

# **Statistical Analysis of Heavy-Duty Vehicle Fleet Sizes for Project E-55**

Warren H. White  
Department of Chemistry  
Washington University  
St. Louis, MO 63130-4899

Richard F. Gunst  
Department of Statistical Science  
Southern Methodist University  
Dallas, TX 75275-0332

Report to  
National Renewable Energy Laboratory  
May 1, 2000

List of Abbreviations and Acronyms

CARB:	California Air Resources Board
CBD:	Central Business District (test cycle)
CO:	carbon monoxide
CO <sub>2</sub> :	carbon dioxide
CRC:	Coordinating Research Council
DOE:	Department of Energy
E-55:	CRC project number
EMFAC7G:	CARB mobile source emissions computer model
g/mile:	grams per mile
HC:	hydrocarbons
HDD:	heavy-duty diesel
hp:	horsepower
ID:	identification code
NO <sub>x</sub> :	nitrogen oxides, NO and NO <sub>2</sub>
NREL:	National Renewable Energy Laboratory
PM:	particulate matter
VMT:	vehicle miles traveled
WVU:	West Virginia University

## Abstract

A planning study has been conducted in advance of a substantial field campaign of chassis dynamometer measurements of heavy duty vehicles ("E-55"). The projected measurements are intended to characterize particulate matter and nitrogen oxides emissions in urban areas by in-use, heavy-duty diesel vehicles, in a manner that can be used to assess the accuracy of the current emission factor estimates from mobile source emission models. A key consideration in the design of the dynamometer campaign is the quantity of testing required to overcome the random variability of emissions, between different vehicles in the same weight class and between different tests of the same vehicle. The objective of this planning study has been to provide guidance on the number of vehicles and retests needed to provide any given level of confidence in the results.

Our analysis is based on data compiled by West Virginia University (WVU) from 775 individual chassis dynamometer runs they had carried out on 63 different heavy-duty diesel vehicles between April 1992 and June 1999. We first identified a relatively homogeneous subset of 247 runs on 18 Class 8 trucks that had been performed using a common set of test procedures. We then used analysis of variance techniques to apportion variability within this subset into contributions from: (1) differences in the mean emissions of different vehicles, (2) differences in emissions from the same vehicle tested at different times, and (3) differences recorded in various runs with the same vehicle at the same time, or "measurement error." Nitrogen oxides emissions from individual vehicles were relatively stable over time, most of the observed variability being accounted for by differences between vehicles. Particulate matter emission rates were much less stable, with much of the observed variation attributable to changes over time in the emission rates of individual trucks. Measurement error proved a negligible contributor to observed variability for all pollutants.

The results of our statistical modeling yield estimates for the expected levels of statistical uncertainty in test fleet emissions, as a function of the number of vehicles tested and the number of tests on each vehicle. There are substantial decreases in the uncertainty limits as fleet sizes increase from 10 to 120 vehicles, but the number of repeat tests per vehicle has little effect on uncertainty. Thirty vehicles are sufficient to determine nitrogen oxides emissions to within -15% and +18%, carbon dioxide to within -4% and +5%, and particulate matter to within -21% and +27%.

## Introduction

In late January, 2000, the National Renewable Energy Laboratory (NREL) initiated an expedited program (Phase 1) of statistical analysis to support planning for the U.S. Department of Energy's Department of Heavy Vehicle Technologies, California Air Resources Board (CARB), and Coordinating Research Council (CRC) project E-55. The objective of E-55 is to characterize current in-use, heavy-duty diesel (HDD) emissions in urban areas, in a manner that can be used to assess the accuracy of the current particulate matter (PM) and nitrogen oxides (NO<sub>x</sub>) emission factor estimates from mobile source emission models. It is anticipated that E-55 will include a substantial field campaign in southern California of resource-intensive chassis dynamometer measurements of in-use heavy-duty diesel vehicles.

It is known that HDD emissions vary significantly, between different vehicles in the same weight class and between different tests of the same vehicle. This variability is an important determinant of the number of vehicle tests, and hence the scale of resources, that will be required to attain the project objective. The objective of Phase 1 is to provide a rational basis for estimating the number of chassis dynamometer tests needed in E-55.

The present report outlines the analyses that were done in Phase 1, and summarizes the expected dependence of statistical uncertainty on the size of the test fleet. Results are based on an examination of statistical variability in a unique emissions data base compiled by West Virginia University (WVU) from their past chassis dynamometer tests. More comprehensive results and details of the statistical analysis are documented in three attached appendices. Phase 1 did not address other important aspects of E-55 design, such as choice of test cycle and vehicle recruitment strategy, which will be examined by the contractor chosen to carry out the actual measurements.

## Data

In February, 2000, WVU provided us with an electronic data file summarizing the results of 775 individual chassis dynamometer runs they had carried out on 63 different heavy-duty diesel vehicles between April 1992 and June 1999. Each run yielded emission rates for NO<sub>x</sub>, CO, HC, CO<sub>2</sub>, and PM, which were expressed consistently as g/mile. The runs involved an assortment of vehicle types, fuels, and driving cycles; factor combinations were not systematically varied throughout the data set as a whole. Table 1 lists the information included in the file.

<u>Parameters and Variables</u>
Test ID
Vehicle Reference No.
Test Run ID
Primary Fuel ID
Vehicle Type
Vehicle Manufacturer
Manual/Auto Transmission
Vehicle Model Year
Odometer Reading
Gross Vehicle Weight
Engine Manufacturer
Engine Model
Engine Power
Testing Weight
Cycle Name
Test Date
NO <sub>x</sub> , g/mile
CO, g/mile
HC, g/mile
PM, g/mile
CO <sub>2</sub> , g/mile

Table 1. Information in the WVU data file, BOOK4NEW.XLS.

All but six of the trucks tested by WVU had gross vehicle weights (GVW) of at least 48,000 lb, placing them in weight class 8, and these “heavy-heavy-duty” trucks account for over 90% of the individual dynamometer runs. Class 8 diesels represent only 18% of the heavy-duty (Classes 2-8) gas and diesel fleet in southern California, but are estimated to account for 35% of this fleet’s vehicle miles traveled (VMT). Recent modeling with CARB’s EMFAC7G indicates that Class 8 diesels are responsible for 52% of all NO<sub>x</sub> emissions by heavy-duty vehicles, and 61% of all PM emissions by heavy-duty vehicles. The WVU data thus characterize the trucks responsible for a majority of the NO<sub>x</sub> and PM emissions by heavy-duty vehicles in southern California.

The WVU test data are in several respects well suited to the study of variation in vehicle emissions. Most testing involved repeated measurements during a single day’s experimental set-up, using the same vehicle, test cycle, operator, and fuel. These replicate determinations allow us to characterize the measurement precision associated with the test procedure itself. Moreover, several of the individual vehicles were revisited in subsequent years, the same vehicle being tested at different points in its life. These retests allow us to determine the variability associated with such unobserved factors as vehicle deterioration and maintenance.

As with any real set of measurements, the WVU data impose certain constraints on our analysis. Although the total number of vehicles tested represents a heroic experimental effort, it

remains small from some important perspectives. The data are not suited to characterizing “high-emitters” and other potentially important sub-populations such as poorly maintained vehicles, because they were obtained from a reasonably well maintained and recruitable fleet. Moreover, vehicles of different model years and technologies were often tested on different cycles or with different fuels. Emissions differences associated with vehicle characteristics, which will vary in the population sampled by E-55, are thus confounded with those attributable to shifts in testing procedures, which will be standardized in E-55. We must restrict our attention to subsets of relatively homogeneous test conditions if we are to minimize contamination by avoidable and thus irrelevant variability, and this further reduces the number of vehicles available for analysis.

### Analysis

The WVU data include results for a wide variety of vehicles and test conditions, from transit buses simulating downtown stop-and-go operation to long-haul tractors simulating suburban deliveries. Our first task was to select a relatively homogeneous subset in which differences between comparable vehicles’ emissions would not be unduly confounded with uncontrolled variations in vehicle categories and testing conditions.

Three engine power ratings are most common in the WVU data: 275 hp, 300 hp, and 350 hp. Vehicles with these engines were most often tested on one of three cycles: the 5-mile route, the 5-peak cycle, or the central business district (CBD) cycle. Almost half of the WVU data (352 runs on 28 vehicles) come from tests of these engine ratings on these operating cycles. All vehicles in this subset are tractor trucks or refuse trucks, and there was little overlap in testing between the two types: only one tractor was tested on the CBD cycle, and only one refuse truck was tested on either of the other two. We restricted our final consideration to the 247 runs from 18 trucks (17 tractors, 1 municipal refuse) tested on the 5-mile route and/or 5-peak cycle. These two cycles were chosen because seven of the 18 trucks were tested on both, more than any other pair of cycles in the data file. Although they do not necessarily represent actual truck behavior, these cycles provide a reproducible basis on which to compare emissions levels from truck to truck.

The 18 trucks in our analytic subset represent a limited sample of the real-world vehicle population, and one that was not selected at random from that population. Any inferences about emissions variations in the real-world vehicle population require a model that could extrapolate from the observed variations in our sample.

The replicate tests that WVU carried out on individual vehicle/cycle combinations allow us to distinguish and quantify the particular component of variance that arises from irreducible variability in the test procedure itself, which we may think of as measurement error. The observable behavior of measurement error across different vehicles and cycles offers us empirical guidance on the proper form of a statistical model for vehicle emissions.

The simplest modeling assumption is that the expected variability of emissions within a collection of vehicles, in terms of g/mile standard deviation, is independent of our choice of vehicles and test cycle. Since the mean rate of emissions from our collection does depend on the selection vehicles and test cycles, this assumption implies that emissions variability is uniform across the range of mean emissions. Measurement error is a component of emissions variability. We found that the observed measurement error increases with measured average emissions in the WVU data, indicating that variability is not uniform. An alternative assumption commonly applied to vehicle emissions is that variability is proportional to average emissions, in which case the variability of logarithmically transformed emissions is uniform. Exploratory plots of the WVU data show satisfactory fidelity to this assumption for the measurement error component, and our statistical analysis is accordingly performed on the logarithms of the emissions rates, conforming to the constant relative variability model.

Individual runs occasionally yielded anomalous emissions, well beyond the expected range of variation for the given vehicle and testing conditions. To avoid undue influence by a few unrepresentative results, a total of 16 emissions measurements were excluded from the statistical modeling because they yielded Studentized residuals greater than 3 for one of the pollutants. (The chance occurrence of a Studentized residual greater than 3 is less than 1% for a normal population of emissions. On the other hand, Studentized residuals can easily exceed 3 if a pollutant is moderately or extremely deviant from a population of normal emissions.) Our forecasts of variability accordingly presuppose a similar trimming of the E-55 results.

We used analysis of variance techniques to apportion the variability in log-emissions into contributions from: (1) differences in the mean emissions of different vehicles (the “vehicle” component), (2) differences in emissions from the same vehicle tested at different times (the “odometer” component), and (3) differences recorded in various runs with the same vehicle at the same time (the “error” component). The three sources of variability are assumed to be statistically independent, so that their contributions to the total variance of log-emissions is additive:  $\sigma_{\text{total}}^2 = \sigma_{\text{vehicle}}^2 + \sigma_{\text{odometer}}^2 + \sigma_{\text{error}}^2$ . The estimated log-scale standard deviations  $\sigma_{\text{vehicle}}$ ,  $\sigma_{\text{odometer}}$ , and  $\sigma_{\text{error}}$  are summarized in Table 2.

Pollutant	Log-scale standard deviations			
	Vehicle	Odometer	Error	Total
NO <sub>x</sub>	0.44	0.13	0.05	0.47
CO	0.59	0.37	0.12	0.70
HC	0.62	0.37	0.07	0.73
CO <sub>2</sub>	0.08	0.09	0.04	0.13
PM	0.47	0.46	0.12	0.66

Table 2. Method of moments estimates of variation in individual chassis dynamometer runs. Values are based on analysis of existing WVU data from multiple tests of 18 heavy-heavy-duty diesel vehicles on the 5-mile route and/or 5-peak cycle.

Table 2 shows emissions of  $\text{NO}_x$  from a given vehicle to be relatively stable over time (*i.e.*,  $\sigma_{\text{odometer}}$  is small), leaving most of the variability in our 247 runs (less the 3 extreme  $\text{NO}_x$  emissions measurements) accounted for by differences between vehicles. PM emission rates are much less stable, and much of the observed variation in PM is attributable to changes over time in the emission rates of individual trucks. Measurement error is a negligible contributor to observed variability for all pollutants.

## Results

Given the parameter estimates in Table 2 for the variation in individual dynamometer runs, we can straightforwardly estimate the sampling variance in means derived from  $r$  replicate runs on each of  $n$  vehicles:  $\sigma_{\text{mean}}^2 = \sigma_{\text{vehicle}}^2/n + \sigma_{\text{odometer}}^2/n + \sigma_{\text{error}}^2/nr$ . The standard error  $\sigma_{\text{mean}}$  characterizes the statistical uncertainty of the tested fleet's mean as an estimator for the larger population from which the fleet was sampled. Because we are working with logarithmically transformed data, the means in question correspond, in the original g/mile scale of the measurements, to geometric rather than arithmetic means. Confidence bounds (95%) for the geometric mean are given by  $\exp(\text{Avg} - 2 \sigma_{\text{mean}})$  and  $\exp(\text{Avg} + 2 \sigma_{\text{mean}})$ , where Avg is the log-scale emissions average.

Table 3 gives approximate 95% confidence intervals calculated for some illustrative fleet sizes. The intervals are specified in terms of percent deviations [ $100\exp(-2 \sigma_{\text{mean}})-100$ ,  $100\exp(2 \sigma_{\text{mean}})-100$ ] from the (geometric) sample mean. Results are relatively insensitive to the number of replicate tests performed on each vehicle; those in Table 3 assume  $r = 3$ . Confidence intervals for arithmetic sample means can be estimated by back-transformation of the log-scale statistics; Tables 3 and 4 of Appendix 3 summarize these for a range of  $n$  and  $r$ .

	Number of vehicles					
	10	20	30	60	90	120
	95% confidence intervals					
$\text{NO}_x$	-25%, +34%	-19%, +23%	-15%, +18%	-11%, +13%	-9%, +10%	-8%, +9%
CO	-36%, +56%	-27%, +37%	-23%, +29%	-17%, +20%	-14%, +16%	-12%, +14%
HC	-37%, +58%	-28%, +38%	-23%, +30%	-17%, +21%	-14%, +16%	-12%, +14%
$\text{CO}_2$	-7%, +8%	-5%, +6%	-4%, +5%	-3%, +3%	-3%, +3%	-2%, +2%
PM	-34%, +52%	-26%, +34%	-21%, +27%	-16%, +19%	-13%, +15%	-11%, +13%

**Table 3.** Confidence intervals for geometric-mean emissions rates from chassis dynamometer tests, as a function of the test fleet size. Values are based on analysis of existing WVU data from multiple tests of 18 heavy-heavy-duty diesel vehicles on the 5-mile route and/or 5-peak cycle. Estimates assume 3 replicate tests on each vehicle.

The logic underlying Table 3, though based on Class 8 data, should be equally applicable to classes 4 to 7 (i.e. all trucks bigger than pickups) because the diesel technology is essentially the same. The logic may also extend to Class 2a and 3 vehicles, but these are generally more overpowered than underpowered and driving style may become an influence on PM emissions.

### Concluding Remarks

Table 3 above and Tables 3 and 4 in Appendix 3 summarize the expected levels of uncertainty for different fleet sizes in estimating the mean emissions for one vehicle class using a test cycle similar to the 5-mile cycle or the 5-peak route. There are substantial decreases in the uncertainty limits as fleet sizes increase from 10 to 120 vehicles, but the number of repeat tests per vehicle has little effect on the uncertainty limits. Repeat tests are necessary for experimental quality assurance, but two repeat tests per vehicle provide sufficient statistical reliability.

The uncertainty limits calculated in this project can be used as a guide to the likely level of uncertainty in fleet average emissions so long as the selected fleet is not far more heterogeneous in emissions than the 18 vehicles used in this study. The use of additional criteria such as model year to select vehicles would be expected to yield more homogeneous subgroupings within the test fleet. Depending on the effectiveness of such stratified sampling, the desired statistical precision might be attained with fewer vehicles for the mean of an individual subgroup.

It is likely that the year of manufacture of the vehicle will influence the emissions, because of changing emissions standards that heavy duty vehicles must meet. However, since standards do not change annually, it may be possible to group trucks manufactured in years with similar standards. This would suggest, for example, that vehicles manufactured in the years 1994-1997, and 1998 and later, might be grouped together. While standards did change prior to 1993, a third pre-1993 group, or two groups representing the older trucks, might be considered. In this way emissions trends may be quantified more reliably without directly requiring representative trucks from every model year. Such stratification could not be attempted here because only 4 of the 18 trucks in our study were produced after 1993, and they were all from the 1994 model year.

It is understood that trucks built in the 1990's may exhibit off-cycle emissions, but the triggering of these emissions is not well documented. It may prove necessary in testing vehicles to ascertain whether the vehicle controller is in original condition or whether the off-cycle emissions have been disabled during a retrofit. Two clear classes of vehicle may become evident in the study due to off-cycle behavior. In Table 2, the variation between vehicles for CO<sub>2</sub> was small, but that for NO<sub>x</sub> was large. Since it is usually observed that the NO<sub>x</sub>/CO<sub>2</sub> ratio for an

engine is fairly steady if there are no excursions in injection timing, the high value for  $\text{NO}_x$  variation might be attributed to the varied presence or absence of "off-cycle" emissions where  $\text{NO}_x$  may increase more than twofold with only a small change in the opposite direction occurring in  $\text{CO}_2$ .

PM emissions, in contrast to  $\text{NO}_x$ , may emerge as being closely tied to vehicle maintenance, and intervals of service and air filter change, so that PM will vary with odometer reading.

### Acknowledgments

This analysis would not have been possible without the generous assistance of Dr. Nigel Clark at WVU, who provided an electronic spreadsheet summarizing the WVU chassis dynamometer test data, extensive documentation of WVU's test procedures, and helpful advice concerning analytic and experimental strategies. Hector Maldonado of CARB also provided helpful advice, together with useful information on current fleet characteristics and emissions estimates for southern California. We acknowledge financial support of this project from DOE's Office of Heavy Vehicle Technologies through the National Renewable Energy Laboratory.

Appendices (by R.F. Gunst)

1. Assessment of Statistical Modeling Assumptions (3/1/00)
2. Selection of Vehicles and Test Conditions (3/14/00)
3. Uncertainty Estimates (3/15/00)