The Use of Mixed Effects ANCOVA to Characterize Vehicle Emissions Profiles

MICHAEL L. DEATON

JAMES J. WINEBRAKE

James Madison University

ABSTRACT

This paper uses a mixed-effects analysis of covariance model (with both fixed and random effects) to characterize mileage-dependent emissions profiles for any given group of vehicles having a common model design. Such profiles are useful for evaluating, for example, how emissions will change over time within a new line of vehicles. The U.S. Environmental Protection Agency uses these types of evaluations to certify whether or not new models conform to existing emissions standards. Given such a group of vehicles, the statistical model introduced in this paper describes both the average emissions profile for that group while also accounting for individual vehicle variability among vehicles within the group. The model can be used to provide realistic confidence bounds for the average emissions deterioration profile within a given group, therefore allowing accurate emissions comparisons of multiple groups. The approach is illustrated with a sample of emissions data from two types of vehicles: natural gas Dodge Ram vans and gasoline Dodge Ram vans (all from the 1992–94 model years). The population profile for nonmethane hydrocarbons is explored. The results indicate the presence of vehicle-to-vehicle variation within each

---

Michael L. Deaton, James Madison University, Integrated Science and Technology Program, MSC 4102, Harrisonburg, VA 22807. Phone: 540–568–2725. Fax: 540–568–2955. Email: deatonml@jmu.edu.

vehicle type. This variation leads to confidence profiles that can be markedly different (but more appropriate) than what would be obtained from a simple fixed-effects regression model. The results highlight the potential for incorrectly characterizing emissions profiles whenever decisionmakers rely on standard regression techniques.

INTRODUCTION

Policymakers who establish emissions standards for new vehicles often focus on both the *baseline* emissions when the automobile is new, as well as on the rate at which those emissions *deteriorate* with vehicle age and use. Unfortunately, the emissions (as well as the emissions deterioration rate) from any individual vehicle after a specified amount of use can vary significantly from the average emissions of all similar vehicles under the same conditions. Hence, when evaluating average emissions across a population of vehicles that are nominally identical (same make, model, design) but utilize new technologies (such as alternative fuels), it is necessary to characterize the emissions profiles (average emissions as a function of mileage traveled) for the population, while also accounting for variation among vehicles within the population.

Over the past several years, many studies have attempted to collect and analyze emissions from in-use alternative fuel vehicles (AFVs) (i.e., AFVs operating in normal, daily driving conditions). Examples of these studies for light-duty vehicles include the work of Gabele (1990, 1995), Kelly et al. (1996a, 1996b, 1996c), Kirchstetter et al. (1996), Norbeck et al. (1998), Durbin et al. (1999), and Whalen et al. (1999). Examples from the heavy-duty literature include Clark et al. (1998), Chandler et al. (1999), and McCormick et al. (1999).

One significant data-collection effort has been funded by the U.S. Department of Energy and managed by the National Renewable Energy Laboratory (NREL). This program has collected emissions data from over 400 AFVs and gasoline control vehicles operating in federal government fleets. These vehicles operate on a variety of fuels, including methanol blends, ethanol blends, compressed natural gas, and propane. Vehicles are operated in various federal agency fleets and represent a variety of driving conditions and operations. The National Alternative Fuels Data Center (AFDC), located in Golden, Colorado, collects and publishes data from these emissions tests.

Policymakers are interested in the results of such studies in order to evaluate the potential impact of AFVs on air pollution. This necessarily requires that researchers develop models for the emissions generated by these vehicles over their useful lifetime. These emissions profiles may then be used to characterize lifetime emissions for those vehicles and to help establish standards for acceptable emissions levels at various points in a vehicle's lifetime.

The goal of this paper is to illustrate one approach that evaluates an assumed functional relationship between emissions and mileage, but also attempts to properly incorporate and account for variation in emissions from one vehicle to another. In doing so, a more complete understanding of the average deterioration in a group of vehicles and of the variation among vehicles and between fuel types is possible.

The statistical model described in this paper is a generalization of the classic analysis of covariance (ANCOVA) model. This approach is more precise than conventional regression models because it accounts for both engine age (as measured indirectly by odometer readings) and variations between vehicles of the same make and model.[1] Furthermore, the generalized ANCOVA allows more realistic estimates of the variation inherent in comparisons between vehicles operating on different fuels and allows more realistic estimates of the size of confidence bands for the average emissions across all vehicles and also for individual vehicle emissions.

The second section illustrates the impact that variations among vehicles can have on estimated emissions profiles and on the width of confidence bands for the average emissions profile. We use a simple example to illustrate the key concepts. We demon-

[1] The statistical model presented in this paper can be generalized to describe emissions profiles in populations containing a variety of vehicle designs, model years, etc. This more generalized model would be useful for characterizing the emissions of a highly diversified population (i.e., a fleet owned by a large corporation or government agency). However, this paper focuses on the more restrictive problem of characterizing the emissions profile in a group of vehicles that are nominally identical with respect to model design, engine type, etc.

strate that evaluations of emissions profiles that fail to properly account for vehicle-to-vehicle variation can lead to confidence bands that give overly optimistic estimates of the precision with which the average emissions profile (averaged across all vehicles in the group of interest) can be determined.

The third section describes a general mixed-effects ANCOVA model that may be used to: 1) estimate emissions profiles in one or more groups of vehicles, and 2) compare emissions profiles among those groups. This model accounts for random variations between vehicles, thereby avoiding the pitfalls illustrated in the second section.

The final section demonstrates the use of the general ANCOVA model described earlier by analyzing nonmethane hydrocarbon (NMHC) emissions from 58 in-use vehicles selected from the AFDC database. All 58 vehicles are Dodge Ram vans with the same engine size, and all from model years 1992–94. Twenty-seven of these vehicles ran exclusively on compressed natural gas, while the other 31 vehicles were dedicated to the exclusive use of California Phase II reformulated gasoline (RFG).

THE IMPACT OF VEHICLE-TO-VEHICLE VARIATION ON ESTIMATED EMISSIONS PROFILES

One seemingly common-sense approach to evaluating emissions profiles over vehicle lifetimes is to express emissions as a simple linear function of mileage (thereby indirectly accounting for deterioration effects). That is, one can fit the simple linear regression model

$$Y_{ij} = \alpha + \beta m_{ij} + \varepsilon_{ij} \qquad (1)$$

where Y_{ij} is the j^{th} emissions reading on the i^{th} car taken at odometer reading m_{ij}. This model assumes that emissions are a linear function of mileage. This model is also based on the important assumption that the only random variation in emissions comes from the error term ε_{ij}.

Such an approach, however, does not adequately account for the inherent variation among individual vehicles within a group.[2] Hence, the resulting confidence bands for the *average group-wide average emissions profile,* as well as the tolerance bands giving estimates of the expected *range* of emissions from individual vehicles, are often too narrow. This failure to account for vehicle-to-vehicle emissions variability may also lead to incorrect statistical testing and estimation procedures, thereby making it difficult to reliably detect differences between groups of vehicles and fuel types.

In order to illustrate these concepts, imagine the case in which one randomly selected new car is used to evaluate the population-wide average emissions profile for all similar vehicles.[3] This vehicle is driven for 100,000 miles on a test track and its NMHC emissions are measured every 10,000 miles. This imaginary study would provide 10 ordered pairs of data (miles driven, NMHC emissions). The common-sense approach described above would use these 10 observations to fit a simple model of the form given in equation (1), where Y_{ij} is the measured NMHC emissions of the i^{th} car after m_{ij} miles of driving; m_{ij} is the miles driven by car i on the j^{th} measurement, and ε_{ij} is the random variation due to unexplained factors.[4] It is typically assumed that the ε_{ij}'s are independently distributed from a normal distribution with a mean of zero and a standard deviation of σ_{ε}. Under this traditional regression model (which does not account for vehicle-to-vehicle variation), the population-wide average emissions $E(Y)$ after m miles of driving is given by

$$E(Y) = \alpha + \beta m. \qquad (2)$$

Conventional least-squares estimation of the above model leads to estimates of α and β, which are designated as $\hat{\alpha}$ and $\hat{\beta}$. Using these well-known results, along with the simplifying assumption that

[2] A *group* of vehicles is defined here as all vehicles that are nominally *identical* with respect to make, model, engine size, year, and fuel type. The analysis reported herein assumes that a random sample of vehicles from this group has been taken and the emissions monitored over an extended mileage range.

[3] It is clear that the use of one vehicle to characterize the emissions profile for an entire group of similar vehicles is not a very sound practice. However, this simple case will be used here in order to simplify the mathematical presentation. Moreover, the Environmental Protection Agency's emissions certification program requires manufacturers to test only one vehicle in order to estimate emissions profiles for an entire population of similar vehicles (Hormes 2000).

[4] Note that the subscript i is not necessary here, but is included to emphasize the fact that the i^{th} car in the population has been selected. The reasons for including this notation will be evident later.

the error standard deviation σ_ε is known, the conventional regression approach will lead to the following quantities of interest (Graybill 1996).

1. The estimated emissions profile:

$$\hat{E}(Y) = \hat{\alpha} + \hat{\beta}m. \quad (3)$$

2. A 95% confidence band for the average emissions (averaged across all vehicles in the population) at mileage m:

$$\hat{E}(Y) \pm 1.96 \cdot \sigma_\varepsilon \cdot \sqrt{\frac{1}{n} + \frac{n(m-\bar{m})^2}{n\Sigma m_{ij}^2 - (\Sigma m_{ij})^2}} \quad (4)$$

3. A 95% prediction band for the emissions of an *individual vehicle* at mileage m:

$$\hat{E}(Y) \pm 1.96 \cdot \sigma_\varepsilon \cdot \sqrt{1 + \frac{1}{n} + \frac{n(m-\bar{m})^2}{n\Sigma m_{ij}^2 - (\Sigma m_{ij})^2}} \quad (5)$$

Note that the quantity $\bar{m}$ in equations (4) and (5) stands for the average mileage odometer reading in the data, and n is the total number of observations in the study (n = 10 in this example).

Now suppose that there is a sizeable difference in emissions levels between vehicles in the population. For simplicity, assume that all the vehicles in the population exhibit the same deterioration rate of NMHC emissions (i.e., the value of β is the same for all vehicles in the population), but that the baseline emissions value is different from one vehicle to another (i.e., the intercept varies between vehicles). In this case, we can generalize the model in (1) to be

$$Y_{ij} = \alpha + \upsilon_i + \beta m_{ij} + \varepsilon_{ij}. \quad (6)$$

Notice that the only difference between (6) and the traditional model in (1) is that quantity υ_i has been added to the intercept. With this model, α is the average value of the intercept (averaged across all vehicles in the population), and the quantity υ_i is the amount that the intercept for vehicle i deviates from the population-wide average (α). Here, all vehicles in the population exhibit emissions profiles that follow the same slope, but these profiles of are offset from one vehicle to the next.

Assuming the vehicle in the study was randomly selected, the value of υ_i is random. Moreover, if the value of α is unknown, the value of α and υ_i cannot be uniquely determined from the data. It is typically assumed that the values of the υ_i in the population are independent and follow a normal distribution with a mean of zero (i.e., the average intercept across all vehicles in the population is α) and a standard deviation of σ_υ (i.e., the intercepts vary randomly from vehicle-to-vehicle, and the standard deviation of intercepts from all vehicles is σ_υ).

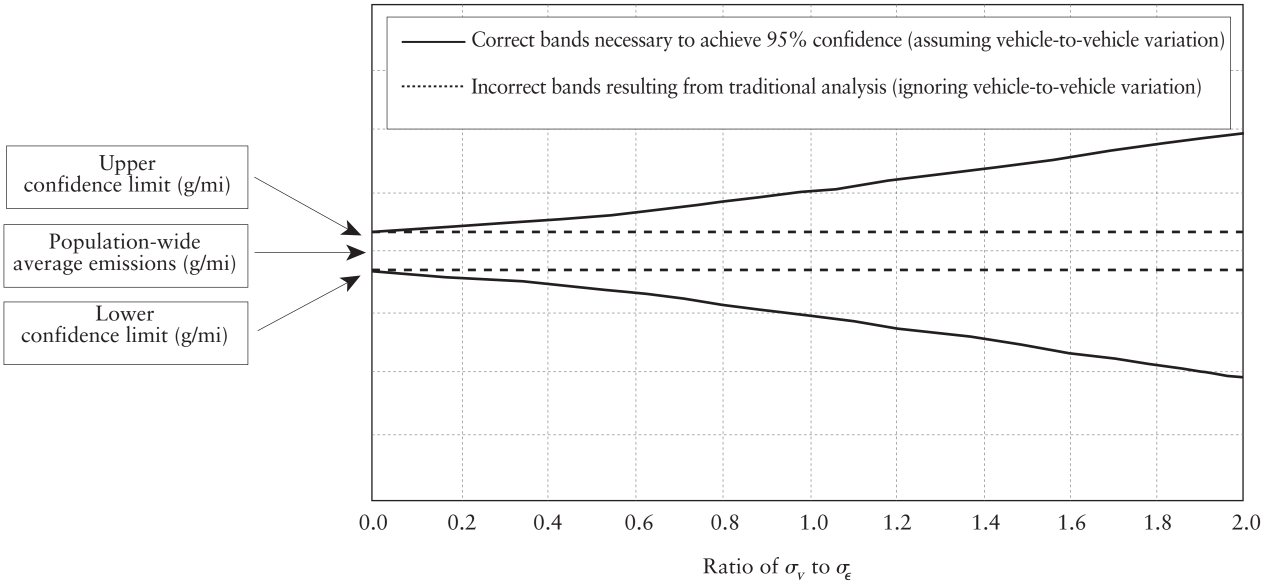
Now suppose that the researcher fails to recognize the structure in (6), and fits the model in (1) using standard least squares techniques; that is, he fits a model that fails to account for the random variation between vehicles.[5] Given these assumptions, Appendix A shows that the following are true:

1. The estimated average profile given in (3) still gives an unbiased estimate of the population-wide average emissions; and
2. A 95% confidence band in (4) for the population-wide average emissions and a 95% prediction band in (5) for predicting the emissions of an individual vehicle after m miles of use are too narrow. The discussion below elaborates on this point.

Statement 2 above is supported by figure 1, which illustrates the width of a 95% confidence band for population-wide average emissions at 55,000 miles in the hypothetical example. The theoretically correct 95% bandwidth is spanned by the outside, solid-line curves. Confidence intervals that have a 95% probability of including the actual population-wide average emissions have an expected bandwidth that corresponds to the solid-line curves. The bandwidth of the traditional interval, as determined from equation (4) above, is spanned by the inside, dashed-line curves. Confidence intervals based on this bandwidth will have less than 95% probability of including the true population-wide average emissions. The x-axis displays the ratio of the vehicle-to-vehicle

[5] The study design in our example would be inadequate for detecting vehicle-to-vehicle variation. If vehicle-to-vehicle variation was believed to be present, care would be taken to collect data from several randomly selected vehicles from the fleet. Using the techniques described later in this paper, the value of the vehicle-to-vehicle standard deviation could then be estimated.

FIGURE 1 Comparison of Confidence Bands on Population-Wide Average Emissions When Vehicle-to-Vehicle Variation is Present



standard deviation (σ_v) to the error standard deviation (σ_ϵ). Hence, when this ratio is zero, there is no vehicle-to-vehicle variation and the traditional approach is appropriate. Notice that when the ratio on the x-axis is zero, the "correct" confidence band and the band from traditional regression are identical.

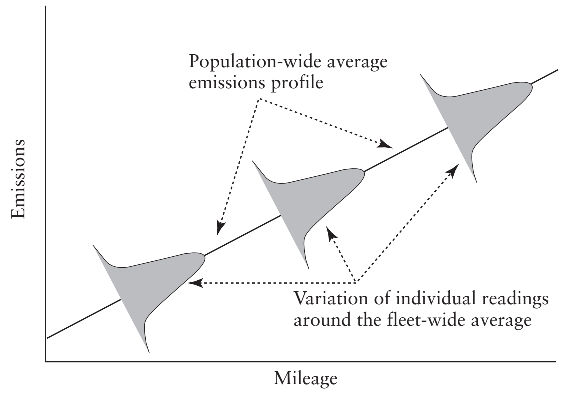
On the other hand, when the ratio on the x-axis is large, the vehicle-to-vehicle variation is also large. In such a case, the traditional regression model fails to account for the additional source of variation between vehicles. For example, consider the case when the vehicle-to-vehicle variation is the same size as the error variation (i.e., the ratio on the x-axis is equal to 1). It is clear from figure 1 that the traditional confidence band is too narrow by a factor of 3 or more. Hence, in this case, the traditional approach leads to a grossly over-optimistic picture of how precisely the population-wide average emissions profile may be estimated. In fact, even if the size of vehicle-to-vehicle variation is small (as when the ratio on the x-axis is 0.4 to 0.6), the error in the confidence bandwidth can be large. In such a case, the use of the conventional simple linear regression model in (1) will lead to confidence bands that are advertised to have a 95% confidence level, but that have a much lower confidence level in reality.[6]

Figure 1 illustrates the practical implications of vehicle-to-vehicle variation. Figures 2, 3, and 4 illustrate this in a slightly different way. Figure 2 illustrates the case in which there is no vehicle-to-vehicle variation. In this case, all the vehicles in the population have an assumed common emissions profile, indicated by the solid line. However, because of random variations from one measurement of emissions to the next (due to imprecision in laboratory methods, etc.), a given vehicle's emissions measurement at a particular mileage will vary randomly around the population-wide profile. This variation is represented by the bell-shaped curves spaced along the line. Each bell-shaped curve represents the distribution of emissions measurements that one could expect to see at the specified mileage reading.

Figure 3 illustrates the case in which each vehicle in the population has its own emissions profile. More specifically, figure 3 represents the case in which all of the profiles are parallel (i.e., the rate of emissions deterioration is constant for all vehicles),

[6] The error in the confidence band will not be as great if multiple cars are included in the sample. Nonetheless, even if multiple cars are sampled, the error in the confidence bandwidth can still be sizeable, provided that the vehicle-to-vehicle variation is large.

FIGURE 2 Illustration of the Case in Which All Vehicles Have the Same Emissions Profile



while the intercept of the emissions profile varies from one vehicle to the next. This corresponds to the model in (6). Notice that each individual line in figure 3 also displays several bell-shaped curves that represent the distribution of actual emissions measurements from each individual car at a given mileage. Figure 4 superimposes on figure 3 the population-wide average emissions profile, along with a corresponding set of bell-shaped curves along that profile. Notice that the bell-shaped curves in figure 4 are much wider than in figure 2 where no vehicle-to-vehicle variation is present. This is because the collection of emissions readings from a randomly selected car at a fixed mileage will vary from the population-wide average due to random error variation (σ) and because of variations between vehicles (σ_v).

Hence, if vehicle-to-vehicle variation is present in the form indicated in equation (6), then regression analysis that is based on the simple linear model in (1) will lead to confidence bands and prediction intervals that can be highly inefficient and possibly even deceptive. Policymakers who rely on such estimates to make comparisons between different groups of vehicles (e.g., vehicles operating on different fuels) run a sizeable risk of making decisions that do not realistically reflect the actual capabilities of those populations.

FIGURE 3 Illustration of the Case in Which Individual Vehicles' Emissions Profiles Have Different Intercepts from One Vehicle to the Next

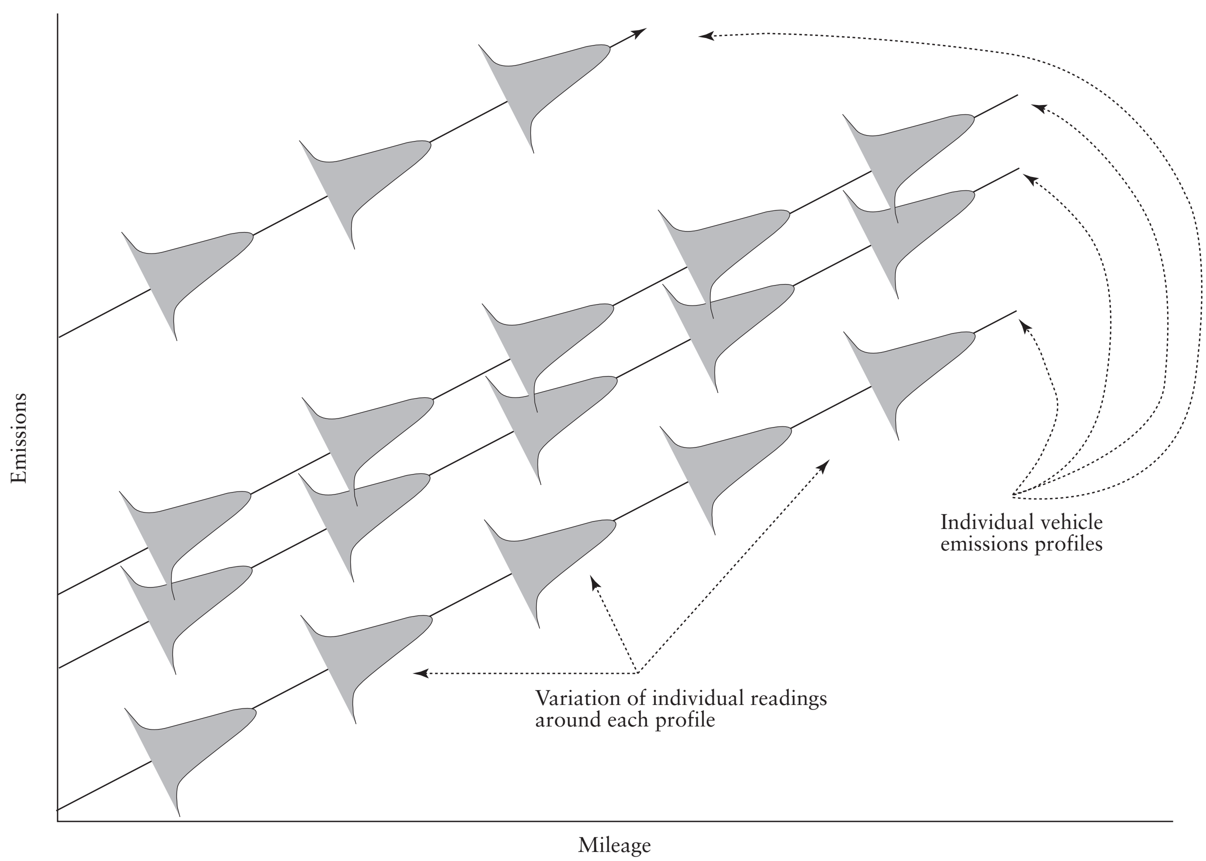
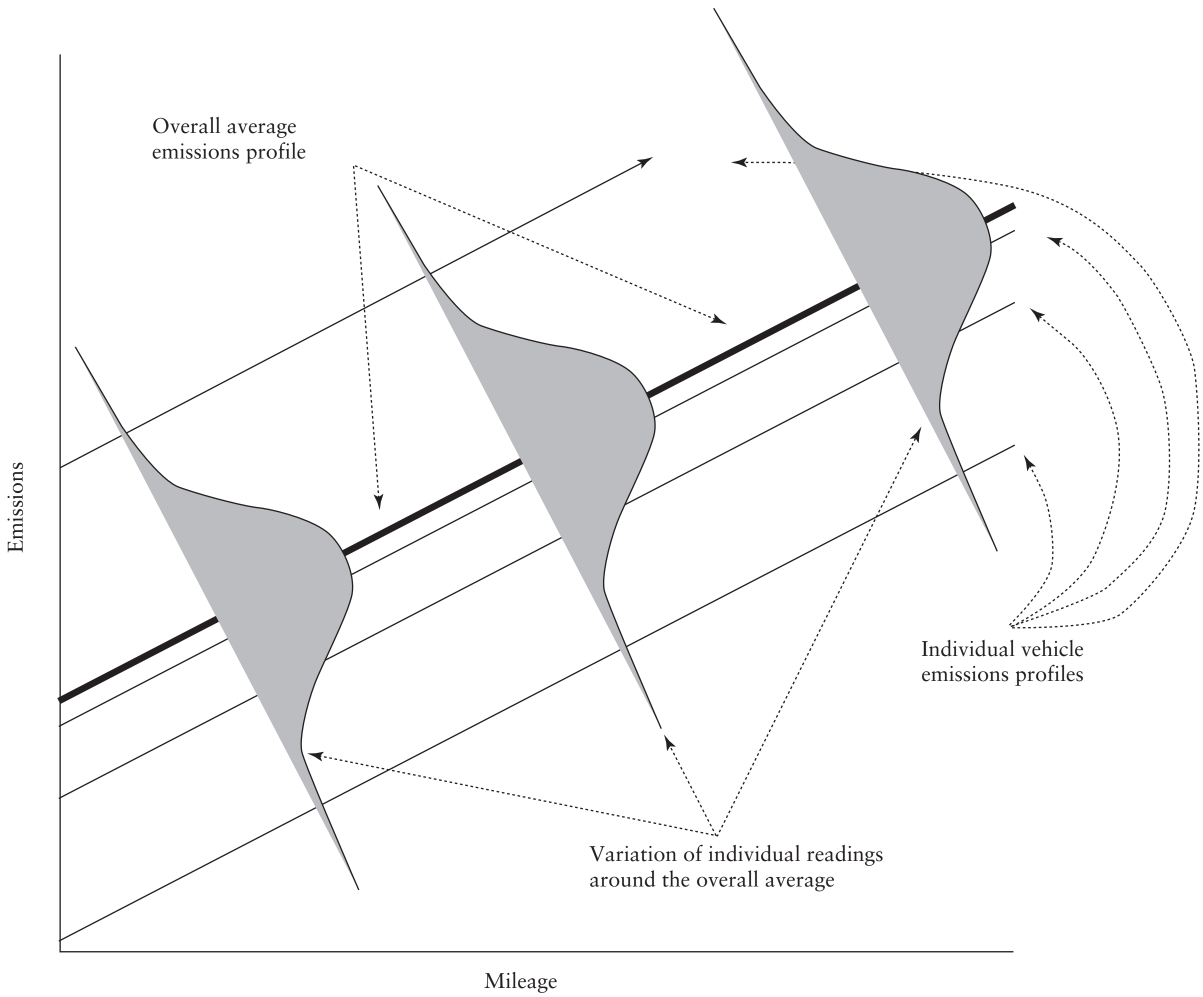


FIGURE 4 Average Emissions Profile in the Presence of Vehicle-to-Vehicle Variation



Consider, for example, the case in which the vehicle-to-vehicle standard deviation is the same size as the error standard deviation. Hence, for this case, the ratio on the horizontal axis of figure 1 is 1.0. In such a case, if model (1) is used to characterize the population-wide emissions profile, then the resulting 95% confidence bands for average emissions at 55,000 miles will be too narrow by approximately 70%, and the confidence level for those bands will in fact be much lower than 95%. Such an error can lead policymakers to have an overly optimistic picture of how variable emissions will be from vehicles in this population. This misunderstanding can lead to emissions standards that are unreasonably tight.

Figure 1 also suggests that the type I error rate (i.e., the α level) associated with traditional hypothesis testing procedures can be much greater than the advertised level whenever vehicle-to-vehicle variability is present and is not properly accounted for in the analysis. This means that chances of spurious statistically significant results can be much greater than the advertised α-level when vehicle-to-vehicle variability is ignored. For example, suppose that the vehicle-to-vehicle standard deviation is the same size as the error standard deviation and a two-sample t-test with an α-level of 0.05 is used to compare a group of alternative fuel vehicles to a corresponding group of conventional fuel vehicles. Further suppose that the analysis did not properly account for the vehicle-to-vehicle variation. Based on the earlier argument, the resulting hypothesis test may in fact have an α-level that is much greater than the advertised α-level of 5%. This means that the researcher has much more than a 5% risk of incorrectly finding a difference between the two groups of vehicles when no such difference really exists.

Model (6) is more realistic than model (1) because it allows for potential variations between vehicles in a population. However, (6) can be further improved by allowing for different deterioration rates as well as different baseline emissions from one vehicle to another. In such a case, one would expect that the problems with the confidence bands and prediction bands from a traditional regression model would be even more acute than illustrated here. The next section introduces this more general model and also incorporates terms that allow for statistical comparison of different populations or fuel types.

DESCRIPTION OF A GENERAL MODEL

The statistical model used in this study relies on the general methodology of *analysis of covariance* (ANCOVA) discussed in Searle (1971). This model can be used to compare two or more "treatments" that have been applied to a group of individuals. In the present study, the "group" consists of individual vehicles assumed to be nominally "identical" with respect to make, model, engine size, fuel type, etc. The treatments are the different fuels under which these vehicles are operated. The response of interest is the emissions of a given pollutant. The simplest ANCOVA model accounts for the fact that the response (i.e., emissions) depends on a "covariate" (i.e., mileage driven), which can change from one observation to the next. In this sense, the ANCOVA model is a general application of the standard analysis of variance (ANOVA) in which one or more treatments are compared, but in which there is no covariate.

The model illustrated here generalizes the simplest ANCOVA model to also account for the random variation between vehicles within the population. By doing so, the analyst is afforded accurate tests for comparing emissions profiles among fuel types and for comparing emissions at any specified mileage. The approach is well established in the statistical literature (see, e.g., Searle 1971 and Federer and Meredith 1992), but it has received little attention in the field of emissions modeling (one exception is a study conducted by Battelle Memorial Institute (1995)).

Let Y_{ijk} represent a specific emissions constituent as observed on the k^{th} test on the j^{th} vehicle that is operating on fuel type i. Let $m_{k(i,j)}$ stand for the k^{th} mileage reading on car j operating on fuel type i. It is assumed that only one emissions result is obtained at each mileage reading on each vehicle (but the model can be generalized to handle multiple measurements). The model has the form:

$$Y_{ijk} = [\alpha + \beta \bullet m_{k(i,j)}] + [\phi_i + \delta_i \bullet m_{k(i,j)}] + [\upsilon_{j(i)} + \varpi_{j(i)} \bullet \mathrm{m}_{\kappa(i,j)}] + \varepsilon_{ijk} \qquad (7)$$

The first two terms $[\alpha + \beta \bullet (m_{k(i,j)})]$ represent the *average* dependence of the emissions on vehicle mileage, regardless of which fuel type is used or the variation that is inherent among individual vehicles. The next two terms $[\phi_i + \delta_i(m_{k(i,j)})]$ represent how this average dependence is affected by fuel type i. The next two terms $[\upsilon_{j(i)} + \varpi_{j(i)}(m_{\kappa(i,j)})]$ represent how the average dependence is affected by the unique characteristics of vehicle j that operates on fuel type i.

This model allows for the realistic situation in which there is an overall population-wide deterioration curve that describes the average emissions for all vehicles in the group of interest that are using fuel type i. The group-wide emissions curve when operating on fuel type i is defined by the expression $\alpha + \beta \bullet (m_{k(i,j)}) + \phi_i + \delta_i(m_{k(i,j)})$. However, the model also accounts for the fact that each vehicle in the group may have an emissions curve that differs slightly from the average curve for all similar vehicles. This variation from the average curve can occur in either the intercept (through $\upsilon_{j(i)}$), the slope (through $\varpi_{j(i)}$), or through both the intercept and slope. The final term (ε_{ijk}) represents the random variation in emissions that are not accounted for in the model. This variation may be attributed to such things as variation from the test method used, differences between laboratories (if each car is tested at multiple labs), or any number of other factors.

The assumptions behind this model are stated as follows:

- Assumption 1: At a fixed mileage, emissions follow a normal distribution.
- Assumption 2: The quantities α, β, ϕ_i, and δ_i in the model in equation (7) are fixed, but unknown parameters. Moreover, since the ϕ_i and δ_i represent deviations from the mean intercept and slope, respectively, it assumed that $\sum \phi_i = \sum \delta_i = 0$. If the study is aimed at characterizing the emissions profile of a fixed or specified group of vehicles and for a fixed set of fuel types, then this fixed-effects assumption is

reasonable. However, if the study's goals are to characterize emissions across a wide variety of vehicles and fuel types, but data have been collected on only a random sample of vehicles and a random sample of fuel types, this assumption must be relaxed. The present study (and many other studies of practical interest) are consistent with this fixed effect's assumption.

- Assumption 3: $\upsilon_{j(i)}$, $\overline{\omega}_{j(i)}$, and ε_{ijk} are all random quantities. Each of these terms is assumed to follow a normal distribution having a mean of zero. The standard deviations of these distributions are σ_υ , $\sigma_{\overline{\omega}}$, and σ_ε, respectively. The standard deviations σ_υ and $\sigma_{\overline{\omega}}$ measure how much individual vehicle emissions profiles will vary around the population average emissions profile; that is, the larger σ_υ and $\sigma_{\overline{\omega}}$ are, the more individual vehicle emissions profiles may vary from the population average profile. It is also assumed that $\upsilon_{j(i)}$, $\overline{\omega}_{j(i)}$, and ε_{ijk} are mutually independent.

The reader should note that this model does not explicitly account for variation between the laboratories conducting the tests. The AFDC data analyzed in this paper were collected across three different laboratories, one of which was located at a high altitude. Lab-to-lab variation can be a dominant source of variation in these types of measurements. However, the model will provide a reliable test for comparing emissions from the two fuel types provided that (i) each car was tested at only one lab, and (ii) within each lab, vehicles from both fuel types were tested. Both requirements were satisfied by the data analyzed in this paper. Furthermore, under these assumptions, the lab-to-lab variation will be accounted for in the model, but will be indistinguishable from vehicle-to-vehicle variability. Hence, if the analysis suggests a large variation between vehicles within each group of interest, we cannot conclude that this source of variation is found only in differences between vehicles. It may partly be caused by variations between testing labs.

EXAMPLE APPLICATION: 58 DODGE RAM VANS FROM THE AFDC DATABASE

The ANCOVA model presented here was applied to emissions values from the AFDC database for 27 compressed natural gas (CNG) Dodge Ram vans and 31 gasoline counterparts (henceforth referred to as "RFG" for "reformulated gasoline"). Data was extracted on August 11, 1998. Several pollutants were measured on each car. Results for nonmethane hydrocarbons are analyzed and reported here.

Emissions tests on these vehicles were conducted at three commercial laboratories in various locations in the United States. A competitive bidding process was used to select the labs. A panel of experts (including U.S. Environmental Protection Agency—EPA—personnel) conducted site visits to ensure that standardized testing methods were used across all three labs and that appropriate quality assurance procedures were in place. Each vehicle was tested using the EPA's Federal Test Procedure (FTP) protocol at accumulated mileage readings of approximately 4,000 miles, 10,000 miles, and every 10,000 miles thereafter. Because of obvious logistical reasons, it is not the case that all the vehicles were tested at these exact mileage specifications. The general test procedures, emissions test driving profiles, and hydrocarbon specification procedures, along with other facts about the AFDC testing program and vehicles are reported elsewhere (Kelly et al. 1996a, 1996b, and 1996c).

Table 1 provides information about the vehicles, their fuels, and the number of vehicles per fuel (sample sizes). Note that all the CNG vehicles were original-equipment-manufactured Dodge Ram vans (i.e., none of the vehicles was an aftermarket conversion). Although no data are available on exactly how each vehicle was used, it is assumed that all the vehicles experienced similar driving conditions. This assumption may not be valid, and thus should be considered when interpreting the results of this analysis.

As shown in table 1, the alternative fuel vehicles come mostly from model year (MY) 1992, with fewer coming from MY 1994. The reverse is true for the RFG vehicles in the study. This discrepancy could jeopardize the ability to make comparisons of the CNG and RFG emissions if different emissions control systems had been installed on the

TABLE 1 Information on Vehicle Types and Fuels

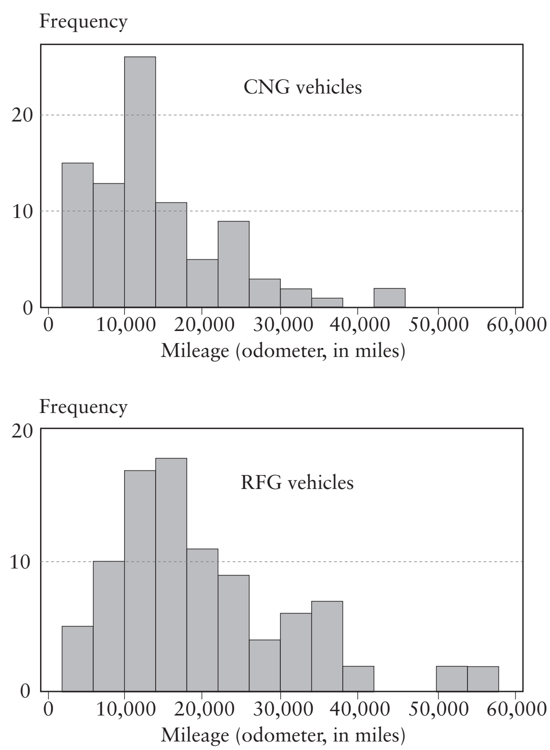
Vehicle type	N (by model year)
Dedicated original equipment manufactured CNG Dodge Ram B250 Van (CNG/Ram)	22 (1992) 5 (1994)
■ 5.2 liter V-8 engine configuration ■ Multi-point fuel injection ■ 4-speed automatic ■ 11.1–15.7 equivalent gallon fuel capacity ■ 6,400 lbs gross vehicle weight ■ LEV-certified	
RFG Dodge Ram B250 Van (RFG/Ram)	11 (1992) 20 (1994)
■ 5.2 liter V-8 engine configuration ■ Multi-point fuel injection ■ 4-speed automatic ■ 35 gallon fuel capacity ■ 6,400 lbs gross vehicle weight	

1992 vehicles as compared with the 1994 vehicles. This, however, is not the case: emissions control systems in MY 1992 and MY 1994 vehicles are identical for the Dodge Ram vans in this study.[7] It is also important to recognize that these vehicles are now 6 to 8 years old and that they represent emissions control technologies that may have been modified or even replaced. The reader is encouraged to keep in mind the fast pace at which emissions control technologies may change (especially for new AFVs), and to take the potential for new technological advancement into account when interpreting the results reported here. Beyond this issue, MY is given no further consideration in the modeling and analysis.

These NREL-tracked vehicles were FTP tested several times at each of several different mileages. The AFDC database contains weighted FTP (WT) test results for each vehicle at each mileage, which were used in the present study. The original AFDC database included data on over 450 vehicles and 13 different models. In order to provide a sample of vehicles that represented a uniform population with respect to model (body design and engine) and model year, only the data for Dodge Ram vans was used. This original sample included 108 such vehicles. Vehicles were eliminated that were tested at only one mileage reading or if the difference in mileage between the first test and last test was less than 4,000 miles. In addition, emissions tests at mileages less than 3,000 miles were eliminated due to the possibility of a "green catalyst" effect. These selection criteria left the final sample of 58 vehicles (27 CNG and 31 RFG vehicles).

A comparative frequency distribution of the collective mileages with all tests on all 58 vehicles is shown in figure 5. The average mileage for all tests on all CNG vehicles is 14,159 miles, with a median of 11,397 and a maximum of 45,159. The average mileage for all tests on all RFG vehicles is 20,217 miles, with a median of 17,206 and a maximum of 57,099. It is impossible to determine from the available data whether these differences are due to variations in trip duration, trip frequency, or both. It should be noted that the original experimental design specified that all vehicles be tested at the same mileage readings through the course of the study. This allows emissions profiles to be equitably monitored across all vehicles, thereby simpli-

FIGURE 5 Mileage Frequency Distribution for Natural Gas (CNG) and Reformulated Gasoline (RFG) Vehicles



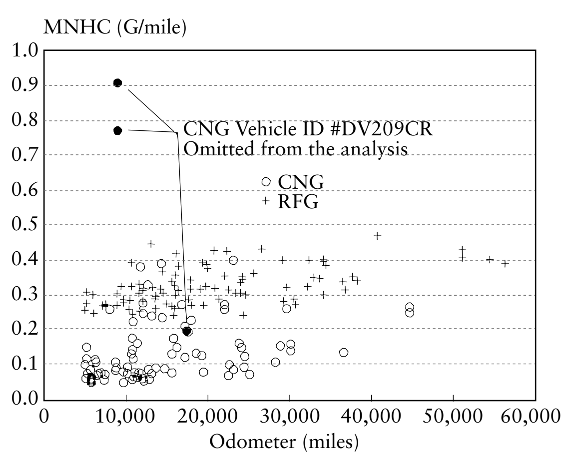
[7] Note, however, that the emissions control equipment for the CNG vehicles is designed for operation on CNG and is different from the equipment used in RFG vehicles.

fying the interpretation of the analysis. Unfortunately, due to the logistical limitations and the large scope of the study, this ideal was not strictly achieved (as illustrated by the non-uniform distribution of mileages in figure 5). While this departure from the intended design complicates the analysis somewhat, it does not invalidate the approach described here. Furthermore, the statistical model discussed above characterizes emissions deterioration only for the specific range of mileages covered in the data. At the outer limits of this range, the precision of the estimated profile is less than at the center of the range where more data are available. This is reflected in wider confidence bands around predicted emissions at high mileages.

Figure 6 visually displays the raw data for all 58 vehicles. A difference in NMHC emissions between the fuel types is suggested in this plot. In addition, the rate of increase in NMHC emissions does not exhibit any sizeable difference between the two fuel types. Both of these features are formally addressed and tested in the analysis.

Figure 6 also exhibits two outliers. These both came from one CNG vehicle that yielded much higher NMHC emissions in its first readings than in subsequent readings. That vehicle's data were omitted from the analysis.

FIGURE 6 Plot of NMHC Emissions vs Odometer Reading for 58 Dodge Ram Vans



RESULTS

As previously noted, the ANCOVA model presented in equation (7) is used to determine whether statistically significant differences exist in the average emissions profile between vehicles operating on different fuels (CNG and RFG), while also accounting for the variations that are inherent from one vehicle to another. The emissions profiles generated by this model estimate the average emissions values that can be expected for a group of vehicles operating on each particular fuel type at any given mileage.

Average emissions values for each fuel type were determined by fitting the complete model in equation (7) using the PROC MIXED procedure in SAS, version 6.12. A listing of the appropriate SAS code is provided in Appendix B. Parameter estimates and their variances were found, allowing the generation of predicted values and confidence bands for the average population-wide emissions component of the model when operating on a particular fuel type. In other words, values and confidence bands were determined for E_i, where E_i is the average emissions from vehicles when operating on fuel type i at a specific mileage m, as follows

$$E_i = \alpha + \beta \bullet m + \phi_i + \delta_i \bullet m \quad (8)$$

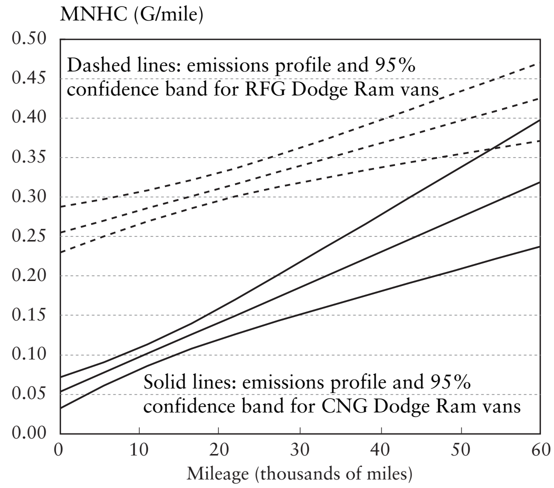
The NMHC emissions profiles in equation (8), along with their 95% confidence intervals, are plotted in figure 7.

The analysis also provides estimates of the error variance (σ^2_ε) and the two variances associated with vehicle-to-vehicle variation (σ^2_ν and σ^2_ω). Table 2 displays these estimates for NMHC. Recall that figure 1 demonstrates that when the vehicle-to-vehicle variation is large relative to the error variation, a model that fails to account for such variation will lead to confidence intervals that are too narrow for the stated level of confidence. Figure 1 shows that the larger the *ratio* of vehicle-

TABLE 2 Estimated Vehicle-to-Vehicle Variation and Error Variation for NMHC Among 58 Dodge Ram Vans

Variance component estimate		
σ^2_ε	σ^2_ν	σ^2_ω
0.0033	0.0000	0.0012

FIGURE 7 NMHC Emissions Profiles with 95% Confidence Bands for Dodge Ram Vans (gasoline vs. natural gas models)



to-vehicle standard deviation to error standard deviation, the more misleading are the confidence intervals (or hypothesis tests) derived from using an incorrect model. From table 2, the ratio in figure 1 is calculated by

$$\sqrt{\frac{\text{vehicle-to-vehicle variance}}{\text{error variance}}} = \sqrt{\frac{\sigma^2_{\nu} + \sigma^2_{\varpi}}{\sigma^2_{\varepsilon}}} = 0.61$$

Using this value on the x-axis in figure 1 suggests that a traditional analysis that fails to account for this large variation between vehicles can lead to confidence bands that are too narrow by a factor of approximately 50%. If emissions standards for in-use vehicles are based on such analyses, those standards may in fact provide an unrealistic picture of the range of emissions to be expected over the lifetime of any group of vehicles.

Table 3 summarizes the results of standard ANCOVA F-tests used to compare the average emissions profiles between the two fuel types. The *F-test for different slopes* in table 3 indicates whether the *rates* of emissions deterioration are the same for both fuel types. If this test is significant, there is strong evidence that the slopes of the NMHC emissions profiles differ between the two fuel types. If the first F-test is not significant, the second F-test (*F-test for a common nonzero slope*) and third F-test (*F-test for a common intercept*) should be examined. If the second test is significant (and the first F-test is not significant), it is safe to conclude that NMHC emissions do change with mileage and that the two groups of vehicles exhibit parallel (and possibly identical) profiles. If the third F-test is significant (and the first F-test is not significant), it is safe to conclude that the two groups of vehicles exhibit parallel, but distinct emissions profiles. Those profiles may be "flat" (unchanging with mileage) or they may exhibit a common nonzero trend, depending on whether or not the second F-test is significant.

TABLE 3 ANCOVA F-Test Results for Comparing NMHC Emissions Profiles Between CNG and RFG Vehicles

F-test for different slopes (p-value)	F-test for a common nonzero slope (p-value)	F-test for a common intercept (p-value)
Not significant	Significant	Significant
(0.1394)	(0.0001)	(0.0001)

Figure 7 displays the estimated emissions profiles for NMHC in both types of vehicles. With respect to NMHC, the CNG vehicles in the study appear to be cleaner than their RFG counterparts across all mileages. This is supported by the F-tests in table 3. The *F-test for slope* and the *F-test for a common nonzero slope* jointly indicate that there is a common nonzero slope in the NMHC emissions profiles for both groups of vehicles. The *F-test for a common intercept* in table 3 indicates that, while the two profiles appear to have a common slope, they are distinct. Combining these results with figure 7, it can be seen that the CNG Rams represented in this study indeed have lower average NMHC emissions than the RFG Rams throughout the mileage range covered and that this difference is statistically significant.

CONCLUSIONS

This paper motivates and describes a generalized analysis of covariance (ANCOVA) model for characterizing emissions profiles among populations of vehicles operating on different fuel types. The approach is illustrated on a data set comprised of 27 CNG and 31 RFG Dodge Ram vans operating in the U.S. federal fleet. The analysis and discussion emphasizes that a proper analysis of emissions must consider: 1) the emissions deterioration that occurs over the lifetime of a vehicle; 2) the emissions variability that is prevalent for individual vehicles; and 3) the emissions variability from one vehicle to another. Conventional regression analyses fail to properly account for 2 and 3. The ANCOVA model used in this study explicitly accounts for all of these factors and can be readily applied to more precisely characterize the emissions of any alternative or conventional fuel vehicles.

Moreover, by properly accounting for variation between vehicles, one can develop a more realistic understanding of the *range* of emissions values that are possible from any randomly chosen vehicle in the population. This range may, in fact, be considerably different from what would be obtained from more classical regression models that fail to account for variations between individual vehicles. This type of understanding can be critical to policymakers and researchers.

The confidence bands displayed in figure 7 are based on the model in equation (7) that accounts for variation among vehicles in the same population. While common sense suggests that such variation does exist, its impact on analyses aimed at characterizing emissions profiles has not generally been appreciated. Whenever the vehicle-to-vehicle variation is large (compared with the error variation), then any analysis that fails to account for variation between vehicles can lead to confidence bands around the emissions profile that are misleading (and may even be seriously misleading). In such a case, comparisons of emissions profiles from different populations or different fuel types are suspect.

APPENDIX A: FORMULAS USED FOR GENERATING FIGURE 1

This section outlines the statistical theory behind the confidence bandwidths displayed in figure 1. It is assumed that the reader is familiar with probability theory and the theory of general linear statistical models as described in Graybill (1976).

Recall that the context for interpreting figure 1 is as follows. Data is collected on some emissions constituent (e.g., NMHC) from a single vehicle after 10,000, 20,000,….,100,000 miles of use. Least squares analysis is then used to fit the model given in equation (1) and to calculate traditional confidence bands for the average emissions after 50,000 miles (using equation (4)). Now suppose that there is some unknown vehicle-to-vehicle variation among the cars in the population of interest. In particular, the intercept of equation (1) varies randomly from one vehicle to the next, so that the correct model for these data is actually equation (6). The question to be answered is this: how misleading is the confidence interval calculated from equation (4)? Figure 1 attempts to provide one way of answering that question.

Note that figure 1 displays the *95% confidence bandwidth* for the traditional confidence interval (from equation (4)), along with the corresponding bandwidth that would be necessary to achieve 95% confidence (assuming that the model in equation (1) is correct). Given the relative size of the vehicle-to-vehicle variation (σ_v) with respect to the error variation (σ), the expected width of the traditional confidence interval can be compared with the width that would be necessary to achieve true 95% confidence (i.e., in order that the probability that the interval covers the true average emissions is truly equal to 95%). The x-axis specifies the ratio $\sigma_v \div \sigma$ and the y-axis displays the expected size of the ± bounds of the traditional interval and the theoretically correct interval. Figure 1 clearly illustrates that as $\sigma_v \div \sigma$ increases, the disparity between the confidence intervals increases.

In order to demonstrate how the bandwidths in figure 1 are calculated, a matrix representation of the general regression model will be used (Graybill 1976). Suppose one new vehicle is randomly selected from the population of interest. This vehicle will be operated for a fixed number of miles (e.g., 100,000 miles), and one or more emissions constituents (e.g., NMHC) will be measured at fixed mileages along the way. Suppose n emissions values are obtained from the vehicle during the life of the study. Further suppose that the relationship between emissions and mileage for each car is correctly represented by equation (6); that is,

$$Y_{ij} = \alpha + \nu_i + \beta m_{ij} + \varepsilon_{ij}, \tag{9}$$

where $i = 1$, and $j = 1,\ldots, n$. Assume that the error terms (ε_{ij}) are independent and identically distributed according to a $N(0,\sigma^2)$ distribution, and that the ν_i terms are independent and identically distributed according to a $N(0,\sigma_\nu^2)$ distribution.

Now suppose that vehicle-to-vehicle variation (as represented by ν_i in (9)) is mistakenly assumed to be absent, and traditional regression methods are used to fit the model in (1), i.e.,

$$Y_{ij} = \alpha + \beta m_{ij} + \varepsilon_{ij}, \tag{10}$$

Using the traditional ordinary least squares estimate of the model in (10), the goal is to calculate the average bandwidth of the 95% confidence interval for the model in (10) (which is based on the assumption of no vehicle-to-vehicle variation) and compare its bandwidth with the correct bandwidth that would be required in order to assure 95% confidence (when vehicle-to-vehicle variation is correctly incorporated).

Following Graybill (1976), matrix notation can be used to represent the model in (10). Define the following matrices.

$$X_{n\times 2} = \begin{bmatrix} 1 & m_{1,1} \\ 1 & m_{1,2} \\ \cdots & \cdots \\ 1 & m_{1,n} \end{bmatrix} \quad {}_{n\times 1} = \begin{bmatrix} Y_{1,1} \\ Y_{1,2} \\ \cdots \\ Y_{1,n} \end{bmatrix} \quad B_{2\times 1} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad E_{n\times 1} = \begin{bmatrix} \varepsilon_{1,1} \\ \varepsilon_{1,2} \\ \cdots \\ \varepsilon_{1,n} \end{bmatrix}$$

$$\Sigma_{n\times n} = var\,(Y_{n\times 1}|X).$$

The model in (10) can then be written in matrix notation as follows:

$$= XB + E.$$

Using ordinary least squares, the estimate of the regression coefficients, $\hat{B}$, is given by

$$\hat{B} = (X'X)^{-1}\,X'Y,$$

and the estimated population-wide average emissions at mileage m is given by

$$M'\hat{B} = M'(X'X)^{-1}\,X'Y,$$

where $M' = (1, m)$. It is easily shown (Graybill 1976) that this estimate is normally distributed with a mean of $M'B$ (i.e., the estimate is unbiased) and a standard deviation equal to

$$\sqrt{M'(X'X)^{-1}\,X'\Sigma X(X'X)^{-1}\,M}.$$

Hence, assuming that the covariance matrix Σ is known, the theoretically correct 95% confidence interval for the estimated emissions at mileage m is given by

$$M'\hat{B} \pm 1.96 \cdot \sqrt{M'(X'X)^{-1}\,X'\Sigma X(X'X)^{-1}\,M} \tag{11}$$

Whenever there is no vehicle-to-vehicle variation, then $\Sigma = \sigma^2 I$, where I is the identity matrix, and expression (11) simplifies to

$$M'\hat{B} \pm 1.96\,\sigma \cdot \sqrt{M'(X'X)^{-1}\,X'IX(X'X)^{-1}\,M} =$$

$$M'\hat{B} \pm 1.96 \cdot \sqrt{M'(X'X)^{-1}\,M}. \tag{12}$$

This last expression is the matrix representation of the confidence band in equation (4).

On the other hand, if vehicle-to-vehicle variation is present, then $\Sigma = \sigma^2 I + \sigma^2_{\nu} J$ where J is a matrix of all 1s. Under these conditions, expression (11) does not simplify to the form in (12). Hence, if it is mistakenly assumed that there is no vehicle-to-vehicle variation and expression (12) (or, equivalently, expression (4)) is used to calculate confidence intervals, the resulting confidence bands will be based on incorrect error terms, and the confidence interval will be less than 95%. The correct 95% bounds are instead given by (11).

The error term in expression (12) (applied to the hypothetical example discussed in section 2) corresponds to the traditional confidence bandwidth displayed in figure 1. The error term in expression (11) corresponds to the correct 95% bandwidth displayed in figure 1.

APPENDIX B: SAS CODE FOR FITTING THE ANCOVA MODEL AND OBTAINING 95% CONFIDENCE BANDS FOR THE POPULATION AVERAGE EMISSIONS DISPLAYED IN FIGURE 6

/*

SAS code to get "best" variance component estimates and predicted emissions separately within each fuel type. These predictions and standard errors correctly account for the covariance structure imposed by the random effects.

A separate call to PROC MIXED is required for each response.

Variables are:

VID = vehicle ID code (unique for each vehicle)

FUEL = type of fuel used by the vehicle

```
              (model assumes only one fuel type
              is used on each vehicle)
    ODOM =    odometer reading
    NMHC =    nonmethane hydrocarbon reading
              on the vehicle at the specified
              mileage
*/
PROC MIXED DATA = SASUSER.FINALRAM
METHOD=ML;
CLASSES VID FUEL_TYP;
/*
The MODEL statement specifies only the "fixed
terms" in the model (i.e., the fuel type and odome-
ter reading). The FUEL*ODOM crossproduct
term instructs SAS to fit a separate slope for each
FUEL type.
*/
MODEL NMHC = FUEL ODOM ODOM*FUEL
/ SOLUTION DDFM=SATTERTH;
/*
The RANDOM statement identifies those terms in
the model that are random. Any terms identified in
the RANDOM statement are automatically includ-
ed in the model and are therefore not explicitly
named in the MODEL statement.
*/
RANDOM VID(FUEL) ODOM*VID(FUEL);
/*
The LSMEANS statements instruct SAS to calcu-
late the predicted mean emissions for each fuel type
at the specified mileage reading. This corresponds
to the quantity given in equation (8) of the paper.
The LSMEANS statement also provides the stan-
dard error that can be used to calculate the 95%
confidence interval for the mean emissions at the
specified odometer reading.
*/
LSMEANS FUEL/AT ODOM = 5000 PDIFF;
LSMEANS FUEL/AT ODOM = 10000 PDIFF;
LSMEANS FUEL/AT ODOM = 15000 PDIFF;
LSMEANS FUEL/AT ODOM = 20000 PDIFF;
LSMEANS FUEL/AT ODOM = 25000 PDIFF;
LSMEANS FUEL/AT ODOM = 30000 PDIFF;
LSMEANS FUEL/AT ODOM = 35000 PDIFF;
LSMEANS FUEL/AT ODOM = 40000 PDIFF;
LSMEANS FUEL/AT ODOM = 45000 PDIFF;
LSMEANS FUEL/AT ODOM = 50000 PDIFF;
LSMEANS FUEL/AT ODOM = 55000 PDIFF;
LSMEANS FUEL/AT ODOM = 60000 PDIFF;
RUN;
```

REFERENCES

Battelle Memorial Institute. 1995. *Clean Fleet: Vehicle Emissions, Statistical Analysis Report No. 6*. Columbus, OH.

Chandler, K.L., P. Norton, and N. Clark. 1999. *Interim Results from Alternative Fuel Truck Evaluation Project,* SAE Report 1999–01–1505. Warrendale, PA: SAE International.

Clark, N.N., B.L. Rapp, M. Gautam, W. Wang, and D. Lyons. 1998. *A Long Term Field Emissions Study of Natural Gas Fueled Refuse Haulers in New York City,* SAE Report 982456. Warrendale, PA: SAE International.

Durbin, T., J.R. Collins, J.M. Norbeck, and M.R. Smith. 1999. *Evaluation of the Effects of Alternative Diesel Fuel Formulations on Exhaust Emission Rates and Reactivity,* Report #99-VE-RT2P-001-FR, South Coast Air Quality Management District Contract #98102. Sacramento, CA: College of Engineering–Center for Environmental Research and Technology, University of California–Riverside.

Federer, W.T. and M.P. Meredith. 1992. *The American Statistician* 46:155–62.

Gabele, P.A. 1990. Characterization of Emissions from a Variable Gasoline/Methanol Fueled Car. *Journal of the Air and Waste Management Association* 40:296–304.

Gabele, P.A. 1995. Exhaust Emissions from In-Use Alternative Vehicles. *Journal of the Air and Waste Management Association* 45:770–77.

Graybill, F.A. 1996. *Theory and Application of the Linear Model.* North Scituate, MA: Duxbury Press.

Hormes, L., U.S. Environmental Protection Agency, Ann Arbor, MI. 2000. Personal communication.

Kelly, K.J., B.K. Bailey, T.C. Coburn, W. Clark, L. Eudy, and P. Lissiuk. 1996a. *FTP Emissions Test Results from Flexible-Fuel Methanol Dodge Spirits and Ford Econoline Vans,* SAE Technical Paper 961090. Warrendale, PA: SAE International.

_____. 1996b. *Round 1 Emissions Test Results from Compressed Natural Gas Vans and Gasoline Controls Operating in the U.S. Federal Fleet,* SAE Technical Paper 961091. Warrendale, PA: SAE International.

_____. 1996c. *Federal Test Procedure Emissions Test Results from Ethanol Variable-Fuel Vehicle Chevrolet Luminas,* SAE Technical Paper 961092. Warrendale, PA: SAE International.

Kirchstetter, T.W., B.C. Singer, R.A. Harvey, G.R. Kendall, and W. Chan. 1996. Impact of Oxygenated Gasoline Use in California Light Duty Vehicle Emissions. *Environmental Science and Technology* 30:661–70.

McCormick, R.L., M.S. Graboski, T. Allemean, A.M. Herring, and P. Nelson. 1999. *In-Use Emissions from Natural Gas Fueled Heavy-Duty Vehicles*, SAE Report 1999–01–1507. Warrendale, PA: SAE International.

Norbeck, J.M., T.J. Truex, M.R. Smith, and T. Durbin. 1998. *Inventory of AFVs and AFV Comparison: OEM vs. Retrofits,* Report #98-VE-RT2W-008-FR, South Coast Air Quality Management District Contract #97029. Sacramento, CA: College of Engineering–Center for Environmental Research and Technology, University of California–Riverside.

Searle, S.R. 1971. *Linear Models*. New York, NY: John Wiley & Sons Publishing.

Whalen, P., K. Kelly, and M. John. 1999. *Barwood CNG Cab Fleet Study: Final Results,* NREL/TP-540–26035. Golden, CO: National Renewable Energy Laboratory.