specific Power: A Useful Parameter for Remote Sensing and Emission Studies." presented at the 9th CRC On-Road Vehicle Emissions Workshop, San Diego, April 21, 1999.

(23) US-EPA OTAQ. "Development of Emission Rates for Light-Duty Vehicles in the Motor Vehicle Emissions Simulator (MOVES2010)." Final Report EPA-420-R-11-011. Washington, DC, USA.: US Environmental Protection Agency., 2011. http://www.epa.gov/otaq/models/moves/documents/420r11011.pdf.

(24) Sjödin, Å, Borken-Kleefeld, J., Carslaw, D., Tate, J., G.-M. Alt, J. Dela Fuente, Y., Bernard, P McClintock, R Gentala, N Vescio, and S Hausberger. "Real-Driving Emissions from Diesel Passenger Cars Measured by Remote Sensing and as Compared with PEMS and Chassis Dynamometer Measurements - CONOX Task 2 Report." C 294. Gothenburg, Sweden: IVL Swedish Environmental Research Institute, 2018.

(25) Borken-Kleefeld, J., Bernard, Y., Carslaw, D., Sjödin, Å. "Contribution of Vehicle Remote Sensing to In-Service/Real Driving Emissions Monitoring - CONOX Task 3 Report." Gothenburg, Sweden: IVL Swedish Environmental Research Institute, 2018.

(26) Grange, S.K., Farren, N.J., Vaughan, A.R., Rose, R.A. and Carslaw, D., 2019. Strong temperature dependence for light-duty diesel vehicle NOx emissions. Environmental science & technology.

(27) Bishop, G.A. Stedman, D.H., 2008. A decade of on-road emissions measurements. Environmental Science & Technology, 42(5), pp.1651-1656.

(28) Claudio Mazzoleni, Hampden D. Kuhns, Hans Moosmüller, Robert E. Keislar, Peter W. Barber, Norman F. Robinson, and John G. Watson, Djordje Nikolic. On-Road Vehicle Particulate Matter and Gaseous Emission Distributions in Las Vegas, Nevada, Compared with Other Areas. J. Air & Waste Manage. Assoc. 54:711–726, 2004.