**United States
Environmental Protection
Agency**

Guidance on Use of Remote Sensing for Evaluation of I/M Program Performance

Guidance on Use of Remote Sensing for Evaluation of I/M Program Performance

Certification and Compliance Division
Office of Transportation and Air Quality
U.S. Environmental Protection Agency

1. Introduction

This document is intended to provide guidance for performing I/M program evaluations using a Remote Sensing Device (RSD). The next section is a background regarding EPA regulation of state I/M programs and a history of methods used to evaluate these programs. Section 3 describes different approaches to evaluate I/M programs, using roadside pullover data or independent remote sensing measurements. Section 4 summarizes equipment specifications and measurement procedures, while Section 5 outlines important design parameters for the collection and analysis of RSD program evaluation data. Section 5 also discusses quality control issues that should be considered in an evaluation. Section 6 describes in detail three alternative methods to perform short-term evaluations of the I/M programs using remote sensing data and discusses the advantages and disadvantages of each. The three methods are the Step Change, the Comprehensive, and the Reference analysis methods. How in-program data can be used to evaluate the long-term, cumulative effect of I/M programs is covered in a separate document. Appendix A contains some simple trouble-shooting methods that can be applied in the field as a first check to determine if an RSD unit is functioning properly.

It is strongly recommended that any state considering the use of RSD for program evaluation purposes work closely with their respective regional EPA office and the Office of Transportation and Air Quality to ensure the most up-to-date practices are incorporated into the evaluation. Furthermore, states interested in using RSD for program evaluation must recognize the need within their own agencies to develop a minimum level of expertise with the technology and procedures to ensure reliable data are collected and analyses performed.

It should also be recognized given the difficulties associated with I/M program evaluations, that an evaluation based on both out-of-program data (e.g. RSD) and in-program data will provide a more accurate estimate of overall program performance than simply relying on one method alone.

2. Background

2.1. History of I/M

The Environmental Protection Agency (EPA) has had oversight and policy development responsibility for vehicle inspection and maintenance (I/M) programs since the passage of the Clean Air Act (CAA) in 1970 (1)*, which included I/M as an option for improving air quality. The first I/M program was implemented in New Jersey in 1974 and consisted of an annual idle test of 1968 and newer light-duty gasoline-powered vehicles conducted at a centralized facility. No tampering checks were performed and no repair waivers were allowed.

I/M was first mandated for areas with long term air quality problems beginning with the Clean Air Act Amendments of 1977 (2). EPA issued its first guidance for such programs in 1978 (3); this guidance addressed State Implementation Plan (SIP) elements such as minimum emission reduction requirements, administrative requirements, and implementation schedules. This original I/M guidance was quite broad and difficult to enforce, given EPA's lack of legal authority to establish minimum, Federal, I/M implementation requirements. This lack of regulatory authority -- and the state-to-state inconsistency with regard to I/M program design that resulted from it -- was cited in audits of EPA's oversight of the I/M requirement conducted by both the Agency's own Inspector General, as well as the General Accounting Office.

In response to the above-cited deficiencies, the 1990 Amendments to the Clean Air Act (CAAA) (4) were much more prescriptive with regard to I/M requirements while also expanding I/M's role as an attainment strategy. The CAAA required EPA to develop Federally enforceable guidance for two levels of I/M program: "basic" I/M for areas designated as moderate non-attainment, and "enhanced " I/M for serious and worse non-attainment areas, as well as for areas within an Ozone Transport Region (OTR), regardless of attainment status. This guidance was to include minimum performance standards for basic and enhanced I/M programs and was also to address a range of program implementation issues, such as network design, test procedures, oversight and enforcement requirements, waivers, funding, etc. The CAAA further mandated that enhanced I/M programs were to be: annual (unless biennial was proven to be equally effective), centralized (unless decentralized was shown to be equally effective), and enforced through registration denial (unless a pre-existing enforcement mechanism was shown to be more effective).

In response to the CAAA, EPA published its I/M rule on November 5, 1992 (5), which established the minimum procedural and administrative requirements to be met by basic and enhanced I/M programs. This rule also included a performance standard for basic I/M based upon the original New Jersey I/M program and a separate performance standard for enhanced I/M, based on the following program elements:

- Centralized, annual testing of MY 1968 and newer light-duty vehicles (LDVs) and light-duty trucks (LDTs) rated up to 8,500 pounds GVWR.

---

* References are denoted by underlined italic numerals in parentheses and are listed in Section 8.

- Tailpipe test: MY1968-1980 - idle; MY1981-1985 - two-speed idle; MY1986 and newer - IM240.

- Evaporative system test: MY1983 and newer - pressure; MY1986 and newer - purge test.

- Visual inspection: MY1984 and newer - catalyst and fuel inlet restrictor.

Note that the phrase "performance standard" used above was initially used in the CAA and is misleading in that it more accurately describes program design. Adhering to the "performance standard" does not guarantee an I/M program will meet a specific level of emissions reductions. Therefore, the performance standard is not what is required to be implemented, it is the bar against which a program is to be compared.

At the time the I/M rule was published in 1992, the enhanced I/M performance standard was projected to achieve a 28% reduction in volatile organic compounds (VOCs), a 31% reduction in carbon monoxide (CO), and a 9% reduction in oxides of nitrogen (NOx) by the year 2000 from a No-I/M fleet. The basic I/M performance standard, in turn, was projected to yield a 5% reduction in VOCs and 16% reduction in CO. These projections were made based upon computer simulations run using 1992 national default assumptions for vehicle age distributions, mileage accumulation, fuel composition, etc., and were performed using the most current emission factor model then available for mobile sources, MOBILE 4.1. That version of the MOBILE model was the first to include a roughly 50% credit discount for decentralized I/M programs, based upon EPA's experience with the high degree of improper testing found in such programs. This discount was incorporated into the 1992 rule, and served to address the CAAA's implicit requirement that EPA distinguish between the relative effectiveness of centralized versus decentralized programs.

The CAAA also required that enhanced I/M programs include the use of on-road testing and that they conduct evaluations of program effectiveness biennially (though no explicit connection was made between these two requirements). In establishing guidelines for the program evaluation requirement, the 1992 I/M rule specified that enhanced I/M programs were to perform separate, state-administered or observed IM240's on a random sample of 0.1% of the subject fleet in support of the biennial evaluation. Unfortunately, the program evaluation procedure for analyzing the 0.1% sample was never developed with sufficient detail to actually be used by the states. In defining the on-road testing requirement, the 1992 rule required that an additional 0.5% of the fleet be tested using either remote sensing devices (RSD) or road-side pullovers. Furthermore, the role that this additional testing was to play -- i.e., whether it was to be used to achieve emission reductions over and above those ordinarily achieved by the program, or whether it could be used to aid in program evaluation -- was never adequately addressed.

At the time the 1992 I/M rule was being promulgated, EPA was criticized for not considering alternatives to the IM240. California in particular argued in favor of the Acceleration Simulation Mode (ASM) test, a steady-state, dynamometer-based test developed by California, Sierra Research, and Southwest Research Institute. In fact, this test had been considered by EPA while the I/M rule was under development, but the combination of IM240, purge, and pressure testing was deemed sufficiently superior to the ASM that EPA dismissed ASM as an option for

enhanced I/M programs. Nevertheless, EPA continued to evaluate the ASM test in conjunction with the State of California and by early 1995, sufficient data had been generated to support EPA's recognizing ASM as an acceptable program element for meeting the enhanced performance standard.

In early 1995, when the ASM test was first deemed an acceptable alternative to IM240, the presumptive, 50% discount for decentralized programs was still in place. Even at that time, however, the practical importance of the discount was waning, in large part due to program flexibility introduced by EPA aimed at allowing enhanced I/M areas to use their preferred decentralized program designs. This flexibility was created by replacing the single, enhanced I/M performance standard with a total of three enhanced performance standards:

* High Enhanced: Essentially the same as the enhanced I/M performance standard originally promulgated in 1992.

* Low Enhanced: Essentially the basic I/M performance standard, but with light trucks and visual inspections added. This standard was intended to apply to those areas that could meet their other clean air requirements (i.e., 15%, post-1996 ROP, attainment) without needing all the emission reduction credit generated by a high enhanced I/M program.

* OTR Low Enhanced: Sub-basic. Intended to provide relief to those areas located inside the OTR which -- if located anywhere else in the country -- *would not have to do I/M at all*.

Despite the additional flexibility afforded enhanced I/M areas by the new standards outlined above, in November 1995 Congress passed and the President signed the National Highway Systems Designation Act (NHSDA) (6), which included a provision that allowed decentralized I/M programs to claim 100% of the SIP credit that would be allowed for an otherwise comparable centralized I/M program. These credit claims were to be based upon a "good faith estimate" of program effectiveness, and were to be substantiated with actual program data 18 months after approval. The evaluation methodology to be used for this 18-month demonstration was developed by the Environmental Counsel of States (ECOS), though the criteria used were primarily qualitative, as opposed to quantitative. As a result, the ECOS criteria developed for the 18-month NHSDA evaluations were not deemed an adequate replacement for the CAAA and I/M rule required biennial program effectiveness evaluation.

In January 1998, EPA revised the I/M rule's original provisions for program evaluation by removing the requirement that the evaluation be based on IM240 or some equivalent, mass-emission transient test (METT) and replaced this with the more flexible requirement that the program evaluation methodology simply be "sound" (7). In October 1998, EPA published a guidance memorandum that outlined what the Agency considered to be acceptable, "sound," alternative program evaluation methods (8). All the methods approved in the October 1998 guidance were based on tailpipe testing and required comparison to Arizona's enhanced I/M program as a benchmark using a methodology developed by Sierra Research under contract to EPA. Even though EPA recognized that an RSD-based program evaluation method may be possible, a court-ordered deadline of October 30, 1998 for release of the guidance prevented EPA from approving an RSD-based approach at that time.

The focus of this document is to address the concerns EPA has concerning RSD-based program evaluation methods with regard to equipment specifications, site selection, and data collection, as well as outline and explain the advantages and limitations of each RSD analysis methodology. As its operating premise, EPA recognizes that every program evaluation method will have its limitations, regardless of whether it is based upon an RSD approach or more traditional, tailpipe-based measurements. Therefore, no particular program evaluation methodology is viewed as a "golden standard." Ideally, each evaluation method would yield similar conclusions regarding program effectiveness, provided they were performed correctly. Unfortunately, it is unlikely we will see such agreement among methods in actual practice, due to the likelihood that different evaluation procedures will be biased toward different segments of the in-use fleet. Therefore, it is conceivable that the most accurate assessment of I/M program effectiveness will result from evaluations which combine multiple program evaluation methods.

3. General Approaches to I/M program evaluation

3.1 Defining Program Evaluation

Aside from the technical challenges involved in gathering I/M program evaluation data, there are also subtleties regarding what data is necessary that must be understood. The evaluation of Basic I/M programs is strictly qualitative as per standard SIP policy protocols used to evaluate stationary source emission reductions. Historically, these type of qualitative evaluations have included verification of such parameters as waiver rates, compliance rates, and quality assurance/ quality control procedures, but they have not involved quantitative estimates of emission reductions using in-program or out-of-program data.

The evaluation of Enhanced I/M programs is not as clearly defined and is left to the discretion of the Regional EPA based on the data available. In some instances, it may be possible to estimate the cumulative emission reductions, that is the current fleet emissions are compared to what that same fleet's emissions would be if no I/M program were in existence. However, directly measuring the fleet's emissions to determine the No-I/M baseline is not possible in an area that has implemented an I/M program. Therefore, in order to determine quantitatively whether the level of SIP credit being claimed is being achieved in practice, it becomes necessary to rely on modeling projections to estimate the No-I/M fleet emissions or measure the emissions of a surrogate fleet that is representative of the I/M fleet. The RSD procedures outlined in this guidance provide methods for estimating a fleet's No-I/M emissions using a surrogate fleet.

Two other analyses are also possible that can provide useful information regarding program performance. The first method may be thought of as "one-cycle" since it compares the current I/M fleet emissions to the same I/M fleet's emissions from a previous year or cycle. An analysis such as this would yield information with regard to how the program is improving or declining from year to year. The other method should be considered "incremental" in that it compares the current I/M fleet's emissions to that same fleet's emissions while being subjected to a different I/M program, for instance, comparing a fleet's emissions in an area that has just implemented an IM240 program to that same fleet's emissions the previous year when a Basic Program was in

operation. It should be noted, that there is a small window of opportunity prior to and during the start-up of any I/M program, or program change, to actually measure the fleet emissions that would provide empirical data on the No-I/M fleet emissions. If resources and time permit, it is recommended that these baseline data be gathered in order to reduce I/M program evaluation dependency on modeling projections and provide the most accurate measure of I/M program performance.

3.2. On-Road Data Analysis

Remote sensing measurements can be used as a tool to help achieve the main goal of all I/M programs, namely the reduction of on-road emissions. The general advantages of remote sensing data are the following:

    i)   The testing is unscheduled and measures on-road emissions.
    ii)   A sample of all vehicles driving in an area can be tested.
    iii)   A very large sample of vehicles can be tested for a fraction of the cost of I/M lane testing.
    iv)   Vehicles can be tested over a range of driving conditions, rather than merely the conditions specified in the I/M test[*].
    v)   Vehicles that are often not tested due to condition, size or special dynamometer requirements (heavy duty vehicles, vehicles considered unsafe to test, vehicles requiring four-wheel-drive dynamometers) can be measured.
    vi)   The on-road data can evaluate the extent to which owners are repairing their vehicles prior to emission testing. This is a program benefit that cannot be easily measured by means of in-program data without the use of surveys.
    vii)   RSD can be directly converted to mass emissions per volume or mass of fuel burned and may be used to develop emission inventories independent of models (_9_).

In a well-designed remote sensing program, roadway grade and environmental conditions at the measurement site, as well as vehicle speed and acceleration, will be measured and used to calculate the vehicle load for each individual emissions measurement. Analyses can then be performed on a subset of measurements with a distribution of loads similar to that encountered by a single vehicle on the program's I/M test. In addition, by employing careful site selection criteria, remote sensing has the potential to measure emissions under driving modes not currently incorporated into I/M tests.

Emissions measured by remote sensing instruments, and in idle and ASM tests, are reported in terms of concentration of total exhaust. Remote sensing data, then, can be directly compared

---

[*] By measuring vehicles on-road, RSD has the ability to measure vehicle performance at high power, "off-cycle" conditions that cannot be readily measured on a dynamometer because of tire slip, tire damage, safety concerns, vehicle owner concern and damage claims. Although off-cycle emissions are not regulated by the vehicle certification process, and their measurement may not be desired for I/M evaluation, they may be an important component of estimating the mobile source inventory. Therefore, on-road measurement of high power, "off-cycle" performance may be used to develop a complete emissions inventory and to assess the effectiveness of repairs under "off-cycle" conditions.

with emissions results from I/M programs utilizing idle or ASM testing. However, some enhanced I/M programs measure mass emissions, and report emission results in grams per mile. Remote sensing concentration measurements can be converted to grams per gallon, using combustion chemistry equations, and then grams per mile, using an estimate of the instantaneous fuel economy (miles per gallon) of the vehicle at the time of measurement. The accuracy of the conversion from emissions concentration to grams per mile depends on the accuracy of the estimate of instantaneous fuel economy. Fuel economy varies by vehicle type, technology and age, as well as by vehicle load, thus complicating the conversion. Areas conducting IM240 or ASM testing should plot mean RSD emissions against mean initial IM240 or ASM emissions by vehicle type and model year. These plots typically show a linear relationship with high correlation coefficients and can be used to establish a direct relationship between the RSD measurements and the I/M test results. IM240 program data also includes CO_2 emissions and thus can be directly converted to emission per gallon and compared to on-road data. These comparisons have been published and show R^2 generally greater than 0.95, although the slopes and intercepts are not 1.0 and 0.0 ([10]).

In particular, remote sensing data can be used in several ways to evaluate the effectiveness of an I/M program:

i) Remote sensing programs measure vehicles at different times relative to their last I/M test. Therefore, remote sensing data can be used to estimate how quickly repair effectiveness diminishes over time and how much repair is made just prior to the I/M test, as well as track changes in fleet emissions due to changes in test procedures.

ii) Remote sensing programs measure almost every vehicle that drives by the instrument, regardless of whether it is participating in the I/M program. Remote sensing data therefore can be used to estimate the number and emissions of vehicles legally exempted from, or illegally avoiding, the I/M program, as well as estimating their emissions. In addition, remote sensing data can identify individual vehicles that never complete the current I/M cycle, or that do not report for testing in a subsequent test cycle, but are still being driven in the I/M area.

However, as with in-program data, there are inherent limitations to RSD data.

i) The primary objection raised by opponents of RSD is that it must be assumed that a one second snapshot of the vehicle's emissions is characteristic of that vehicle's emission profile.

ii) Fleet coverage is also a very realistic concern as it is often difficult to obtain readings on more than 50% of the fleet, which means that there may not be any emission readings for half of the vehicle population.

iii) The quality control and quality assurance aspects of RSD data collection and analysis have not been as well documented as those for traditional tailpipe testing.

Random roadside pullover testing has similar advantages to remote sensing; the test is unscheduled, and vehicles can be tested at different times relative to their last I/M test. However, roadside testing programs may be more expensive and time-consuming than some remote sensing programs, and so many fewer vehicles can be tested. California has operated a roadside pullover testing program for several years. An advantage of roadside testing is that the vehicles can be tested using the same test methods as those employed in the I/M program. They can also be inspected for visual or functional failures. However, the sample of vehicles participating in the California roadside testing program may not reflect the on-road fleet, since participation in the program is not mandatory, and it is also difficult to verify that vehicle selection is unbiased. Furthermore, roadside pullovers are politically unacceptable in many areas.
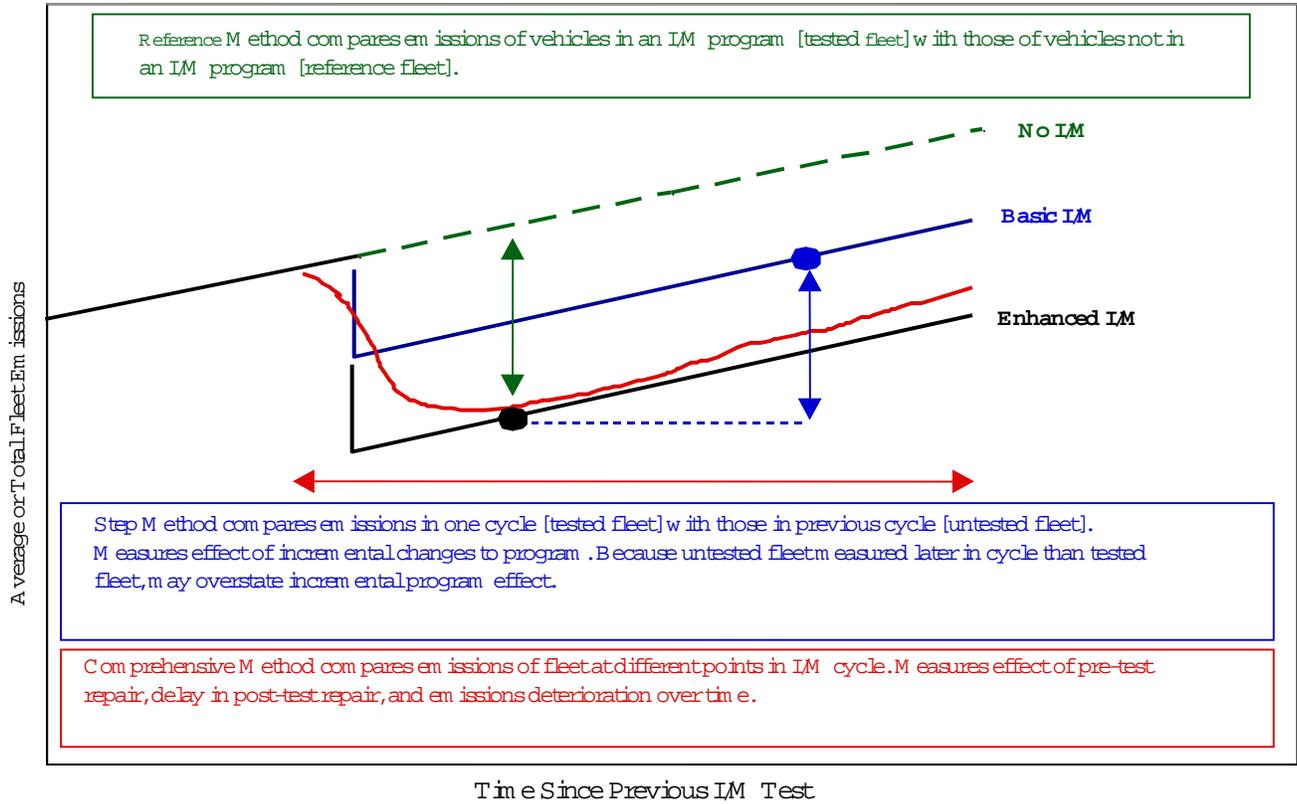
3.4. Three RSD Program Evaluation Methods

In this document three methods, not necessarily exclusive, of using remote sensing data to analyze I/M program effectiveness are discussed. These are the Step Change, the Comprehensive, and the Reference Methods. The Step Change and Comprehensive evaluation methods are quite similar. Remote sensing measurements are made on a fleet of vehicles in an I/M area. The fleet is then divided into two sub-fleets, based on whether or not individual vehicles have been tested under the current I/M program. The emissions of the two sub-fleets are then compared, after accounting for differences in vehicle type and age. The difference in the emissions of the tested fleet and the untested fleet is the apparent benefit of the I/M program in reducing emissions.

The primary difference between the two methods is the number of remote sensing measurements required. The Step Change Method can be performed using a relatively small number of measurements, on the order of 20,000 to 50,000. The Comprehensive Method requires many more remote sensing measurements (several million in the Phoenix example) in order to perform the detailed analyses of program effectiveness. Collecting this much remote sensing data can be relatively expensive; however, if such data are already being collected as part of another program (such as a Clean Screen program), the additional cost of analyzing the data is minimal. The drawback of the Step Change and Comprehensive Methods (aside from the general concerns with regard to RSD mentioned above) is that they only measure the effect of incremental changes in I/M programs unless repeated year after year.

The Reference Method is designed to measure the full effect of an I/M program on a vehicle fleet, by comparing the emissions of a fleet subject to I/M with estimated fleet emissions if no I/M program were in place. The accuracy of the Reference Method hinges on the ability to find a fleet in a non-I/M area as similar to the I/M area fleet as possible. Because vehicle emissions are quite variable, both between vehicles and within an individual vehicle, and because many differences between vehicle fleets and their environment can affect vehicle emissions, finding a suitable reference area can be challenging. One way to determine the degree of bias in the reference fleet is to obtain data from a second reference fleet; if there are few biases, the two reference fleets should look the same. The Reference Method can also be used to compare the impact of two I/M programs in different locals. Although this will provide a relative comparison between two programs it will not provide any data to compare an I/M program to a No-I/M fleet.

Figure 3.1 below illustrates some of these differences.

Figure 3.1. I/M Program Evaluation Methods Using Remote Sensing Data

Reference Method compares emissions of vehicles in an I/M program [tested fleet] with those of vehicles not in an I/M program [reference fleet].

No I/M

Basic I/M

Enhanced I/M

Average or Total Fleet Emissions

Step Method compares emissions in one cycle [tested fleet] with those in previous cycle [untested fleet]. Measures effect of incremental changes to program. Because untested fleet measured later in cycle than tested fleet, may overstate incremental program effect.

Comprehensive Method compares emissions of fleet at different points in I/M cycle. Measures effect of pre-test repair, delay in post-test repair, and emissions deterioration over time.
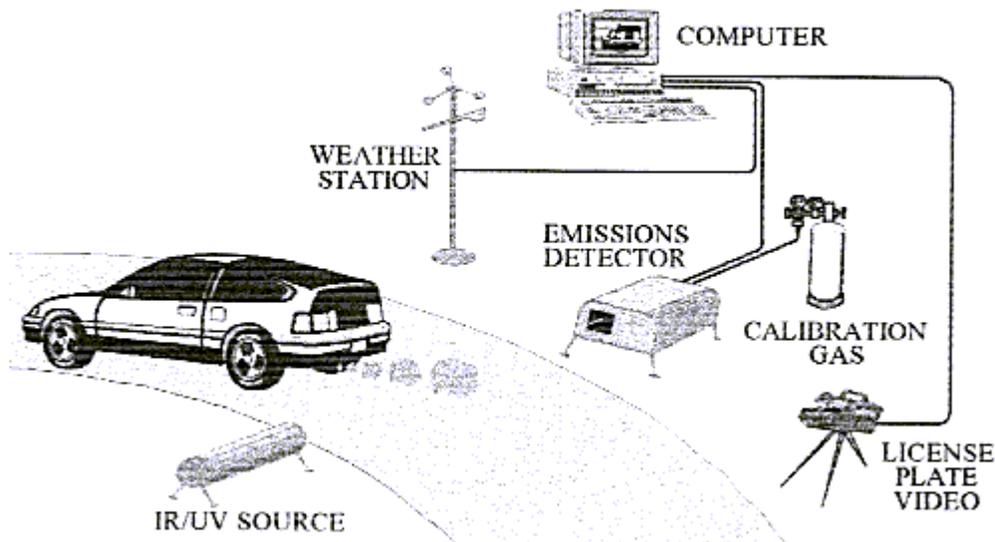
Time Since Previous I/M Test

4. Equipment Specifications and Measurement Procedures

4.1. The Remote Sensing System

Figure 4.1 shows a generic diagram of an RSD system which measures CO, CO2, HC, NO, and smoke opacity set up along a single lane of road. The make and model year of the vehicle are identified from the video picture.

Figure 4.1: RSD Operational Diagram

4.2. Theory of Operation

Remote Sensing Devices have been designed to emulate the results one would obtain using a conventional exhaust gas analyzers. Because the effective plume path length and amount of plume seen depend on turbulence and wind, one can only determine ratios of CO, HC, or NO to CO2. Assuming complete and instantaneous mixing, these ratios, Q for CO/CO2, Q' for HC/CO2, and Q" for NO/CO2 are constant for a given exhaust plume. By themselves, Q and Q' are useful parameters with which to describe the combustion system. When the corresponding combustion equations are solved many components of the vehicle operating characteristics can be determined including the instantaneous air/fuel ratio and the % CO,% HC, and % NO which would be read by a tailpipe probe. The equations given below are based upon a carbon mass balance and make use of the fact that the IR HC analysis method only measures about one half of the carbon which would be measured by means of an FID for instance.

$$\% \, CO2 = 42/(2.79 + 2Q + 0.84Q')$$
$$\% \, CO = Q * (\% \, CO2)$$
$$\% \, HC = Q' * (\% \, CO2)$$
$$\% \, NO = Q'' * (\% \, CO2)$$

To derive mass emissions in g/gal of fuel from Q and Q' a fuel density of 0.75 g/mL and the carbon-hydrogen ratio of 1:2 are assumed to yield:

$$CO2 \text{ mass emission (g/gal)} = 8922/(1 + Q + 6Q')$$
$$CO \text{ mass emission (g/gal)} = 5678*Q/(1 + Q + 6Q')$$
$$HC \text{ mass emission (g/gal)} = 8922*2*Q'/(1 + Q + 6Q')$$
$$NO \text{ mass emission (g/gal)} = 6083*Q''/(1 + Q + 6Q')$$

The vehicle's instantaneous air to fuel ratio is

$$A/F \text{ by mass} = 4.93(3 + 2Q)/(1 + Q + 6Q')$$

All diesel and most gasoline powered vehicles show a Q and Q' near zero since they emit little to no CO or HC. To observe a Q greater than zero, the engine must have a fuel-rich air/fuel ratio and the emission control system, if present, must not be fully operational (_11_).

In the case of diesel combustion, misfire causes high HC readings. Since the overall air/fuel ratio is very lean, even when over-fueling and sooting are taking place, CO emissions only arise from pockets of incomplete combustion, and are limited to about 3% CO, compared to a broken gasoline-powered vehicle which can exceed 12% CO.

Recently, the ability to measure nitric oxide (NO) has been added to the existing IR capabilities. The light source, across the road, now contains a deuterium or xenon arc lamp and IR/UV beam-splitter which is mounted in such a manner that the net result from the source is a collimated beam of UV and IR light. As with CO and HC measurements, the NO measurements are possible by ratioing to the CO2 measured in the plume. All pollutants except HC are a specific gas which can unambiguously be measured and calibrated. Exhaust HC is a very complex mixture of oxygenated and unoxygenated hydrocarbons. The filter chosen measures carbon-hydrogen stretching vibrations which are present, but not equally in all HC compounds. This system can easily distinguish gross polluters from low emitters, but the results on an individual vehicle cannot be expected to correlate perfectly with a flame ionization detector, with ozone-forming reactivity, or with air toxicity, since the three are not correlated to one another. For large sample sizes the fleet average emissions correlate well with IM240 g/mi measurements (_12_).

Newer technologies may also be used in place of the UV/IR detectors described above, such as tunable diode lasers.

4.3. Operation

When a motor vehicle passes through the beam of a calibrated instrument on the road, the computer notices the blocked intensity of the reference beam. This causes the previous 200 ms of data (20 points) to be stored in memory as the "before car" buffer. The blocked voltages are continuously interrogated both to remember the lowest values (zero offset) and to look for a beam unblock signal. When an unblock signal is recognized, the video picture is frozen into the video screen memory and thus goes to the image recorder, and the next 50 data points (1/2 sec of

exhaust) are placed in a data table. The zero offsets are subtracted from all data. The data stream is interrogated for the highest CO2 voltage. This is the least polluted 10 ms average seen during the 0.7 sec. of data devoted to this vehicle. This set of data (often, but not always, in the before car buffer) then becomes the "clean air reference" (CAR) against which all other data are compared. After all signals have been ratioed to the reference channel, and ratioing the results to the CAR result for that channel, one now has a set of 50 postcar, corrected, fractional transmissions which are converted to gas concentrations such as would have been observed in the gas analyzer. These concentrations are then correlated to CO2 and the slope and error of the slope determined. These slopes (the ratios of the pollutants to CO2) are corrected by the correction factors determined for that time by means of roadside calibration. These slopes now are the Q, Q' and Q" described earlier.

The data obtained for each vehicle provide three pollutant ratios. The RSD software now solves the combustion equation for the measured pollutant ratios, compares the errors to preset error limits, and, if acceptable, reports the measurements as % CO, % CO2, % HC, and % NO such as would be measured by a tailpipe probe with the results corrected for water and for any excess air which may not have participated in combustion. In view of the fact that the instrument is calibrated with propane, percent HC is reported as propane; however, other HC species such as hexane or 1,3 butadiene could be used for this purpose as well. The four derived concentrations, % CO, % HC, % NO, and % CO2, are placed on the video output together with the vehicle image (which has been waiting without results for about 0.7 sec.).

This image now stays on the screen until the next vehicle comes by to repeat the process. If these results are to be compared to vehicles of known emissions, or gas cylinders puffed into the beam, it is important to compare the three ratios and not the four derived concentrations since there are not actually four independent pieces of information. For example, if a person blocks the beam and exhales into it during the 1/2 sec. after they have unblocked the beam, the computer sees the exhaled CO2, finds no CO, HC, or NO, and reports zeros for those pollutants and about 15% CO2. Exhaled breath rarely contains even 2% CO2, but the system only measures the ratios, and assumes (incorrectly in this case) that the emissions are from a fully stoichiometric automobile using gasoline as fuel. A puff from a cylinder which contains 50% CO and 50% CO2 would be read as 8.6% CO and 8.6% CO2 because the ratio is what is measured not the absolute concentrations.

Special software traps should be employed to deal with two cars traveling very close together. In this case, the before car buffer from in front of the first is used as a potential source of clean air reference for the exhaust of the second. The video picture of the first is replaced by the second before any data are overwritten. High pickup trucks thus often get two pictures, only the last of which has emissions data.

Other software traps reject data when the slope errors are too large, and when there is no sign of any significant exhaust plume (such as behind 18-wheel trailers whose tractors have elevated exhausts).

For the interested reader, Appendix A contains a brief description of some trouble-shooting procedures that can be performed quickly in the field as a first step to verify if an RSD unit is operating properly or if it is in need of service.

4.4. Operational Difficulties

4.4.1 Signal/Noise Considerations

Remote emissions measurements would all be very straightforward if one were able to measure directly behind the tailpipe of each passing car. Absorptions would be large, and the system signal/noise (S/N) would not be limiting. In fact, vehicle tailpipes are not in standardized configurations, vehicle engine sizes are not uniform, and there is very rapid turbulent dilution of the exhaust behind vehicles moving faster than about 5 mph. Thus, one is forced to make engineering tradeoffs between the desire to measure all vehicles and the necessity to have an adequate S/N so as not to report incorrect exhaust emissions values.

The detection of gas absorption is based upon the reduction of signal on one detector versus the reference detector. Thus, the average car measured at an uphill freeway ramp in Denver shows an exhaust plume already diluted by a factor of about 10. This situation gives rise to an easily measurable 14% reduction in the CO_2 voltage. Because the average CO content is about 1/20th of the CO_2 and the HC 1/10th of the CO, the average total changes in CO and HC voltages are only 3 and 1 part in 1000, respectively. The NO channel shows a similar response as HC. Thus, the instrument builder's challenge is to build a system in which part per thousand changes in IR and UV intensity are accurately measured in all weather conditions beside a normal road at a measurement frequency of 100 Hz. At other locations, the plume dilution factor is 100 and a decision must be made whether the individual instrument's S/N is adequate for readings to be reported or if the data should be reported as invalid. This bleak outlook is somewhat mitigated by the fact that the source need only maintain a stable intensity for about two seconds for a complete measurement series and the fact that the data reduction process intrinsically "averages" all the 1/2 sec. data to only three ratios.

Newer technologies having improved S/N ratios may be available and used over greater distances.

4.4.2. Weather

Measuring light intensities over a 10 m path to better than a few parts per thousand can be inhibited by bad weather. Ambient temperature and humidity variations are not a problem, but snowflakes and heavy rain add too much noise to all data channels. Wet or very dusty roadways cause a plume of spray or dust behind vehicles moving above about 10 mph. These plumes also add noise to the system, and generally increase the data rejection rate to an unacceptable level.

At the most productive sites, the remote sensor can gather data on 10,000 vehicles in a working day; thus, it often generates data faster than the operator can handle. In such cases, taking the day off to analyze data when the weather conditions are not appropriate may be beneficial. Gross polluting vehicles are thought to be the same vehicles on dry as well as on wet days.

4.4.3. Interference
The HC wavelength suffers from some interference from gas phase, and certainly from particulate phase, water (so-called "steam" plumes from colder vehicles operating at low ambient temperatures). When steam plumes are so thick that you cannot see through them (Fairbanks, AK., at forty below zero) the system no longer operates since all wavelengths are absorbed or scattered too much for useful data to be acquired.

4.4.4. Optical Alignment
If the instrument is not perfectly optically aligned, the voltages are likely to be very sensitive to equipment vibration. Since moving vehicles both shake the roadway and generate wind pulses, rigid instrument mounting is as important as perfect internal and external optical alignment. Software is written so that these noise sources generate "invalid" flags. Proper alignment at a well characterized RSD-site can yield 95% valid RSD readings on passing vehicles using UV/IR detector technology.

The system is designed to operate on a single-lane road. Freeway ramps, turn lanes, and the inevitable road closures for sewer, gas, water, telephone, and road maintenance are often good candidates for RSD emission measurement sites. Multiple-lane operation has been reported but is not recommended.

4.4.5 Emissions Variability
Emissions of motor vehicles are not constant from second to second or from day to day. Broken vehicles in particular often seem to have a large random component to their emissions irrespective of what test is used to make the measurement (*13*). Some vehicle emission variability has known causes such as the initial operation of cold vehicles before the engine control system stabilizes and the catalyst begins operation, or when the vehicle is accelerated at full throttle. Both situations give rise to large CO and HC emissions from even well-maintained vehicles, but can be minimized through careful site selection.

4.5 Instruments

4.5.1. Calibration Checks
Two separate calibration procedures should be performed on every remote sensing unit. The first is conducted in a laboratory and should be performed by the equipment manufacturer. It may consist of exposure in the laboratory at a path length of about 22 ft to known absolute concentrations of NO, CO, CO2, and propane in an 8 cm IR flow cell with CaF2 or other IR transmitting windows. The calibration curves are used to establish the fundamental sensitivity of each detector/ filter combination to the gas of interest. The results of this calibration should be provided to the state or contracting party upon request.

The second calibration should be every hour (*14*) during operation until the stability of the individual system is quantified and characterized using statistical process control methods. Once control charts have been established, the calibration frequency may be reduced appropriately. Several puffs of gas designed to simulate all measured components of the exhaust are released from a cylinder containing certified amounts of NO, CO, CO2, and propane into the optical beam path. The ratio readings from the instrument are compared to those certified by the

cylinder manufacturer. In this way the system never actually measures exhaust emissions; it basically compares the pollutant ratios in a known standard gas cylinder and those measured in the vehicle exhaust.

The gases used for the second calibration shall by certified to +/-2% of a known NIST standard and be in the following ranges:

CO	1-9%
HC as C3	300-4100 ppm
NO	1500-3600 ppm
CO2	5-14% (with the balance oxygen free nitrogen)

Additionally, some quick checks are provided in Appendix A that may be useful in trouble-shooting equipment in the field.

4.5.2. Other Instrument Parameters

At a minimum the following parameters shall also be recorded in all RSD program evaluation studies for each RSD site in a stations log. The log may be kept electronically or in hardcopy format.

i)   A description of the RSD equipment including light source, make/model of instrument, and detector type.
ii)   The name of the operator and the van. If more than one operator or van are used, key and record which operator and/or van was used for each measurement.
iii)   Complete description of the calibration procedure.
iv)   Audit check results
v)   Calibration check results
vi)   Any equipment changes
vii)   Verification of speed and acceleration measurement devices

4.6. Site Description

A site description for each RSD data collection site shall be generated that shall include the following information.

i)   Road map with features affecting traffic flow.
ii)   Note any change in the position of the light source, detector, etc. from previous RSD studies
iii)   Note any change in traffic patterns from previous RSD studies.
iv)   Note the altitude of the site and the road grade. Include a field in the database showing the road grade in percent for all measurements.
v)   Digital picture of the site including all cones, etc, that would influence motorist driving patterns.
vi)   Global Positioning Satellite coordinates based on the NAD86 reference standard.

4.7. Measurements

4.7.1. Data Collection

The following measurements shall be recorded at each site where RSD program evaluation data are collected.

    i)   %CO2, %CO, %NO, %HC, maximum CO2, all error terms, restarts, and negative emission numbers. Include a field showing whether HC is reported as propane or hexane.

    ii)   Speed and acceleration. Vehicle Specific Power shall be calculated as described below. Valid VSP values shall be between 0-20 kW/ton.

$$VSP_{kW/t} = 4.39*\sin(slope)*v+0.22*v*a+0.0954*v+0.0000272*v^3$$

        where "a" is vehicle acceleration in mph/s, "v" is vehicle speed in mph, and slope is the road grade in degrees[*].

    iii)   Location of speed measurement relative to emission measurement. It is recommended that vehicle speed be measured 5-10 m prior to the emissions measurements.

    iv)   Time and date of measurement

    v)   License plate. Record all plates including in-state, out-of-state (OS), dealer (D), paper plate (PP), obscured plate (OP), and no-plate-visible (NPV)

    vi)   Hourly temperature, barometric pressure, and relative humidity

    vii)   Describe how plume strength is determined and flagged, as well as the criteria for rejecting measurement attempts.

    viii)   Site reference label

    ix)   RSD unit number or unique identifier

4.7.2. Multiple Measurements

Multiple measurements made on the same vehicle shall be treated in one of the following ways; however, the program evaluation report will clearly state which method has been chosen and the rational behind this choice. A multiple measurement is not restricted by the timeframe over which it is collected. Therefore, it may be hours, days, weeks or months. Option (iv) below is recommended, although there may be circumstances when another option may be more appropriate.

    i)   Multiple measurements are treated as independent readings

    ii)   Multiple measurements are averaged and treated as a single reading

    iii)   Multiple measurements are discarded and only the first reading is used

    iv)   The maximum, minimum and average values are reported to provide as comprehensive a snapshot of a vehicle's emission profile as possible.

4.7.3. Operators

Care must be taken to ensure operators are properly trained in the routine operation of the equipment and fully understand and implement the QA/QC required procedures. Furthermore, it

---

[*] This equation should be considered generic in that it may be applied to all types of vehicles. More accurate equations dependent on MY and/or vehicle type may be developed in the future.

is imperative that daily vehicle quotas do not compromise the operators judgments or actions with regard to QA/QC and the data collection process.

4.8. Database Format

The RSD data collected shall be made available in an ASCI text file that may be easily ported into a standard commercially available database software package such as Access, Oracle, SAS, etc. If special procedures are required to port the data into such a software package the software code or procedures shall be provided upon request.

4.9. Department of Motor Vehicle Data

Department of Motor Vehicle data shall be reported as follows.

    i)   Date DMV data received from DMV
    ii)   Information indicating how current the most recent DMV data in the file are.
    iii)   VIN, Model Year, Make, Model, Fuel Type, Vehicle Type, Zip Code
    iv)   I/M test date.
    v)   I/M test results in g/mi, ppm or percent.

4.10. Note Any Changes that Could Affect the Analysis

Any changes to the I/M program which would impact the analysis shall be recorded and reported in the program evaluation report. Such changes may include, but are not limited to, changes in the operational details of the I/M program itself, or the use of a seasonal fuel program to reduce mobile source emissions.

5. Design Parameters and Quality Assurance/Quality Control Protocols

5.1. Overview

This section outlines a number of critical issues that must be addressed to perform a program evaluation using RSD technology. These issues include data collection design parameters, equipment specifications, calibration procedures, quality control, and several known sources of bias in vehicle emissions measurements that can affect any evaluation of an I/M program. Some of these are unique to remote sensing data, while others apply to evaluations based on in-program data as well. The issues or types of bias that must be considered in a remote sensing program evaluation have been broadly grouped into the following categories and discussed under the appropriate headings below: vehicle population, vehicle load, vehicle identification, instrument calibration, measurement method, socioeconomics, seasonal effects, program avoidance, regional differences, program details and emissions distributions.

The importance of five issues (vehicle load, program avoidance, vehicle identification, program details and emissions distributions) are roughly similar for each of the three evaluation methods. Because the Reference Method relies on measurement in two different geographic regions, it is

most sensitive to all of the remaining types of bias. The likelihood of bias can be minimized if multiple reference sites are chosen and the sites are well-characterized with common load characteristics. Because the Comprehensive Method requires large numbers of measurements, multiple vans and sites can increase a bias due to instrument calibration, socioeconomics, and seasonal effects. In collecting data at a single site over a short time period, the Step method eliminates the potential for socioeconomic and seasonal bias between the two measured subfleets; however, the estimate of program effectiveness may be biased if the site chosen or the time of testing does not capture the distribution of driver socioeconomics or environmental variables representative of the I/M area. This potential source of bias can be tested by comparing the measured fleet numbers by model year to other data[*], bearing in mind that on-road measurements are expected to measure newer, higher annual mileage vehicles more than older, lower annual mileage vehicles.

5.2. Vehicle Population
Goal: Account for differences in vehicle fleet distributions in the program evaluation analysis.

Perhaps the most common source of bias when comparing emissions of two fleets of vehicles is the vehicle distribution of the two fleets. Older, higher mileage, vehicles tend to have higher emissions than newer, lower mileage, vehicles. Light duty trucks were built to less strict emissions standards than passenger cars, and are observed to have higher in-use emissions. In addition, there is a wide range in average emissions by vehicle model, even for vehicles of the same age (15).

Differences in vehicle fleets can be determined by comparing vehicle distributions of the two fleets by type and model year. (Note: The Step, Comprehensive and Reference Methods all compare fleet averages; however, the composition of these sub-fleets is different for each method.) Average emissions by type and model year should be calculated for each fleet and compared to determine any emissions differences between the two fleets. The average emissions for each fleet should then be weighted by a single distribution of vehicles by type and model year (preferably that of the I/M program area), to determine the overall fleet emissions and the percent difference between the two fleets.

Table 5.1 displays examples CO emissions by model year from samples of vehicles measured in a reference area and an I/M area. The composite fleet averages of 0.86% CO for the reference area and 0.58% CO for the I/M area suggest the I/M area vehicles are 32% cleaner. This is not a fair comparison, however, because it is evident from the fleet fraction percentages (Columns D and G) that the I/M area sample contains a greater proportion of newer vehicles.

To overcome this, the I/M area model year CO contributions are re-weighted according to the reference area fleet fraction percentages. This is shown in column H. The adjusted composite emissions level for the I/M area is now 0.76% CO, resulting in an apparent 12% (1-0.76/0.86) benefit from the program. It should be noted that this 12% apparent benefit should be converted

---

[*] This data may be obtained from Department of Motor Vehicle (DMV) records or modeling defaults, although empirical DMV data would be preferred.

to a mass basis to be more meaningful and allow more direct comparisons to other I/M program evaluation results as well as results from other air pollution control programs.

Of course, this raises the question as to what extent the greater proportion of newer vehicles in the I/M fleet is the result of the I/M program. Addressing this question is difficult. No current analyses of in-program or out-of-program data provides information in this regard. At this time, further studies are needed to address this issue.

Because the Step Change and Comprehensive Methods compare fleets of vehicles from the same I/M area, there is likely to be little difference between the two fleets with respect to fleet distribution. However, when using the Reference Method, vehicle populations can be significantly different between different geographical areas, as can fuel composition, environmental factors, and motorist socioeconomic status (discussed below).

Table 5.1: Average RSD Readings by Model Year

A	B	C	D	E	F	G	H
Model	Reference Area			I/M Area			
Year	Avg CO	Count	Fleet %	Avg CO	Count	Fleet %	E x D
Pre-60	3.45	70	0.04%	1.60	16	0.03%	0.00
Y60-65	4.12	390	0.20%	3.61	39	0.08%	0.01
Y66-70	3.50	1333	0.69%	3.24	137	0.28%	0.02
Y71-75	2.74	2661	1.37%	2.50	310	0.64%	0.03
Y76-80	2.42	10259	5.27%	2.19	1173	2.44%	0.12
Y81	2.24	2818	1.45%	1.64	373	0.78%	0.02
Y82	1.94	3430	1.76%	1.34	470	0.98%	0.02
Y83	1.71	5440	2.80%	1.36	707	1.47%	0.04
Y84	1.64	8424	4.33%	1.23	1203	2.50%	0.05
Y85	1.39	10322	5.31%	1.18	1654	3.44%	0.06
Y86	0.99	12067	6.20%	0.83	2172	4.51%	0.05
Y87	0.83	12532	6.44%	0.77	2497	5.19%	0.05
Y88	0.72	14410	7.41%	0.70	2853	5.93%	0.05
Y89	0.68	14803	7.61%	0.61	3059	6.36%	0.05
Y90	0.56	14479	7.44%	0.53	3366	7.00%	0.04
Y91	0.50	14666	7.54%	0.50	3717	7.72%	0.04
Y92	0.43	12977	6.67%	0.42	3645	7.57%	0.03
Y93	0.37	14617	7.51%	0.36	4350	9.04%	0.03
Y94	0.28	13222	6.80%	0.30	4507	9.37%	0.02
Y95	0.23	15055	7.74%	0.26	5435	11.29%	0.02
Y96	0.17	9668	4.97%	0.20	4320	8.98%	0.01
Y97	0.11	876	0.45%	0.21	2116	4.40%	0.00
Avg/Tot	0.86	194519	100.00%	0.58	48119	100.00%	0.76

5.3. Vehicle Loads
Goal: Ensure that RSD measurements are made under known vehicle operating conditions.

Another important source of potential bias is the load under which the vehicle is operating when the emissions measurement is made. Emissions per gallon are very much less speed and load dependent than emissions per mile, nevertheless load is an important variable. Researchers use vehicle specific power (VSP equation given earlier) which is a function of vehicle speed, acceleration, drag coefficient, and tire rolling resistance, and roadway grade, to characterize the load the vehicle is operating under at the time the measurement is made (_16,17_).

On-road remote sensing units measure tailpipe exhaust plumes for a fraction of a second as vehicles pass by the unit. HC, CO and NOx pollutant emissions are estimated by comparing the

ratio of their concentrations to the concentration of CO2 seen in the vehicle exhaust plume. Although, the remote sensing unit does not measure the volume of exhaust gases produced, a number of vehicle load conditions can elevate the remote sensing observed emission levels:

i) When a motorist lifts his/her foot off the gas pedal, the volume of air and fuel flowing through the vehicle engine and exhaust system is suddenly reduced. Under these circumstances, the ratio of HC and CO to the now reduced level of CO2 is often increased. Although the volume and mass of emissions are substantially reduced when a driver lifts off the gas, to the remote sensing unit, the ratio of the concentrations of HC and CO to CO2 are actually higher and a higher emissions value is recorded. This effect is greatest for HC.

ii) When a motorist presses sharply on the accelerator, the vehicle may go into what is termed an 'off-cycle' condition. The current generation of vehicles have been certified using the Federal Test Procedure; however, this test does not cover the full power range of the vehicle. Consequently, vehicles were designed to minimize emissions only over the power range tested in the certification cycle. At higher powers, so called "off-cycle" or power enrichment emissions often increase dramatically although the vehicle is functioning as designed. Under these circumstances, a vehicle can have high emissions when measured by remote sensing but may meet the I/M inspection requirements. This effect is greatest for CO and NOx.

For these reasons, multiple remote sensing measurements for the same vehicle can vary considerably if the site is such that the operating mode of the vehicle at the time of the measurement is not consistent. As stated earlier (Section 4.7.2), it is recommended that in the case of multiple measurements, all data are retained, or the maximum, minimum and average values are reported to provide as comprehensive a snapshot of a vehicle's emission profile as possible. For broken vehicles, the variability and the likelihood of high readings is extreme. For low emitting, new or well-maintained vehicles, variability caused by driving mode changes under normal operating circumstances is very small.

The load under which each individual vehicle is driving, or VSP, should be calculated based on vehicle speed, acceleration, and roadway grade, as described earlier. The distribution of VSP should then be compared between different remote sensing sites to determine if vehicles are being driven differently at different sites. If there are enough remote sensing measurements, average emissions by vehicle type and model year can be weighted by a common VSP distribution to remove any bias introduced by different vehicle loads at different remote sensing sites.

With regard to repair effectiveness, it is important to recognize that not only are absolute emission levels sensitive to vehicle load; the percent change in emissions from vehicle repair is as well. An analysis of repair effectiveness on a sample of vehicles given a full IM240 test before and after repair indicates that the percent reduction in emissions over the moderately loaded portion of the IM240 was only half that of the reduction over the entire IM240 (*18*).

Therefore, it is critical that any analysis of remote sensing data used to characterize fleet emissions in general or estimate repair effectiveness include the calculation of vehicle load. To minimize the possibility of a driver making sudden throttle changes it is recommended remote sensing units be sited in locations such as highway on or off ramps. In addition, analyses that rely on data from more than one remote sensing site should re-weight average emissions at different sites by a similar distribution of vehicle loads, to allow proper comparison of emissions data collected at each site. There is some evidence that older vehicles behave differently than newer vehicles with respect to VSP. In the future, vehicles designed to meet supplemental FTP certification requirements can be expected to behave differently than today's vehicles. Consequently, adjusting calculations, if required, should probably divide the populations into several ranges of model years. Table 5.2 illustrates the various loads vehicles are subject to during emission tests or accelerations.

Table 5.2 Examples of VSP Values

Activity	VSP (kW/metric ton)
Maximum Rated Power	44-120
0-60 in 15 seconds	33
60 mph up 4% grade	23
FTP or IM240 max	23
Typical RSD site	10-15
Average IM240	8
ASM5015	6
ASM2525	5

Figures 5.1, 5.2 and 5.3 illustrate the relationships between emission and VSP for various vehicle MY groupings. Maintaining as narrow a VSP window as possible will help minimize variability between site measurements, although there may be practical limitations of how tight the VSP operating window can be held. The data presented in the following three figures indicate relatively constant CO and HC emissions for VSP values between 5 and 20 kW/metric ton, while NO emissions are more variable even if the VSP window is reduced to 10 to 20 kW/metric ton. Therefore, for this data set it would appear that a VSP range of 15 +/- 5 kW/metric ton would be the recommended target to minimize site-to-site load variability.

Figure 5.1 RSD %CO vs VSP
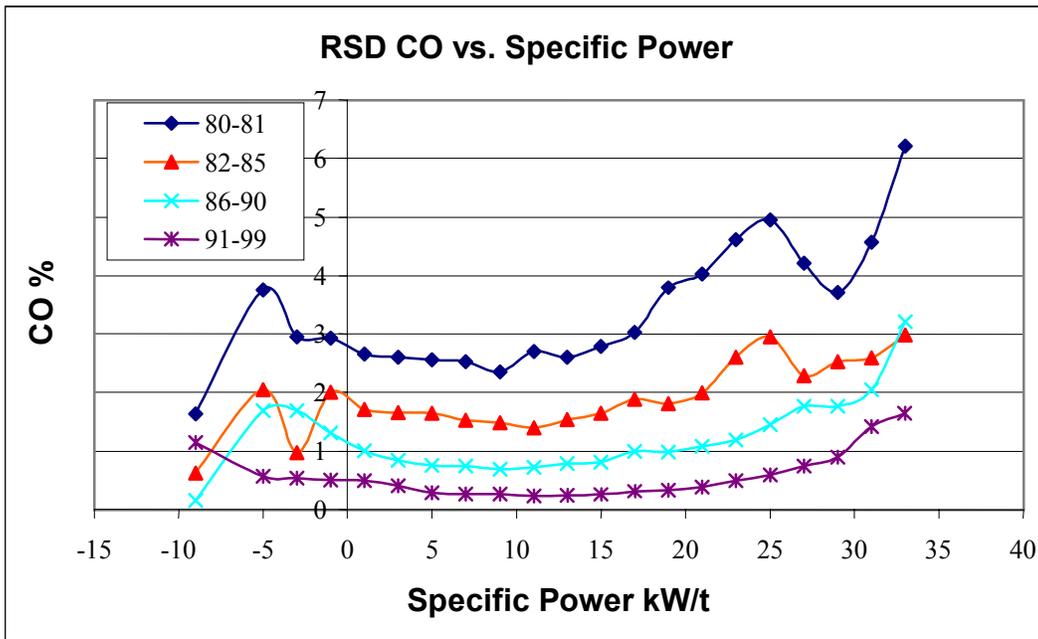(Denver Remote Sensing Clean Screen Pilot 12/99)

Figure 5.2 RSD %HC (C6) vs VSP
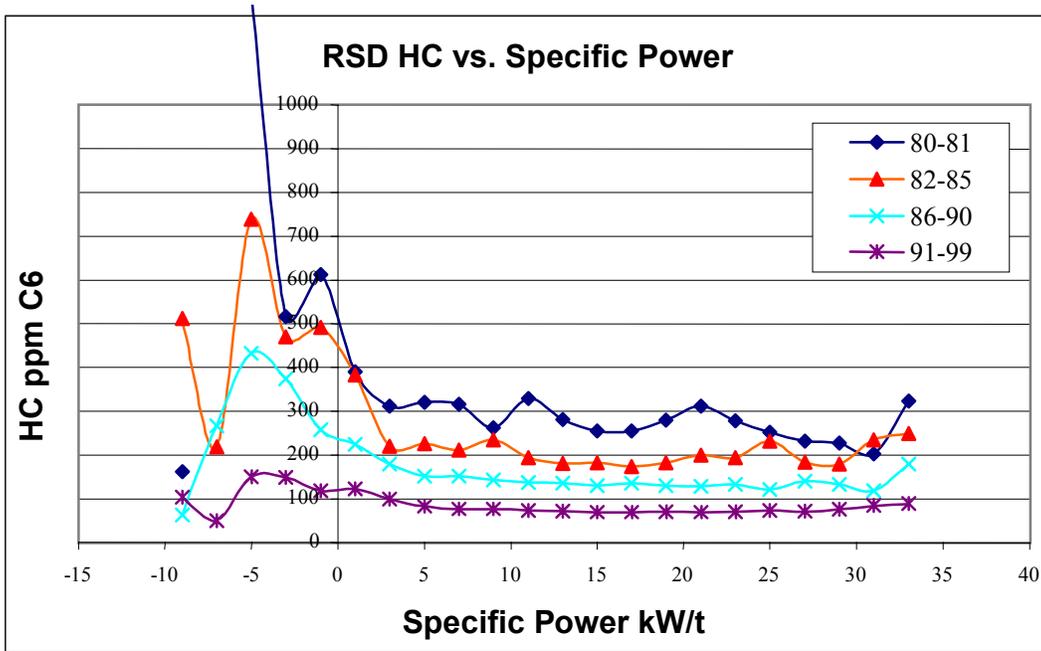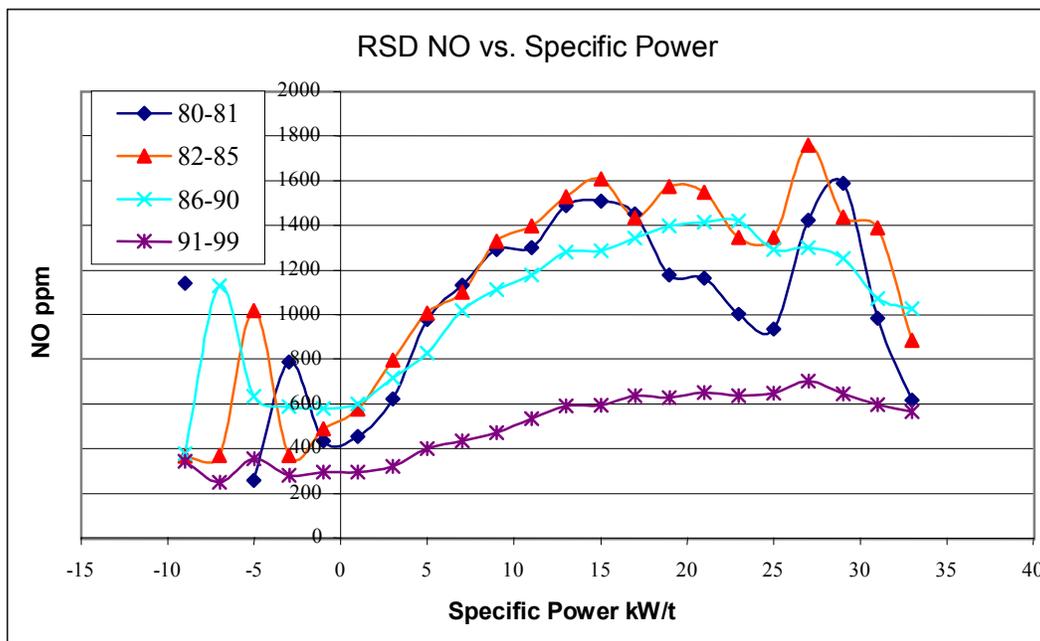(Denver Remote Sensing Clean Screen Pilot 12/99)

Figure 5.3 RSD %NO vs VSP
(Denver Remote Sensing Clean Screen Pilot 12/99)

5.4. Vehicle Identification

Goal: Identify vehicle license plate so RSD emissions may be linked to specific vehicle and I/M test result if available.

Optical character recognition is commonly used to read license plates in RSD studies; however, car must be taken to ensure these data are accurate. The license plate's design or color scheme may adversely affect the accuracy of the data, and this would obviously result in errors in linking the RSD reading with the correct I/M test result. If manually entry is to be used to enter license plate data into a database, procedures should be developed to identify and correct transcription errors.

It must also be understood that depending on a state's infrastructure regarding vehicle registration tracking and ease of access to the I/M test database, matching the RSD data with the appropriate I/M test result can be more difficult than anticipated.

5.5. Instrument Calibration
Goal: Ensure RSD units are calibrated using standardized procedures.

More detailed calibration specifications are provided in Section 4.5; however, it should be noted that the accuracy specifications on instruments may have a greater range than the differences between fleets, so the instruments may meet specifications but still give significantly different results.   For example, if the CO specification is +/- 0.25%, at a typical fleet average of 1% CO, one system could be centered at 1.05% and another at 0.95%. Both are well within specification but would report a 10% difference in two identical fleets.

Several approaches are possible for identifying and correcting this problem. Not all may be feasible:

     i)   Examine unit certification and audit data to determine offsets.
     ii)   Run the units side by side to obtain comparative results.
     iii)   Compare emission distributions for new model years of vehicles whose emissions profiles are expected to be the same in both fleets.

5.6. Measurement Methods
Goal: Convert concentration based RSD measurements on individual vehicles to mass based fleet estimates.

Remote sensing measures emissions in terms of concentration ratios in the total exhaust, while I/M programs that use idle or ASM testing measure emissions concentrations. However, programs that use IM240 or IM240-derivatives use concentration readings, air flow and miles driven on a dynamometer to calculate mass emissions. Therefore, fuel consumption data for an area may be used with fleet average RSD or ASM measurements taken in units of g/kg fuel to determine the fleet average emissions or the fleet average emissions could be converted to g/mi values by using instantaneous vehicle fuel economy estimates.

Also, as mentioned earlier (Section 3.3) areas conducting IM240 or ASM testing should plot mean RSD emissions against mean initial IM240/ASM emissions by vehicle type and model year. These plots typically show a linear relationship with high correlation coefficients and can be used to establish a direct relationship between the RSD measurements and the I/M test results.

5.7. Socioeconomics
Goal: Minimize the socioeconomic influence on data collection so that the I/M program benefits are quantified and not the socioeconomic differences that exist between fleets due to income.

It is believed that the vehicles owned by relatively low-income drivers tend to have higher emissions, from a combination of vehicle age and mileage, model, and historical maintenance practices.   Researchers have found that vehicle owner socioeconomics can affect vehicle emissions independent of even vehicle type, age, and model (_19_). Specifically, in one study CO and HC emissions were found to be roughly 25% higher in Lynwood CA than in El Monte CA[20].

The socioeconomic background of the drivers of vehicles measured by remote sensing can be quite different depending on where the instrument is located.

The effect of driver socioeconomics on remote sensing emissions can be identified by graphing average emissions by vehicle type and age for each measurement site, after correcting for different load conditions at each site. Driver socioeconomics must be considered when selecting sites for remote sensing measurement. If measurements from different sites are to be compared, such as under the Reference Method, sites with similar driver socioeconomics should be used. One method to determine if a true cross section of vehicles is being sampled is to plot the percentage of RSD measurements vs. ZIP code[*].

If it is discovered that the differences in fleet emissions between two sites are due primarily to socioeconomic factors, there is no easy way to deconvolute the existing data. Therefore, this issue should be addressed in the planning phase before any data is collected.

5.8. Seasonal Effects
Goal: Minimize the influence of seasonal variables on data collection.

Since no existing I/M programs vary their cutpoints vary by season, seasonal effects may influence a vehicle's measured emissions and therefore whether it passes its I/M test. However, the seasonal effects impact vehicle operations independently of whether emissions are measured by in-program analyzers or RSD. Therefore, a seasonal effect may introduce a bias when comparing, for instance, remote sensing measurements taken during two distinct time periods.

Vehicle emissions as measured by the Arizona program vary by season as depicted in Figures 5.4-6. Figure 5.4 shows the daily average CO of initial IM240 tests of Arizona passenger cars over a three year period (filled circles, left scale). Emissions of cars that are fast-passed or fast-failed are extrapolated to their full IM240 equivalents. The trend in the maximum daily temperature is also shown (gray lines, right scale). The solid vertical lines denote the calendar years, whereas the dashed vertical lines denote the changes in fuel. CO, and HC, are higher in warmer summer months; while NOx shows the opposite seasonal trend, and is higher in winter months. Colorado IM240 data show similar seasonal patterns.

It is unclear whether the seasonal variation is due to a combination of seasonal temperatures and changes in fuel composition, or to inadequate conditioning of vehicles prior to testing. The seasonal variation in Arizona remote sensing (Figure 5.5) and loaded idle (Figure 5.6) data appears to mirror that of the Arizona IM240 emissions, suggesting that vehicle conditioning is not the cause of the variation. However, the seasonal variation in CO and HC in the Wisconsin IM240 program (Figure 5.7) and the Minnesota idle program (Figure 5.8) are in the opposite direction, that is, CO and HC are higher in winter months. (The trend in Wisconsin NOx follows that of Arizona and Colorado.) More analysis is needed to better understand these seasonal

---

[*] Other parameters may be used to segregate the data such as IM area, previous IM test result or MY. A specific example in which IM area was used may be found in the June 19, 2000 Inspection & Maintenance Review Committee Report, "Evaluation of the Enhanced Smog Check Program", Appendix F.

trends, and why they differ by area; however, these trends can be identified using RSD and should be discussed as a component of an IM program evaluation.

Average emissions can be plotted by time periods (preferably weeks or days) and compared with average temperatures and fuel seasons to determine if there is a seasonal variation in remote sensing and/or I/M emissions. To reduce any seasonal effect on emissions, remote sensing measurements for the Reference Method should be made during roughly the same time period.

Figure 5.4. Daily Average CO, Arizona IM240

Daily Average CO (adjusted), Initial Tests of Passenger Cars
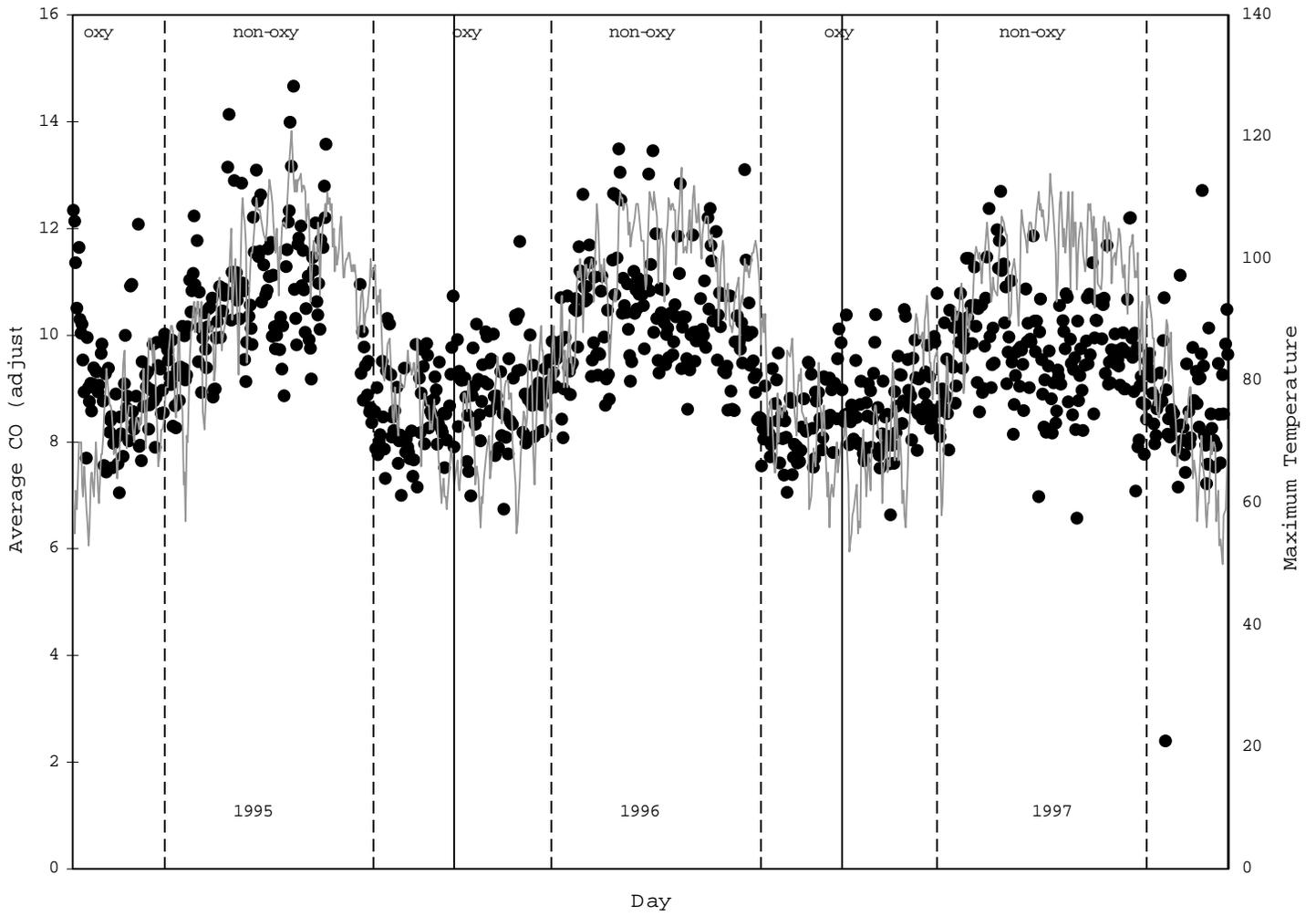1995-97 Arizona IM240

Figure 5.5. Daily Average CO, Arizona Remote Sensing

Average Remote Sensing CO, by Day
1996-1997 Arizona
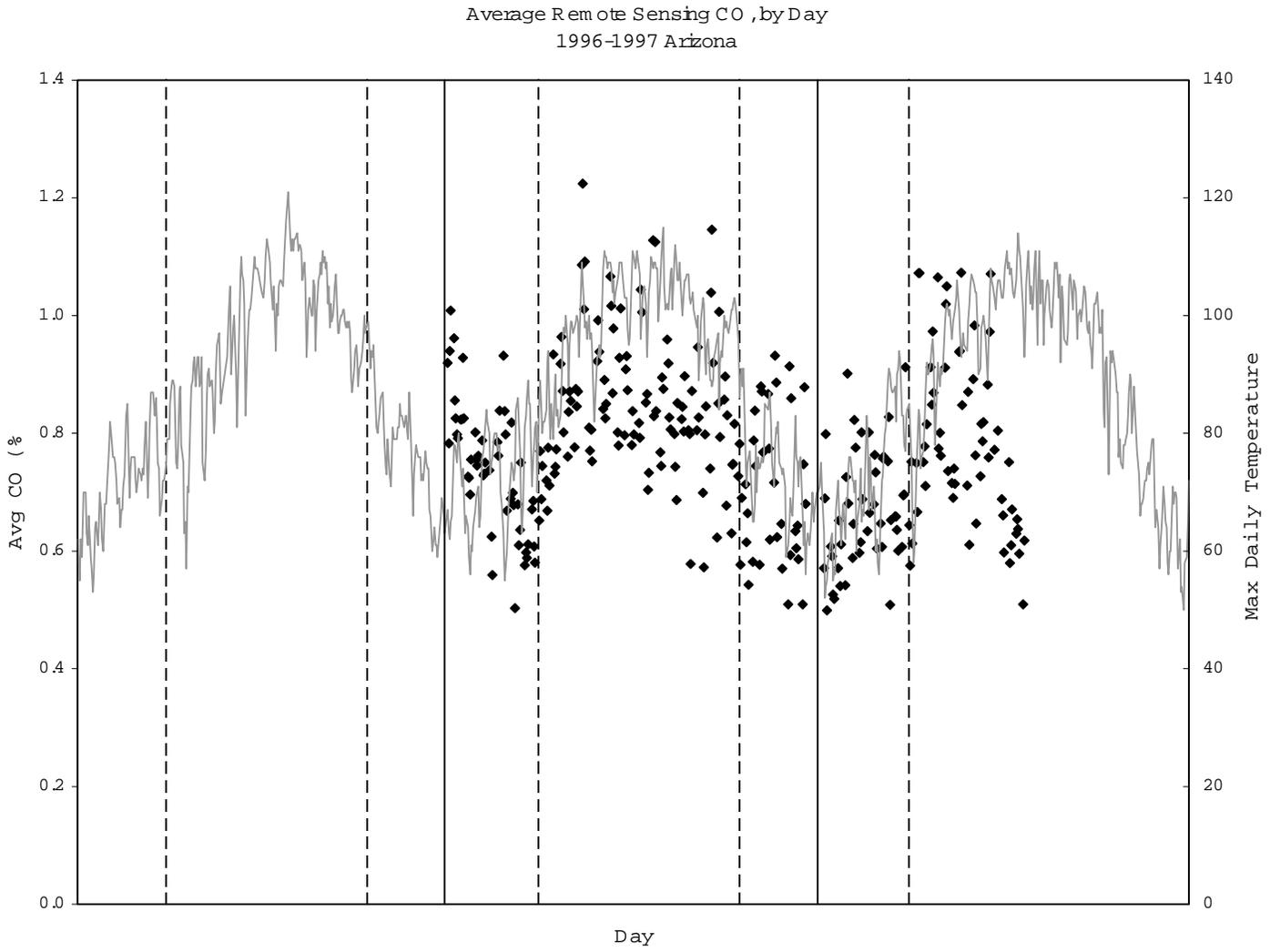
Figure 5.6. Daily Average CO, Arizona Loaded Idle (Pima County)

Average Loaded Idle CO, by Day
1995-97 Arizona

Figure 5.7. Daily Average CO, Wisconsin IM240

Daily Average CO, Initial Tests of Passenger Cars
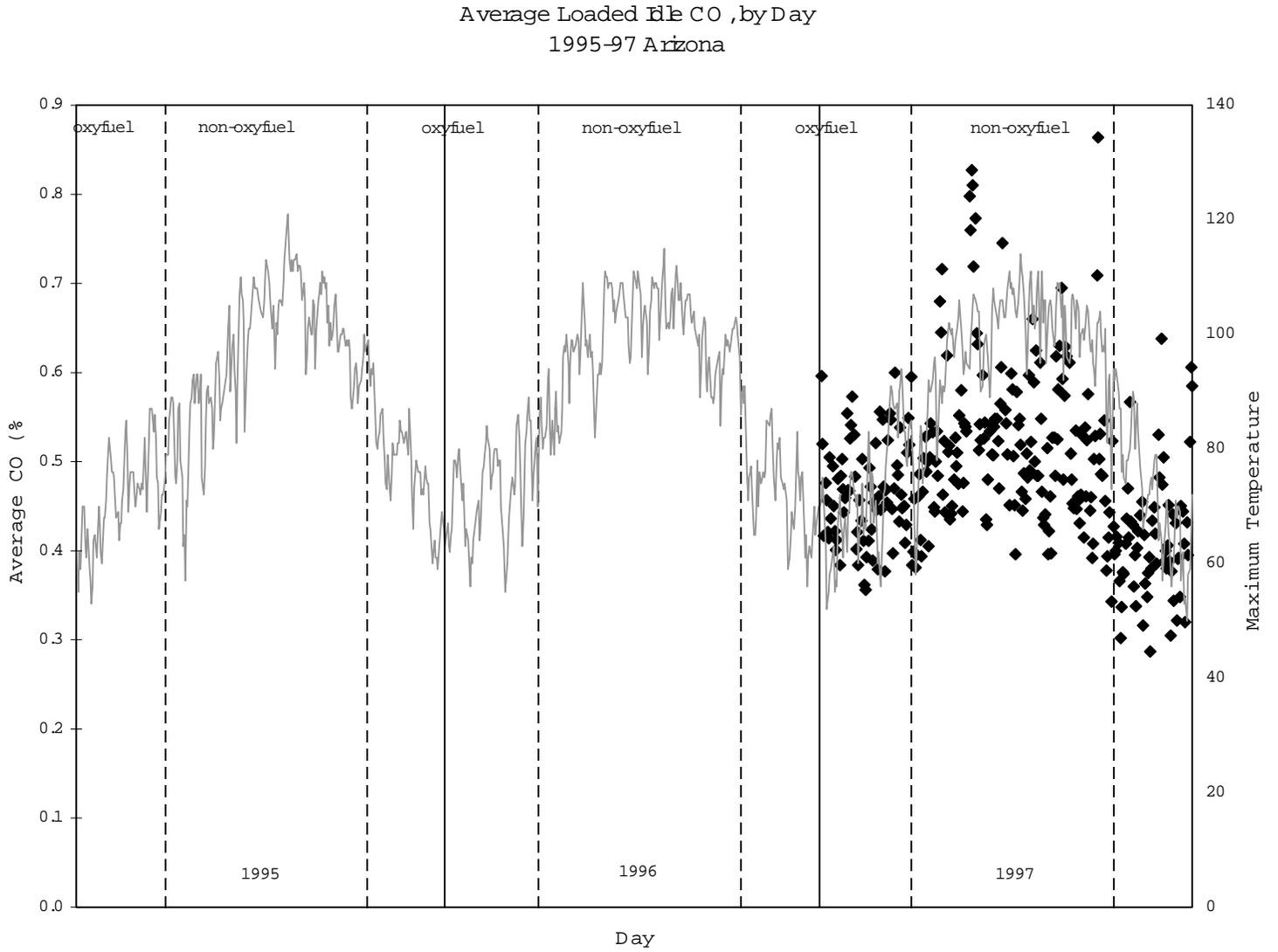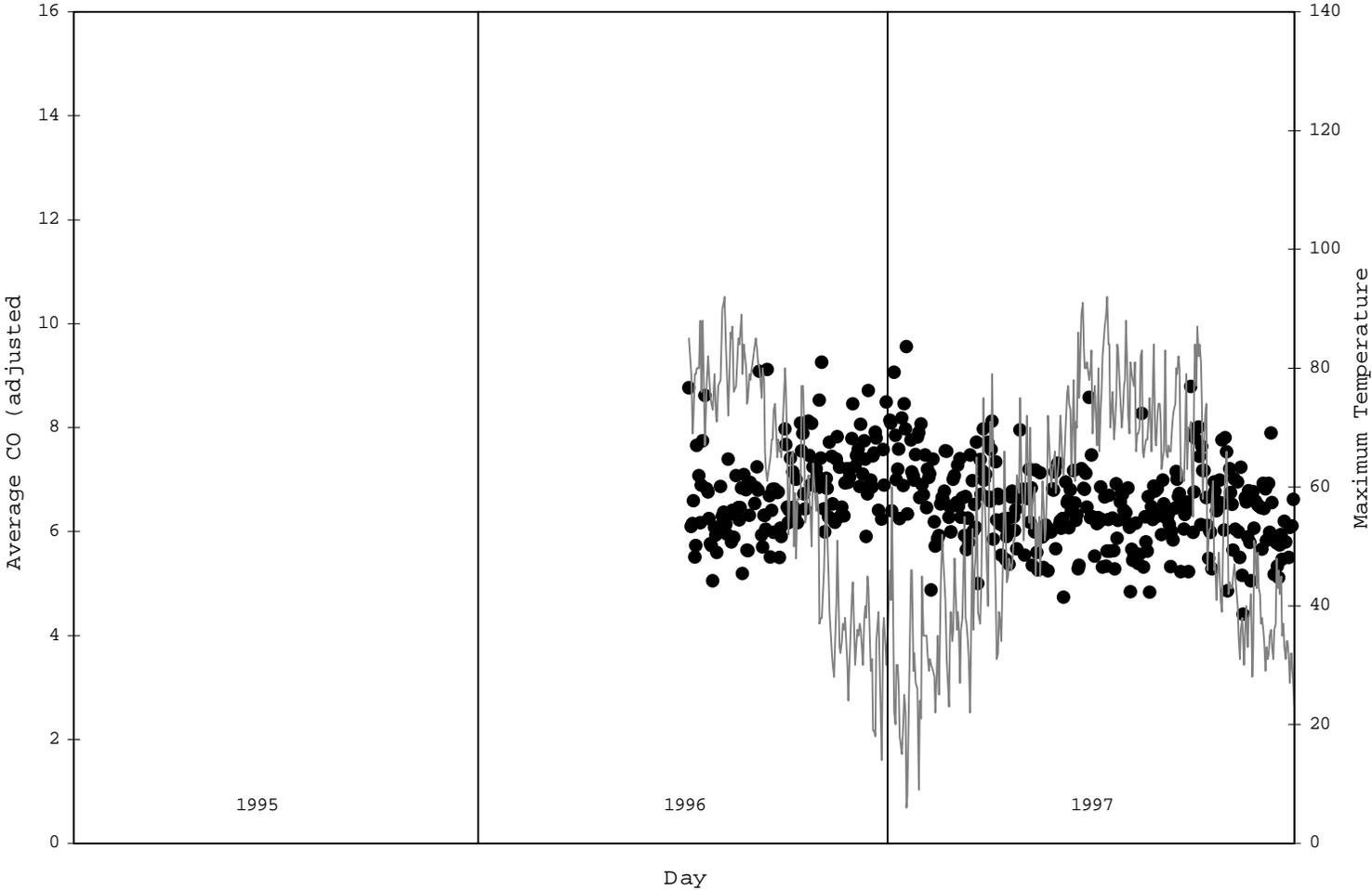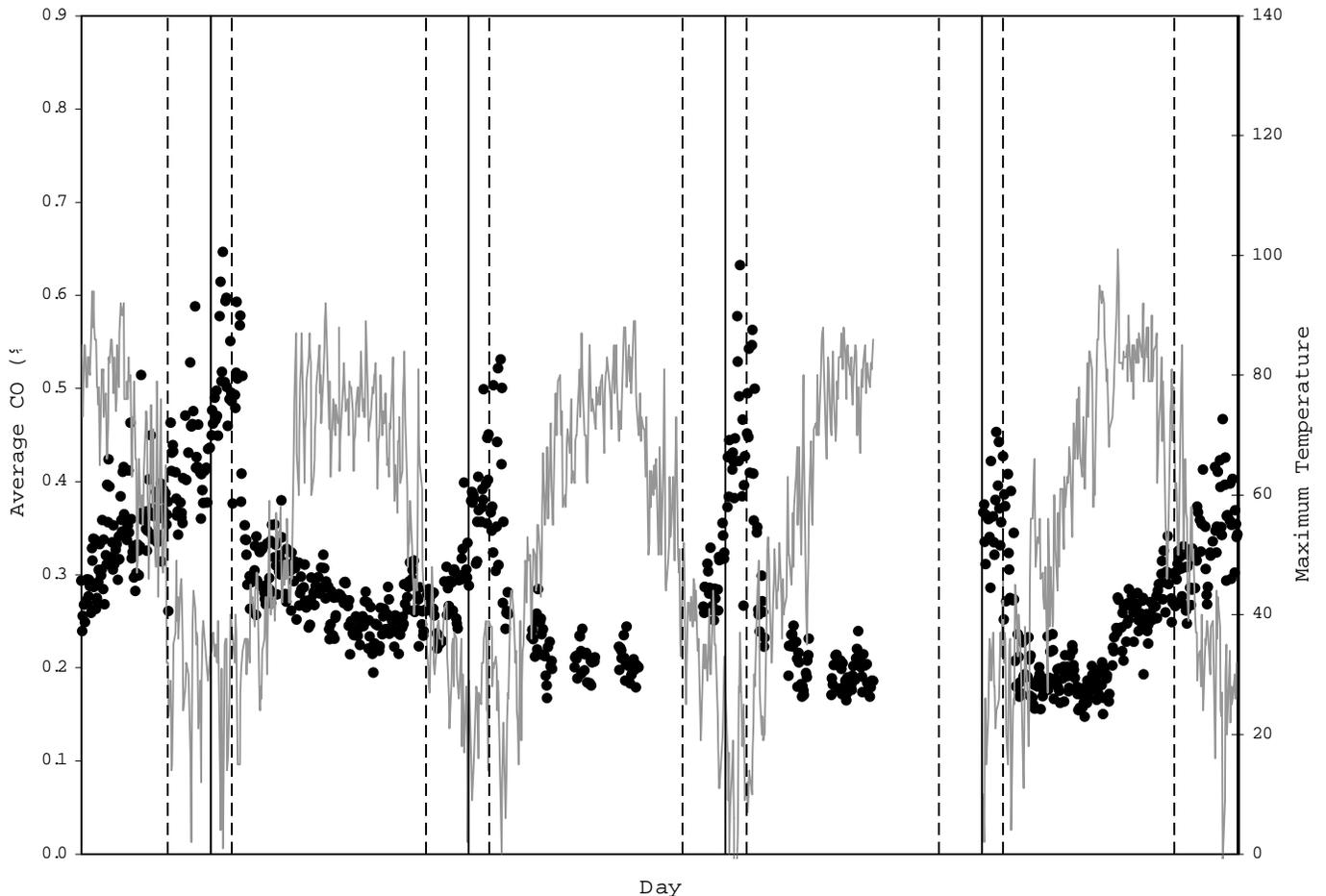1996-97 Wisconsin IM240

Figure 5.8. Daily Average CO, Minnesota Idle

Daily Average CO, Initial Tests of Passenger Cars
1991-95 Minnesota Idle

5.9. Program Avoidance
Goal: Account for emissions from motorists who are avoiding the I/M program.

There is evidence that I/M programs are inducing owners to re-register their vehicles outside of I/M areas (*10*, *21*). If these re-registrations are legitimate, i.e. drivers relocating their residences or selling their vehicles to new owners outside of the I/M area, then the program has helped to reduce emissions in the I/M area. However, there is evidence that a portion of these re-registrations are attempts to avoid I/M testing and many of these vehicles continue to be driven in the I/M area (*12*, *22*). Studies have estimated that program avoidance can lower the apparent CO reductions on the order of 2% (*10*, *12*). This program avoidance complicates any evaluation of an I/M program, in that analysis of I/M data would indicate emissions reductions (vehicles

leaving area) that are not occurring on the road. As discussed above, remote sensing data can include such vehicles in their estimate of fleet emissions. In addition, remote sensing data can be used to identify the subset of vehicles that are no longer registered in the I/M area but continue to be driven in the area.

The design of a remote sensing program itself can influence which vehicles are measured under the program. A program which provides a negative incentive, such as additional I/M testing, for driving past a remote sensor may encourage drivers to avoid having their vehicle measured by a remote sensor. On the other hand, a program that is intended for research purposes only, or provides only a positive incentive (the possibility of being exempted from the next I/M test), will result in a more representative sample of vehicles measured.

The distribution of vehicles (by type and model year) measured by remote sensing should be compared with the distribution of vehicles registered in the area, or reporting for I/M testing. Any differences between the two distributions may indicate a bias in one of the samples and suggest a possible program avoidance issue that needs to be addressed. However, care must be taken when performing this comparison because RSD measurements reflect on-road driving distributions while traditional I/M testing is registration based. Therefore, it is possible that RSD will over-sample newer vehicles relative to a registration-based I/M program.

5.10. Regional Differences (policies, environment, fuel composition, etc.)
Goal: Account for differences in fleet emissions not attributable to I/M across geographic regions.

A number additional variables, such as environmental conditions, fuel composition, vehicle registration, safety inspection, public attitudes, and tax policies, etc., can result in biases in emissions measurements made in different regions. These biases would have the biggest impact on an evaluation using the Reference Method. Some of these potential biases and methods for minimizing their impact are discussed in more detail in the Reference Method section below.

5.11. Program Details
Goal: Identify and account for I/M program operation details in the program evaluation analysis.

Biennial I/M programs use a simple technique to determine which vehicles are to be tested in which year. For example, Colorado requires vehicles of even model years to be tested in even calendar years, and vehicles of odd model years to be tested in odd calendar years. Arizona bases a vehicle's test year on whether the last digit in the vehicle identification number is odd or even. Different states have different policies regarding whether I/M tests are required when a vehicle changes ownership, and the circumstances under which a vehicle's registration date and year changes. Additionally, vehicles that are newly registered in AZ or CO must be tested when they are first registered, regardless of their model year or last digit in their VIN. These factors complicate the determination of whether a particular vehicle has been tested under the current I/M program or not. Therefore, it is essential that the date of each vehicle's last I/M test be used to determine whether the vehicle has been tested under the current I/M program.

States may also often have different policies regarding vehicle license plates. The license plate of a car sold in Colorado stays with the original owner, whereas the license plate is transferred to the new owner in Arizona. These policies may complicate the matching of license plates observed by remote sensing units with the correct vehicle and I/M test result information. These and similar details of registration and I/M programs should be understood to minimize the misidentification of the tested and untested vehicle fleet.

5.12. Emissions Distributions
Goal: Identify possible sources of bias in the measured emissions.

One way of determining whether emissions measurements are biased is to compare average emissions by vehicle type and model year, as described in Section 5.2. If average emissions by model year are consistently higher for one group of vehicles than another, then the emissions of that group of vehicles may be biased by some of the factors discussed above. Another approach is to compare the distribution of emissions of a subset of similar vehicles in the I/M-tested and untested fleets. Because vehicle emissions are highly skewed, with a majority of vehicles with low emissions and a small number of vehicles with very high emissions, differences between groups of vehicles will be more readily apparent if the distribution is plotted on a log-normal scale. Three ways to compare emissions distributions are outlined below.

i) One way of looking for changes in the shape of emissions distributions is to look at the contribution of the dirtiest 10% of vehicles which contribute a large percentage of the total emissions. Table 5.3 illustrates the contribution of the dirtiest 10% of vehicles in each model year. The vehicles' CO emissions were measured at multiple RSD sites that have been divided into three groups based on the mean vehicle specific power of vehicles measured at the site. The percentages in Table 5.3 show that the percentage of emissions concentrated in the dirtiest 10% of vehicles is greatest among the newest model vehicles that have a smaller number of high emitters.

ii) Another method is to divide vehicles into equal groups (quintiles or deciles), and plot the average emissions of each group. Decile plots focus attention on the majority of vehicles that have relatively low emissions. Figure 5.9 is a decile plot using the same data as Table 5.3 for 2000 model vehicles; however, now it becomes easier to distinguish between the low and high emitting vehicles.

iii) The third method is to plot the full distribution of vehicles, rather than quintiles or deciles; the full distribution allows closer examination of the differences in the small number of high emitters in two samples of vehicles. Figure 5.10 is a full distribution plot of the data shown in Table 5.3 and Figure 5.9 for 2000 model vehicles. The use of a logarithmic scale highlights the difference among the few high emitters in each data set.

Table 5.3. CO Emissions 10% dirtiest by MY

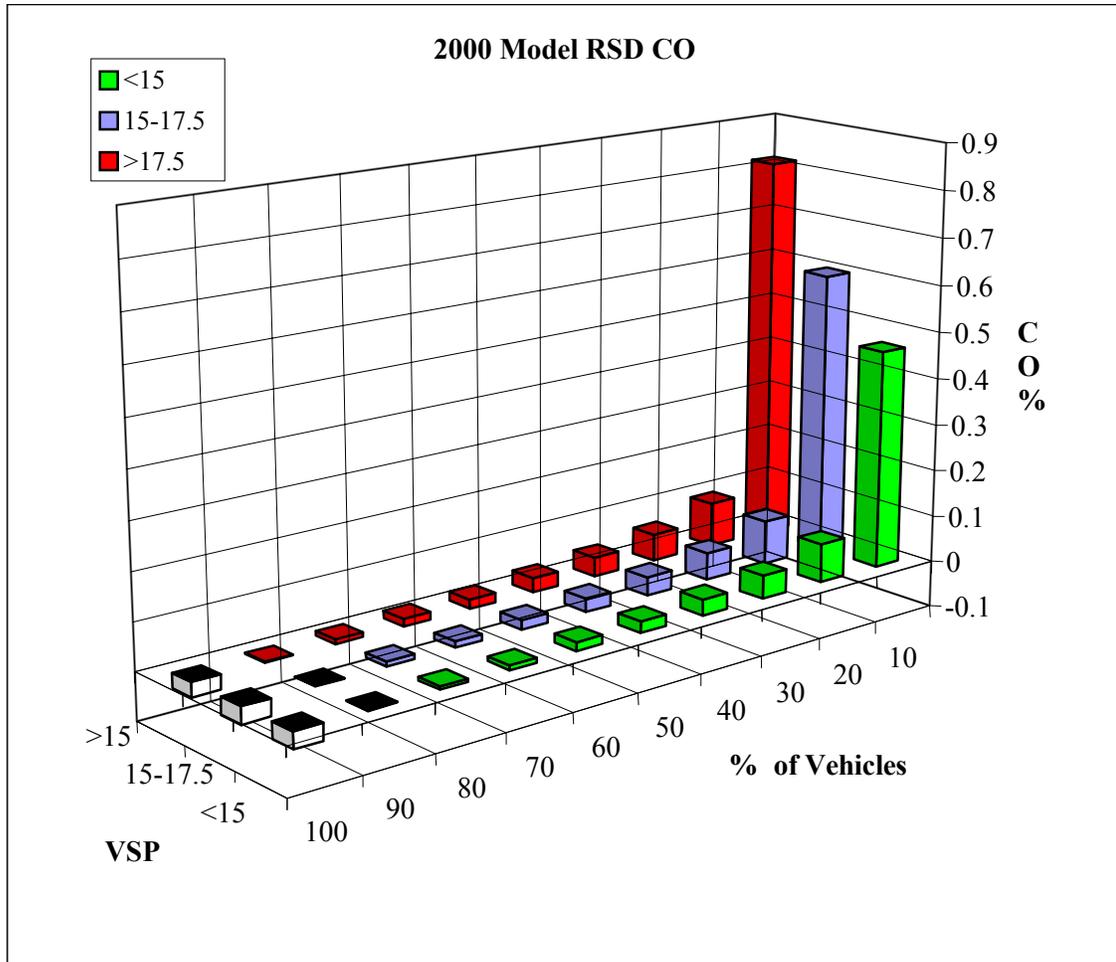| Year	Site VSP (kW/t)		
	<15	15-17.5	>17.5
1976-80	46%	43%	39%
1981	47%	44%	42%
1982	50%	44%	44%
1983	51%	48%	48%
1984	54%	48%	48%
1985	54%	50%	48%
1986	60%	56%	53%
1987	61%	57%	55%
1988	62%	60%	59%
1989	62%	60%	59%
1990	60%	60%	61%
1991	61%	62%	63%
1992	61%	61%	64%
1993	60%	60%	63%
1994	59%	60%	65%
1995	61%	63%	67%
1996	61%	61%	67%
1997	63%	63%	68%
1998	65%	66%	72%
1999	68%	68%	74%
2000	71%	73%	77%
2001	72%	75%	77%

Figure 5.9. CO Emissions Decile Plots

2000 Model RSD CO

Figure 5.10. CO Emissions Full Emissions Distributions
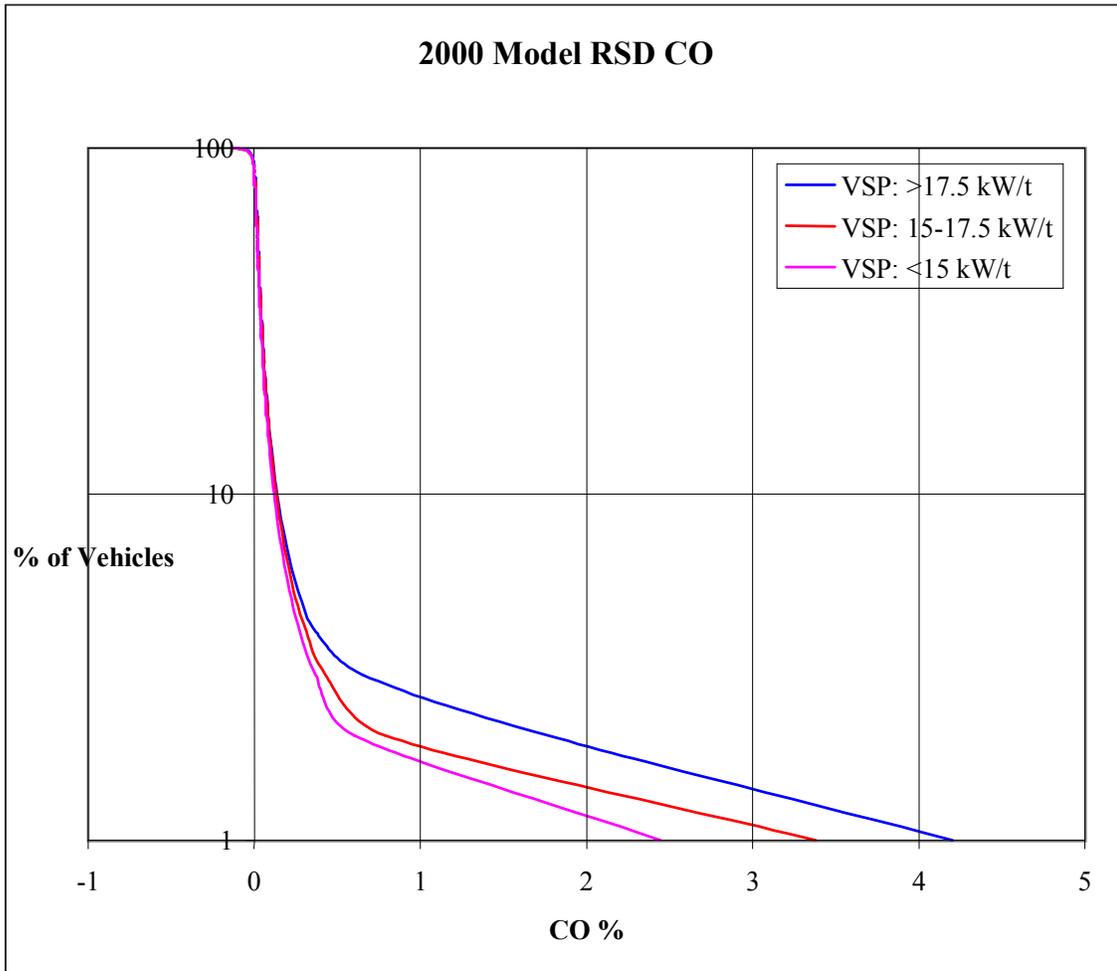
6. Evaluation Methods

This section outlines three methods to use remote sensing data to analyze I/M program effectiveness over the short term. The first two methods, the Step Change and Comprehensive methods, involve remote sensing measurements collected in an I/M area; the final method, the Reference Method, compares remote sensing measurements collected in an I/M area with measurements collected in an external, or reference. Each of these methods is described in more detail below.

6.1. Step Change Method

6.1.1. Description

There are several reasons for performing on-road emission reductions independent of an I/M program. New technology vehicles are lower emitting for a given fleet age than older technology vehicles. Depending on local and national economic factors, the fleet age itself may be changing (newer vehicles are lower emitting) and it is possible that public education or willingness to carry out required maintenance is less compliant than anticipated, and the auto repair industry capability are improving irrespective of the presence or absence of an I/M program. All these factors make it important not only to measure the on-road emission reductions of the I/M fleet, but also to measure the emissions of a well matched control fleet, preferably differing only in I/M status.

The Step Method is an on-road evaluation of new or changed I/M programs using a built-in representative control group. On-road emissions are the parameter which I/M programs are intended to control[*]; however, most I/M programs emphasize testing of fully warmed-up exhaust emissions. If I/M exhaust emissions failure is followed up by successful repair, scrapping the vehicle or relocating it to a region from which it is rarely driven in the program area, then the program should show on-road exhaust emission reductions. When a new I/M program starts or when there is a major program change, then there is a window of opportunity to evaluate the effectiveness of that change in reducing on-road emissions. That window arises when the new (or changed) program has impacted about 50% of the local fleet. If an annual program starts, then the window is after about six months. In a biennial program the window is after the first year. The concept behind this evaluation is that the untested fleet serves as the representative control group for the tested fleet. Ideally, data collection should be carried out at a sufficient number of sites in the area to ensure appropriate representation and sampling should include surface streets as well as highway on/off ramps; however, a single well-traveled site can be representative of an I/M area. As mentioned in Section 5.6, one method to determine if a true cross section of vehicles is being sampled is to plot the percentage of RSD measurements vs. ZIP code.

---

[*] It is tacitly assumed on-road emissions are controlled by linking the I/M standards to certification standards, as vehicle emissions shouldn't be expected to be reduced below their certification levels. Whether this strategy is appropriate or valid with respect to reducing on-road emissions and improving air quality is a discussion beyond the scope of this document.

6.1.2. Application Examples

Colorado had various versions of decentralized idle/2500 tests since the early 1980s and switched in the Denver metro area to a biennial centralized IM240 based program on January 1 1995. Because the program is biennial, by January of 1996, roughly half the measured fleet (odd MY) had been through the new I/M program and the other half (even MY) had missed a year of their old annual program. On-road monitoring was carried out for five days in January of 1996 at a single heavily trafficked site. Approximately 26,000 valid, plate-matched records were obtained.

Data were collected at a freeway off-ramp to eliminate cold-start vehicle emissions. Vehicle load was not measured as it was assumed tightly curved uphill ramps have little off-cycle power-enrichment, and the tested and untested MY are randomly interspersed and subject to the same loads thus making for a valid comparison independent of load. Additionally, the VSP concept was at best in the developmental stage. However, EPA strongly would recommend vehicle load be characterized using VSP as described earlier for all program evaluation studies.

DMV records provided county of registration, I/M eligibility and most recent I/M status (pass, fail, or waiver). Individual emission data bases are not normally distributed; however, if one treats the means from each measurement day as an independent sample then these sub-samples can be analyzed using normal statistics. This resulted in 5 means (1 for each day) per fleet. For a fleet of about 26,000 vehicles it was found that the uncertainty in the apparent emissions benefits is $\pm$2%. This error would be reduced with a larger fleet size provided that approximate equality between tested and untested vehicles could be maintained.

The first analysis was "eligible and certainly tested" versus " eligible in the future but not tested" giving an apparent 7$\pm$2% CO benefit. During this first analysis it was recognized that many vehicles should have been tested but were not, so a second analysis was "should have been tested" versus "not tested". This reduced the apparent benefit to 6$\pm$2%. Approximately 1300 vehicles registered in locations not required to take the I/M test were also measured at one site and these vehicles showed higher average on-road emissions. However, they also showed an alternation of emissions by MY as if the I/M program had caused failing vehicles to be reregistered to outlying counties but yet continue to be driven in Denver. A follow up study a year later confirmed that indeed this effect is happening and, when included for that site, reduces the apparent benefit by 2%. The contribution of these "repair avoidance" cheaters to the basin wide fleet emissions cannot be determined from one freeway interchange site, but their emissions were large enough that at the measurement site the 6$\pm$2% apparent I/M CO benefit was reduced to 4$\pm$2% (*10, 12*).

The same database actually allowed for two other I/M benefit tests of lower precision. Using only the even MY vehicles, the on-road emissions of those tested versus those untested was evaluated. This resulted in a 5$\pm$3% apparent I/M CO benefit. Evaluation of the difference in on-road emissions between vehicles of all MY tested within four months before the measurement time and two months after indicated an apparent 8$\pm$6% benefit for CO. On-road benefits for HC and NO were insignificant. The analyses discussed above were published in the literature (*10*).

Several factors obscured the clean 50/50 split between untested even MY and tested odd MY in these studies. For instance, many 1994 MY vehicles were tested in 1995, 1995 and later MY new vehicles obtained a four-year I/M waiver, and all vehicles had to take the I/M test upon change of ownership regardless of MY. However, many of these potentially confounding factors can be corrected.

6.1.3. Potential Systematic Errors

A major advantage of a single-site, single time I/M evaluation study is that instrument calibration and vehicle load/speed are irrelevant since both fleets are subject to the same measurement system. A second advantage is the measured and the control fleets are perfectly matched socioeconomically. A third advantage is that the evaluation can be carried out with only a single week of work to within 2% accuracy levels, and the fleet average remote sensing data has been shown to correlate very well with fleet average IM240 data (*10*).

However, three disadvantages are apparent; one that the window of opportunity is only when a new program starts up or a program change which is predicted to have measurable effect is initiated; the second that the reference group of untested vehicles may not be a correct reference; the third is added diligence is needed to ensure a representative sample is obtained.

There is some evidence that change of ownership vehicles have higher emissions than the average of the same MY. This effect would cause the average of the untested even MY vehicles (the control group) to be biased low and thus cause an underestimate of apparent I/M benefit. It is possible to attempt to correct for this bias*23*. This study eliminated the large sample of 1994 MY vehicles which had been tested because they were very numerous and certainly a few months older than the untested (last quarter) of the 1994 MY. These two effects both lower the apparent emission of the untested fleet, thus increasing the apparent I/M benefit from the previous 4%-7% range to 8%-11% with the same $\pm$2% error. The last two analyses are not effected by these corrections and remain at 5 ± 3% and 8 ± 6% apparent I/M benefit for CO (*12*).

There had been an annual I/M program in place in Denver for more than ten years. The odd MY fleet took the old test in 1994 and the new in 1995. The untested even MY fleet skipped testing in 1995 because their scheduled IM240 was in 1996. If the old program had no benefit, then this skip introduces no bias. If the old program had emissions benefits which last a long time (long repair lifetimes as in the EPA Mobile model) then no bias is introduced, but, the apparent benefit is that of the new program relative to the older one; not relative to a "no I/M" baseline. To the extent that repair lifetimes are not as long as modeled by EPA and the old program did lead to reduced emissions, then the skipped annual test moves the control group back toward the no I/M line, thus overestimating the I/M benefit relative to the previous program but with the upper limit being relative to no I/M.

To correct for this bias, one needs to estimate both the emission reductions from the previous (idle/2500) program and the apparent repair lifetime, but this is not straightforward. If from the DMV records one can determine which tested odd MY vehicles were not changing ownership, then the even MY bias is removed and the study measures the apparent I/M benefit for the fleet which does not change ownership.

6.2. Comprehensive Method

6.2.1. Description

The Comprehensive Method involves comparing remote sensing emission measurements of a fleet of vehicles measured prior to initial I/M testing with those of a fleet of vehicles measured after final I/M testing. The difference in fleet average remote sensing emissions is the initial percent reduction due to the I/M program. Sufficient numbers of measurements are made so that emission reductions can be evaluated by vehicle type and model year, and by I/M result. Important observations about repair effectiveness and program avoidance can be made if enough vehicles are measured.

One of the main reasons for using remote sensing measurements to evaluate the effectiveness of I/M programs is that remote sensors measure emissions of vehicles that may not be participating in an I/M program. The Comprehensive Method differs from other remote sensing methods, in that it explicitly compares emissions reductions of the I/M tested fleet as measured by the program and as measured independently by remote sensing. The Comprehensive Method can also be used to compare the emissions of the I/M-tested fleet with those of the non-I/M-tested fleet, as can the other methods.

6.2.2. Application Examples

The Comprehensive Method concept was first applied by Doug Lawson, using unscheduled roadside idle testing of randomly selected vehicles from CARB's 1989, 1990, and 1991 random roadside surveys. Lawson found that average emissions levels of vehicles tested prior to their I/M test were about the same as those of vehicles tested after their I/M test. The emissions levels measured during the scheduled I/M tests were 60% less than the emissions measured during unscheduled testing either before or afterwards (*24*). The analysis was limited, in that fewer than 5,000 vehicles were analyzed in any given year.

Radian International was the first to apply this method using remote sensing data, in a 1997 evaluation of California's I/M program for the California Bureau of Automotive Repair (*25*). For their analysis Radian had access to over 3.5 million RSD measurements from the Statewide On-Road Emissions Measurement System. Because of concerns regarding the accuracy of some of the RSD instruments, the first 6 months or so of RSD data were not included in the analysis (the report gives no indication of how many measurements, or vehicles, were involved in the analysis). Radian also excluded RSD measurements taken at sites that had a relatively high percentage of high emitting vehicles from the newest model years. Radian grouped the RSD measurements into two time periods: 30 to 90 days prior and 0 to 90 days after. Radian also grouped vehicles by model year group and I/M outcome (initial pass, initial fail/final pass, initial fail/no final pass). However, despite the large sample size, Radian did not have enough remote sensing measurements to compare pre- and post-I/M remote sensing emissions of the same vehicles (that is, a total of three emissions measurements on each vehicle), so the cost associated with such a study should not be underestimated.

More recently the Comprehensive Method was used by Lawrence Berkeley Laboratory in analyzing 4 million remote sensing measurements on 1.2 million vehicles in the Phoenix I/M area (*18*, *22*, *26*). It was found that initial emissions reductions as measured by remote sensing were roughly half that as measured by the initial and final IM240 tests; IM240 data indicated a 15% reduction in fleet-wide CO and HC emissions due to the program (g/mi units), while the remote sensing data indicated only a 7% reduction in CO and an 11% reduction in HC emissions (g/kg fuel units). Because there is a small gas mileage benefit to CO and HC emissions reductions the per mile emission reduction as measured by RSD would be slightly higher. For instance, assuming a 10% gas mileage improvement and a 10% emissions reduction after repair would increase the 7% CO and 11% HC g/kg fuel RSD measurements to 8% and 12% respectively. However, these values are still below the 15% reductions determined using IM240 data. Part of this discrepancy may be due to the different loads vehicles are subjected to under IM240 testing and remote sensing measurement. As in the earlier Step Method study, the VSP concept was still under development and not available as a tool to reduce measurement bias due to vehicle load.

The Comprehensive analysis found that average remote sensing emissions increased as vehicles got further from their I/M test; the initial 12% reduction in fleet-wide CO emissions less than one month after I/M testing declined to only a 6% reduction in fleet-wide CO emissions one year after I/M testing. In other words the repair benefits did not last nearly as long as they do in the I/M models. The Comprehensive analysis also found that average RSD emissions increase as vehicles get closer to their scheduled I/M test; this is especially true for vehicles that fail their initial I/M test. An analysis of emissions trends in the weeks prior to their initial I/M test indicates that the average emissions of these initial fail vehicles do decline slightly immediately prior to I/M testing, suggesting that pre-test repairs and/or adjustments are being made.

6.2.3. Steps

Under the Comprehensive Method, a large number of remote sensing measurements are taken at suitable sites throughout an I/M area. License plates from the remote sensing measurements are then matched with license plates either in a registration database, or in the I/M testing database. How remote sensing measurements are matched with vehicle information depends on how each state registers vehicles. For instance, some states (such as Arizona) assign license plates to vehicles; when a vehicle is sold, the license plate stays with the vehicle. In contrast, other states (e.g. Colorado) assign license plates to a driver and when a vehicle is sold the license plate stays with the driver and can be affixed to a new vehicle. It is critical that license plates obtained from remote sensing programs be matched to the correct vehicle, and in some cases this will require tracking a vehicle's VIN to link it with the appropriate I/M test record and then match it to the RSD measurement. It must be understood that depending on a state's infrastructure regarding vehicle registration tracking and ease of access to the I/M test database, this task can be very difficult.

The result is a large database of vehicles, some with multiple remote sensing measurements and multiple I/M tests (vehicles that fail their initial I/M test and return for subsequent testing). Vehicles are then classified into several groups, based on the results of their I/M test(s): 1) vehicles that pass their initial I/M test; 2) vehicles that fail their initial test but pass a subsequent

test; 3) vehicles that fail their initial test and do not receive a subsequent I/M test; and 4) vehicles that fail their initial test and fail a subsequent I/M test. Vehicles can be further categorized into more groups, based on the time between initial and final I/M test, or the results of their emissions test vs. the results of visual or functional I/M tests.
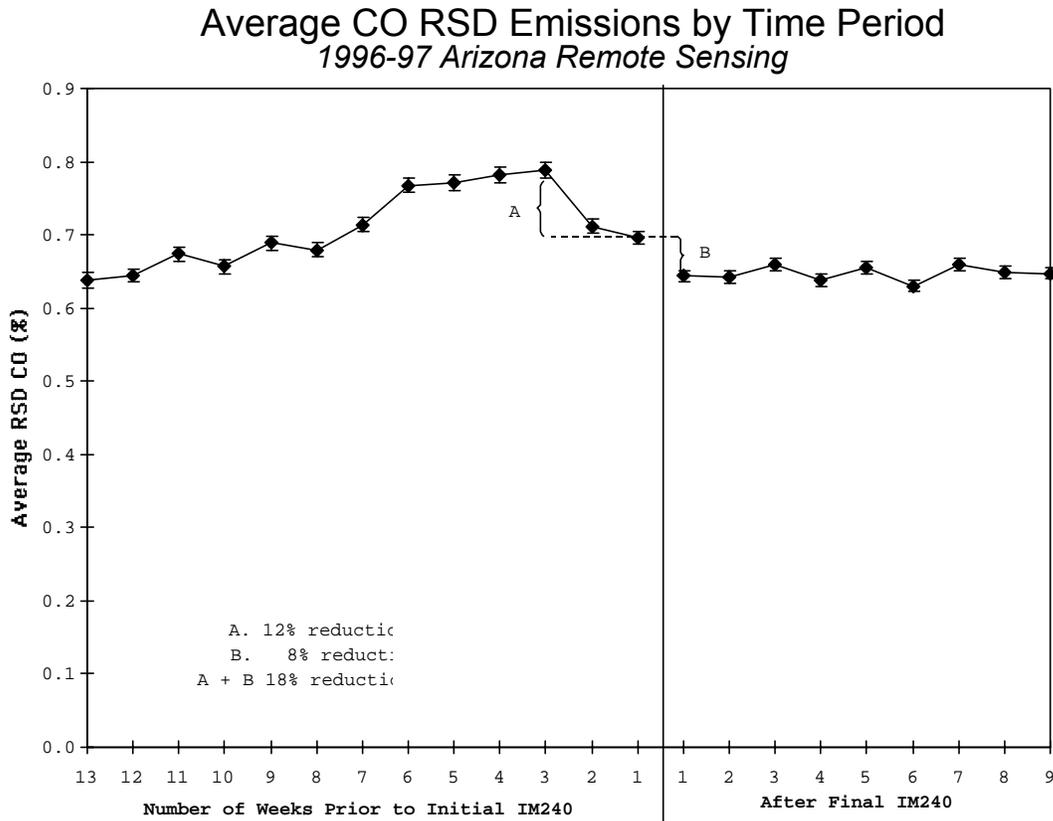
Individual records are then categorized based on the time between the remote sensing measurement and the initial or final I/M test. For example, individual remote sensing measurements of vehicles can be grouped into 3 month time periods (0 to 3 months, 3 to 6 months, 6 to 9 months) prior to the vehicle's initial I/M test and after the vehicle's final I/M test. These time periods can be shortened to as little as one month or one week, depending on the number of remote sensing measurements. Remote sensing measurements of individual vehicles with multiple measurements in a given time period can be averaged to obtain a single measurement for that vehicle in that time period, or can be treated as independent observations (meaning that some vehicles are "double-counted" in some time periods).

Given the size of the database collected in the Comprehensive Method, valuable insight into repairs and repair durability can also be estimated. Analyses should include calculating average emissions as a function of time period, I/M result, vehicle type and model year and plotting the results. To determine the initial effectiveness of the I/M program, only remote sensing measurements over a relatively short period should be used, to minimize the impact of changes to the vehicle on the results. For example, average emissions of up to 3 months prior to initial I/M test can be compared with average emissions of up to 3 months after final I/M test. The difference in average remote sensing emissions is the initial emissions reduction due to repair of many vehicles identified by the I/M program as high emitters (some of the emission reduction may also be due to vehicles passing a subsequent I/M test without any repairs being made.) The emission reduction can be calculated for the entire tested fleet to determine the overall impact on the fleet, as well as for subsets of the fleet with different I/M results, to determine the impact of the program on, say, vehicles that fail initial I/M testing. The initial emissions reductions as measured by remote sensing can then be compared with the initial emissions reductions as measured by I/M testing. The analysis can also be extended to time periods further after I/M testing to analyze the short-term durability of any repairs made under the I/M program.

There is evidence that some vehicles are repaired or receive maintenance just before their scheduled I/M test; these pre-test repairs may result in the initial I/M test underestimating the average emissions prior to I/M testing. This underestimation of the baseline emissions may in turn result in an underestimation of the effectiveness of the I/M program. The data in Figure 6.1 provide evidence that owners do perform maintenance prior to an I/M test and survey data indicated that 35% of vehicle owners brought their vehicle in for a tune-up prior to their initial test. The figure shows average weekly remote sensing CO emissions in different time periods before the initial, and after the final, I/M test of each vehicle. The figure indicates that emissions increase as vehicles get closer to their I/M test; however, emissions decrease substantially (12%) about three weeks prior to the initial I/M test. An evaluation based only on measurements taken immediately before and after I/M testing would estimate an 8% reduction in emissions. If the effect of pre-test repairs and adjustments are included, however, the reduction attributable to the program increases to 18%. To minimize the effect of pre-test repairs on baseline emissions, remote sensing measurements made within a month before a scheduled I/M test can be excluded

from the analysis (i.e., remote sensing measurements from 1 to 3 months prior to the initial I/M test can be compared with remote sensing measurements from 0 to 3 months after the final I/M test; Radian used this approach in their analysis of California RSD data).

Figure 6.1. Average CO RSD Emissions by Time Period, 1996-97 Arizona Remote Sensing

Average CO RSD Emissions by Time Period
1996-97 Arizona Remote Sensing

(chart: x-axis "Number of Weeks Prior to Initial IM240" (13 down to 1) and "After Final IM240" (1 to 9); y-axis "Average RSD CO (%)" from 0.0 to 0.9. Annotations: A. 12% reduction; B. 8% reduction; A + B 18% reduction)

6.2.4. Advantages/Disadvantages

There are several advantages to using the Comprehensive Method:

i)   The initial emissions reductions attributable to the program can be independently measured, and can be compared with those measured by the program itself.

ii)   The repair effectiveness over the short-term (i.e., up to 2 years after final I/M testing) can be independently measured. Short-term repair effectiveness can be compared with long-term repair effectiveness as measured using multiple years of in-program data on the same vehicles.

iii) The effect of pre-test repairs on average emissions can be measured.

iv) Because large numbers of remote sensing measurements are made, the Comprehensive Method allows the identification of vehicles that do not report for, or do not complete, I/M testing, yet are still being driven in the I/M area. Video camera surveillance can also be used to identify non-compliant vehicles, at less expense than remote sensing measurement; however, video cameras will only provide information on registration avoidance without any air quality data on high emitting vehicles

The primary disadvantage of the Comprehensive Method is that it requires a large number (on the order of millions) of remote sensing measurements. The method can be applied on smaller sample sizes (20,000 or more), but the error on the fleet average emissions estimate will increase. Since RSD measurements made up to roughly 3 months prior to and after I/M testing are most representative of the condition of the vehicles when they were tested under the I/M program, only these measurements can be used to estimate initial program effectiveness. In a biennial (24-month) I/M program, therefore, only about a quarter of the vehicles measured by RSD will have been measured within 3 months of their I/M test. However, the remaining RSD measurements can be used to estimate short-term repair effectiveness, and the effect of pre-test repairs on fleet emissions.

6.2.5. Potential Systematic Errors

Because the Comprehensive Method relies on large numbers of remote sensing measurements, the remote sensing program will likely have to occur over several months or possibly a year. Vehicle emissions as measured by the Arizona and Colorado IM240 programs vary by season; HC and CO are higher in warmer summer months, while NOx is higher in winter months. It is unclear whether this variation is due to a combination of seasonal temperatures and changes in fuel composition, or to inadequate conditioning of vehicles prior to testing (the seasonal variation in the Wisconsin IM240 program data, Arizona remote sensing data, and the Minnesota idle program data are in the opposite direction of the variation in the Arizona and Colorado IM240 program data). No existing I/M programs vary their cutpoints by season to account for seasonal effects on emissions. There is a possibility that seasonal variation in emissions measured by remote sensing and the I/M program may introduce a systematic bias in the analysis.

The efficiency of remote sensing sites in identifying unique vehicles decreases over time; that is, many vehicles drive by the same sites every day. So concentrating the remote sensing program on a handful of sites, measured throughout the year, may limit the total number of vehicles measured. More sites may be used to increase the number of vehicles measured; however, this may increase any effect of site bias (either due to the fleet of vehicles or the roadway configuration at individual sites) on the evaluation results. The vehicle specific power of individual remote sensing readings can be calculated, using roadway grade at the remote sensing site as well as speed and acceleration measurements, and used to minimize any site bias attributable to site characteristics, as discussed in Section 5.3.

6.3 Reference Method

6.3.1. Description

The Reference Method for evaluating I/M programs involves comparing remote sensing data from vehicles registered in an I/M program area to vehicles registered in a non-I/M program area. (The Reference Method may also be used to compare the fleet average emissions from one I/M program to the fleet average emissions of another I/M program; although this section focuses on the I/M to No-I/M comparison.) Obtaining an adequate sample size of non-I/M program vehicles will typically require conducting measurements in a separate geographic area, or the "reference" area. The reference area, by virtue of its absence of an I/M program, serves as a surrogate untested fleet. The difference in fleet emissions between the I/M program area being evaluated and its "reference" area represents the emission reductions attributable to I/M program effectiveness. Additionally, this difference can then be compared with that predicted by mobile models, such as MOBILE, to determine an overall effectiveness rating. The validity of this approach depends upon selecting a reference area without distinctive characteristics that will systematically bias the evaluation, as well as the accuracy of the model if such an approach is used. This section provides general guidance for conducting such an evaluation, including selection of a reference area, data needs, and data analysis approaches.

6.3.2. Application Examples

The Air Quality Laboratory of Georgia Institute of Technology used the Reference Method to evaluate the effectiveness of the basic I/M program in place in Atlanta in 1994. At that time, I/M was required for vehicles registered in only four counties of the Atlanta 13-county metropolitan area: Fulton, DeKalb, Cobb and Gwinnett. The remaining nine counties, which were not tested until enhanced I/M was implemented, served as the reference fleet. The results of the evaluation indicated that Atlanta's basic I/M program was more effective for cars than predicted by the MOBILE model, but less effective than predicted for trucks. The Georgia Department of Natural Resources used this result to support the mobile source emission reduction credit claimed in the State of Georgia's 1996 State Implementation Plan. The Reference Method was also be used to evaluate Atlanta's enhanced I/M program in October 2000.

6.3.3. Applying the Method

Using the Reference Method for I/M program evaluation involves three major tasks: selecting a reference area, gathering the necessary data, and analyzing that data.

6.3.3.1. Reference Area Selection

There are 6 key criteria to consider in selecting a reference area and they are presented below.

   i) Distance
      Perhaps the most critical criterion for selecting a reference area is suitable geographic distance from the reference area. Recent analyses of Denver and Ohio registrations suggest that I/M programs motivate vehicles to migrate out of an area to adjacent non-

I/M counties (*12*, *27*). Thus, if an agency were to select an adjacent area to evaluate its I/M program, higher-emitting vehicles may migrate to the reference area, making for an artificially dirtier untested fleet. Therefore, reference areas should be chosen at a significant distance from the I/M program area to lower the probability of vehicle migration.

ii) Fleet Age
The age of the fleet is another critical factor in selecting a reference area. Vehicle age is a well-documented contributor to automobile emissions. Consequently, fleet age is a critical consideration in selecting a reference area for an I/M program evaluation. To illustrate, comparisons between an older fleet within an I/M area and a younger fleet in a reference area will underestimate I/M program effectiveness. Isolating emissions by model year between the older and younger fleet will improve the comparison, but such controls will not account for the affects of higher annual vehicle miles traveled (VMT) or potentially higher maintenance rates of the older fleet that influence emissions. VMT data are not readily available in all jurisdictions, but may be inferred using traffic count data and vehicle population information from the state department of transportation. While VMT may be estimated from other data sources, maintenance rates are generally unobservable. Thus, the reference fleet should be roughly the same age as the I/M area fleet. Comparable fleet age can be determined most easily by a bar chart that plots the percentage distribution of vehicles within each model year for the I/M program area and its reference area.

iii) Climate
Climate is another key consideration in selecting a reference area. A variety of factors related to climate affect automobile emissions, and thus the selection of a reference area. For example, salt may be applied to roads in colder climates, potentially resulting in higher rates of catalytic converter rusting, which in turn influences vehicle emission control capacity. At the other extreme, high temperatures, such as those found in Arizona, may more rapidly dissolve the polymer used in emission components, adversely affecting their functioning. Altitude is another climatic factor which may result in differential emissions through potentially faster deterioration rates of emission control systems. A wealth of resources – including National Weather Service data – are available to assist policymakers in identifying areas within their region that provide comparable climatic conditions.

iv) I/M Program Policies
Differences in policy programs between an I/M evaluation area and its reference area may bias program evaluation. For example, a safety inspection program that requires functional lights and brakes may speed fleet turnover by denying registration to poor-condition vehicles. While the emission profiles of these vehicles is uncertain, it is a reasonable hypothesis that they are higher than average emitters and that a safety inspection program will weed some of them out, thus shifting fleet emissions downward. Thus, the presence of a safety program in a reference area might underestimate the effectiveness of the I/M program being evaluated by providing an artificially low baseline for comparison.

v) Motor Vehicle Tax System
The tax system for motor vehicles is another source of variance in the fleet distribution. To illustrate, an ad valorem tax that declines rapidly with vehicle age may have the affect of slowing fleet turnover by making ownership of older vehicles more affordable. Conversely, ad valorem taxes in a reference area that are onerous among all model years may shift the income level of older vehicle owners upward such that the socioeconomic characteristics of vehicle owners are not equivalent by model year between the comparison areas. State policies on antique vehicles can also influence fleet age and condition. For example, Georgia vehicles 25 years and older receive permanent tags with no further requirements for taxation, emissions testing or registration. This exemption may result in a concentration of very old vehicles compared with other areas that offer no such exemption. These are just a few examples of how public policies seemingly unrelated to air quality can nonetheless influence fleet emissions. Consequently, policymakers should research policy programs in candidate states to rule out the potential for systematic emission biases that could result from their presence.

vi) Socioeconomic Factors
Finally, socioeconomic conditions are the least studied of the influences on automobile emissions. Most of the evidence regarding the influence of socioeconomics on fleet condition and emissions is anecdotal, relying on conventional wisdom that less affluent people will drive older vehicles (an assertion for which there is some evidence) and that they cannot afford to properly maintain their vehicles (for which there is little evidence). Another assumption is that older motorists drive their cars infrequently but maintain them well. While socioeconomic conditions have received relatively little scholarly attention in comparison with physical influences on automobile emissions, it is nonetheless wise to consider them in selecting a reference area because they may represent the unobserved influences of maintenance practices, driving behavior, and culture.

6.3.3.2. Data Needs

In addition to remote sensing data from the I/M evaluation area and its comparison fleet, the Reference Method requires registration data, I/M records, and model outputs, assuming it is desired to include the model as a part of the analysis protocol. Remote sensing data should be collected from the I/M program area and its reference area under similar physical conditions and within roughly the same timeframe. Simultaneous data collection prevents differences that may occur due to temperature affects on emissions or seasonal policy changes such as fuel changes.

Registration data are needed to generate the characteristics of remotely sensed vehicles, such as registration address, model year and vehicle type. The registration address is particularly critical for identifying whether a vehicle is located in the I/M area or reference area. For example, if the I/M program area and reference area are located near one another, then it is possible to measure inspected vehicles in the reference area and reference area vehicles in the I/M program area. Registration address can also be used to generate demographic characteristics for the registration area. This process, known as geocoding, locates the census block group of the registration address. The census block group, in turn, can be used to generate demographic data from the

most recent national census on its residents. These demographic data include median household income, median family income, and the number of households receiving social security, retirement and public assistance. Given the inverse relationship between a census block group's median household income and the average age of its registered fleet, these data provides additional controls for fleet age, as well as safeguard controls for the unobservable influences of maintenance practices, driving habits, and cultural effects (*28*).

I/M records provide two optional pieces of information for the Reference Method. The first is odometer data. Odometer data can be used to extract annual vehicle miles traveled (VMT), which contribute to wear and tear and ultimate deterioration of a vehicle's emission control system. VMT is typically calculated by subtracting odometer readings for two consecutive years, dividing by the number of days between inspections, and multiplying that figure by 365. (The daily mileage must be multiplied by 730 for states with biennial testing.)

I/M records can also be used to identify "invalid" reference area vehicles and non-compliant inspection area vehicles. If a significant number of reference area vehicles have recently migrated from the inspection area, it is possible that the evaluation be biased high or low depending on the average emission level of the migrating vehicles. I/M records can also be used to estimate noncompliance in the I/M program area by identifying vehicles whose emissions inspections have lapsed. This information prevents the I/M fleet from appearing artificially dirty, while contributing valuable information to the compliance aspect of program performance. Finally, emission factor modeling output from MOBILE or another model that predicts emissions of the inspected and non-inspected fleets can then be used to compare with real-world differences in inspected and non-inspected fleets measured by the remote sensing data.

RSD data can also be combined with exhaust emission factors for cars and light-duty trucks extracted from the model tailpipe emission factors. These emission factors project average grams/mile by model year and are the product of a range of inputs, including program design (testing technology, model-year coverage, and emissions standards), fleet characteristics (fleet VMT and age distribution), and operating modes (hot stabilized emissions to correlate with the condition of in-use vehicles). Inputs for the I/M-county fleet will include the design elements of the current program such as the emissions analyzer, range of model years required for inspection, and the testing mode, e.g. one-speed idle testing. I/M program elements for the non-I/M fleet are simply omitted. The modeling process will also require the model year distribution of the evaluation and reference fleets.

It should be noted that use of MOBILE may introduce analytical complexity as well as increased technical uncertainty in the results due to the internal coding of the model that will inherently make comparisons and computational assumptions the user may not fully appreciate.

6.3.3.3. Data Analysis

The Reference Method can involve a variety of analytical approaches to assess the effectiveness of an I/M program. The raw emissions of an I/M program area and its reference area can be compared with histograms to determine any differences in the distribution of high emitters, low emitters, and median points. The significance of emissions differences by model year can be

determined through error bar charts that plot the mean emissions plus the associated uncertainty. Regression modeling can be used to determine the influence of registration in the I/M program area versus its reference area on emissions. RSD emission differences in inspected and reference fleets can be compared to the differences predicted by EPA mobile models to determine an I/M program effectiveness rating (*29*).

6.3.4. Advantages and Disadvantages

The Reference Method has strengths and weaknesses for the evaluation of I/M programs. Most importantly, it is a quantitative estimate of I/M effectiveness that is easy to calculate given adequate data, although incorporating modeling output into the analysis will certainly add a layer of complexity. As an external reference point for evaluating I/M programs, it provides ongoing opportunities for evaluation whether a program is within a year of implementation or five years into operation is irrelevant. However, a significant amount of information is required beyond remote sensing data, including registration records, I/M records and model outputs. Furthermore, no reference area will completely match the I/M area profile, thus there is always the risk that some characteristic will systematically bias the I/M program evaluation higher or lower than it should be. Finally, the method will not work in some states (such as California), where there are no reference fleets because the entire state is included in the I/M program area. The Reference Method can also be used to compare on-road emissions in the region to be evaluated to those in another region, such as Arizona, where I/M effectiveness has been estimated by other methods.

7. Summary
Three methods for estimating I/M program effectiveness using RSD data were outlined in this guidance. Every effort was made to provide as much detail as possible with regard to data collection procedures, QA/QC protocols, analysis methods, and sources of error or possible bias associated with a given method; however, it is recognized that improvements to those methods outlined in this document will continue to evolve. Therefore, it is strongly recommended that any state considering the use of RSD for program evaluation purposes work closely with their respective regional EPA office and the Office of Transportation and Air Quality to ensure the most up-to-date practices are incorporated into the evaluation. Furthermore, states interested in using RSD for program evaluation must recognize the need within their own agencies to develop a minimum level of expertise with the technology and procedures to ensure reliable data are collected and analyses are performed properly.

It should also be recognized given the difficulties associated with I/M program evaluations, that an evaluation based on both out-of-program data (e.g. RSD) and in-program data will provide a more accurate estimate of overall program performance than simply relying on one method alone.

8. References

1 Clean Air Act, 1970
2 Clean Air Act Amendments, 1977
3 EPA Inspection/Maintenance Policy Guidance, 1978.
4 Clean Air Act Amendments, 1990.
5 57 FR 52950 or 40 CFR Part 51, IM Program Requirements; Final Rule, November 5, 1992.
6 National Highway System Designation Act of 1995 (23 U.S.C. 101).
7 62 FR 1362 or 40 CFR Parts 51 and 52, Minor Amendments to Inspection Maintenance Program Evaluation Requirements; Amendment to the Final Rule, January 9, 1998.
8 "Guidance on Alternative IM Program Evaluation Methods, EPA Memo, Office of Mobile Sources, Regional and State Programs Division, October 30, 1998.
9 Singer, Harley, Littlejohn, Ho and Vo, "Scaling of Infrared Remote Sensor Hydrocarbon Measurements for Motor Vehicle Emission Inventory Calculations", ES&T (32)21, p.3241, 1998.
10 Stedman, Bishop, Aldrete, Slott, "On-Road Evaluation of an Automobile Emission Test Program" ES&T., 31, p.927, 1997.
11 Stedman and Bishop, "Measuring the Emissions of Passing Cars", Accounts of Chemical Research, 29(10), p.489, 1996.
12 Stedman, Bishop, Slott, "Repair Avoidance and Evaluating Inspection and Maintenance Programs", ES&T, 32, p.1544, 1998.
13 Wenzel, Singer, and Slott, "Some Issues In the Statistical Analysis of Vehicle Emissions", J. Transportation and Statistics (3)2, p.1, September 2000.
14 Mann and Jones, CRC Report, "On-Road Remote Sensing of Automobile Emissions in the Research Triangle Park, North Carolina Area: 1997 and 1998", p.5, March 2000.
15 Wenzel and Gumerman. "In-Use Emissions by Vehicle Model", Presented at 8th CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1998.
16 McClintock, "The Colorado enhanced I/M Program 0.5% Sample Annual Report", Remote Sensing Technologies Inc., Prepared for Colorado Department of Public Health and Environment, 1998.
17 Jimenez, McClintock, McRae, Nelson and Zahniser "Vehicle Specific Power: A Useful Parameter for Remote Sensing and Emission Studies." Presented at 8th CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1998.
18 Wenzel, Reducing Emissions from In-Use Vehicles: An Evaluation of the Phoenix Inspection and Maintenance Program using Test Results and Independent Emissions Measurement, Environmental Science and Policy, (4), p.359, 2001.
19 Wenzel, "I/M Failure Rates by Vehicle Model", Presented at 7th CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1997.
20 Stedman, Bishop, Beaton, Peterson, Guenther, McVey and Zhang, "On-Road Remote Sensing of CO and HC Emissions in CA", Final Report to Air Resources Board, AO32-093.
21 McClintock "The Denver Remote Sensing Clean Screening Pilot" , Prepared for the Colorado Department of Public Health and Environment, 1999.

22 Wenzel, "Reducing Emissions from In-Use Vehicles: An Evaluation of the Phoenix Inspection and Maintenance Program using Test Results and Independent Emissions Measurement", Environmental Science and Policy, (4), p.377, 2001.

23 Slott, "The Use of Remote Sensing Measurements to Evaluate Control Strategies: Measurements at the End of the First and Second Year of Colorado's Biennial Enhanced I/M Program", Presented at the 8th CRC On-Road Vehicle Emissions Workshop, San Diego CA, April 1998.

24 Lawson, "'Passing the test'—Human behavior and California's Smog Check program," J. Air Waste Manage. Assoc., 43, p.1567, 1993.

25 Klausmeier and Weyn, "Using Remote Sensing Devices (RSD) to Evaluate the California Smog Check Program", Report to the California Bureau of Automotive Repair, October 2, 1997.

26 Wenzel, "Human Behavior in I/M Programs," Presented at the 15th Annual Mobile Sources/Clean Air Conference, Snowmass, CO, September, 1999.

27 McClintock, "I/M Program Avoidance and Enforcement", Presented at the 15th Annual Mobile Sources/Clean Air Conference, Snowmass, CO, September, 1999.

28 Leisha DeHart-Davis private communication with Jim Lindner.

29 Rodgers, Lorang, DeHart-Davis, "Measuring IM Program Effectiveness Using Optical RSD: Results of the Continuous Atlanta Fleet Evaluation", Atmospheric Environment, submitted for publication.

Appendix A: On-Road Evaluation of a Remote Sensing Unit

All on-road remote sensors carry out at least a measurement of the CO/CO_2 ratio in the exhaust of a passing vehicle. It is possible for an interested party to carry out a quantitative evaluation of the precision of this measurement. This evaluation can be done without going to the expense and complexity of an on-road audit using a vehicle of known emissions (wet gas audit), or a vehicle designed to puff surrogate compressed gas mixtures of known ratios (dry gas audit).

The measurement of exhaust CO/CO_2 ratio is obtained by estimating the slope of a graph of CO versus CO_2 (or more properly delta CO versus delta CO_2). The evaluation is carried out by observing the quality of the individual data points which are used to derive this slope. Several on-road remote sensors operate for 0.5 seconds at 100 hz, thus obtaining 50 data points for this correlation. Several on-road remote sensors use a puff of gas of known CO/CO_2 ratio as a field calibration. For these sensors, the system operator can display the CO/CO_2 graph from a calibration, whether the calibration was considered valid or not.

EVALUATION OF A CALIBRATION PUFF:

Figure 1 shows a valid CO/CO_2, HC/CO_2 and NO/CO_2 on-road calibration puff (FEAT 3002, Sept. 27, 2001, Casa Grande, AZ). When evaluating a remote sensor, the first parameter to note is the quality of the data and the fit. In the case shown, all 50 points are almost touching the straight line and $r^2 = 0.99$. The next parameter to note is the extent of the data spread on the CO, HC, NO and CO_2 axes. Different instruments use different units. These graphs show the gas concentrations %CO, %HC (propane), %NO and CO_2 in an 8cm cell[*]. These units are chosen to correspond approximately to what would be measured were one to directly probe a tailpipe. The units however do not matter, but the spread of both gases in a plot such as Figure 1 is important to note.

Figure 2 shows a CO/CO_2, HC/CO_2 and NO/CO_2 on-road calibration puff (FEAT 3002, August 29, 2001, Phoenix AZ). This was not a valid calibration. In this case, the calibration gas appears to be mixed with exhaust from a vehicle which had recently passed through the optical beam. It is not important that occasional invalid calibrations look bad. It is important that the instrument is able to obtain valid calibrations, which look like Figure 1, and are carried out with a data spread comparable to a typical automobile at the same site. This parameter also must be determined at the roadside in order to evaluate the instrument.

---

[*] It should be noted that air spectroscopy gas optical absorption data often are given in strange units because what is measured is the product of concentration and path length. Thus, atm.cm or %.cm or ppm.cm, or even % in 8cm are all units which may be used and all can be inter-converted. In fact the CO2 plume from a typical car as measured by an on-road sensor can be as large as 1atm.cm, but more often is 0.01 or 0.01 atm cm which could also be rendered as 1%.cm. CO is typically 1/10 of that and HC and NO 1/100.

Another noise evaluation which one should ask any instrument to be able to perform is a calibration but without any added calibration gas. The graphical evaluation is uninteresting, namely a cluster of points at the origin. However, the spread of these points along each of the axes is a direct measure of the noise which the instrument will see from all passing vehicles. Again, the spread should be compared to the spread expected from a typical motor vehicle in a realistic roadway situation using the same remote sensing unit.
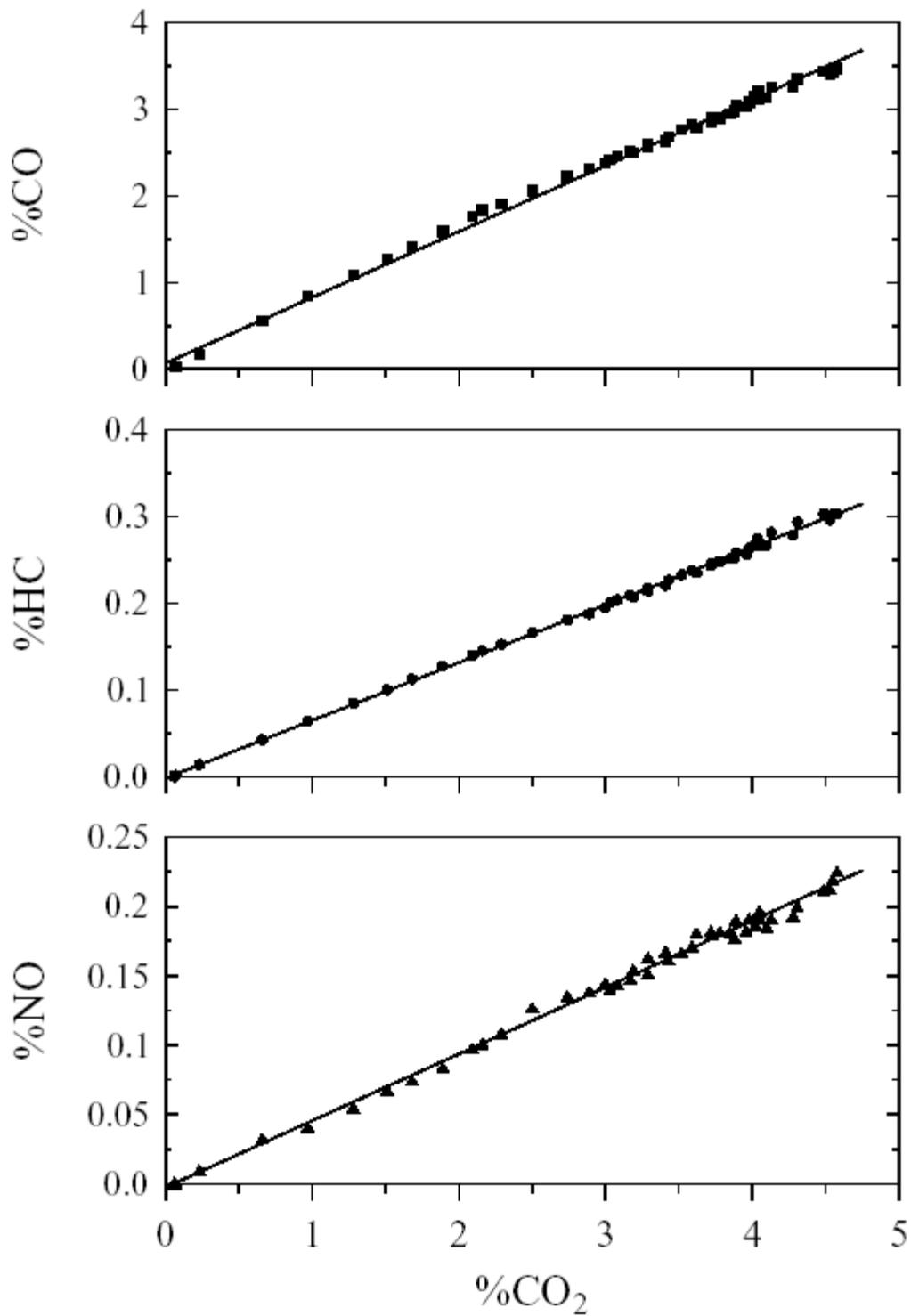
Figure 1. Half-second puff calibration plots for CO, HC and NO. The straight lines are linear least squares regressions of the data**.**
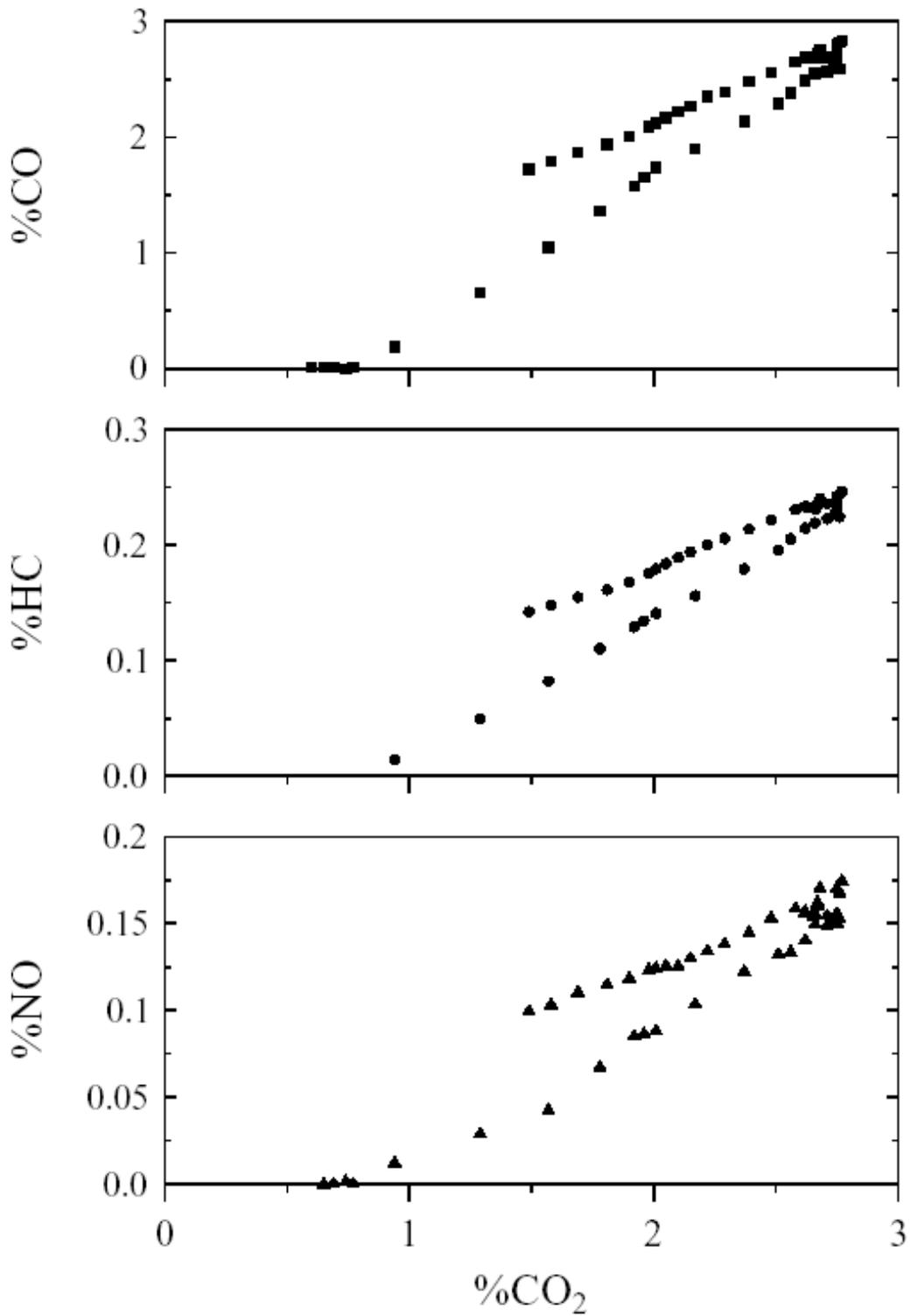
Figure 2. Half-second calibration gas puff for CO, HC and NO which has been contaminated with exhaust from a passing vehicle.

EVALUATION OF INDIVIDUAL MOTOR VEHICLE EMISSIONS:

At the roadside, when the instrument is operating and calibrated, call up and observe CO/CO_2 ratio graphs from about three randomly chosen vehicles. The skewed distribution of emissions implies that these are all likely to be low emitting cars with very small CO/CO_2 slopes. The parameter to observe on these graphs is the range (spread) of the CO_2 data. If the CO axis is auto scaling, the noise may look very bad but actually be very good. Note the CO_2 spread. It should be comparable to the calibration, or at least not less than about 10x smaller.

Figure 3 shows typical data from a passing vehicle. The CO_2 readings are from about 0.3% to 1.3%, for a total spread of 1% CO_2 in 8cm. The spread for the calibration shown in Figure 1 is about 4.5% and in Figure 2 about 2.2%. In both cases the calibrations are at a comparable, although larger spread than the on-road data. Now it is necessary to evaluate the CO/CO_2 graph on a vehicle with higher than zero CO/CO_2 ratio. If the raw data are stored and can be recalled and graphed from each vehicle, then wait for a vehicle with $CO/CO_2 > 0.25$ (about 3.5% CO on the video screen). Now observe this CO/CO_2 graph. The CO_2 spread should be comparable to the three low CO emitters observed earlier. The CO spread should be comparable to the CO spread on the calibration puff, or at least not less than about 10x smaller. If these criteria are met and this graph looks "good", for instance, $r^2 > 0.9$, then you have an instrument likely to provide precise and accurate measurements ,if the calibration gas supplier is trustworthy, data.

Figure 4 shows on-road CO/CO_2 data from a cold-start vehicle measured at the University of Denver. A similar evaluation analysis can be carried out for HC and NO; however, if the CO/CO_2 data do not pass muster, then HC/CO_2 and NO/CO_2 are much less useful because the readings are missing a major component of the carbon balance. Note also that HC emissions are smaller and harder to measure than CO, so more (relative) noise is to be expected. If the data you see at roadside are of similar or better quality then you are observing a good instrument. If they are not up to this quality, then your should think twice about accepting the data until the operator/vendor can convince you that the instrument is functioning properly.

The ability to read vehicle exhaust independently of vehicle type should also be verified. This may be done by making a note of the valid reading rate from normal sedans and from SUV's and pickups while observing roadside operations. In a perfect world all vehicles with ground level exhaust should be measured. In reality some are not, but this should be observed to be a random process or a systematic one caused by driving mode (noticeable decelerations) not one caused by vehicle type or body height.
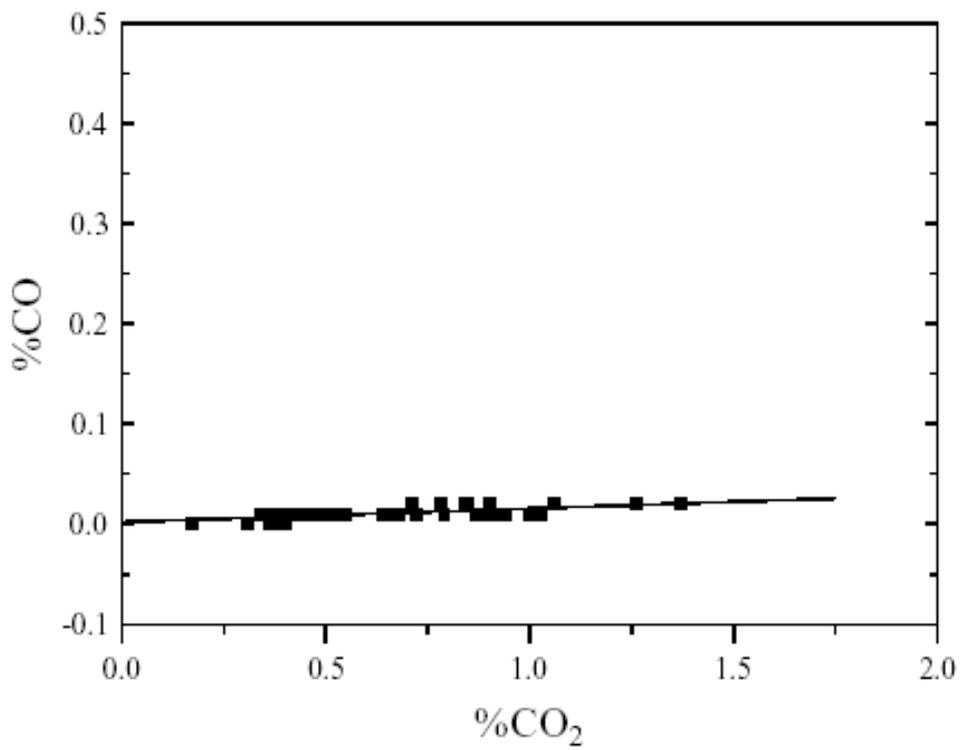
Figure 3. In-use data for a low CO emitting vehicle.
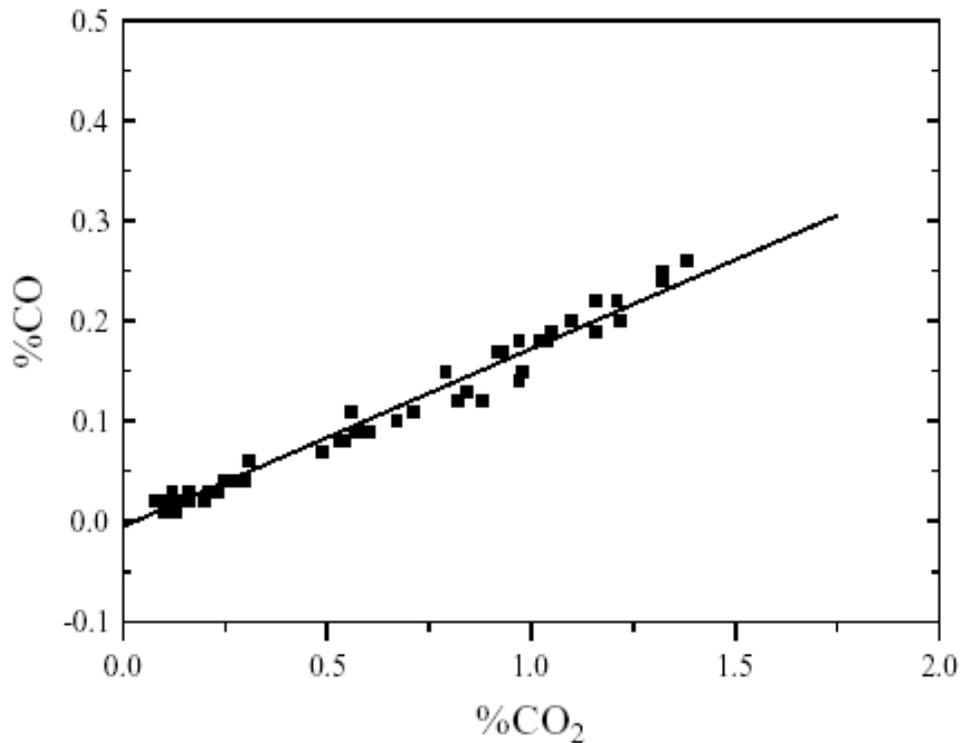
Figure 4. In-use data from a cold-start vehicle with elevated levels of CO.

EVALUATION USING EXHALED BREATH:

A non-smoking human exhales CO_2 and negligible amounts of CO, HC and NO. The remote sensor should be able to read human breath as a passing car, as long as it is accompanied by a blocked and unblocked optical beam. Fifteen readings of breath with the FEAT instrument in the laboratory yielded a mean CO reading of 0.07% with a standard deviation of 0.04%. HC read a mean of 39 ppm propane with a standard deviation of 50 and NO a mean of -3 ppm with a standard deviation of 18 ppm.