Some Issues in the Statistical Analysis of Vehicle Emissions

TOM WENZEL

BRETT C. SINGER

Environmental Energy Technologies Division

Lawrence Berkeley National Laboratory

ROBERT SLOTT

Private Consultant

ABSTRACT

This paper presents some of the issues that complicate the statistical analysis of real-world vehicle emissions and the effectiveness of emissions control programs. The following issues are discussed: 1) inter- and intra-vehicle emissions variability, 2) skewness of the distribution of emissions from in-use vehicles, 3) the difficulty of obtaining statistically representative vehicle samples, 4) the influence of repeat testing on only a subset of the vehicle fleet, and 5) differences among common test methods and pollutant measurement devices. The relevance of these issues is discussed in light of three regulatory purposes: testing the compliance of in-use vehicles with certification standards, evaluating the effectiveness of vehicle inspection and maintenance programs, and estimating emissions inventories for air quality modeling and compliance planning. A brief history and description of common vehicle emissions tests is also provided.

INTRODUCTION

In recent years, emphasis on the measurement of vehicle emissions has shifted from laboratory testing towards the analysis of "real-world" emissions. The term "real-world" is used to differentiate

Tom Wenzel, Mail Stop 90–4000, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720. Email: TPWenzel@lbl.gov.

between the carefully controlled and limited environmental conditions and driving patterns associated with laboratory testing and those encountered on road. The "real-world" vehicle fleet is composed of new and aging vehicles with widely varying maintenance and operational histories and includes unregistered and out-of-state vehicles. By contrast, laboratory testing is often performed solely on new or well maintained vehicles that represent only a portion of the on-road fleet. Since the ultimate goal of vehicle emissions control devices and programs is to improve ambient air quality, analyses of program and technology effectiveness should focus as much as possible on real-world emissions reductions. Likewise, motor vehicle emissions inventories developed for air quality modeling and planning should accurately represent real-world fleets and conditions. This paper describes five major statistical issues that complicate the development of real-world vehicle emissions inventories, the evaluation of emissions control program effectiveness, and the process by which manufacturers certify that their vehicles are in compliance with emissions standards. Examples are given of how each of these statistical issues can complicate the analysis of emissions data. The presentation begins with a summary of the primary method used to measure vehicle emissions, the Federal Test Procedure, and alternative measurement techniques that have been developed in the last two decades.

MEASUREMENT TECHNIQUES

Several different techniques have been developed to measure vehicle emissions. Each of these techniques has strengths and weaknesses which should be considered when analyzing emissions measurements.

Federal Test Procedure

The first large-scale sampling of vehicle emissions was for the purpose of certifying manufacturer compliance with new-car emissions standards prescribed in the Clean Air Act Amendments (CAAA) of 1970. The U.S. Environmental Protection Agency (EPA) established an elaborate testing protocol, called the Federal Test Procedure (FTP), so that all vehicles could be tested under identical preparation and driving conditions. The FTP begins with overnight storage of the vehicle at a prescribed temperature in order to ensure that the engine and catalytic converter begin the test at this temperature. The vehicle is then rolled onto a treadmill-like device called a "dynamometer," where the vehicle is driven through a standard 30-minute speed/time trace, or "driving cycle." The FTP was designed in the early 1970s to simulate combined highway and city driving in urban Los Angeles. A top speed of only 57 mph and a top acceleration of only 3.3 mph per second were set to accommodate limitations of the dynamometers available when the test was developed. Tailpipe exhaust is mixed with a specified amount of dilution air and collected in large bags over three distinct portions of the driving cycle. The first bag captures the initial "cold start." "Hot stabilized" operation is captured in the second bag, and emissions following a "warm start" are measured in the third bag. Gas analyzers measure the concentrations of hydrocarbons (HC), carbon monoxide (CO), oxides of nitrogen (NO_x), and carbon dioxide (CO_2) in each bag. Concentration units relate the amount of each pollutant to the amount of total air collected (e.g., in percent or part per million (ppm) units). Mass emissions during each portion of the driving schedule are calculated as the product of the molecular mass and measured concentration of each pollutant and the total volume of air collected. Mass emissions are then related to the simulated distance traveled to yield gram per mile (gpm) emissions factors for each bag. The bag gpm emissions are then averaged together, weighted by the relative amount of driving under each section of the cycle, to achieve a composite gpm exhaust emissions rate. The FTP includes measurement of fuel evaporation during the driving cycle (running losses), for a short period after driving ceases (hot soak), and as the vehicle sits in an enclosed chamber during a multi-hour temperature cycle (diurnal).

Idle Testing

An idle emissions test measures pollutant concentrations in the tailpipe exhaust of a stationary vehicle. The test was proposed in the 1970 CAAA as a quick and inexpensive means to identify in-use vehicles with irregularly high emissions. Unlike the

FTP, idle testing includes no transient vehicle operation and no engine load. Idle testing is not used for NO_x emissions testing since NO_x emissions are always low during idle. HC and CO emissions during idle also may not be representative of emissions when a vehicle is driven under load. The 1977 CAAA required that all urban areas with poor air quality use idle testing in vehicle inspection and maintenance (I/M) programs. The first I/M programs used tailpipe probes to measure the concentrations of HC, CO, and CO_2 in the exhaust of idling vehicles. An enhancement of the basic idle test involves putting the car in neutral and revving the engine to 2500 rpm in an attempt to simulate the vehicle's emissions under loaded conditions.

IM240

The IM240 test uses 240 seconds of the FTP driving schedule to measure hot stabilized emissions during transient and loaded mode vehicle operation. It is the centerpiece of guidelines developed by the EPA to meet the Enhanced I/M program mandate of the 1990 CAAA. Enhanced I/M was designed to address several shortcomings of original I/M programs by 1) measuring emissions, including NOx, during loaded mode vehicle operation and 2) separating vehicle testing from vehicle repair by requiring a centralized network of contractor-run test-only facilities. Although desired for Enhanced I/M, no practical tests are available to measure evaporative HC emissions in an I/M setting. In the IM240 test, exhaust emissions are run directly through gas analyzers and can be quantified on a test-composite or a second-by-second basis. It was envisioned that the capability of analyzing second-by-second emissions would assist mechanics in properly diagnosing and repairing malfunctions of the emissions control system. Of the alternative emissions measurement techniques, the IM240 most closely resembles FTP testing. However, it is also the most time-consuming and costly test.

Acceleration Simulation Mode (ASM)

Many states resisted the use of centralized IM240 testing, citing the length of the test and the inconvenience to motorists of driving further to a small number of centralized test stations.[1] The California Bureau of Automotive Repair (BAR) developed an alternative test method to the IM240 called the Acceleration Simulation Mode (ASM) test. During an ASM test the vehicle is placed on a dynamometer and run at one or more distinct operating modes. These modes are defined as a certain vehicle load at a given speed; for instance, the California program gives each vehicle a 2525, 25% of the maximum vehicle load encountered on the FTP at 25 miles per hour, and a 5015, 50% of the maximum vehicle load encountered on the FTP at 15 miles per hour, ASM test. Emissions are measured in exhaust concentration using a tailpipe probe, just as in the idle test. The ASM test can be considered an improvement over the idle test in that emissions are measured when a vehicle is under load. However, the ASM does not measure emissions under varying loads and speeds, as does the IM240. In addition, NO_x emissions, which are not measured during idle testing, are measured under the ASM test. Eventually, EPA relaxed its requirement of centralized IM240 testing and allowed states to use alternative test methods such as the ASM if they could demonstrate that their alternative method would achieve the same reduction in emissions as the IM240.

Remote Sensing

In the late 1980s, researchers at the University of Denver developed a device to remotely measure the emissions of a vehicle as it is driven on the road (Bishop et al. 1989; Zhang et al. 1993). Remote sensors measure the changing intensity of a light beam directed across a roadway as the beam interacts with a passing vehicle's exhaust plume. The first generation sensors used an infrared source and a series of filters to isolate specific wavelengths that are absorbed by the CO, HC, and CO_2 in vehicle exhaust. A video camera placed alongside the remote sensor records each vehicle's license plate information, which is stored together with the emissions measurement. The license number can be

[1] Another factor was the political power wielded by the test and repair industry, which foresaw a centralized system displacing independent service stations that relied on I/M testing for a large portion of their business.

used to retrieve information about each vehicle (age, type, and perhaps mileage) from registration records. Remote sensors measure pollutant ratios, such as CO/CO_2 and HC/CO_2, but cannot measure absolute concentrations because the amount of exhaust dilution is not known. However, since more than 99% of fuel carbon atoms are emitted as CO, HC, or CO_2, the emissions ratios can be combined with known fuel properties (e.g., fuel carbon content) to calculate the mass of each pollutant emitted per gallon of fuel burned (Bishop et al. 1989; Zhang et al. 1993). Fuel-normalized emissions factors can be calculated for any emissions test, including the FTP, IM240, and ASM, as long as measurements of both CO and CO_2 are available. In recent years, remote sensors have been developed for the measurement of on-road emissions of NO and individual hydrocarbons or other emissions gases such as ammonia (Zhang et al. 1996; Jimenez et al. 1999b; Popp et al. 1997).

A single remote sensing instrument can measure emissions of thousands of vehicles per day for a fraction of the cost of conducting a similar number of idle, ASM, IM240, or FTP tests. In addition, the testing is unscheduled, so with an appropriately designed monitoring program actual on-road emissions can be measured from a large fraction of vehicles regularly in use without drivers taking steps prior to testing that would lower their vehicle's emissions. Remote sensors thus provide valuable data for estimating actual on-road emissions. Fuel-normalized emissions factors have been measured for tens of thousands of vehicles throughout the Los Angeles area. These factors have been combined with fuel sales data to estimate total exhaust emissions of the on-road vehicle fleet (Singer and Harley 1996; Singer and Harley 2000). However, there are limitations to remote sensing. The instrument accurately measures the emissions of a given vehicle as it is being driven for a fraction of a second only, and, therefore, overall emissions for the measured vehicle may differ considerably from those measured by one remote measurement. As a result, a single remote sensing measurement should not be regarded as indicative of typical emissions for any individual vehicle. In general, remote sensing is most valuable at providing data on fleet-average emissions or typical emissions from a certain vehicle model or type. Repeat measurements of individual vehicles can be used to identify high- or low-emitting vehicles.

One concern about the use of remote emissions data is that the vehicle driving condition (or load) at the time of measurement is unknown. To address this issue, remote sensors have been sited to measure emissions from vehicles under a known driving condition, often while driving uphill under moderate load. It is also becoming commonplace to measure roadway grade, along with vehicle speed and acceleration, at the time of each remote emissions measurement. The driving mode can be estimated by a calculation of the physical load encountered by the vehicle as a result of aerodynamic drag, tire rolling resistance, inertial and gravitational acceleration forces, and engine friction (Ross 1994; Jimenez et al. 1999a; Singer 1998). To address concerns about measuring emissions during cold start driving, remote sensors are sometimes located on highway off-ramps or on surface thoroughfares that cannot be accessed directly from residential streets.

On-Board Diagnostics

A new technology that has the potential to contribute important information about vehicle emissions is the on-board diagnostic (OBD) computer system required on all new cars sold after 1995. These systems were designed by manufacturers in response to regulations by the California Air Resources Board (CARB) and EPA. The OBD system is designed to monitor over 50 parameters of vehicle and engine operation. If the on-board computer detects malfunctions or operations that would lead to tailpipe emissions greater than 1.5 times the certification standard, the system stores a "fault" code in the computer and turns on a "malfunction indicator light" (MIL) on the dashboard to alert the driver. The intent of the OBD regulations is twofold: to encourage drivers to bring their vehicles in for inspection and repair as soon as problems are detected and to record engine parameters to assist mechanics in diagnosing and repairing malfunctions. The regulations have had additional beneficial results. The OBD systems have encouraged manufacturers to design better and more durable engine and emissions controls, including more extensive monitoring and backup systems.[2] In addition, OBD systems are identifying manufacturing flaws on individual vehicles before they leave the plant. In the next few years, EPA is expected to require that all states operating Enhanced I/M programs fail vehicles with illuminated MILs. CARB anticipates that OBD will eventually replace the periodic emissions testing in conventional I/M programs. A drawback to OBD systems is that they do not measure tailpipe emissions directly; rather, they predict when emissions are likely to exceed standards, based on extensive monitoring of engine and emissions control parameters. Therefore, the usefulness of OBD data is currently limited to determining failure rates of the vehicle fleet. However, there is some discussion about eventually requiring later generations of OBD systems to directly measure tailpipe emissions.

STATISTICAL ISSUES

This section describes and discusses five major statistical issues that complicate the analysis of in-use vehicle emissions.

Inherent Variability in Vehicle Emissions

Real-world vehicle emissions are highly variable. Emission variability from vehicle to vehicle spans several orders of magnitude, while the emissions of most vehicles will vary substantially with environmental and driving conditions. Emissions of some vehicles are unrepeatable: different emissions occur from one test to another, even when test conditions are carefully controlled.

Vehicle-emission variability is a consequence of the way emissions are generated and how they are controlled. Exhaust emissions are formed in the engine as a result of unburned fuel, HC, and partially burned fuel, HC and CO, and from undesirable side reactions, NO_x. Emissions control systems are designed to reduce pollutant formation in the engine and to chemically convert engine-out pollutants to less harmful products in the catalytic converter. When functioning properly, modern vehicle-emissions controls reduce tailpipe emissions levels to five percent or less of those observed from pre-control vehicles produced in the late 1960s. However, if the engine or the emissions control system fails to operate as designed, exhaust emissions may rise by orders of magnitude.

■ ■ ■

There are numerous factors affecting the variability in emissions across different vehicles. Several of these are discussed below.

Vehicle Technology. The increasingly stringent new-car emissions standards specified in the CAA Amendments of 1970, 1977, and 1990 have been met primarily through technological improvements. Emission control technologies incorporated into vehicles over the past 30 years included the use of exhaust gas recirculation to reduce NO_x formation in the engine, the addition of catalytic converters for exhaust gas treatment, the replacement of carburetors with throttle-body and port fuel injection, and computer control of air-fuel mixing and spark timing. In most cases, these and other vehicle-emission control improvements have been introduced to the entire new car fleet over just a few model years. Real-world emissions are sensitive to vehicle technology independent of vehicle age.

Vehicle Age and Mileage Accumulation. As vehicles age and accumulate mileage, their emissions tend to increase. This is both a function of the normal degradation of emissions controls of properly functioning vehicles, resulting in moderate emissions increases, and malfunction or outright failure of emissions controls on some vehicles, possibly resulting in very large increases in emissions, particularly CO and HC.

Vehicle Model. Some vehicle models are simply designed and manufactured better than others. Some vehicle models and engine families are observed to have very low average emissions while others exhibit very high rates of emissions control failure (Wenzel 1997). The design of a particular emissions control system affects both the initial

[2] Manufacturers wish to prevent the MILs from turning on since they want to maintain customer satisfaction and are responsible for the cost of repairing emissions malfunctions in new cars under warranty.

effectiveness and the lifetime durability of the system, which in turn contributes to a model-specific emissions rate.

Maintenance and Tampering. The degree to which owners maintain their vehicles by providing tune-ups and servicing according to manufacturer schedules can affect the likelihood of engine or emissions control system failure and therefore tailpipe emissions. Outright tampering with vehicles, such as removing fuel tank inlet restrictors to permit fueling with leaded fuel that will degrade the catalytic converter or tuning engines to improve performance, can have a large impact on emissions. Early I/M programs relied on visual inspection to discourage tampering. The advent of sophisticated on-board computers and sensors has greatly reduced the incentive to improve vehicle performance through tampering. In fact, tampering with the sophisticated electronics installed on today's vehicles will likely reduce performance as well as increase emissions. Requirements for extended manufacturer warranties have led to vehicle designs that are less sensitive to maintenance, at least within the warranty period. Nonetheless, there is evidence that maintenance can still affect real-world emissions from new vehicles, at least on some models (Wenzel 1997). Improper maintenance or repair can also lead to higher emissions.

Misuse. The cumulative effect of hard driving, or "misuse," of a vehicle can also increase emissions. For example, prolonged high power driving, such as repeated towing of a trailer up mountain grades, leading to high engine temperatures can cause premature damage to the catalytic converter, resulting in dramatic increases in emissions.

Type of Malfunction. There are many emissions control components that can malfunction or fail. Some of these malfunctions are interrelated; for instance, the onboard computer of a vehicle with a failed oxygen sensor may command a constant fuel enrichment, which can eventually lead to catalyst failure. Different component malfunctions result in very different emissions consequences. In general, malfunctioning vehicles with high CO emissions tend also to have high HC emissions, while vehicles with high NO_x emissions tend to have relatively low CO and HC emissions (Wenzel and Ross 1998).

Socioeconomics. Correlations have been observed between average vehicle emissions and socioeconomic indicators, such as the median household income in the zip code where vehicles are registered (see Singer and Harley forthcoming). This relationship results in part because the vehicle fleet is older in lower income areas. However, even after accounting for vehicle age, average emissions are higher in lower income areas than higher income areas. Even vehicles of the same age and engine family exhibit different failure rates and average emissions when tested at I/M stations located in lower vs. higher income areas (Wenzel 1997). There are three possible explanations for this phenomenon: 1) individual vehicles that have been poorly manufactured (i.e. perform poorly or frequently require repairs) are selectively sold by higher income individuals and eventually wind up in lower income areas, 2) less money is spent on maintenance and repairs in lower income areas, and 3) vehicles with higher mileage are more likely to "migrate" to lower income owners.

■ ■ ■

Different factors account for the variability in an individual vehicle's emissions.

Intermittent Emissions Control Failure. While some emissions control failures, such as a completely degraded catalyst, can lead to high emissions during all vehicle operation, other failures can be intermittent. For example, a vehicle with a partially degraded catalyst may have lower emissions under higher loads because the catalyst may be effective only at very high temperatures. Oxygen sensor, fuel delivery system, and computer malfunctions can also be intermittent. Intermittent control system malfunction can cause large changes in emissions from test to test, even when all of the factors listed below are held constant. This results in uncertainty in the average emissions from such a vehicle.

Driving Mode/Engine Load. Vehicle emissions can vary greatly with changing engine load. The relationship between emissions and load depends on the fuel-delivery and emissions-control technology, but as a general rule NO_x emissions almost always increase with increasing load. Under high speed and acceleration requirements, today's vehicles are designed to have excess fuel injected into the engine cylinder. This "enrichment" of the air/fuel mixture leads to elevated CO and HC formation during combustion, with no oxygen available for pollutant conversion to CO_2 and water in the catalyst. The result is a temporary "puff" of high tailpipe CO and HC emissions (Goodwin and Ross 1996). In some vehicles, fuel injection is cut off during rapid decelerations. This can lead to cylinder misfire and a temporary "puff" of high HC emissions (An et al. 1997). Roadway grade and accessory use, such as air conditioning and heaters, put additional loads on the engine and can affect emissions. Small changes in how a vehicle is driven can also affect emissions. For instance, how a driver shifts gears on a vehicle with a manual transmission or how smoothly a driver depresses and releases the accelerator, may affect emissions rates (Shih et al. 1997).

Engine and Catalyst Temperature. When a vehicle is initially started after more than a few minutes of nonoperation, emissions are temporarily high because both the catalytic converter and oxygen sensor are ineffective at low temperatures. Heated by vehicle exhaust, the devices reach the high temperatures required for their operation after one to four minutes of driving. The temporary control system ineffectiveness at start-up is exacerbated by higher pollutant formation in "cold" engines and commanded fuel enrichment designed to facilitate ignition. The magnitude of cold start emissions depends on the time since the vehicle was last operated, ambient temperature, and the operation of the vehicle after starting.

Ambient Temperature and Humidity. Ambient temperature has a large direct effect on evaporative HC emissions. Very low ambient temperatures (e.g., below 20 degrees Fahrenheit) can influence emissions at ignition and cause the catalysts of some vehicles to cool during short stops. Very high ambient temperatures can have a secondary influence on exhaust emissions because engine load is increased by air conditioner use. Effects can include higher NO_x and an increase in the frequency of commanded enrichment. The amount of water vapor in air can affect NO_x emissions in older and malfunctioning vehicles, but it appears to have less effect on new vehicles with computer engine control.

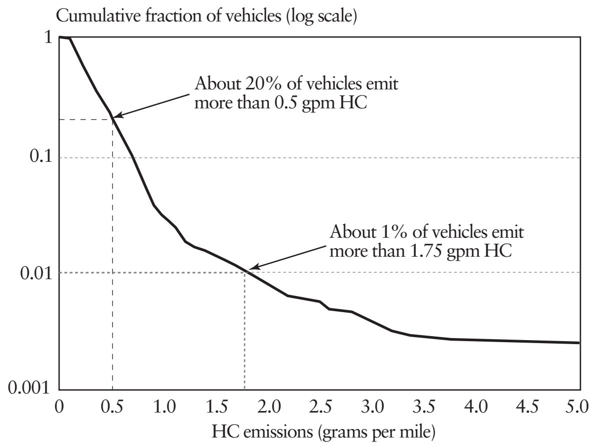
Fuel Quality. Fuel composition can have a substantial impact on vehicle tailpipe and evaporative emissions. Regulations may require changes in fuel composition by season within a region as a strategy to reduce emissions. For instance, some urban areas introduce oxygenates in fuel to reduce CO emissions in the winter and decrease the volatility to reduce evaporative HC emissions in the summer. Fuel composition can vary spatially since some regions in the country have been required or have chosen to adopt year-round reformulated gasoline standards as an emissions-control strategy.

■ ■ ■

The FTP calls for careful control of fuel and the conditions under which vehicles are tested to control for each of these factors (some factors are more tightly controlled than others). Even under these carefully controlled conditions, vehicle emissions can be quite variable (Bishop et al. 1996). A study of repeat FTP tests on the same vehicles found that CO and HC emissions from malfunctioning vehicles can change by over a factor of seven on independent FTP tests although the uncertainty is much less for properly functioning vehicles (Knepper et al. 1993). The emissions of vehicles exhibiting high uncertainty are difficult to characterize for regulatory and modeling purposes.

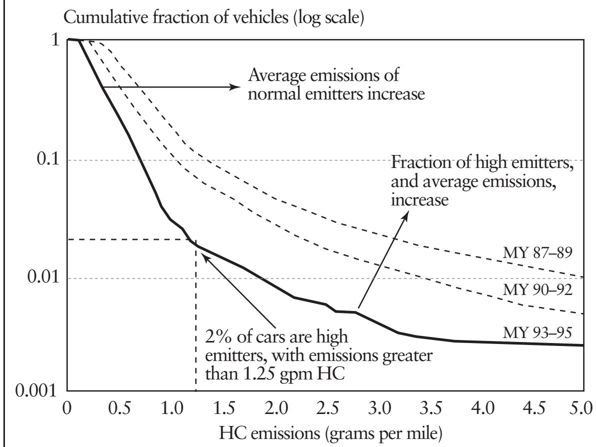
Most of the factors affecting variability and uncertainty in vehicle emissions are widely recognized. However, the degree to which some of these factors affect emissions has not been adequately quantified.

FIGURE 1 An Example of the Distribution of Vehicle HC Emissions



HC emissions from MY 1993–95 fuel-injected cars with 100,000 to 150,000 miles, from 1996–98 Wisconsin IM240 tests (fast pass emissions adjusted to full IM240 emissions).

FIGURE 2 Trends in HC Emissions Distributions as Vehicles Age



HC emissions from MY 1993–95 fuel-injected cars with 100,000 to 150,000 miles, from 1996–98 Wisconsin IM240 tests (fast pass emissions adjusted to full IM240 emissions).

Distributional Assumptions about Emissions Data

The distributions of emissions from large numbers of vehicles are highly skewed. The majority of vehicles have relatively low emissions, while a relatively small number of malfunctioning vehicles have extremely high emissions (Lawson et al. 1990; Stephens 1994; Bishop et al. 1996; Barth et al. 1999; Schwartz 2000). To overcome this difficulty, analysts have typically used the forms of the log-normal (Stephens 1994) and gamma (Zhang et al. 1994) distributions to model vehicle emissions data.

One graphical tool for analyzing this kind of data is to plot emissions as a function of the cumulative fraction of vehicles, as shown in figure 1. The figure shows the fraction of vehicles, on the y-axis, with emissions above a given level on the x-axis. For example, in figure 1, about 20% of the vehicles have HC emissions greater than about 0.5 grams per mile (gpm), while 1% of the vehicles have HC emissions greater than 1.75 gpm. With degradation of emissions controls, the average emissions of normal emitters increase, as shown in figure 2; that is, the upper/left part of the distribution shifts to the right. An increase in the fraction of high emitters, as well as an increase in the average emissions of high emitters, causes the lower/right segment of the distribution to shift upward and become flatter. The change in the shape of the distribution, shown in figure 2 approximately at 1.25 gpm for MY 93–95, can be taken as a cut point for dividing vehicles of the same age and model year into "normal emitters" and "high emitters." The shape of the emissions distribution may vary by pollutant, vehicle type, vehicle age, and so forth.

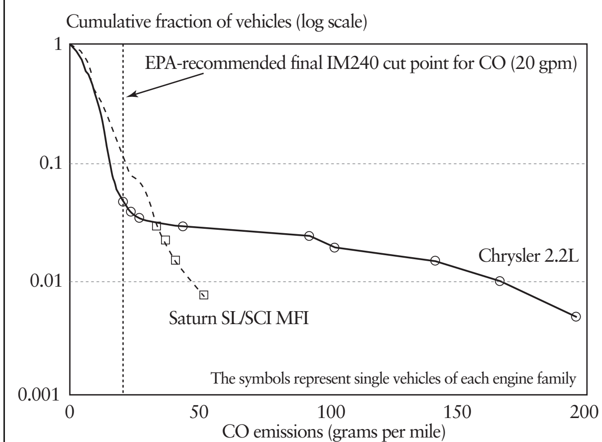
Since, in many cases, vehicle emissions approximately follow a log-normal or gamma distribution, confidence intervals on the mean emissions level are not symmetric. Also, statistical tests, such as t-tests, which depend on normality cannot be used to determine whether the difference in mean emissions from two groups of vehicles is statistically significant unless sample sizes are large. Further, the emissions of different pollutants, or different samples of vehicles, may not necessarily follow the same type of distribution.

As previously suggested, the logarithmic transformation is frequently used to account for the non-normality of the data; yet this may not be the appropriate approach to take. For example, emissions inventory models developed by EPA and CARB multiply estimates of the mean emissions of a group of vehicles by estimates of activity, such as miles driven and number of starts, of that group of vehicles. However, if the mean emissions are calculated based on the logarithmic transformation, then the emissions of any high emitting vehicles in the sample are given much less weight in the estimated mean emissions level, and the models tend to underestimate fleet emissions (Pollack et al. 1999b).

Other approaches to the problem of non-normality have been taken, with varying degrees of success. One way to construct an approximately normal distribution is to consider a collection of average values representing fairly large, unbiased subsets of emissions measurements. Stedman et al. (1997) demonstrated the usefulness of this method in the context of remote sensing measurements taken over a five-day period. First, the average emissions measured by remote sensing for each day were calculated. On the basis of the well-known Central Limit Theorem, the five averages should be approximately normally distributed if the samples measured over each of the five days were unbiased and sufficiently large. The five averages were then averaged to obtain an estimate of fleet-average emissions about which a symmetric confidence interval could be constructed. Normal statistical tests, such as the t-test, were then applied (Stedman et al. 1997). Some researchers are beginning to use nonparametric techniques, such as bootstrap sampling (Pollack et al. 1999a; Frey et al. 1999), since such techniques do not require an assumption regarding the distribution of the underlying population.

Although the skewed nature of vehicle emissions distributions is generally acknowledged, proper statistical tools are not always used to characterize the uncertainty associated with mean emissions levels.

FIGURE 3 CO Emissions Distributions of Two Car Models



142 Saturn CL/SCI cars with multiport fuel injection and 217 Chrysler cars with 2.2-liter engine. From 1996–97 Arizona IM240 tests (fast pass emissions adjusted to full IM240 emissions).

Representativeness of Test Vehicles

The skewed nature of vehicle emissions also has important implications for drawing a representative sample of vehicles from a population for testing. Because vehicle emissions vary by the factors discussed above (vehicle age, technology, make and model, owner socioeconomic characteristics, etc.), a representative sample of vehicles would account for all of these factors. There are two issues to consider when seeking a vehicle sample that is representative of the in-use fleet: the number of vehicles and selection/response bias.

Number of Vehicles

Because there are relatively few high emitters in the population, the sample needs to be large enough so that a number of high emitters is included. As noted above, the number of vehicles required depends on the constituent of interest and the shape of its distribution, as well as the statistical hypothesis to be tested.

The issue of inadequate sample size is demonstrated in EPA's in-use compliance program, which attempts to identify vehicle engine families that have high average in-use emissions for recall and repair by the manufacturer. Under this program, a very small sample (not more than a dozen) of three-to five-year-old in-use vehicles is recruited and its emissions tested under FTP conditions.[3] An engine family with average emissions in excess of the new-car certification standards may be subject to an emissions-related recall.

One limitation of the program is that not enough vehicles of a particular engine family are tested to identify models with a small number of extremely high emitters. Figure 3 demonstrates this situation, using cumulative vehicle distributions of CO emissions for two model-year 1991 engine families, the Saturn SL/SCI with multi-port fuel injection and the Chrysler 2.2 liter engine. The figure shows the emissions distribution of at least 100 individual vehicles from each of these engine families, tested on the IM240 in the Arizona I/M program. Both the Saturn and the Chrysler engine families have the same average CO emissions, 12 grams per mile. However, the figure indicates that the Saturns have relatively high emissions across all

[3] To satisfy legal requirements, vehicles must be "well-maintained and used." This requirement introduces a sampling bias into the test program because manufacturers consult the service history of individual vehicles, using data supplied by their dealers, before agreeing to include vehicles in the test sample.

vehicles tested but no extremely high emitters. The Chryslers, in contrast, have relatively low emissions on the low end of the distribution but several vehicles with extremely high emissions. The Saturn engine family shown in figure 3 failed EPA in-use compliance testing for CO and was recalled, whereas the Chrysler engine family did not. However, the potential reduction in emissions from repairing the high-emitting Chryslers is greater than that of repairing the Saturns. The design of the in-use compliance program identifies for recall engine families with marginally high emissions across all vehicles, rather than engine families with a small number of vehicles with extremely high emissions.

Selection/Response Bias

For detailed FTP testing, agencies usually recruit vehicles by mailing solicitations to potential participants. Participation in such testing programs typically is voluntary although incentives are frequently provided to encourage participation. Both CARB and EPA primarily use mail solicitation to obtain vehicles for data to feed their emissions inventory models. The output from these models is input into regional air quality models to forecast the effect of emissions control programs on future air quality. One large source of uncertainty in the vehicle emissions inventory models is the potential for selection bias in voluntary vehicle recruitment.[4]

The makeup of the vehicle sample is likely affected by the perceived rewards and penalties for participating in the study. Rewards typically include cash, use of a rental vehicle, and, sometimes, repairs to the vehicle. Penalties may include inconvenience, risk of damage to the vehicle, and possible future requirements to repair the vehicle at the owner's expense. It can be argued that voluntary recruitment programs where high-emitting vehicles are repaired free of charge may attract a disproportionate number of dirty vehicles. Voluntary recruitment programs using mail solicitation typically achieve a response rate of only 10 to 15% (CARB 1997). In fact, one study where the registration of vehicles not responding to mail solicitations was to be suspended still achieved a response rate of only 60% (CARB 1996).

The recruitment of vehicles with high emissions is particularly difficult. Many recruited vehicles likely to have high emissions cannot be tested because the condition of the vehicle (e.g., bald tires or fuel, oil, coolant, or exhaust leaks) would threaten the safety of the technicians performing the test. The degree to which testing programs change the condition of the vehicle prior to testing may also affect the emissions test results. For example, a long list of restorative maintenance procedures (such as replacing spark plugs, air filters, mufflers, distributor caps and rotors, and adjusting ignition timing) is performed on cars to be tested for compliance with in-use emissions standards (CARB 1994). In contrast, very little of this restorative maintenance is performed on vehicles recruited for "as received" emissions testing.

EPA, CARB, and others acknowledge the possibility of selection bias in voluntary vehicle procurement programs, but few studies have been conducted to estimate its effect. An analysis of remote sensing readings of emissions from vehicles whose drivers were asked to participate in a roadside pullover program conducted in California in 1994 found that vehicles whose owners refused to participate had, on average, 2.5 times higher on-road emissions than those of owners who did agree to participate (Stedman et al. 1994). However, a similar experiment conducted as part of a more recent roadside pullover testing program found that vehicles whose owners declined to participate had the same average remote sensing emissions by model year as those whose owners agreed to participate (Wenzel et al. 2000).

One method of estimating the selection bias in vehicle recruitment via mail solicitation is to conduct a formal experiment where the emissions of vehicles whose owners volunteer to participate are compared with emissions of vehicles that are unavailable or whose owners decline to participate. The emissions for this comparison would come from each vehicle's next (or last) regularly scheduled I/M test. The experiment would need to be conducted in a state that has a loaded mode,

---

[4] The National Academy of Sciences recently published a review of EPA's emissions inventory model summarizing this source of bias and other limitations of the model (NAS 2000).

centralized I/M program, such as Arizona, Colorado, Illinois, or Wisconsin. For example, if the expected acceptance rate of a voluntary mail solicitation test program is 10% and 100 vehicles are desired, 1,000 invitations would normally be mailed. To determine the effect of selection bias, one could instead mail 10,000 invitations. Owners of about 9,000 vehicles would not be available or would decline to participate; 1,000 owners would agree to participate, 100 of whose vehicles would actually be brought in for FTP testing. I/M emissions from the 900 vehicles that were volunteered but not chosen would be compared with a randomly selected subset of the 9,000 vehicles that were not volunteered by their owners for testing.

Other methods can be used to obtain emissions measurements of relatively unbiased samples of vehicles. As discussed above, CARB has conducted roadside testing of vehicles randomly pulled over by law enforcement officers. Vehicle I/M programs are designed to measure the emissions of virtually all vehicles registered in an urban area. Finally, remote sensing instrumentation allows the unscheduled testing of nearly all vehicles that drive by the sensors. However, none of these methods is entirely free from sample bias. For legal reasons, roadside testing in California must be voluntary. Roadside and remote sensing studies only measure emissions of vehicles that happen to drive by the measurement sites, as noted previously, and each method has siting limitations. Finally, not all registered vehicles report for I/M testing.[5] In addition, these alternative methods for measuring vehicle emissions have other drawbacks, as discussed earlier. Researchers should consider the advantages and disadvantages of each measurement method when designing an emissions collection program.

[5] Older and newer vehicles often are exempted from I/M testing, and many eligible vehicles do not report for testing. For example, up to 26% of the vehicles in the Phoenix I/M program that failed their initial I/M test between January 1996 and July 1997 did not receive a final passing test within 3 to 15 months of their initial test. Of these vehicles, about one-third were still driving in the I/M area more than two years after their initial test (Wenzel 1999).

The Effect of Repeated Testing

In I/M programs, vehicles that fail initial testing are supposed to be repaired and then retested until they pass the test. Vehicles are therefore characterized by the results of their initial test: vehicles with low emissions are passed and not retested, while vehicles with high emissions are retested, presumably after repairs, until they pass. However, as discussed above, emission levels of many vehicles, particularly those with intermittent malfunctions, can vary substantially from one test to the next, and, consequently, the average emissions of the fleet of failing vehicles may be lower in subsequent testing, even without any repairs, solely due to emissions variability. Likewise, the average emissions of the fleet of passing vehicles may be higher in subsequent testing due to emissions variability. For the same reason, the average emissions of the fleet of vehicles that failed their initial test will be higher if they were retested after their final passing test. Evaluations of I/M programs that do not account for the variability observed in repeated testing of only a portion of the vehicle fleet may overstate the emissions benefits of these programs.

Little effort has been made to determine what the effect of repeated testing has on the estimated effectiveness of an I/M program. The primary reason is insufficient data: it is not often the case that large numbers of vehicles are repeatedly tested under identical conditions. The easiest way to remedy this situation, of course, would be to conduct an experiment using multiple tests on the same vehicles under identical conditions. Many states have the capability of conducting full IM240 tests on a random sample of the vehicles that report for testing, and this capability allows states to collect the type of data needed. For example, Heirigs and Gordon (1996) report the results of an experimental program in Arizona in which back-to-back IM240 tests were conducted on a sample of vehicles that failed their initial I/M test after waiting at least 15 minutes for a test lane to open.

Testing Methodologies

The methodology used to measure vehicle emissions should be taken into account when analyzing emissions measurements. Although the type of test method used is not strictly a statistical issue, it is important to consider when analyzing emissions data, particularly when comparing measurements made with different instrumentation or under different methods.

Units of Measurement

Tailpipe emissions can be reported as exhaust concentrations (e.g., percent or ppm), normalized to the amount of fuel used (e.g., grams per gallon), or normalized to the distance traveled (e.g., grams per mile). The relationship between exhaust concentrations and fuel-normalized emissions factors is approximately linear except for extremely high CO and HC emitters (Singer 1998). In contrast, relating fuel-normalized to mileage-normalized emissions factors requires knowledge or assumptions about fuel efficiency. This issue is relevant when attempting to use non-FTP methodologies to evaluate in-use compliance with new car emissions standards expressed in grams per mile because driving mode directly affects fuel efficiency. At the extreme, mileage-normalized emissions are infinite under idle conditions when the vehicle is not moving; therefore, mileage-normalized emissions can be very high for driving cycles that include significant amounts of idle time. Likewise, the fuel economy measured during one set of driving conditions (e.g., the FTP) may differ from fuel economy under a different set of conditions (e.g., the fixed load conditions of ASM testing). Additional uncertainty thus results when using EPA-reported fuel efficiency values to convert remote sensing, ASM, or idle test results (in concentration or grams per gallon) to grams per mile.

Testing Methodology

The various test methodologies described earlier measure emissions during different vehicle driving modes and potentially widely varying environmental conditions. Even the IM240 and FTP, both dynamometer-based methodologies that include controlled, transient vehicle operation, involve different combinations of engine loads. Also, unlike the FTP, environmental conditions and vehicle preparation are generally not predefined or controlled for IM240 tests. In addition, since the purpose of I/M programs is to merely identify, and eventually repair, high-emitting vehicles and not necessarily to accurately measure every vehicle's emissions, EPA allows the IM240 test to be varied for exceptionally clean or dirty vehicles. For example, clean vehicles may pass their I/M test after only 30 seconds of driving, while exceptionally dirty vehicles may fail after 94 seconds of testing. Application of these "fast pass" and "fast fail" rules vary from state to state. The use of shorter test cycles complicates comparisons of fleet average emissions and emissions reductions because the driving patterns of the shortened tests differ from that of the full IM240, and the effect of uncontrolled environmental conditions and vehicle pretest conditioning are more pronounced. This suggests that great care should be taken when comparing emissions measured using different test methods, and/or under different test conditions. Some researchers have developed factors to convert emissions measured in I/M programs to projected emissions under FTP test conditions. These factors are developed by running regression models on the measured I/M and FTP emissions using a relatively small sample of vehicles tested under both test conditions (see Austin et al. 1997 and DeFries and Williamson 1997). However, such factors are only valid on a fleet-average basis, not for the emissions of individual vehicles (DeFries et al. 1999). Another approach is to compare instantaneous emissions measured during a specified engine load (Jimenez 1999a; McClintock 1999), which would allow remote sensing measurements, for example, to be compared with FTP, IM240, and even to ASM emissions test results.

Pollutant Measurement Equipment

The same basic physical principles and analytical equipment are used to measure CO and CO_2 concentrations in the FTP bags, from the tailpipe probes used in idle and IM240 I/M testing, and by roadside remote sensing. Thus, while the uncertainty of any CO measurement obtained by remote sensing may be higher than that of an FTP bag measurement (due to a lower signal and more interference in the remote measurement), results of the two tests may still be directly compared, all other factors being equal. This is not the case for HC and NOx. For HC, there is an important difference between the infrared (IR) technique used by remote sensors and for tailpipe I/M testing and the flame ionization detector (FID) systems used during FTP testing. FID systems essentially count carbon atoms and provide equivalent results on a per-carbon basis for individual hydrocarbon compounds with different structures. Infrared HC analyzers measure infrared light absorption at a wavelength specific to the carbon-hydrogen bond structure typical of n-alkanes (compounds like propane, butane, and hexane). FIDs and IR analyzers are both typically calibrated with propane standards. However, an infrared analyzer calibrated with propane will report only a fraction of the carbon atoms from hydrocarbon compounds that have different structures than propane (e.g., benzene, toluene, and ethene, all of which are major components of HC emissions in vehicle exhaust). The relationship between IR and FID measurements of exhaust HC depends on the relative amounts of each HC compound in a vehicle's exhaust, known as HC speciation, and the particular wavelength filter used in the infrared analyzer. On a fleet-average basis, infrared analyzers used for vehicle exhaust measurement (including remote sensors) report only about 50% of the HC emissions that would be reported by a FID measurement on the same exhaust sample (Singer et al. 1998). The disparity can vary from 20 to 80% for individual vehicles, depending on the distribution of HC species in the tailpipe exhuast, a function of the driving mode and the condition of the catalyst.

CONCLUSIONS

The applicability of several different methods in measuring real-world vehicle emissions has been described. In addition, several issues that complicate the statistical analysis of real-world vehicle emissions have been presented. For example, selection bias is often apparent in the recruitment of vehicles, but emissions professionals have few means to operationally or statistically remedy the situation. In addition, the data necessary to estimate the direction or degree of any such bias are often unavailable. Consequently, to meet the challenges inherent in analyzing vehicle emissions in such a way as to effectively have an impact on public policy and environmental quality, access to much more information is needed. In addition, statistical rigor and innovative approaches will be necessary in the design of experimental programs and the analysis of resulting emissions data. Better decisionmaking will likely only result through interdisciplinary cooperation among emissions professionals, engineers, scientists and members of the statistical community.

REFERENCES

Austin, T.C., L.S. Caretto, T.R. Carlson, and P.L. Heirigs. 1997. Development of a Proposed Procedure for Determining the Equivalency of Alternative Inspection and Maintenance Programs. Report prepared for U.S. EPA Office of Mobile Sources, Regional and State Programs Division, Sierra Research. Ann Arbor, MI: USEPA.

An, F., G. Scora, M. Barth, and J. Norbeck. Characterization and Modeling of Vehicle Lean-Burn Hydrocarbon Emissions. Paper presented at the Seventh CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1997.

Barth, M., F. An, T. Younglove, C. Levine, G. Scora, M. Ross, and T. Wenzel. 1999. The Development of a Comprehensive Modal Emissions Model. Final report, NCHRP Project 25–11. Washington, DC: Transportation Research Board, National Cooperative Highway Research Program.

Bishop, G.A., J.R. Starkey, A. Ihlenfeldt, W.J. Williams, and D.H. Stedman. 1989. IR Long-Path Photometry: A Remote Sensing Tool for Automobile Emissions. *Analytical Chemistry* 61, no. 10:671–7.

Bishop, G.A., D.H. Stedman, and L. Ashbaugh. 1996. Motor Vehicle Emissions Variability. *Journal of the Air & Waste Management Association* 46: 667–75.

California Air Resources Board (CARB). 1996. Comparison of the IM240 and ASM Tests in CARB's I&M Pilot Program. El Monte, CA.

_____. 1997. Test Report of the Light-Duty Vehicle Surveillance Program, Series 13 (LDVSP13). Mobile Source Operations Division, Report Number MSD-97–01. El Monte, CA.

DeFries, T.H. and H.J. Williamson. 1997. Methodology for Estimating California Fleet FTP Emissions. Report prepared for California Bureau of Automotive Repair by Radian International, Austin, TX.

DeFries, T.H., C.F. Palacios, S.D. Kishan, and H.J. Williamson. 1999. Models for Estimating California Fleet FTP Emissions from ASM Measurements. Draft report prepared for the California Bureau of Automotive Repair by Eastern Research Group, Austin, TX.

Drachand, K.D. of California Air Resources Board (CARB). 1994. Memorandum to Tom Cackette. June 23.

Frey, H.C., J. Zheng, and A. Unal. Quantitative Analysis of Variability and Uncertainty in Highway Vehicle Emissions Factors. Paper presented at the Ninth CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1999.

Goodwin, R. and M. Ross. 1996. Off-Cycle Emissions from Modern Passenger Cars with Properly Functioning Emissions Controls. SAE Technical Paper Series, 960064. Warrendale, PA: Society of Automotive Engineers.

Heirigs, P.L. and J. Gordon. 1996. Preconditioning Effects on I/M Test Results Using IM240 and ASM Procedures. SAE Technical Paper Series, 962091. Warrendale, PA: Society of Automotive Engineers.

Jimenez, J.L., P.M. McClintock, G.J. McRae, D.D. Nelson, and M.S. Zahniser. 1999a. Vehicle Specific Power: A Useful Parameter for Remote Sensing and Emissions Studies. Paper presented at the Ninth CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1999.

Jimenez, J.L., M.D. Koplow, D.D. Nelson, M.S. Zahniser, and S.E. Scmidt. 1999b. Characterization of On-Road Vehicle NO Emissions by a TILDAS Remote Sensor. *Journal of the Air & Waste Management Association* 49:463–70.

Knepper, J.C., W.J. Koehl, J.D. Benson, V.R. Burns, R.A. Gorse Jr., A.M. Hochhauser, W.R. Leppard, L.A. Rapp, and R.M. Reuter. 1993. Fuel Effects in Auto/Oil High Emitting Vehicles. SAE Technical Paper Series, 930137. Warrendale, PA: Society of Automotive Engineers.

Lawson, D.R., P. Groblicki, D.H. Stedman, G.A. Bishop, and P.L. Guenther. 1990. Emisions of In-Use Motor Vehicle in Los Angeles: A Pilot Study of Remote Sensing and the Inspection and Maintenance Program. *Journal of the Air & Waste Management Association* 40:1096–105.

McClintock, P. 1999. Remote Sensing Measurements of Real World High Exhaust Emitters, CRC Project No. E-23-Interim Report. Tuscon, AZ: Remote Sensing Technologies.

National Academy of Sciences (NAS). 2000. *Modeling Mobile Source Emissions*. Washington, DC: National Academy Press.

Pollack, A.K., P. Bhave, J. Heiken, K. Lee, S. Shepard, C. Tran, G. Yarwood, R.F. Sawyer, and B.A. Joy. 1999a. Investigation of Emissions Factors in the California EMFAC7G Model. Report prepared for Coordinating Research Council, Project E-39.

Pollack, A., T. Stoeckenius, and C. Tran. 1999b. Memorandum to Phil Enns, EPA OMS, regarding MOBILE6 deterioration rates. 18 August.

Popp, P.J., G.A. Bishop, and D.H. Stedman. Development of a High-Speed Ultraviolet Spectrophotometer Capable of Real-Time NO and Aromatic Hydrocarbon Detection in Vehicle Exhaust. Paper presented at the Seventh CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1997.

Ross, M. 1994. Automobile Fuel Consumption and Emissions: Effects of Vehicle and Driving Conditions. *Annual Review Energy Environment* 19:75–112.

Schwartz, J. 2000. Smog Check II Evaluation: Part II: Overview of Vehicle Emissions. Final Report. Sacramento, CA: California Inspection and Maintenance Review Committee.

Shih, R., S. Fable, and R.F. Sawyer. Effects of Driving Behavior on Automotive Emissions. Paper presented at the Seventh CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1997.

Singer, B.C. 1998. A fuel-based approach to estimating motor vehicle exhaust emissions. Ph.D. diss., University of California, Berkeley, CA.

Singer, B.C. and R.A. Harley. 1996. A Fuel-Based Motor Vehicle Emissions Inventory. *Journal of the Air & Waste Management Association* 46:581–93.

Singer, B.C., R.A. Harley, D. Littlejohn, J. Ho, and T. Vo. 1998. Scaling of Infrared Remote Sensor Hydrocarbon Measurements for Motor Vehicle Emissions Inventory Calculations. Environmental Science & Technology 53, no. 21:3241–48.

Singer, B.C. and R.A. Harley. 2000. A Fuel-Based Inventory of Motor Vehicle Exhaust Emissions in the Los Angeles Area during Summer 1997. *Atmospheric Environment* 34:1783–95.

Stedman, D.H., G.A. Bishop, S.P. Beaton, J.E. Peterson, P.L. Guenther, I.F. McVey, and Y. Zhang. 1994. On-Road Remote Sensing of CO and HC Emissions in California. California Air Resources Board, Contract No. A032–093.

Stedman, D.H., G.A. Bishop, P. Aldrete, and R.S. Slott. 1997. On-Road Evaluation of an Automobile Emission Test Program. *Environmental Science & Technology* 31: 927–31.

Stephens, R.D. 1994. Remote Sensing Data and a Potential Model of Vehicle Exhaust Emissions. *Journal of the Air & Waste Management Association* 44:1284–92.

Wenzel, T. 1997. I/M Failure Rates by Vehicle Model. Paper presented at the Seventh CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1997.

Wenzel, T. 1999. Evaluation of Arizona's Enhanced I/M Program. Paper presented at the Ninth CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1999.

Wenzel, T. and M. Ross. 1998. Characterization of Recent-Model High-Emitting Automobiles. SAE Technical Paper Series, 981414. Warrendale, PA: Society of Automotive Engineers.

Wenzel, T., B. Singer, E. Gumerman, R. Sawyer, and R. Slott. 2000. Evaluation of the Enhanced Smog Check Program: A Report to the California Inspection and Maintenance Review Committee. Final Report. Berkeley, CA: LBNL.

Zhang, Y., D.H. Stedman, G.A. Bishop, P.L. Guenther, S.P. Beaton, and J.E. Peterson. 1993. On-Road Hydrocarbon Remote Sensing in the Denver Area. Environmental Science and Technology 27, no. 9:1885–91.

Zhang, Y., G.A. Bishop, and D.H. Stedman. 1994. Automobile Emissions are Statistically y-Distributed. Environmental Science and Technology 28:1370–4.

Zhang, Y., D.H. Stedman, G.A. Bishop, S.P. Beaton, P.L. Guenther, and I.F. McVey. 1996. Enhancement of Remote Sensing for Mobile Source Nitric Oxide. Journal of the Air & Waste Management Association 46:25–9.