ABSTRACT

Unal, Alper. Measurement, Analysis, and Modeling of On-Road Vehicle Emissions Using Remote Sensing. (Under the guidance of Dr. Chris Frey).

The main objectives of this research are; to develop on-road emission factor estimates for carbon monoxide (CO) and hydrocarbon (HC) emissions; to collect traffic and vehicle parameters that might be important in explaining variability in vehicle emissions; to develop an empirical traffic-based model that can predict vehicle emissions based upon observable traffic and vehicle parameters.

Remote sensing technology were employed to collect exhaust emissions data. Traffic parameters were collected using an area-wide traffic detector, MOBILIZER©. During the measurements, license plates were also recorded to obtain information on vehicle parameters.

Data were collected at two sites, having different road grades and site geometries, over 10 days of field work at the Research Triangle area of North Carolina. A total of 11,830 triggered measurement attempts were recorded. After post-processing, 7,056 emissions were kept in the data base as valid measurements. After combining with the traffic and license vehicle parameters, a data base has been developed.

Exploratory analysis has been conducted to find variables that are important to explain the variability of the emission estimates. Statistical methods were used to compare the mean of the emissions estimates for different sub-populations. For example, multi-comparison analysis has been conducted to compare the mean emissions estimates from vehicles having different model years. This analysis showed that the mean emissions from older vehicles were statistically different than the mean emissions estimates from the recent model year vehicles.

One of the contributions of the research was developing an empirical traffic-based emission estimation model. For this purpose, data collected during the study were used to develop a novel model which combines the Hierarchical Tree-Based Regression method and Ordinary Least Squares regression.

The key findings from this research include: (1) the measured mean CO emission estimate for Research Triangle park area of North Carolina is estimated as 340 grams/gallon, whereas the mean HC emissions estimate is found to be as 47 grams/gallon (2) inter-vehicle variability in vehicle emissions can be as high as two orders-of-magnitude; (3) intra-vehicle variability is lower compared to the inter-vehicle variability; (4) some vehicle variables such as vehicle model year and vehicle type are important factors in explaining the inter-vehicle variability in emissions estimates; (5) emission estimation model developed in this research can be applied to estimate the emissions from on-road vehicles.

MEASUREMENT, ANALYSIS, AND MODELING OF ON-ROAD VEHICLE EMISSIONS USING REMOTE SENSING

by

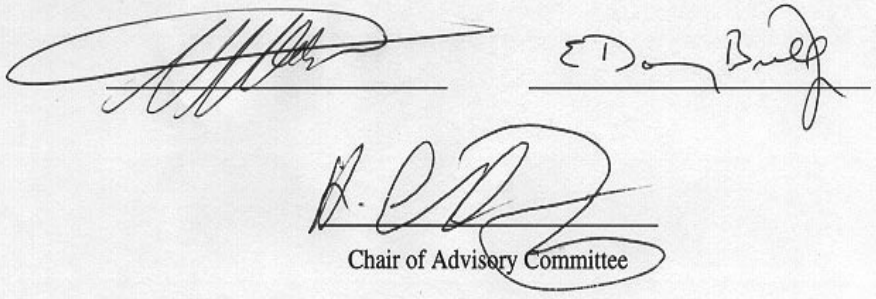
Alper Unal

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

DEPARTMENT OF CIVIL ENGINEERING
Environmental Engineering and Water Resources

Raleigh, North Carolina
1999

APPROVED BY:



Chair of Advisory Committee

Dedicated to my loving wife Gözde and

to my parents in Türkiye

BIOGRAPHY

11th November, 1973	Born to Zeynep and Nadir Unal in Afyon, Turkey
1980-85	Primary school education at Maltepe Ilkokulu
1985-88	Secondary school education at Turk Egitim Dernegi (TED) Ankara College
1988-91	High school education at T.E.D. Ankara College
1992-96	Undergraduate college education at Middle East Technical University (METU), Ankara to study environmental engineering
1997-99	Attended graduate study at North Carolina State University to study environmental engineering

ACKNOWLEDGEMENTS

Family	My wife Gozde, and my family in Turkey for their great support and love
Advisors	Dr. Christopher Frey, Ph. D., and Dr. Nagui Rouphail, Ph.D., for their invaluable guidance
Friends	Maysa, Matt, Junyu, Naveen, Ranjit, Sudeep, and Russ for the friendly working atmosphere
Special Thanks	Dr. Randall Guensler, Ph.D., and William Bachman at Georgia Institute of Technology for their help in obtaining the data and information required for the project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Highway vehicles constitute major emission sources for several air pollutants, including nitrogen oxides (NO_X), carbon monoxide (CO), particulate matter (PM), and volatile organic compounds (VOCs). VOCs are also often referred to as hydrocarbons (HCs). These pollutants have significant adverse effects on human beings and the environment. Carbon monoxide reacts with the hemoglobin in blood to prevent oxygen transfer and depending on the concentration and time of exposure, its effects range from slight headaches to nausea to death (Cooper and Alley, 1994). NO_X and HCs in the presence of sunlight form tropospheric ozone (O_3), a severe eye, nose, and throat irritant can lead to acute respiratory illness. NO_X is a potentially significant contributor to a number of environmental effects such as acid rain and eutrophication in coastal waters. Particulate matter cause visibility reduction, corrosion, acute respiratory illnesses, premature deaths and cancer (USEPA, 1995). The ambient concentrations of NO_X, CO, HC, PM, and O_3 are regulated under the National Ambient Air Quality Standards (NAAQS). Emission rates for individual motor vehicles are regulated by Federal standards (Cooper and Alley, 1994).

There is a need to obtain on-road emissions data that are representative of actual driving conditions. Good information about emissions can then be used to develop necessary emission inventories and transportation plans. The objectives of this study are to:

(1) Conduct on-road remote sensing of pollutant emissions while collecting traffic data using an area-wide traffic detector, MOBILIZER©;

(2) Determine the on-road emission rates of vehicles;

(3) Conduct statistical methods to explain variability in emission estimates; and

(4) Develop an empirical traffic-based model to estimate vehicle emissions based upon measurable traffic and vehicle parameters.

This study is aimed at providing an empirical linkage between traffic parameters and on-road vehicle emission rates. Traffic data were collected using a video-based vehicle tracking system, MOBILIZER©. On-road vehicle emissions were collected using a remote sensing device (RSD). The results of this study are: estimates of on-road CO and HC emissions; explanatory variables including individual vehicle information such as model year; and traffic parameters such as vehicle speed, acceleration, and headway. Statistical methods have been used to analyze the differences in emission behaviors for different data sets. An empirical model has been developed to estimate vehicle emissions from measurable traffic and vehicle parameters.

In the rest of this chapter, background information on current approaches used in vehicle emissions estimation and the motivation for on-road data collection will be discussed. In Chapter 2, the remote sensing technology and video tracking system is explained in detail. In Chapter 3, we discuss the procedure by which sites were selected. In Chapter 4, data analysis and discussion of the findings of exploratory data analysis is

presented. Chapter 5 explains the steps of model development and obtained results. Conclusions and recommendations for future work are provided in Chapter 6.

1.1 Mobile Source Emissions

From 1970 to 1993, there has been a substantial decrease in pollutants from motor vehicles, primarily through technological improvements such as catalytic converters and fuel injection, which are now standard equipment on cars. During this period of time, highway vehicle emissions of CO and HCs declined by 32 percent and 53 percent respectively (Nizich *et al.*, 1994). However, vehicle emissions still comprise a substantial portion of the national emission inventory; they account for 26 percent of HC, 32 percent of NO_x and 62 percent of CO emissions (USEPA, 1993b). These numbers may be misleading due to uncertainties associated with emission factor development (Frey, 1997).

1.2 Emissions Regulations

The harmful effects of air pollution on public health were formally recognized by the requirements of the Clean Air Act Amendments (CAAA) of 1970, which mandated establishment of National Ambient Air Quality Standards (NAAQS) for six criteria pollutants: carbon monoxide; lead; nitrogen oxides; ozone; particulate matter; and sulfur dioxide (Curran *et al.*, 1994). The CAA has been amended three times, in

1970, 1977, and 1990. The 1990 CAAA has motivated the most stringent requirements for noncompliance, especially for O_3 and the particulate matter.

In 1994, according to EPA's estimates (USEPA, 1995), about 62 million people lived in counties that violated one or more of the NAAQS. Thus, the CAAA contains stringent requirements for further reductions in emissions from highway vehicles by having strict monitoring and sanctions for non-performance, and to bring non attainment areas into compliance (TRB, 1995). Areas for which the ambient concentration of a criteria pollutant exceeds the NAAQS are said to be in non attainment for that pollutant. Such areas are subject to severe restrictions on permitting of any new emission sources, and are required to develop plans to reduce emissions to acceptable levels.

One of the most important air pollution regulations that affect mobile sources is the "conformity" rule. Conformity is a determination made by Metropolitan Planning Organizations (MPOs) and Departments of Transportation (DOT) that transportation plans, programs, and projects in non attainment areas are in compliance with the standards contained in State Implementation Plans (SIPs) (i.e., plans that codify a state's CAAA compliance actions) (FHWA, 1992). To demonstrate conformity, a transportation plan or project must improve air quality with respect to one or more of the following: (1) the motor vehicle emission budget in the SIP; (2) emissions that would be realized if the proposed plan or program is not implemented; and/or (3) emissions levels in 1990 (TRB, 1995). Conformity requirements have made air quality a primary constraint on transportation planning (Sargeant, 1994).

1.3 Approaches to Estimating Motor Vehicle Emissions

An emission inventory is a listing and description of air pollutant emitting sources, including a quantitative estimate of pollutant emissions (Stern, 1976). In developing inventories, emission factors and emissions producing activity data are used. An emission factor is the amount of pollutant produced per unit activity. For highway vehicles, emission factors are commonly expressed on grams of pollutant emitted per vehicle-mile of travel basis. Thus, the activity data required for emission inventory development is an estimate of total vehicle miles traveled.

At present, four different methods are used or proposed to calculate motor vehicle emission factors. These methods are: driving cycle-based emission factor models; modal emissions-based models; fuel-based approaches; and on-road emissions data-based models. Driving cycle-based approaches underlie the current practice for vehicle emissions estimates in the U.S.

1.3.1 Driving Cycle-Based Models

The two highway vehicle emission factor models used for regulatory purposes in the U.S. are EMFAC7 in California and MOBILE5a. These models are based upon emissions data for selected driving cycles. A driving cycle is composed of a unique profile of stops, starts, constant speed cruises, accelerations and decelerations and is typically characterized by an overall time-weighted average speed (TRB, 1995). Different driving cycles are used to represent driving under different conditions.

Driving cycle test data are used as the basis for estimating emission factors in these models. For example, in MOBILE5a, base emission rates (BER) are derived by driving new and in-use light-duty motor vehicles through the Federal Test Procedure (FTP). As explained by TRB (1995), the FTP is an emission test composed of a defined cycle of starts, stops, accelerations, and constant-speed cruises conducted on a laboratory dynamometer under standard conditions. These conditions include the use of a specific fuel, control of test cell temperature, and replication of a predetermined speed profile by a test driver (USEPA, 1993b).

During the test, emissions are collected in bags. The cold and hot start mode emissions are collected in Bags 1 and 3 respectively, and hot stabilized mode emissions are collected in Bag 2. For the MOBILE5a model, the emissions from light duty vehicles operating in all three phases are used to calculate the baseline emissions. The baseline emission rate (calculated in grams/mile) for a light duty vehicle is estimated by averaging the result from the three phases of the FTP for that vehicle class operating at an average speed of 31.6 kph (19.6 mph), which is the average test speed of the entire FTP. The instantaneous speeds in the FTP vary from zero to 57 mph (TRB, 1995).

For Light Duty Gasoline Vehicles (LDGV), ten driving cycles other than the FTP, with average speeds ranging from approximately 2.5 mph to 65 mph, were used to develop a speed correction factor. The speed correction factor is the ratio of emissions at a given average speed to that of the FTP. However, a given driving cycle may have instantaneous speed vary from zero to as high as 80 mph. The emissions measured in these cycles depend upon the overall speed and acceleration profile, and

not solely upon average speed. Therefore, the use of average speed as the only descriptor of a driving cycle can be misleading. Furthermore, when speed correction factors are "interpolated" between driving cycles on the basis of average speed, it actually represents a form of "extrapolation" (Kini and Frey, 1997).

The MOBILE and EMFAC driving cycle-based models are used for regulatory purposes by EPA and California Air Resources Board (CARB), respectively. However, both of these models have some disadvantages and weaknesses. The emissions estimates of both models are based on a limited set of driving cycles. Historically, these driving cycles have been limited at representing real driving conditions, which affect the emission factors. Many of the driving cycles were developed years ago using old equipment whose capabilities were limited. For example, the FTP contains no accelerations greater than 3.5 mph per second due to limitations of dynamometer technology when the cycle was first developed. However, real world accelerations may be as high as 7 mph per second. Such accelerations lead to higher engine loads and higher emissions than are accounted for in the FTP. As reported by TRB (1995), the FTP cycle was compared with two cycles developed by Effa and Larsen (1993) having approximately the same average speed of 16 mph. The latter cycles were developed based upon driving on arterial roads and freeways in the Los Angeles area. The results show that FTP cycle underestimates the frequency of driving activities at higher speeds and accelerations. This type of driving conditions can lead to substantial increases in emissions.

Driving cycle-based models do not consider the differences in engine load while calculating the emission factor. For example at higher altitudes air density is decreasing and engines tend to operate at rich air/fuel mixtures, which means more HC and CO emissions. When the load on the engine is high, the temperature of the engine increases significantly. Thus, to prevent over-temperature damage to the engine and catalyst, vehicles are often designed to operate in fuel rich mode under high engine loads. As stated by USEPA (1993b), it was found that HC and CO emissions increased by almost 20 to 100 times during this type of fuel rich operation. The high engine load may be due to high accelerations, high speeds, positive road grades, air conditioning operation, or any combination of any of these items.

1.3.2 Modal Emissions-Based Models

Modal emissions-based models relate emissions directly to the operating mode of vehicles. The operating modes include cruise, acceleration, deceleration, and idle (Barth and Norbeck, 1997).

Several research studies have been performed using dynamometers and instrumented vehicles producing second-by-second emissions data to investigate vehicle emissions associated with modal events. Since the early 1980s, these modal emissions research projects have been conducted at the California Air Resources Board (CARB) (Cicero-Fernandez and Long, 1994). By testing a small set of newer technology

vehicles, these studies found that CO and HC emissions are greatly affected by various acceleration modes.

In addition to these dynamometer tests, researchers have used instrumented vehicles to collect emissions data while they are driven on the road. Kelly and Groblicki (1993) instrumented a GM Bonneville to collect on-road emissions and have performed several such experiments in Southern California.

Emissions during on-road driving were compared to those obtained on dynamometers using the urban dynamometer-driving schedule (UDDS). They found that during moderate to heavy loads on the engine (i.e. at throttle positions greater than 40 percent and engine speeds greater than 2000 rpm), the vehicle ran under fuel enrichment conditions, resulting in CO emissions 2500 times greater than those at normal stoichiometric operation (HC was 40 times as great). It was found that enrichment conditions, which occurred 1 to 2 percent of the time, were responsible for 88 percent of the CO emitted over the study.

Based upon these data, several types of modal-emissions models have been developed. One way of developing a modal-emissions model is to set up a speed-acceleration matrix in order to characterize vehicle operating modes of idle, cruise, and different levels of acceleration/deceleration. The matrix measures emissions associated with each bin or mode. This emissions matrix can be multiplied with a similar matrix that has vehicle activity broken down so that each bin contains the time spent in each driving mode. The result is the total amount of emissions produced for the specified

vehicle activity with the associated emissions matrix. This approach has been used by West and McGill (1997) to develop a velocity-acceleration matrix in the form of look-up tables for light-duty vehicle fuel consumption and emissions. The tables were intended to be used in traffic simulation models such as TRAF-NETSIM, TRANSIMS, and FRESIM. According to Barth *et al.* (1996), the problem with such an approach is that it does not properly handle other variables that can affect emissions, such as road grade or use of accessories. Correction factors can be used but this can be problematic since their effect will typically be based on secondary testing not associated with the core model.

Another type of modal-emissions based model is based on mapping. This approach has been employed since the 1970s for some fuel economy models. The conceptual approach is to translate real-time speed and route information into instantaneous vehicle rpm and load parameters then use an engine map to look-up the instantaneous emission rates for the specific rpm and load conditions, and continuously integrate the instantaneous emission rates to estimate the total emissions from a given set of vehicle activities. The engine map is based upon steady state testing of the engine and control system. That is, the engine is placed at specific RPM and load conditions and the emission result is recorded, creating a 3-dimensional graph of engine RPM, load, and emission rates (Guensler, 1994). Two models are employed in the emission map regime, VEHSIM and VEMISS. The VEHSIM (vehicle performance simulation) model was originally developed by the Transportation systems Center of the US Department of Transportation to support vehicle energy efficiency studies (Zub, 1981). The VEHSIM program translates vehicle activity (instantaneous speed and acceleration)

into horsepower requirements, after losses, in the form of revolutions per minute (RPM) and torque. As pointed out by Guensler (1994) some questions still remain to be answered about this model. VEHSIM's capability to translating time-speed and route information into engine load and RPM parameters may be questionable. In developing engine maps vehicle mileage accumulation is not taken into consideration. Another weakness is that emissions occurring under transient conditions may not be adequately represented by the emissions map that is derived under steady-state conditions.

VEMISS translates RPM and load parameters into instantaneous emission rates (USEPA, 1993a), where engine maps list output as the engine load parameter. The potential of this model is being evaluated by the EPA for application to general emissions modeling. Variations in catalyst efficiency and dynamic response to changes in pollutant loading may cause problems in engine map model development (Guensler, 1994).

Other mapping type of models have been developed by LeBlanc *et al*., (1995); Shih and Sawyer, (1996); and Shih *et al*., (1997).

The dissaggregate modal modeling approach is similar to emission mapping, in which the relationship between second-by-second emissions and modal activities are developed. Statistical analysis is used to overcome the limitations associated with using few number and types of test vehicles to collect data. Modal activity estimates are based on a link profiling approach, which involves the development of vehicle speed-time traces along specific roadway classifications under specific levels of service. These may

be derived from observations, or from statistical techniques (Guensler, 1994). Guensler *et al.*, (1997) are currently working on a project to develop a GIS-based modal-emissions model at Georgia Institute of Technology.

Physical modeling is another type of modal-emissions models. This approach is based on a parameterized analytical representation of emissions production. The entire emissions process is broken down into compartments that correspond to physical phenomena associated with vehicle operation and emissions production. The compartments used in this model are; tractive power demand function, engine power demand function, gear selection, emission control strategy and emission functions (Barth *et al.*, 1996).

Researchers at the University of California at Riverside have completed the development of a modal-emissions model under sponsorship from the National Cooperative Highway Research Program (NCHRP) (Barth and Norbeck, 1997). In the next phase, the interface between the developed modal model-emissions model and existing transportation models will be examined. These transportation models will use velocity, acceleration, position (road grade), and front and rear vehicle spacings as input to the modal model. Examples of the transportation models that will be used may include: CORSIM and TRANSIMS. A similar approach was used by other researchers such as An and Ross (1996), An *et al.* (1996,1997).

1.3.3 Fuel-Based Models

In the fuel-based method, emission factors are normalized to fuel consumption and expressed as grams of pollutant emitted per gallon of gasoline burned instead of grams of pollutant per mile. In order to obtain an overall fleet-average emission factor, average emission factors for subgroups of vehicles are weighted by the fraction of total fuel used by each vehicle subgroup. The fleet-average emission factor is multiplied by regional fuel sales to compute pollutant emissions (Singer and Harley, 1996). Fuel-based emission factors may be calculated from on-road emissions measurements such as those from remote sensors which are discussed later.

It is also possible to estimate g/gallon emission factors using data collected in tunnel studies. In tunnel studies, the net flux of pollutants and the traffic flow are measured. The data are then analyzed to estimate emissions in grams per mile of vehicle travel. Tunnels studies were conducted on the Sherman Way tunnel in Van Nuys, California, the Caldecott tunnel near Oakland, CA, the Fort McHenry tunnel in Baltimore, and the Tuscarora Mountain tunnel on the Pennsylvania Turnpike (Singer and Harley, 1996; Kirchstetter *et al.*, 1996). The Sherman Way tunnel is often highly congested, with highly variable traffic flow. In contrast, the Tuscarora Mountain tunnel is located in a rural area, and is approximately one mile long. Therefore, traffic tends to flow smoothly through the tunnel. Nearly all tunnel study researches have compared their emission factor estimates to predictions of MOBILE5a and EMFAC. However, these comparisons can be highly misleading. This is because the emissions estimates from the driving cycle-based models are based upon an entire trip, which includes

variations above and below the average speed of the driving cycle. The speed-time profiles in the tunnels may reflect steady cruising, as is often the case in the Tuscarora Mountain tunnel, or other profiles that differ from the driving cycles used to develop the models. Therefore, comparisons between the tunnel-based emissions factors and the model predictions can be interpreted only qualitatively. Differences in the speed profiles in each case must be considered and may help explain any differences in emissions estimates. Furthermore, the precision and accuracy of both the tunnel-based estimates and the model predictions must be considered to evaluate whether any apparent differences are statistically significant.

If the fleet composition within the tunnel varies over time, separate emission factors can be derived for different vehicle types (Gorse, 1984). Thus, the fuel based approach is amenable to the use of emissions data collected for on-road vehicles using either remote sensing or tunnel studies, as opposed to relying on laboratory tests in the driving cycle approach. Therefore, this approach may yield a key benefit of being more representative of on-road emissions.

In a fuel-based inventory, vehicle activity is measured by fuel use. Precise fuel sales data are available at the statewide level from tax records. Calculation of the emission inventory for individual air basins requires that fuel use be resolved to an appropriate spatial scale. Fuel use can be apportioned by tracking fuel shipments from major suppliers, through surveys of filling stations, or by considering the breakdowns of population and of the number of registered vehicles among all air basins in the state (Singer and Harley, 1996).

The total emission of any pollutant is calculated by multiplying the overall fleet-average emission factor for the pollutant, in grams per gallon, by total fuel use. Emissions can be calculated by vehicle class by applying the multiplication separately for each class.

The accuracy of a fuel-based model depends on how well the vehicles and driving modes from which emission factors were measured represent the entire area under study. The accuracy of the age distribution used to weight emissions data from each vehicle model year is another important consideration.

1.3.4 On-Road Data-Based Models

Current in-use models, i.e. MOBILE and EMFAC, are insensitive to vehicles modal events such as acceleration/deceleration, cruise speed, and idling. They can not be used effectively to evaluate the traffic control and management strategies that are aimed at reducing vehicle emissions. Other models, discussed in previous section, are in the developing stage. The efficiency of these models for the microscopic emission analysis of advanced traffic networks is still unknown. On-road emissions data can be obtained by using Remote Sensing Device (RSD) or on-board instrumentation. Detailed information on RSD will be given in following sections.

On-board emissions measurement method is widely recognized as the most desirable approach for quantifying highway vehicle emissions. However, on-board

instrumentation has been prohibitively expensive. Recently, a less expensive and portable system is developed by University of Pittsburgh (Vojtisek-Lom and Cobb, 1997).

The system is deployed by inserting a sampling probe to the tailpipe and routing a small portion of the tailpipe gases to the five-gas analyzer. A five-gas analyzer is able to measure concentrations of HC, CO, CO_2, NO_X, and O_2 in the exhaust gas. The engine diagnostic scanner is connected to an engine diagnostic link. The scanner measures engine parameters such as intake manifold absolute pressure, engine speed (RPM), intake air temperature, vehicle speed (MPH), coolant temperature, and intake mass airflow. These data are downloaded onto a laptop computer. The data from both the five-gas analyzer and the engine analyzer are used to calculate emissions of HC, CO, and NO_X on grams per gallon and grams per mile basis.

A study using a remote sensing instrument was conducted for the Texas Department of Transportation to collect on-road emissions data and develop an emission estimation model (Yu, 1997). Data were collected from five highway sites in Houston Texas, using the Santa Barbara Research Center's RSD "Smog Dog". Throughout the experiment ambient temperature, humidity, speed, and acceleration of the vehicles were recorded. CO and HC emissions were recorded as a volume percentage by RSD. In order to convert these values to a grams per mile basis regression fits obtained by the South Coast Air Quality Management District (SCAQMD) were used.

As explained by Yu (1997) SCAQMD developed a linear correlation relationship between the emission concentrations from the smog check, measured by RSD, and IM240 emissions in grams per mile readings. These two tests are implemented by California to enhance their Inspection and Maintenance (I/M) program. The smog test detects the emission concentrations of the exhaust vehicles at idle and at a fast idle speed of approximately 2500 RPM. If the emission concentrations exceed the emission thresholds which are specific for each vehicle type and model year, the vehicle is scheduled for the IM240 test which can identify the emissions in grams per mile to confirm if the vehicle is indeed a high emitter vehicle. The IM240 test is identical to FTP test except it lasts for 240 seconds. Recognizing the problem that the smog check test can not provide grams per mile emission concentrations, the SCAQMD developed a correlation between smog check data and IM240 mass emissions readings. These correlation were based on data from Orange County remote sensing program, the City of Los Angeles remote sensor program, and Hughes remote sensing data. The correlations are as follows:

$$CO(g/mi) = 11.1 \times CO(\%) + 21.3,\ R^2 = 0.52 \qquad (1\text{-}1)$$

$$HC(g/mi) = 63.3 \times HC(\%) + 1.7,\ R^2 = 0.42 \qquad (1\text{-}2)$$

The coefficients of determination (R^2) for the regression fits are given above. The regression fit can only explain 52 percent of the variability for CO and 42 percent variability for HC. Statistically these numbers are not strong evidences of linear correlation. In order to test the significance of the regression fits statistical tests, t-test or

F-test, should be analyzed. Even significant test results may not imply a linear relationship between smog check and IM240 test results. When the formula is analyzed it is seen that the model predicts that no vehicle can emit less than 21.3 g/mi or CO and 1.7 g/mi of HC. This result is unrealistic. Furthermore, smog check is measuring instantaneous emissions whereas IM240 measures mass concentrations of a vehicle running for 240 seconds. These methods are different and their comparison is problematic. Another important point is that only high emitter vehicles were used to develop the correlation relation. That means this equation can not be used for low emitting vehicles, otherwise it would be out of range. Instead of using this conversion a method based upon chemical mass balances, like the one developed by Frey and Eichenberger (1997), should have been used.

The emissions in g/mi were further converted into emission rates in g/sec based on the instantaneous speed of each vehicle when the respective emission data was recorded using the following equation;

$$CO/HC(g/s) = \frac{CO/HC(g/mi) \times Speed(mi/hr)}{3600} \qquad (1\text{-}3)$$

Using g/sec emissions data a linear regression was fit for emissions where the independent variables were speed, speed^2, acceleration, acceleration^2, temperature, and humidity. After statistical analysis, using t-test results, speed, and acceleration were found to be important variables for CO emissions. Speed, acceleration, and temperature were found to be important for HC emissions. Individual regression fits were obtained

for three different types of vehicles for CO and HC. In all these regression fits different variables were used. The coefficient of multiple determination for the final CO regression fit was 0.52. The coefficient of multiple determination for HC regression fit is not given.

After developing the model Yu (1997) made a comparison with in-use emission models; MOBILE and EMFAC and transportation models; TRANSYT and INTEGRATION. For this purpose, on-road emission model was used to emulate the FTP driving cycles. The instantaneous emission rates were estimated based upon instantaneous speed and acceleration values obtained from FTP driving cycle. Then instantaneous emission rates were summed for the entire driving cycle. The results showed that the emissions from on-road model were higher than the emissions estimates obtained from other models.

Yu (1997) concluded that on-road model is an efficient emission model that can be used in ITS applications. Furthermore, he stated that minimum values of emission factors were obtained at optimal speeds between 30 and 40 mph. He even indicated that by influencing drivers to drive at optimal speed values, overall emission amounts would be reduced. Based on the previous comments on the model, this conclusion is not justified.

1.4 Remote Sensing of On-Road Emissions

Remote sensing technology has been used for about nine years to measure the CO and HC emissions from on-road vehicles. Remote sensing data can be used to develop emission estimates that are representative of actual emissions.

There are several applications of remote sensing in mobile emissions determination. These are: monitoring of emissions to evaluate the overall effectiveness of inspection and maintenance programs; identification of high emitting vehicles for inspection or enforcement purposes; and development of emission factors.

Inspection and maintenance (I/M) programs have been developed in order to check whether highway vehicles are in compliance with local programs aimed at reducing vehicle emissions. In the U.S., regions that are in non-attainment or maintenance areas under State Implementation Plans (SIPs) for National Ambient Air Quality Standards (NAAQS) have been required to implement I/M programs. In general, I/M tests involve measurement of tail pipe emissions at idle and at 2500 rpm with no load on the engine. Inspections are performed in either centralized or decentralized stations, annually or biennially. The centralized I/M programs were generally predicted to be more effective than the decentralized programs. These programs have been criticized for several reasons including: non representativenes of the tests, the possibility of cheating, and lack of repeatability (Zhang *et al.*, 1996b).

It has been reported that the reduction of emissions due to a centralized I/M program in Tucson and a decentralized I/M program in Los Angeles were far less than the U.S. EPA's computer model predicted (Zhang *et al.*, 1996b). Lawson (1993) showed that the various centralized, decentralized, biennial, and "enhanced" I/M programs investigated in California and other locations have done little to reduce emissions from motor vehicle fleet. As a result of these studies, several states intend to stop implementation of their I/M programs and seek other alternatives to the traditional approach to vehicle testing (Zhang *et al.*, 1996b).

An alternative to centralized I/M programs may be the use of random roadside pullover surveys. It has been reported that emissions measured in voluntary random roadside surveys are generally higher than measurements on the same vehicle during a scheduled I/M test (Ashbaugh and Lawson, 1991). However, these roadside surveys were not truly random since 30 percent of the drivers refused the test. Remote sensing measurements revealed that the average on-road CO and HC emissions of those vehicles that refused inspection were more than double those of other vehicles. Thus, the results from the roadside surveys may be biased low because this small fraction of high-emitting vehicles is under-represented. Using biased information could significantly underpredict overall on-road fleet emissions.

Considering the above findings, it is clear that a new method is needed for I/M programs to control the vehicle emissions. There have been several projects done using RSD in I/M programs (Zhang *et al.*, 1996b; Lawson *et al.*, 1990; Zhang *et al.*, 1995). In a study done by Lawson *et al.* (1990) results from remote sensing measurements and

I/M testing were compared. They found that out of 60 vehicles measured, 45 vehicles had passed I/M testing. However, remote sensing results showed that CO emissions for these vehicles were higher than the allowable limit.

In another study, Zhang *et al.* (1996a) evaluated the efficiency of I/M programs in Tucson, AZ, and in rural Colorado. Thousands of data were collected by remote sensing and compared with I/M test results using license plates that are recorded during the measurement. The results showed that vehicles that had passed I/M testing had higher CO emissions than expected. The authors concluded that the I/M programs are less effective than expected in determining and controlling highway vehicle emissions and suggested including methods such as remote sensing in I/M programs to estimate on-road emissions with a better accuracy.

From these studies it is clear that I/M programs need to be improved in order to estimate and control the emissions from highway vehicles. Remote sensing is one way to improve the emission estimates by measuring the emissions while the vehicles are in-use. These measurements are representative of on-road emissions. Thus, remote sensing can be used to monitor the efficacy of I/M programs. In some states, it may be possible to use remote sensing, for assessing and enforcing compliance by individual vehicles, as allowed by state laws.

Several studies have claimed that remote sensing can be useful in identifying high-emitting vehicles. In a project by Stedman (1989), it was found that 8.6 percent of the measurements of on-road vehicles accounted for half of the CO emissions, whereas

71 percent of the measurements were sufficiently low as to have negligible influence on total fleet air pollutant emissions. Stephens *et al.* (1997) stated that 8.9 percent of the measured vehicles were emitting 50 percent of the CO in their study. They also concluded that pre-1980 vehicles emitted 57 percent of the CO and that 1988-1989 vehicles contributed only 2 percent of the CO emitted. In another project, Zhang *et al.* (1995) stated that the majority of the measured on-road exhaust emissions are from the highest emitting 10 percent of the vehicles for CO and 15 percent of the vehicles for HC. Thus, many investigators believe that remote sensing can be used to help enforce motor vehicle emissions regulations. However, the ability to reliably identify a high emitting vehicle is limited by the precision and accuracy of remote sensing, and by the inherent variability in emissions for individual vehicles (Frey and Eichenberger, 1997). These factors are further discussed in Section 2.5.

Remote sensing has been used to develop emission factors for highway vehicles. For example, in a remote sensing study done by Singer and Harley (1996), vehicle emission factors were calculated on a grams per gallon basis. Another study by Bishop and Stedman (1990) estimated the reduction in the CO emissions from the use of oxy-fuels in Colorado. In this study, CO emissions from approximately 117,000 vehicles were measured by remote sensing. The measurements were used to calculate grams of CO emitted per gallon of fuel consumed.

A similar approach was used by Frey and Eichenberger (1997) to develop emission factors for school and transit buses.

2.0 INSTRUMENTATION

In this chapter information regarding the equipment that was used in this study is presented. On-road vehicle emissions measurements were collected using a remote sensing device (RSD) and traffic data were collected using a video tracking system, MOBILIZER©. First the theoretical background and operating principles of the RSD will be given. Then MOBILIZER© system will be explained.

2.1 Emission Data Collection: Theory and Operation

The University of Denver introduced remote sensing devices (RSDs) in 1987 for measuring exhaust carbon monoxide (CO) (Stedman, 1989), followed by General Motors (GM) in 1988. Simultaneous determination of hydrocarbon (HC) emissions was added to the sensors in 1990 (Bishop *et al.*, 1989; Stedman, 1989; Stephens and Cadle, 1991). More recently, a capability to measure NO emissions has been developed (Zhang *et al.*, 1996a; Popp *et al.*, 1997).

2.1.1 Remote Sensing Device (RSD) Equipment

A schematic diagram of a remote sensor deployed at a measurement site is shown in Figure 2.1. The components shown in the diagram are generic to highway vehicle remote sensing technologies. In this study the North Carolina Department of

Environment, and Natural Resource's (NC DENR's) RES-1 "Smog Dog" is utilized for emission measurement. Typical remote sensing equipment is housed in a cargo van, which contains an on-board computer, and calibration, and video systems (including video monitors and videocassette recorder) as shown in Figure 2.2. An infrared (IR) source is used to transmit an IR beam that is detected by an IR receiver. Remote sensing systems have a receiver-alignment module which is used for alignment of the beam from the receiver to the source, a camera for viewing the license plates of passing vehicles, and a video camera for recording images of the back of each passing vehicle. A portable electric generator and an exhaust tube for the on-board generator in the van are located downwind of the IR beam, to avoid interference of the generator exhaust plumes with the vehicle emissions measurement. Traffic cones are used on both sides of the road to warn traffic away from the equipment.

The IR measurement system for remote vehicle emission measurements operates using a technology known as Non-Dispersive Infrared Absorption (NDIR) which is described in detail in Section 2.1.2.

The IR beam is set at a height where it will be obstructed when a vehicle passes in front of it. The height and initial alignment of the beam is determined manually through an iterative process. A low power laser is mounted inside the IR source. Both the source and receiver are initially setup based upon an estimate of what the road-side elevation of each device must be in order for the beam to be at a specified height from the pavement at the location of a typical vehicle exhaust. With experience, it is possible

to initially align the source so that the laser beam can be seen on the front edge of the box containing the IR receiver. When the laser beam is located at the top right corner of the receiver box, then a measurement of the beam height over the pavement is made. This is done by first ensuring that there is no vehicular traffic, and then by manually placing a card or piece of paper in the path of the laser beam at the location over the pavement where the exhaust pipes are typically expected. Using a yardstick, the distance from the pavement surface to the laser beam is measured. This distance is the estimate of the beam height.

The height of the beam must be such that it will both be blocked by the passing vehicle and pass through the subsequent exhaust plume. For automobiles, the beam height is typically 13 to 17 inches.

The final alignment of the beam is done using an electronic alignment module. This module provides a visual representation of the strength of the IR signal. The final alignment is typically accomplished by making minor adjustments to the height and angle of the receiver. The strength of the IR signal is also displayed on the computer console as a voltage. Figure 2.3 illustrates deployment of IR source and receiver at one of the measurement sites.

Information on individual vehicles and fleet composition greatly assist in the effective interpretation of measurements from remote sensing technology. Therefore, license plates of the passing vehicles are recorded by a video camera. A frame grabber captures the image of the license plate. Using a character recognition program, referred

to as Automatic License Plate Recognition (ALPR), the license plate number is recorded. The license plate data are integrated with the remote sensor output. Specific vehicle information such as vehicle's age, make, and model can be obtained from a state vehicle registration database. Database development activities will be presented in Section 4.1.

In our study, a speed-acceleration measurement device has been incorporated into the RSD. This device consists of two pneumatic tubes placed across the roadway in parallel with each other, separated by six feet. The pressure changes from tubes are analyzed by a computer to estimate the speeds of the first two axles of each vehicle. The difference in the speeds of the two axles are used to estimate the vehicle acceleration. Both the two speeds and the acceleration data are included within the RSD output.

In the following section the theory behind remote sensing technology will be discussed.

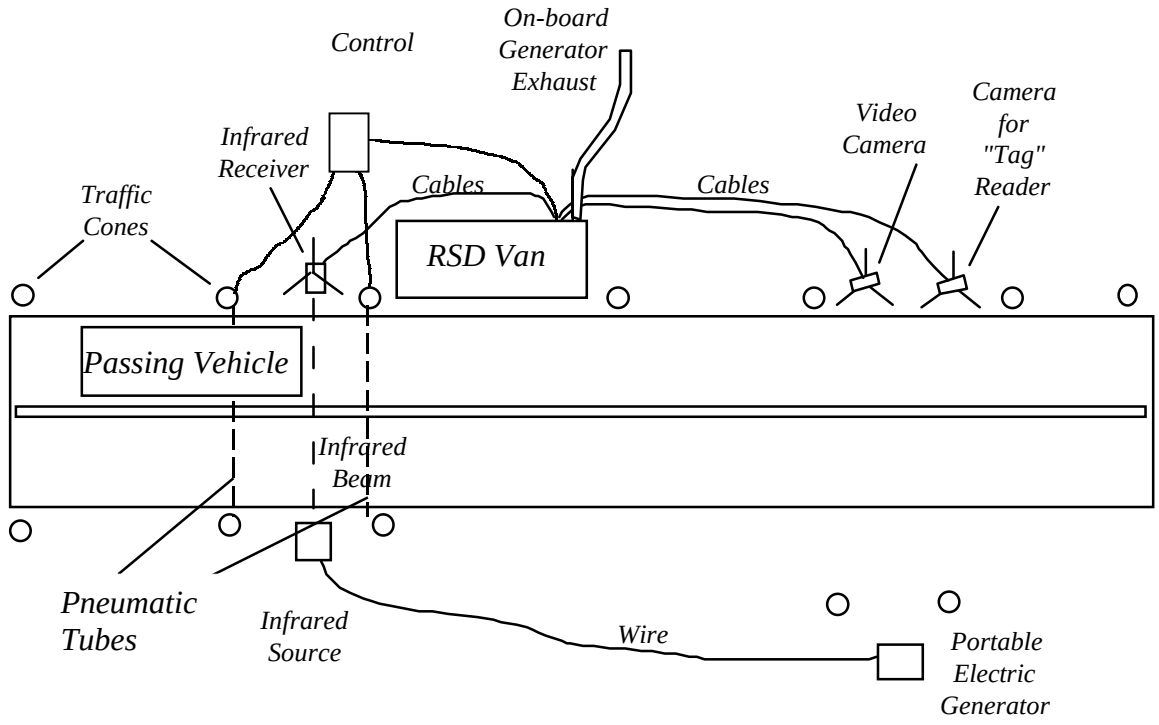


Figure 2.1. Simplified Schematic of a Deployed Remote Sensing Device (not to scale).

Figure 2.2. Photograph of interior of the RES-1 “Smog Dog”. Shown are the computer console (left), video monitor (center), and monitor for the license plate tag reader (left).

Figure 2.3. View of the infrared source (black box on left) and receiver (white object on tripod on right), with the pneumatic tubes for speed measurement, deployed at site 1, junction of State Highway NC 86 and Interstate I-40.

2.1.2 Non-Dispersive Infrared Absorption

NDIR photometry is a method for determining the concentration of a compound in a gas mixture based upon the absorption of IR energy by that compound over a specific spectral range. Specifically, the technique involves determining the difference in IR energy absorption in a beam with a narrow range of wavelengths through which a gas sample is passed containing the compound of interest (Lodge, 1991).

When IR light is exposed to a particular compound, that part of the spectrum which has the same wavelength as the vibrational frequencies of the atomic bonds is absorbed (Brady and Humiston, 1978). By carefully observing how much specific regions of the IR spectrum are absorbed, one can deduce the structure and relative concentration of molecules responsible for the absorption.

RSDs operate on the same generic principle. An IR beam is continuously directed across a lane of traffic to a detector. The detector works on principle of selective absorption of IR radiation by CO, CO_2, and HC gases. It uses band pass filters to isolate the HC, CO, and CO_2 absorption regions of the spectrum. RSD equipment has another non-absorbing region, which is used to monitor the beam intensity. The digital acquisition system, which is triggered as the vehicle breaks the IR beam, acquires, displays, and stores the transient attenuation of the IR beam due to the concentrations of CO, HC, and CO_2 in the dispersing exhaust cloud for a user-selected period. A computer analyzes the data and determines the molar CO/CO_2 and HC/CO_2 ratios (SBRC, 1994).

RSD calibration methods are explained in the following section.

2.1.3 Calibration

In order to insure the accuracy of the RSD system, three separate calibration or verification methods are applied. The system is first calibrated at the factory. It is then calibrated in the field before and during use. The system calibration can also be verified during use by taking a measurement of a calibration gas during 'Puff In Vehicle Mode' (PIV).

The instrument is calibrated at the factory using a commercially available mixture of CO, CO_2, and HC gases whose nominal concentrations are known within one

percent. These gases are measured with different diluent concentrations by an optical cell of known transmittance. The resulting photovoltaic detector voltages versus gas concentration data are stored. A curve-fitting polynomial equation is then generated for each detector, and this equation is stored in the on-board computer system (SBRC, 1994).

Field calibration must be performed daily, prior to data collection, to take the changes in ambient conditions into account. Field calibration compensates primarily for variations in the CO_2 background by using the CO sensor and the CO factory calibration as an absolute standard. The calibration gas is introduced to the path of the infrared beam via a tube from the on-board calibration gas cylinders to the infrared receiver. Typical calibration gas corresponds to a molar CO/CO_2 ratio of 1.0 and a molar HC/CO_2 ratio of 0.167 (SBRC, 1994). In our study, we used calibration gas having a CO/CO_2 ratio of 1.0 and a HC/CO_2 ratio of 0.133.

The resulting RSD measurements of the calibration gas are required to be within a specified tolerance of these ratios of gas concentrations. This forces the concentration-versus-time curves for CO and HC to be in the correct ratio at all times. The resulting distribution of voltage-versus-concentration data is used to modify the factory-generated polynomial curve-fit equations for CO_2 and HC to account for variations in CO_2 levels.

PIV mode measurements allow verification of accurate system operation while data collection is underway. When manually selected, this procedure causes a sample of the field calibration gas to be 'puffed' through the receiver optical cell. The resulting

CO/CO_2 and HC/CO_2 ratios must agree with the calibration gas to within plus or minus 10 percent. If not, then a new calibration must be done. We have separately analyzed all PIV measurements obtained during ten days of field collection in our study. In our study, the average PIV CO/CO_2 ratio was found to be 1.009. The average for the PIV HC/CO_2 ratio is 0.132. These averages are nearly identical to the true ratios of the calibration gas, and in fact are not statistically significantly different from the true values. Therefore, the system is deemed to be accurate. The standard deviation of the PIV measurements was 0.066 for CO/CO_2 and 0.007 for HC/CO_2.

The cumulative distributions of the PIV measurements for the CO/CO_2 and HC/CO_2 ratios are given in Figures 2.4 and 2.5 respectively.

The dotted lines in Figures 2.4 and 2.5 represent the plus or minus 10 percent range within which the PIV measurements should be. When a PIV measurement is outside of this range, a field calibration is done. The new calibration is then verified by taking another PIV to make sure that the calibration is accurate.

The PIV calculations offer a means by which the precision of measurements may be ascertained, since they are based upon repeated readings of a gas of known composition. As shown in Figure 2.4, 95 percent of CO/CO_2 measurements in PIV mode were in the range of 0.88 to 1.13, approximately within plus or minus 12 percent of the mean value. For the HC/CO_2 PIV measurements, as shown in Figure 2.5, 95 percent of the values were within a range of 0.119 to 0.144, approximately plus or minus 10 percent

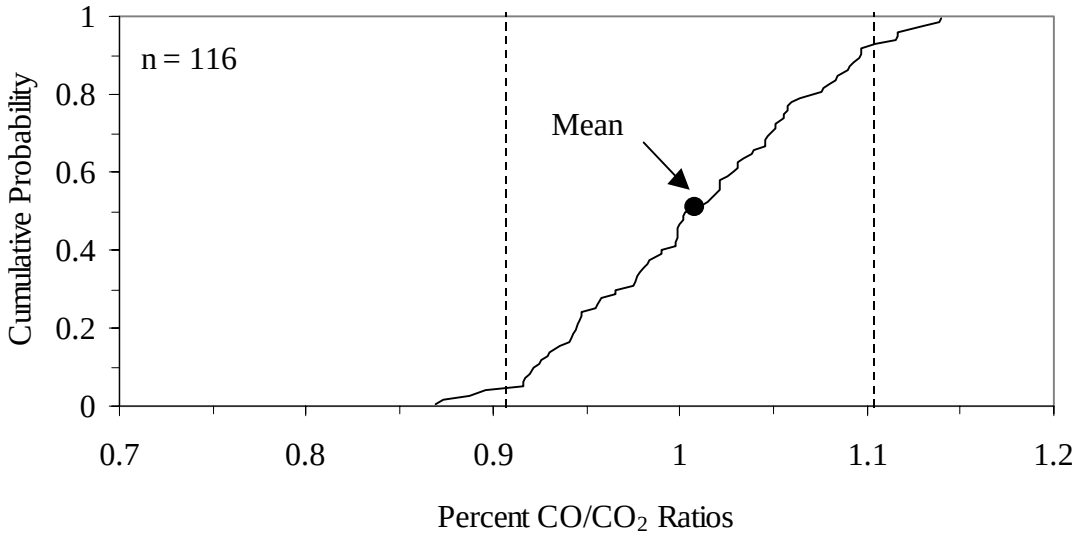


Figure 2.4. CO/CO_2 ratios measured for "Puff in Vehicle" mode.

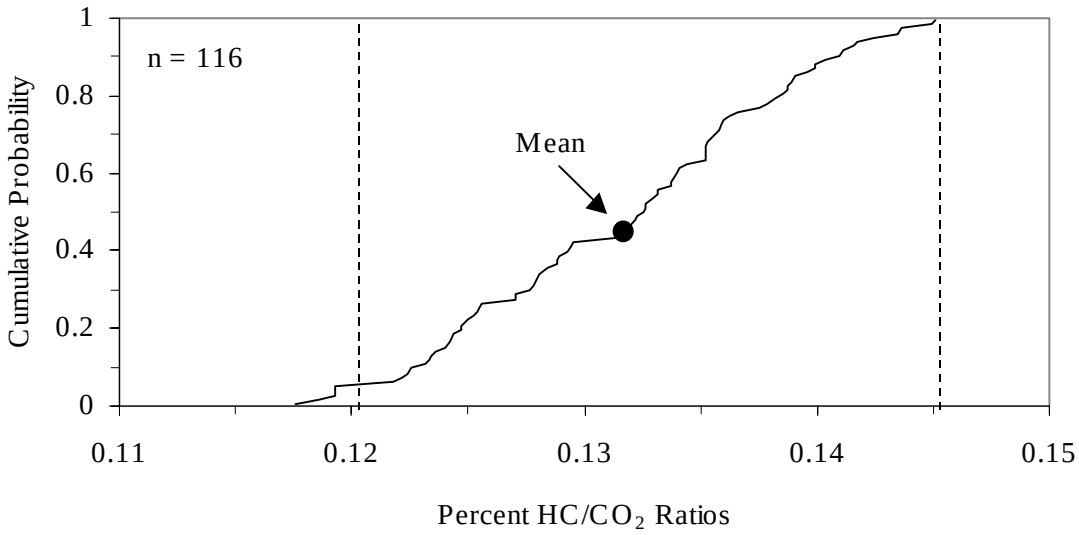


Figure 2.5. HC/CO_2 ratios measured for "Puff in Vehicle" mode.

of the mean value. Of course the distribution of the PIV measurements is influenced by the fact that the system is recalibrated if a PIV measurement is outside the plus or minus 10 percent range. The observed precisions of the measurements of the calibration gas are approximately plus or minus 10 percent.

The following section will explain the steps of RSD operation in detail.

2.1.4 Data Acquisition

In this section, the major steps of RSD operation will be explained in detail. First an overview of the methodology will be given. Then each step of operation will be explained with the help of an example RSD data output file for measurement of a single vehicle.

The steps of RSD operation can be summarized as follows:

(1) When a vehicle enters the path of the IR beam, the beam is broken. When operated in a mode involving automatic collection of background measurements, the RSD will store the most recent background levels of the IR beam obtained over a user-specified pre-trigger interval, such as 0.4 seconds. The vehicle-specific background readings are then used to help calculate the net transmittance of each channel of the IR beam. Alternatively, the system can be operated in a mode by which background readings are taken manually by the operator. The advantage of manual background readings is that the system resets more quickly, enabling more frequent measurements of vehicles. For the manual background reading option the user decides when to collect background readings.

(2) When the beam is reestablished after a vehicle passes, a second signal is sent to the computer indicating that the vehicle has exited from the beam path. At this time, the absorption of the IR beam by CO, CO_2 and HC gases in three

different channels are recorded. Another channel is used as a reference to evaluate any signal attenuation due to changes in atmospheric conditions that systematically affect all channels.

(3) An on-board computer calculates CO/CO_2 and HC/CO_2 ratios based upon the measured decrease in the beam strengths and information available from the calibration file. The details of this calculation will be described.

(4) Molar CO/CO_2 and HC/CO_2 ratios for each vehicle are stored in a file and combined with speed, acceleration, and license plate information obtained from the speed measuring device and license plate reader.

The computer constantly receives data from the four detector channels, at every one-thousandth of a second. As a vehicle enters and exits the IR beam, trigger pulses are sent to the computer for timing purposes. The actual time when data is collected from the vehicle exhaust is user-selectable and is adjustable from 0.4 to 1.5 seconds after the vehicle exits the beam. When processing the data, the RSD computer compares the transmittance of the beam through the exhaust plume with the transmittance of the beam measured during background conditions.

When operating in an automatic background reading mode, there is a limited time during which the vehicle must pass through the beam in order to obtain a valid measurement. The total sampling time for a single vehicle is 2 seconds, which must be allocated to background reading, time for the vehicle to pass through the beam, and time to collect data regarding the vehicle exhaust. Typically, the most recent 0.4 seconds of background readings are saved, and 0.6 seconds is allocated to vehicle

exhaust measurements. This then leaves approximately one second for the vehicle to interrupt the beam as it passes by. If the vehicle is very long and/or moving too slowly, it can be difficult to obtain a measurement in automatic background mode. For this reason, manual background mode is often preferred, which imposes no time constraint on how long the vehicle blocks the beam.

In the case of manual background reading operating mode, a trade-off is that the time between the manual background readings can be large. Therefore, the representativeness of the background reading for the ambient atmospheric conditions at the time a vehicle passes by may be unknown. This problem can be eliminated by taking background readings at a regular interval. Also, a new background is obtained and stored anytime that a new calibration is done.

In our study, we took automatic background readings for four of the data sets collected at Site 2 on May 26, May 27, June 8, and June 9. For data collected on April 23 at Site 1, background readings were taken automatically. The automatic background reading option was turned off for data collected on three days at Site 1, involving April 6, April 7, and April 24. For two data sets collected at Site 2, automatic background readings were not taken including data collected on May 11 and May 12. The data from all sites were analyzed to identify possible drift associated with calibrations and changes in background. The results are given in section 2.1.5.2.

2.1.5 Calculation of CO/CO_2 and HC/CO_2 Ratios

Data are collected and analyzed by the on-board computer to estimate the transmittance of the CO, CO_2, and HC channels of the beam. With the help of calibration curve data, the computer first calculates and displays a plot of the relative percentages of CO versus tailpipe exhaust data points obtained over time from the IR beam. Also displayed are computer generated plots of the relative percentages of concentrations for CO_2 and HC versus time. Then, estimated volume percentages of CO versus CO_2 and HC versus CO_2 are plotted. Linear regression is fitted to data points in each of these plots. The slopes of the regression equations are reported as the CO/CO_2 and HC/CO_2 ratios (SBRC, 1994). In this section, we review in detail these major steps for calculating the ratios.

Details of the steps for the CO/CO_2 and HC/CO_2 ratio calculations by the RSD on-board computer are not reported by the manufacturer or other resources in the literature. In order to understand the procedure, we saved the individual vehicle emission data files during the measurements. Then these files were run via a software called “Regraph”, which is also used in the RSD on-board computer, to calculate the molar CO/CO_2 and HC/CO_2 ratios. The outputs of this program are plots for ratio calculations and an ASCII file which includes the data recorded at different steps of the calculations. An example of the graphical output of this program is given in Figure 2.6 and will be explained. In this section, an example of an individual vehicle emission data file will be utilized to explain the steps of CO/CO_2 and HC/CO_2 ratio calculations. The

data file is given in Appendix A and is for vehicle number 34 observed at Site 2 on June 8.

The data file first reports the polynomial curve-fit equations obtained from the field and factory calibrations. The coefficients of the polynomial curve-fit equations for our example are given in Table 2.1 below:

Table 2.1. Coefficients of Polynomial Fit of Calibration Curves for the Example Data

	CO	CO_2	HC
Constant	-1973.627686	-235.521	-252.009
X	91.2831192	13.04732	11.55001
X^2	-1.555695057	-0.25198	-0.19633
X^3	0.011631903	0.002084	0.001474
X^4	-3.22964E-05	-6.3E-06	-4.1E-06

Source: RSD File As60898a.034

The polynomial curve-fit equation coefficients for CO are constant for all measurements since they are from factory calibration measurements. The CO_2 and HC polynomial curve-fit equation coefficients are obtained from field calibration measurements and vary for each field calibration measurement. Each polynomial is used to calculate the volume percent of each pollutant based upon the percent transmittance of the respective IR beam channel. The calibration curve for CO for this example is given in Figures 2.6. Figures 2.7 and 2.8 illustrates the calibration curves for CO_2 and HC respectively.

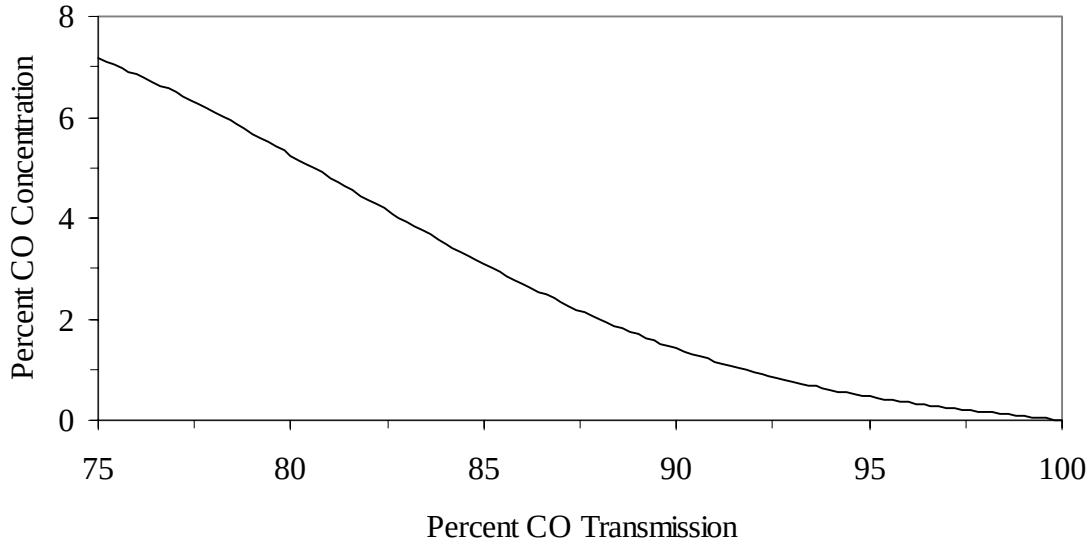


Figure 2.6. Calibration Curve for CO.

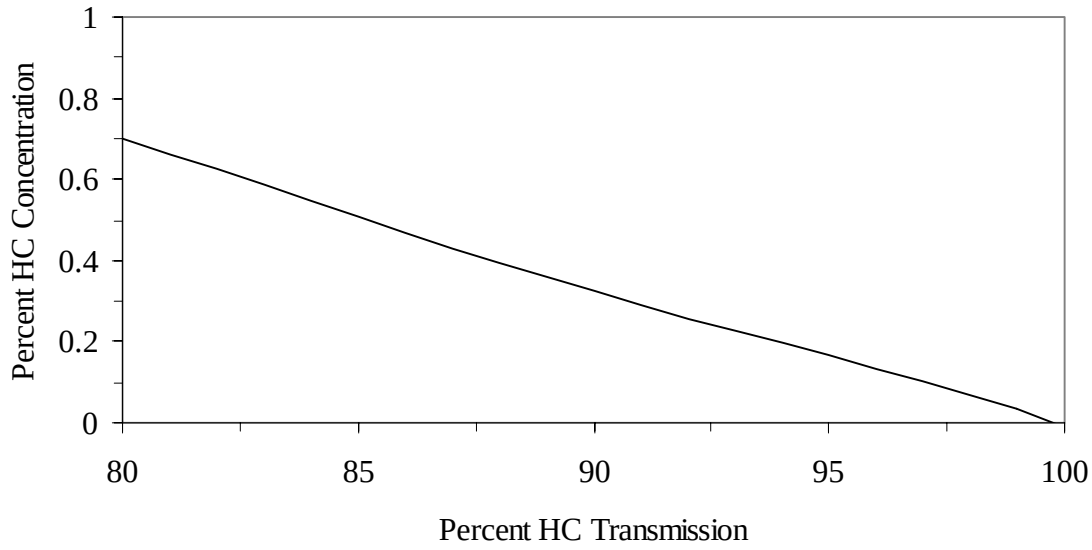


Figure 2.7. Calibration Curve for HC.

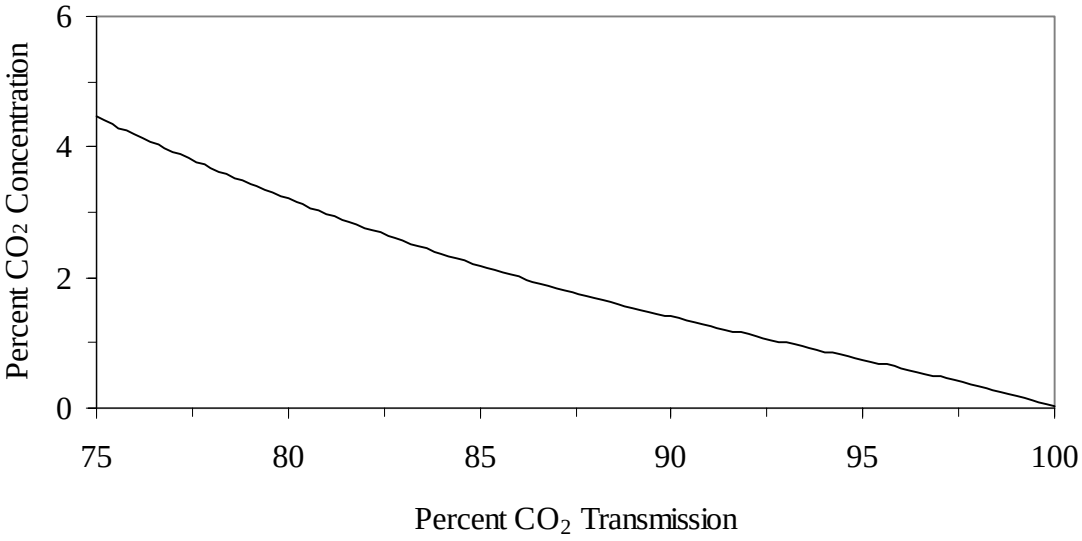


Figure 2.8. Calibration Curve for CO_2.

The calibration curves should be a monotonically decreasing function, otherwise erroneous results might be obtained for the volume percentage of pollutants. If calibration curves show erroneous results, a new calibration must performed. For our example the calibration curves are monotonic and there are no negative values in the estimated percentage of pollutants.

"Pre-car" and "post-car" transmittance data for four IR beam channels, reference, CO, CO_2, and HC, are given in the data file. "Pre-car" refers to background measurements, whereas "post-car" is the data collected after the vehicle exits the beam. Table 2.2 displays a sample of selected pre- and post-car data. Conceptually, these data can be thought of as measures of the strength of the signal of the beam. The exact units of measurement used here are not reported. However, these data are subsequently

normalized, so measurement units are not needed. Table 2.2 contains 10 example rows of data values. Each row represents an increment in time. Typically, 70 points of data are obtained for the post-car measurements. Only 10 are shown for simplicity. All 70 volumes for this example are given in Appendix A.

Table 2.2. Transmittance Data for Pre- and Post-Measurements from the Sample Data

Data Point	Pre-Ref	Post-Ref	Pre-CO	Post-CO	Pre-CO_2	Post-CO_2	Pre-HC	Post-HC
1	19769	19914	22159	22051	12944	11919	16367	16012
2	19759	19905	22160	22045	12945	11762	16358	16000
3	19758	19909	22159	22057	12938	11762	16359	16004
4	19756	19899	22161	22053	12940	11776	16355	15998
5	19740	19893	22142	22057	12937	11920	16345	15990
6	19754	19892	22148	22052	12943	11942	16358	15992
7	19761	19902	22161	22058	12948	12228	16356	15997
8	19756	19906	22159	22066	12945	12402	16356	16004
9	19759	19905	22151	22061	12949	12484	16369	15995
10	19758	19903	22146	22069	12949	12464	16364	15994

Source: RSD File As60898a.034

In order to calculate the transmittance of the beam of each pollutant, a multi-step procedure is followed. Once the transmittance is known, then the estimated volume percent of the pollutant in the beam path is calculated using the calibration polynomials. The first step is to calculate the average beam strength for the background measurements of each beam. Thus, in order to calculate the average background beam strength for CO, the following calculation is made:

$$\text{Average of Pre} - \text{CO} = \frac{1}{\text{n}} \sum_{i=1}^{\text{n}} (\text{Pre} - \text{CO})_{\text{i}} \qquad (2\text{-}1)$$

where i = data point number from Table 2.2. For example, the result for the average Pre-CO measurement is:

$$\text{Average of Pre} - \text{CO} = 22{,}154 \tag{2-2}$$

Based upon averaging of 10 example values shown in Table 2.2, the background-adjusted post-car signal is calculated by dividing each post-car data point by the average pre-car beam strength. This results in a preliminary estimate in the attenuation of the beam due to the presence of the pollutant in the post-car signal compared to the relative absence of the pollutant in the pre-car signal. For the example of CO, a general equation for this calculation is:

$$(\text{Post} - \text{CO})_i^b = \frac{(\text{Post} - \text{CO})_i}{\text{Average of Pre} - \text{CO}} \tag{2-3}$$

For example, the background-adjusted post-car CO beam attenuation for the 10^{th} post-car measurement is:

$$(\text{Post} - \text{CO})_{10}^b = \frac{(\text{Post} - \text{CO})_{10}}{\text{Average of Pre} - \text{CO}} \tag{2-4}$$

$$(\text{Post} - \text{CO})_{10}^b = \frac{22{,}069}{22{,}154} \tag{2-5}$$

$$(\text{Post} - \text{CO})_{10}^b = 0.99616 \tag{2-6}$$

The beam attenuation is adjusted for systematic changes in the strength of the beam associated with changes in atmospheric conditions or equipment operation that affect the beam strength in all channels. In order to make this adjustment, the average strength of the pre-car reference beam is calculated and is then used to calculate a background-adjusted reference beam attenuation factor. The average Pre-Ref beam strength is given by:

$$\text{Average of Pre} - \text{Ref} = \frac{1}{n}\sum_{i=1}^{n}(\text{Pre} - \text{Ref})_i \tag{2-7}$$

$$\text{Average of Pre} - \text{Ref} = 19{,}756 \tag{2-8}$$

The background-adjusted reference beam attenuation is given by:

$$(\text{Post} - \text{Ref})_i^b = \frac{(\text{Post} - \text{Ref})_i}{\text{Average of Pre} - \text{Ref}} \tag{2-9}$$

For the example of an individual post-car measurement, the background-adjusted reference beam attenuation for the 10th post-car measurement is:

$$(\text{Post} - \text{Ref})_{10}^b = \frac{(\text{Post} - \text{Ref})_{10}}{\text{Average of Pre} - \text{Ref}} \tag{2-10}$$

$$(\text{Post} - \text{Ref})_{10}^b = \frac{(19{,}903)}{19{,}756} \tag{2-11}$$

$$(\text{Post} - \text{Ref})_{10}^b = 1.00744 \tag{2-12}$$

This result indicates that the beam generally got stronger for the post-car measurement than for the pre-car. Therefore, to correct for this, it is necessary to make a downward adjustment to the post-car measurement for each pollutant. This is done by dividing the background-adjusted post-car measurement by the background-adjusted change in reference beam strength:

$$(\text{Post} - \text{CO})_i^R = \frac{(\text{Post} - \text{CO})_i^b}{(\text{Post} - \text{Ref})_i^b} \tag{2-13}$$

For the example of the 10th data point in Table 2.2, this becomes:

$$(\text{Post} - \text{CO})_{10}^R = \frac{0.99616}{1.00744} \tag{2-14}$$

$$(\text{Post} - \text{CO})_{10}^R = 0.98880 \tag{2-15}$$

A similar set of calculations is done for all data points and for all pollutants. The result is a table of normalized transmittances for each of the three pollutants, as summarized in Table 2.3. Values such as these are reported directly in the individual vehicle data files written by the RSD on-board computer, as illustrated in Appendix A.

Table 2.3. Normalized Gas Ratios for the Sample Data

Data Points	CO	CO_2	HC
1	0.9874	0.9128	0.9709
2	0.9876	0.9011	0.9706
3	0.9879	0.9010	0.9706
4	0.9882	0.9025	0.9707
5	0.9887	0.9138	0.9705
6	0.9885	0.9155	0.9707
7	0.9883	0.9370	0.9705
8	0.9885	0.9501	0.9708
9	0.9883	0.9565	0.9703
10	0.9888	0.9550	0.9703

Source: RSD File As60898a.034

In order to use the polynomial equations for the calibration curves, which convert beam transmittance to an estimate of the percentage of each gas in the beam path, the transmittance must be multiplied by 100 to convert from fractions to percentages. Thus, for example, the reference-adjusted CO attenuation associated with the 10th data point becomes 98.880 percent instead of 0.98880. The value 98.880 is entered as the x_{CO} value into the CO calibration curve polynomial:

$$Y_{CO,i} = -1973.627686 + 91.2831192\,x_{CO,i} - 1.555695057\left(x_{CO,i}\right)^2 + 0.011631903\left(x_{CO,i}\right)^3 - 3.22964\times10^{-5}\left(x_{CO,i}\right)^4 \quad (2\text{-}16)$$

$$Y_{CO,10} = -1973.627686 + 91.2831192(98.88) - 1.555695057(98.88)^2 + 0.011631903(98.88)^3 - 3.22964\times10^{-5}(98.88)^4 \quad (2\text{-}17)$$

$$Y_{CO,10} = 0.084 \quad (2\text{-}18)$$

For the polynomial equation the full precision of each coefficient in the polynomial model should be retained to obtain the correct estimate of volume percentage. For example, if one uses three significant figures for the coefficients, a result of approximately –34.09 is obtained when in fact the correct answer is 0.0840, for our case. Thus, numerical round-off error is of potentially significant concern when using these high order polynomial curve fits.

This procedure is followed to calculate the percentage of all three pollutants in the beam path, with the results of the calculations for the sample data set shown in Table 2.4.

Table 2.4. Percent Pollutant Values for the Sample Data

Data Points	CO (%)	CO_2 (%)	HC (ppm)
1	0.0949	1.2207	0.0996
2	0.0935	1.3781	0.1006
3	0.0908	1.3807	0.1004
4	0.0882	1.3596	0.1000
5	0.0844	1.2068	0.1006
6	0.0858	1.1840	0.1001
7	0.0876	0.9064	0.1007
8	0.0863	0.7379	0.0999
9	0.0877	0.6556	0.1015
10	0.0840	0.6743	0.1014

Source: RSD File As60898a.034

After calculating the concentrations, the RSD on-board computer plots the percent pollutant concentration versus data points as illustrated in Figure 2.9 for a larger data set than the example. These plots give the variation of pollutant concentrations over time, since each data point is collected at a different time. The computer also displays a scatter plot of percent CO versus percent CO_2 and percent HC versus percent CO_2.

In Figure 2.9 the three plots given at the top section shows the variation of the volume percentage of the pollutants, CO, CO_2 and HC respectively, for the 70 collected data points. As seen in the plots, CO varies between 0.083 to 0.093 percent. The range for CO_2 is between 0.37 to 1.35 percent. For HC data the range is from 0.099 to 1.01 percent. Under each plot the estimated signal to noise ratio (S/N) is given for that pollutant.

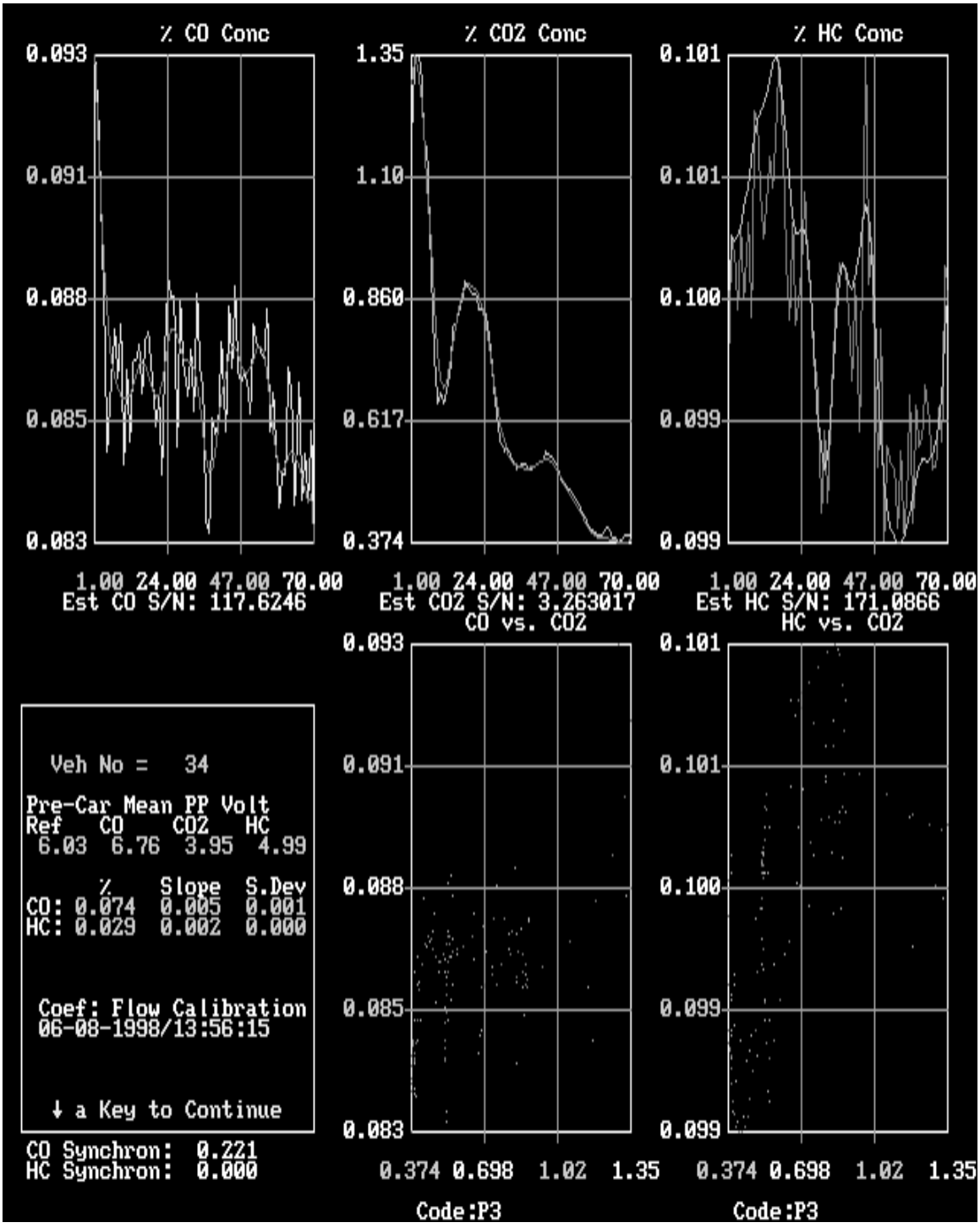


Figure 2.9. RSD output for vehicle number 34 measured on June 8 1998.

2.1.5.1 Statistical Methods for Calculating CO/CO_2 and HC/CO_2 Ratios: Linear Regression

After calculating the pollutant concentrations, the RSD on-board computer plots CO versus CO_2 and HC versus CO_2, as shown in Figure 2.9. These plots are used to calculate the CO/CO_2 and HC/CO_2 ratios. The actual levels of CO, CO_2, and HC in the exhaust plume are continuously changing as the exhaust plume enters the path of the beam and, subsequently, as the plume disperses. However, the ratio of the pollutants should remain approximately constant for a given vehicle. Therefore, linear regressions are fit to the CO versus CO_2 and HC versus CO_2 data to estimate the slope, which yields the ratios for CO/CO_2 and HC/CO_2.

The predictive variable in the regression is the calculated percent of CO_2 for each of the approximately 70 data points obtained from analysis of the IR beam transmittance. The dependent variable in the regression is the calculated percent of CO or HC for each of the approximately 70 data points obtained from analysis of the IR beam transmittances. A linear regression of the form:

$$y = a\,x + b \tag{2-19}$$

is calculated by the RSD computer for each individual vehicle for both CO and HC versus CO_2. For example, in the case of a regression for CO, the slope, a, represents the ratio of the change during the sampling time in the CO volume percentage in the path of the IR beam with respect to a unit change in the CO_2 volume percentage. Thus, the basic assumption in making the measurement is that the relative amounts of both CO

and CO_2 in the plume are not changing, but that their concentrations are both changing by the same dilution factor as the plume disperses over an increasingly larger volume during the time interval of the sample.

The intercept of the regression equation does not provide any useful information. However, as discussed in Section 2.1.5.3, the intercept does play an important role in the calculation of the slopes.

2.1.5.2 Statistical Significance of CO/CO_2 and HC/CO_2 Ratios

Linear regression is a statistical method for analyzing the relationship between two or more quantitative variables which are assumed to be random samples from their distributions. Statistical inferences from random samples are subject to random sampling error. Any statistic estimated from a random statistic is itself a random variable. The probability distribution for a statistic is referred to as a sampling distribution. From the sampling distribution, confidence intervals on the statistic may be inferred.

The parameters of a linear regression are statistical estimates, subject to randomly sampling error. It is possible to obtain values of the slope from regression analysis that are negative, but are not statistically different from zero or a small positive value. In these cases, the measurement would be considered to be valid but an exact

numerical value would not be assigned. The options are to set these values equal to zero, which is the lowest physically-possible value, or to employ a statistical method to randomly assign values between zero and a detection limit.

It is also possible to obtain values of the slope that are negative and that are significantly different than zero. In these cases, the negative slope is likely to be indicative of a problem with that particular measurement. It might represent an anomaly in the emissions or dispersion of the vehicle exhaust plume, failure to adequately sample the vehicle exhaust plume, or an unexplained fluctuation in the operation of the equipment. Whatever the reason, the significantly negative slope can not be accepted as a valid measurement.

Based on the considerations here, it is possible for a negative slope to represent a valid measurement of a very low emission ratio or to represent an invalid measurement. Information regarding the standard error of the slope is not reported in the summary report file, by the RSD on-board computer. Therefore, it is not possible to distinguish which negative slopes are not significantly different from zero versus those that represent an invalid measurement attempt. In order not to introduce into the final data set any invalid measurements, a conservative approach of setting aside all of the negative emissions estimates was used. In future work, a method could be developed to analyze individual vehicle emissions data files and recalculate all of the regression equations to obtain both the slope and the standard error of the slope. Those slopes which are negative but not significantly different from zero could be assigned a value of

zero or a very small positive number, and then be included in subsequent analyses of the data set. However, this effort is beyond the scope of the current study. We note that other practitioners have discarded negative data without considering the statistical issues involved.

The implication of this conservative approach is evaluated in a sensitivity analysis of the data set in Chapter 4. It is known that discarding valid measurements that are not significantly different from zero will lead to an overestimate of the mean values of emissions. The potential magnitude (upper bound) of this bias is quantified in Chapter 4.

2.1.5.3 Evaluation of Linear Regression Procedures

The presence of negative slopes from RSD measurements was recognized early in the project. Alternative methods for performing regression calculations were evaluated in an attempt to develop a statistical method that would produce only positive emission slopes. This work was done in collaboration with National Institute of Statistical Sciences (NISS) and Dr. Chong Gu of Purdue University.

Because the RSD computer uses an approach in which the intercept of the linear regression is allowed to vary, it is possible to obtain negative slopes. In contrast, an alternative method was considered in which the intercept was forced to be zero for each

regression analysis. We refer to this alternate method as the “zero-intercept” method, and to the approach used by RSD on-board computer as the “floating intercept” method.

In order to obtain a positive slope with the zero-intercept method, it is critical that all calculations of the volume percentages obtained from the RSD computer be positive values. Thus, it is important that calibration curves produce only positive numbers for the calculated volume percentages of each of the three gases. Because the calibration curves are developed and implemented by the RSD on-board computer using a fourth order polynomial, there is the potential for significant numerical round-off errors. The discussion of numerical errors due to round-off is discussed in Section 2.1.5.

It is also possible for a fourth order polynomial to inadequately fit some data sets, such that some portion of the domain of the polynomial may result in negative values of concentrations. Therefore, a key premise of the zero-intercept approach is to use a procedure for representing the calibration curves that cannot produce negative volume percent estimates. For this purpose, Dr. Chong Gu employed a procedure based upon the use of splines.

Several data sets obtained from field measurements were analyzed with the zero-intercept approach. It was noticed that when background readings were taken manually on a periodic basis, there is a systematic drift in the slopes calculated from the zero-intercept method, as illustrated in Figures 2.10 and 2.11. The data used in these plots were collected on April 24 at Site 1.

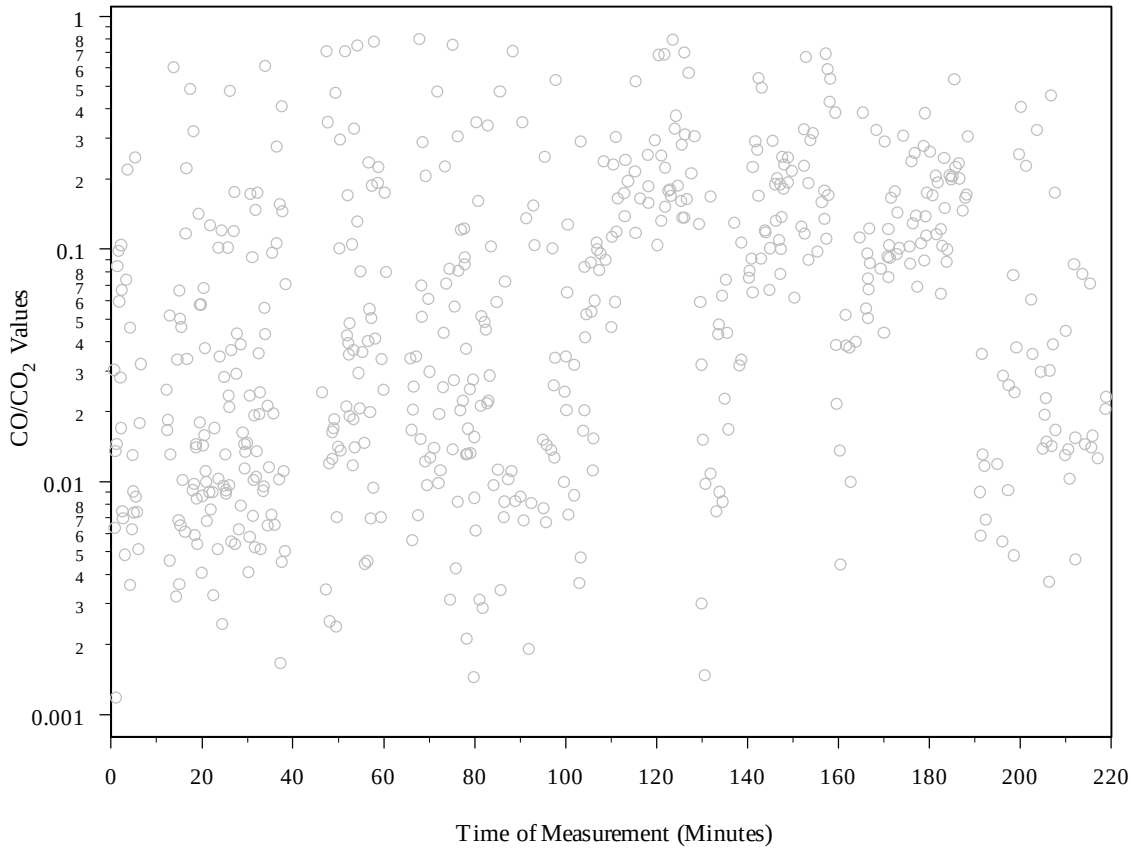


Figure 2.10. Scatter plot of CO/CO_2 values versus data collection time for zero intercept approach.

As seen in Figure 2.10, from the 100th minute to 130th minute there is a substantial increasing trend in the CO/CO_2 data. After the 130th minute, the CO/CO_2 values drop to nearly 0.001 and then follow the same increasing trend through the 160th minute. The pattern repeats again from the 160th to the 190th minute. For this data set, automatic background readings were turned off after about the 100th minute. However, calibrations were done at 30 minute intervals. Similar results were obtained for HC/CO_2 as shown in Figure 2.11.

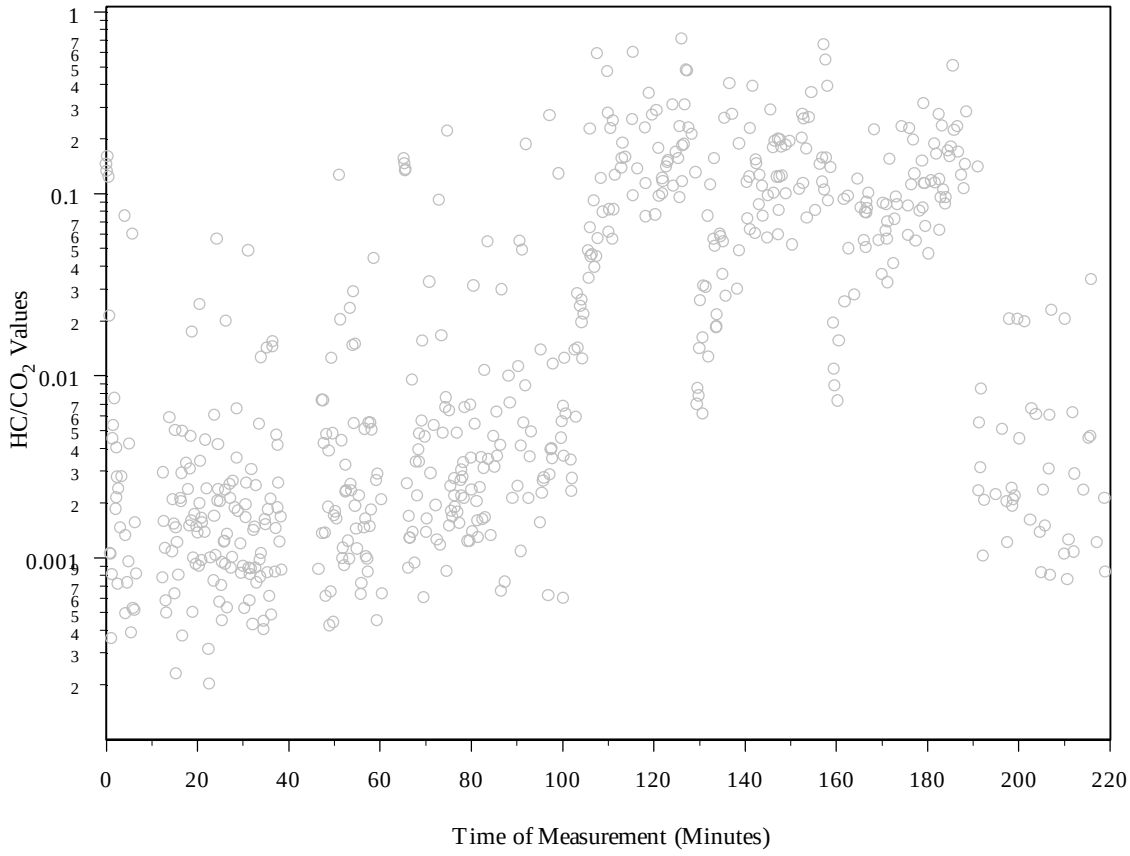


Figure 2.11. Scatter plot of HC/CO_2 values versus data collection time for zero intercept approach.

When background readings are taken automatically for each vehicle, there is no systematic drift. One implication of these observations is that in order to use the zero-intercept approach, automatic background readings must be taken for each measurement.

For comparison, the emissions ratios were calculated using the ”floating intercept” method for the same data set, and the results are shown in Figures 2.12 and 2.13. In contrast to the case for the “zero-intercept” method, there is an absence of the 30 minute cycles of systematic drift.

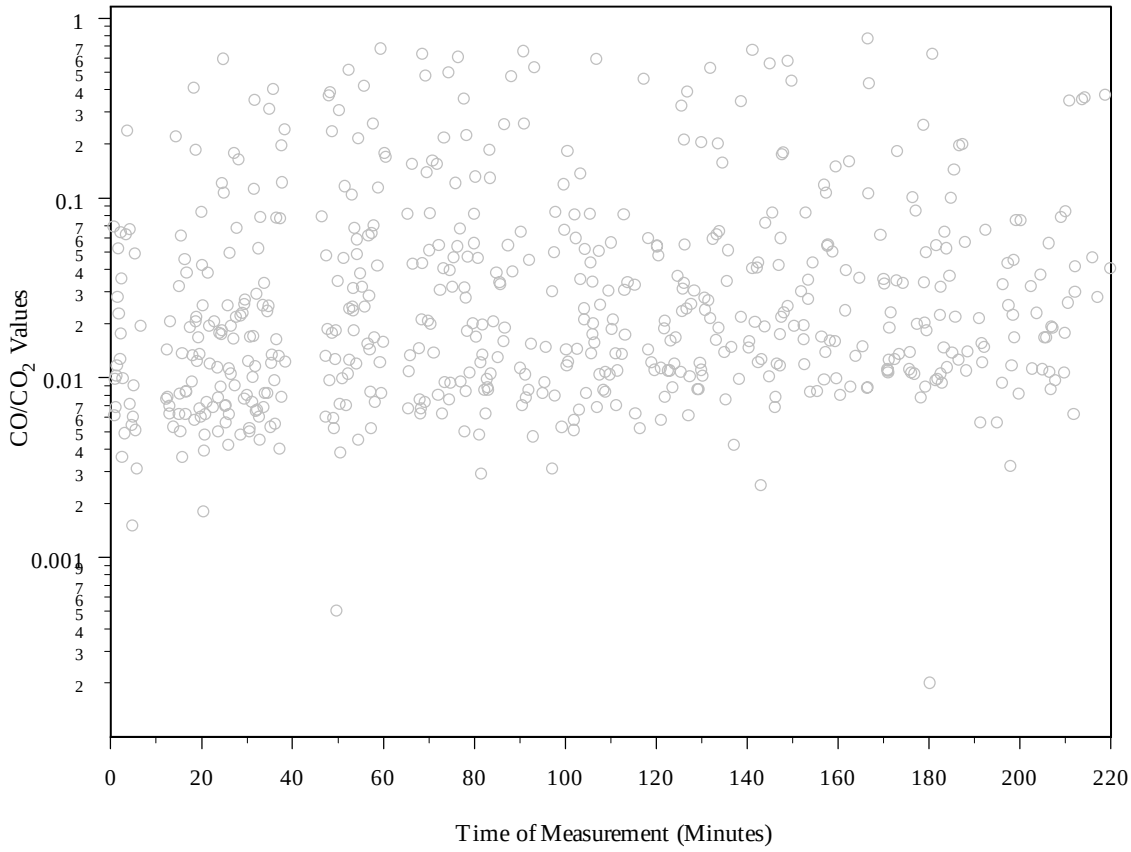


Figure 2.12. Scatter plot of CO/CO_2 values versus data collection time for floating intercept approach.

The main consideration in calculating the emissions ratio is to estimate the change in the concentration of either CO or HC relative to the change in concentration of CO_2 as the plume disperses. It is not essential that the concentration for any of these three pollutants be accurate on an absolute scale. Hence, it is not as critical to the floating-intercept method that the origin be exactly specified or even known. In contrast, the origin used in the "zero-intercept" method must be exactly and accurately zeroed with respect to background levels. Thus, all measurements of the three pollutants must be volume concentrations above background. If background levels change during a sampling

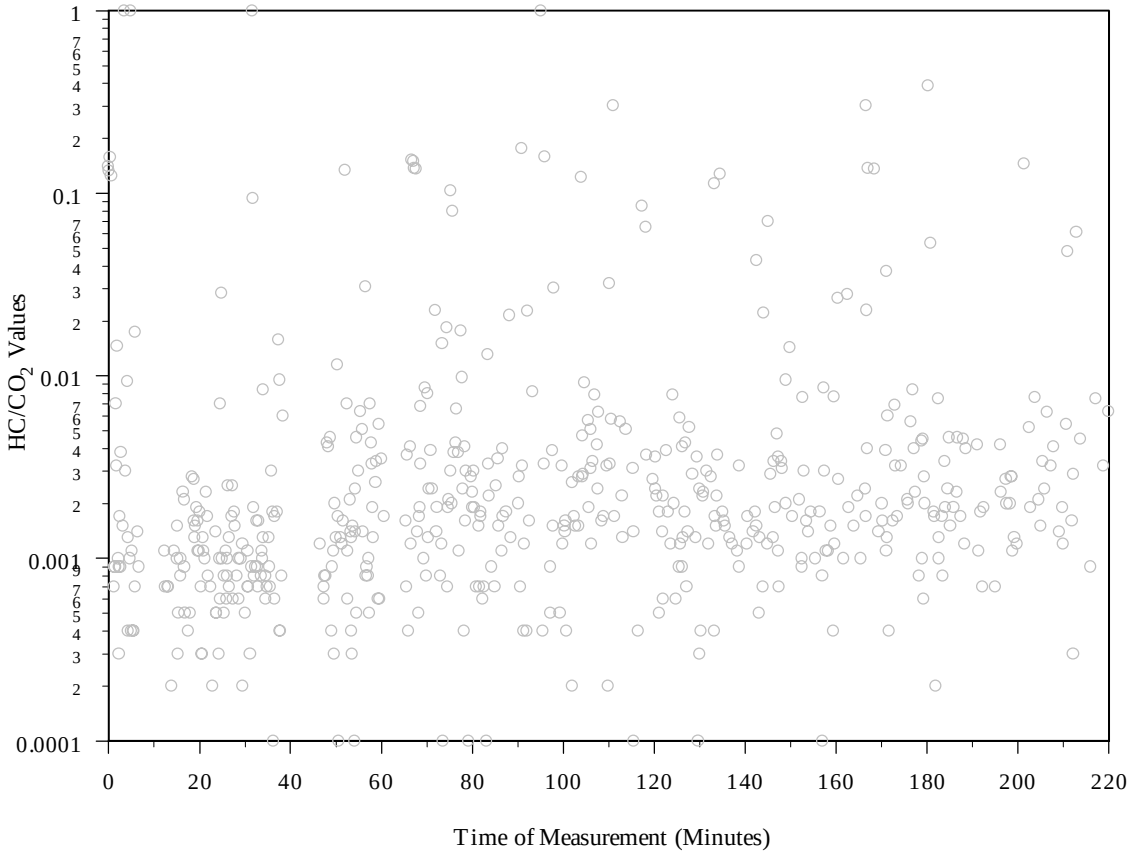


Figure 2.13. Scatter plot of HC/CO_2 values versus data collection time for floating intercept approach.

period in which the background is not updated, then the fixed origin used in the zero-intercept approach becomes inaccurate, leading to systematic errors in the estimates of the slopes.

The instrument has limited sensitivity, from which a detection limit can be inferred, as described in Section 2.1.6.3 . Thus, it is not possible to exactly specify the origin of the CO versus CO_2 or HC versus CO_2 concentration plots.

For these reasons, it was decided to use the floating-intercept method employed by the RSD on-board computer as the basis for the estimates of emission ratios. An advantage of the floating-intercept method is that it is generally more robust to changes in background levels. This enables data collection with manual background readings taken periodically. Hence, another key advantage is that the reset time between measurements of each vehicle is smaller than for when background readings are taken automatically for each vehicle, enabling collection of more measurements in a given period of field data collection, can capture of more measurements of vehicles that are in platoons.

2.1.5.4 Recording of Data by the RSD On-Board Computer

The RSD on-board computer records data for each measurement attempt in an ASCII file. The on-board computer performs a calculation to screen out invalid emission slope estimates. For these cases, an error code of "9.99"or similar is reported for the slope, instead of a value obtained from the regression equations. There is little documentation of the basis by which the on-board computer invalidates a measurement. One reported criteria pertains to the use of automatic background readings. Emissions ratios are reported as invalid if insufficient data are collected for individual vehicle backgrounds while sampling in automatic background mode. However, other reasons for invalidating an emissions measurement are not reported.

Bishop *et al.* (1989), who have developed a different RSD than the one used in this study, state that they consider a measurement to be invalid if the standard error of the estimated emission ratios is greater than plus or minus 20 percent of the estimate ratio. This criteria is not indicated in the user's manual for the RES-1 RSD, although it is possible that some type of criteria similar to this might be employed in the on-board computer.

After calculating the CO/CO_2 and HC/CO_2 ratios for each vehicle, the RSD on-board computer saves them in a file where they are combined with speed, acceleration, and license plate information. This summary data file includes information on: vehicle number; date; time; number of detector channels used (4 channels); license plate of the vehicle; the estimated volume percent of CO and CO_2 value in the tailpipe exhaust; the estimated volume ppm of HC in the tailpipe exhaust; the CO/CO_2 and HC/CO_2 ratios; two speed measurements; and acceleration value for each of the measurement. A portion of a vehicle emissions data file is given for data collected on April 6 in Appendix B as an example of this type of information.

2.1.6 Precision, Accuracy, and Variability of RSD

Like many other measurement methods, remote sensing has some limitations. In this section, factors pertaining to the precision, sensitivity, and accuracy of the RSD measurements and to the inherent variability in vehicle emission measurements are summarized.

2.1.6.1 Precision and Accuracy

In order to evaluate the precision and accuracy of remote sensing, several studies have been made. In one study, developed by Asbaugh *et al*. (1992), the precision of CO emissions measurements were within plus or minus 5 percent and the HC emissions were measured within plus or minus 15 percent. Precisions have also been reported by General Motors (GM) as plus or minus 15 percent for the CO/CO_2 and HC/CO_2 ratios (Cadle and Stephens, 1994).

Stephens *et al.* (1996b) compared emissions of carbon monoxide and hydrocarbons measured using a gas chromatograph, a flame ionization detector (FID), a Fourier transform infrared spectrometer (FTIR), a non-dispersive infrared analyzer (NDIR) and two remote sensors. The vehicle exhaust samples were obtained from dynamometer operation of two gasoline-fueled vehicles using three different fuels. The FID was taken as the benchmark instrument. The ratio of CO/CO_2 and HC/CO_2 emissions obtained by each instrument divided by the corresponding CO/CO_2 and HC/CO_2 measurements using FID was defined as the response factor.

It was found that remote sensing typically performed well in measuring the CO/CO_2 ratio, with a response factor very close to one, indicating that the measurement of CO is accurate. The precision of the RSD was within plus or minus 10 percent (Stephens *et al.*, 1996b). Thus, measurements of CO from RSD are considered to be accurate and acceptably precise.

In contrast, the overall response factors for hydrocarbons were found to be between 0.23 to 0.68. Remote sensors operate by measuring infrared absorption within a narrow range of wavelengths, so they are better able to detect some hydrocarbon species than others. RSDs are typically calibrated to measure propane. Stephens *et al.* (1996b) reported that the RSD response factor for straight chain alkanes varied from 0.94 to 1.11. This indicates that the RSD can accurately measure straight chain HCs. However, the response factor decreased for measurements of chain-branched alkanes. For example, the response factor for iso-octane varied from 0.70 to 0.85. The response factor for olefinic compounds was low for the shortest chain molecules (e.g., 0.05 to 0.1 for ethylene) and increased for the longer chain olefins to values near 0.5 for i-butylene. The response factor for aromatic compounds was also relatively low, ranging from 0.07 to 0.21 for toluene and o-xylene. The overall response factor for remote sensing depends on the distribution of these different types of compounds in the exhaust and upon the individual response factor for each type. Thus, RSD measurements systematically underrepresent emissions of HCs other than straight chain alkane compounds.

In a recent study, Singer *et al.* (1998) developed response factors for remote sensing measurements of HC based upon comparison of measurements between FID and an infrared spectrometer. Their study included measurements of laboratory samples and of actual vehicle exhaust. A set of measurements were taken from prepared laboratory samples for 31 different organic compounds which are found in automobile exhaust. In these measurements, the optics of on-road remote sensing were simulated using a Nicolet Magna 760 FTIR spectrometer with a 3.4 μm filter. Measurements were

also taken from nine high-emitting vehicles which were tested on four different driving cycles, including the urban dynamometer driving schedule, CARB's LA92 (Unified) cycle, the stop-and-go New York City cycle, and a high-speed highway driving cycle developed by CARB (SCF58). For these measurements, a dynamometer FID and a Horiba PIR2000 NDIR analyzer were used. Both measurements were taken with two interference filters, at 3.4 μm and 3.45 μm, to observe the difference in measurements that can be obtained from different RSD technologies.

They found that as alkanes become more highly branched, their response factors decreased. Response factors for aromatic compounds, such as benzene, are very low, compared to the response factors for straight-chain hydrocarbons. An increase in the number of alkyl group on an aromatic ring increases the infrared response, which is similar to the findings of Stephens *et al.* (1996b). Response factors for hydrocarbons range from 0.01 for benzene to 0.98 for C_3-C_5 n-alkanes.

Singer *et al.* (1998) also estimated fleet-average infrared response factors. For this purpose they measured 20 in-use vehicles which had high HC remote sensor readings. Total and speciated exhaust HC emissions were measured for each vehicle by dynamometer testing. A composite HC speciation profile for all vehicles was calculated as the emissions-weighted average of the measured speciation profiles of the individual vehicles. Using this composite speciation profile and individual response factors for each HC, a fleet-average response factor of 0.52 was calculated. This implies that the HC emissions value obtained from remote sensing should be multiplied by a factor of

1.92 to produce an average estimate of total HC emissions. The response factors varied from 0.40 to 0.67 for each of the 20 vehicles, which corresponds to scaling factors from approximately from 1.49 to 2.5, over a 95 percent probability range. In addition, typical fleet-average response factors were calculated using HC speciation profiles measured in roadway tunnel studies. An average scaling factor of 2.2 was estimated for light duty vehicles using conventional gasoline and a value of 2.0 was estimated for vehicles using reformulated gasoline for an RSD using filter of 3.4 μm. For a filter of 3.45 μm, the estimated average scaling factors were found to be 1.9 for conventional gasoline and 1.7 for reformulated gasoline. The standard errors of the scaling factors were estimated to be 0.1.

Stephens *et al.* (1996b) suggest that improvements in RSD technology, to allow for hydrocarbon measurements at multiple infrared wavelengths, may be a necessary but perhaps not perfect solution in order to overcome the measurement biases of the current technology. Based on the limited data collected in their study, one may infer that HC emissions factors developed from remote sensing should be multiplied by correction factors of 1.5 to 4.4 in order to represent total hydrocarbon emissions. In another study, a correction factor of 3 is suggested (Stephens *et al.,* 1996a). Based upon the work of Singer *et al.* (1998) and Stephens *et al.* (1996a), we infer that the HC emissions measurements obtained in this project should be multiplied by approximately a factor of 2 to obtain a more accurate indication of total HC emissions.

2.1.6.2 Variability in Measurements

Bishop *et al.* (1996) report similarities in the observed test-to-test variability in emissions measurements using several types of tests, including FTP and IM240 driving cycles, remote sensing and idle tests. Their conclusions are that individual vehicles may exhibit substantial variability in repeated measurements for individual vehicles on all types of tests. Thus, the authors conclude that the observed variability is not primarily due to limitations of the test methods themselves. For light duty gasoline vehicles, factors such as inconsistency in control of the air-to-fuel ratio, due perhaps to malfunctioning oxygen sensors, may be a dominant source of variability. Mercer and McArragher (1997) have stated that a single remote sensing measurement can not distinguish between normal and gross emitters since there are many variations and uncertainties.

Frey and Eichenberger (1997) found that repeated measurements on individual vehicles displayed a similar range of variation as did measurements obtained for a fleet of vehicles. These findings were replicated in specific studies of school buses and transit buses using remote sensing. These findings imply that it is not possible for example, to identify a high emitting vehicle based upon a single RSD measurement.

2.1.6.3 Sensitivity and Detection Limits

The RSD used in this study, RES-1, has been shown by the manufacturer to provide sensitive and highly repeatable results (SBRC, 1994). According to a study conducted by Jack *et al.* (1995), the minimum detectable concentration (assumed in a 10-cm path length) for hydrocarbons is approximately 3 ppm. Accounting for dilution, a minimum detectable tailpipe concentration is about a factor of 5 to 10 higher. Thus, 30 ppm is suggested as the minimum detectable tailpipe concentration. Similar analysis for CO indicates a minimum detectable tailpipe concentration of 200 ppm. Repeated measurements of standard gas mixtures and measurements of vehicles in the field were conducted to obtain and confirmed these minimum detectable concentration values (Jack *et al.*, 1995).

The information regarding the sensitivity of the RSD suggests that there is a detection limit below which the measurements from the RSD may not be statistically different from a value of zero. In order to estimate possible detection limits in terms of grams/gallon emission estimates, we used a simple combustion model which is given in equation 2.5 below:

$$CH_y + m[0.21O_2 + 0.79N_2] \mapsto aCO + bH_2O + cC_3H_6 + dCO_2 + eN_2 \qquad (2\text{-}20)$$

The equivalent formula for gasoline is assumed to be $CH_{1.95}$ based on Frey and Eichenberger (1997). Mass balance equations for each elemental species can be written as follows:

Table 2-5. Calculation of stoichiometric coefficients

Element	Reactants	=	Products
Carbon (C)	1	=	a + 3c + d
Hydrogen (H)	Y	=	2b + 6c
Oxygen (O)	0.42 m	=	a + b + 2d
Nitrogen (N)	1.58 m	=	2e

In our case y is 1.95. In addition to the four equations in Table 2.5, we can introduce two more equations based upon the concentrations of CO and HC at the detection limits of 200 ppm and 30 ppm respectively.

$$CO = 200\times10^{-6}\frac{\text{gmoleCO}}{\text{gmole exhaust gases}} = \frac{a}{a+b+c+d+e} \tag{2-21}$$

$$HC = 30\times10^{-6}\frac{\text{gmoleHC}}{\text{gmole exhaust gases}} = \frac{c}{a+b+c+d+e} \tag{2-22}$$

By using Equations 2.6 and 2.7 and the equations given in Table 2.5, we can solve for the mass balance coefficients.

The calculated amount of exhaust gases released is 7.57 gmole per gmole of fuel burned. The amount of CO emitted per gmole of fuel burned is:

$$CO = \left(200\times10^{-6}\frac{\text{gmole}CO}{\text{gmole exhaust gases}}\right)\times\left(\frac{7.57\text{gmoles exhaust gases}}{\text{gmole of fuel burned}}\right) \tag{2-23}$$

$$CO = \frac{1.51\times10^{-3}\text{gmoleCO}}{\text{gmole fuel burned}} \tag{2-24}$$

For HC this value is calculated as:

$$HC = \left(30\times10^{-6}\frac{\text{gmole}HC}{\text{gmole exhaust gases}}\right)\times\left(\frac{7.57\text{gmoles exhaust gases}}{\text{gmole of fuel burned}}\right) \tag{2-25}$$

$$HC = \frac{2.27 \times 10^{-4}\,\text{gmoleHC}}{\text{gmole fuel burned}} \qquad (2\text{-}26)$$

This number can then be used to calculate the grams/gallon emissions estimate. The details of this type of calculation are given in Section 4.1.1.

The results for the case of the estimated detection limits are:

$$\text{EF}_{\text{HC}} = \left(201\frac{\text{gmole}}{\text{gallon}}\right)\left(0.0015\frac{\text{gmole CO}}{\text{gmole}}\right)\left(28\frac{\text{g CO}}{\text{gmole CO}}\right) = 8.4\frac{\text{g CO}}{\text{gallon}} \qquad (2\text{-}27)$$

$$\text{EF}_{\text{HC}} = \left(201\frac{\text{gmole}}{\text{gallon}}\right)\left(0.00023\frac{\text{gmole HC}}{\text{gmole}}\right)\left(42\frac{\text{g HC}}{\text{gmole HC}}\right) = 1.9\frac{\text{g HC}}{\text{gallon}} \qquad (2\text{-}28)$$

The values estimated here are much lower than the average emissions estimates of the data collected in this study, as summarized in Chapter 4. Therefore, we do not expect that the detection limit is a significant factor in the interpretation of the results of this study. However, it should be noted that in future work the detection limit could be inferred for the RSD by taken measurements of a gas containing CO_2 but not CO and HC. Repeated measurements of such a gas could indicate the precision with which a zero emission ratio is inferred by the data collection and analysis procedures of the RSD.

2.2 Traffic Data Collection

Video surveillance was selected for use as a representative real-time traffic data collection technology during the course of the study. The details of the traffic data collection effort are given by Dalton (1999). Key aspects of the traffic data collection

pertinent to the development of an integrated vehicle activity and emissions data base are described here.

Major parts of the video surveillance system include: surveillance camera; camera mounting unit; portable lighting tower; camera and pan/tilt remote controllers; color television; video-cassette recorder; line conditioner; and gasoline-fueled electric power generator. Figure 2.14 shows the schematic of the portable surveillance system. The camera was mounted on a 14-foot stable tripod and is intended for use atop an overpass or nearby hill to ensure the best field of view for a portable system. Since no information would be processed in real-time during the study, the camera was connected to a video cassette recorder. Figure 2.15 illustrates the video surveillance equipment: color TV; video-cassette recorder; and line conditioner with power outlets. In preparation for analysis, distance calibration cones were spaced in regular intervals of 20 feet apart along both sides of the study zone.

Processing of the video was performed using the MOBILIZER© video tracking system. MOBILIZER© is composed of a computer connected to a VCR. MOBILIZER© processes vehicle images based on a user-defined lane geometry. The geometry includes: lane location and centerline; distance calibration lines perpendicular to the lane; lines for registering vehicle flow; speed and headway; vehicle length lines along the zone length; and an area for tracking changes in light conditions. MOBILIZER© uses the user-defined geometry file to interpret the video image while the videotape (or real-time video feed) is running. The information files obtained in the study include the vehicle type, a speed estimate and two registration time stamps for each vehicle tracked. Based on the

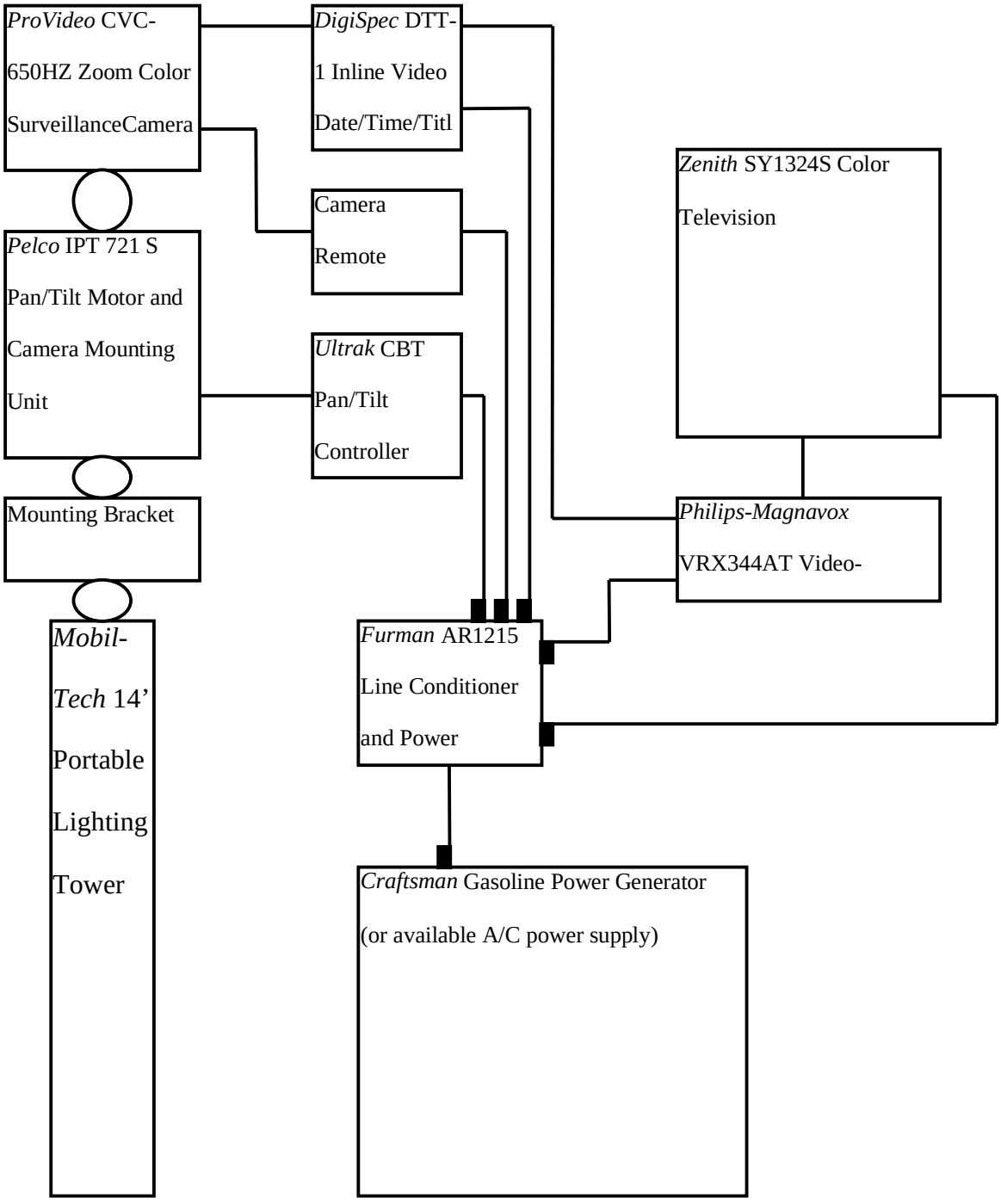


Figure 2.14. Schematic diagram of the portable surveillance system.

Figure 2.15. Video Surveillance System.

information inferred from the video image, time spacing between consecutive vehicles, distance spacing between consecutive vehicles, and platoon order can be calculated (CMS, 1997)

.

2.3 System Setup and Operation

Proper site selection is the most important consideration in remote sensing. Because optical path lengths of greater than 13-m decrease the instrumental sensitivity due to increased background CO_2 (Cadle and Stephens, 1994), the site should have a

single lane of traffic. There must be sufficient shoulder room for the source, detector, video camera, and data acquisition equipment. Current remote sensors use a van or a trailer for housing the data acquisition instrumentation, supplies, and the operator.

Remote sensing is a fair weather technology. It is most effective on dry pavement. Rain, snow, and wet pavement cause scattering of the infrared beam. These interferences cause the frequency of invalid readings to increase.

The importance of safe work practices must be emphasized. Setup of this equipment requires the operator to cross traffic lanes. The operator must ensure compliance with local safety regulations. With that in mind, the first step in setting up the equipment is to deploy safety cones around the front and back of the van. Also traffic cones should be positioned around equipment deployed at the site. The receiver is placed at the road site within fifty feet of the van unless longer equipment cables are obtained.

The following is an outline of the equipment setup:

Work Zone Setup

(1) Remove and deploy " Survey Crew" orange signs several hundred feet before traffic reaches the van

(2) Deploy traffic cones in work zone to direct traffic away from equipment and personnel

Receiver Setup

(1) Remove receiver and receiver tripod from van.

(2) Position receiver on the roadside (in front or back of the van).

(3) Locate receiver on the four mounting studs provided on the receiver tripod and secure with wingnuts (included on tripod).

(4) Adjust the receiver until it is at the desired height above the roadbed and tighten the tripod legs to insure a level, stabilized base.

(5) Position the receiver so that it points in the general direction of the source

(6) Connect the alignment module cable to the telephone jack on the rear of the receiver.

Video System Setup

(1) Remove video camera and tripod from van.

(2) Position camera within 30 feet of the receiver and adjust it so that the field of view includes the target area between the source and receiver.

Source Setup

(1) Remove the source from the van and position within the four cones located earlier (Note: the source and source generator are typically located on the opposite side of the roadway from the van).

(2) Adjust the supporting tripod so that the source is at the desired height above the road surface, and position it to face the receiver.

Source Generator Setup

(1) Remove the source generator from the van and position it between the cones positioned earlier. When positioning the generator care must be taken so that the generators exhaust is directed down the traffic lane, away from the test side

Speed-Acceleration System Setup

(1) Remove pneumatic tubes and control box from the van.

(2) Place pneumatic tubes across the roadway in parallel with each other, separated by six feet.

(3) Connect the pneumatic tubes to the control box, which is connected to the computer system in the van.

Video Surveillance System Setup

(1) Position the tripod at the optimal remote viewing location.

(2) Position the camera onto the tripod.

(3) Adjust the camera so that it covers the area of interest.

(4) Make necessary connection between camera, VCR, and TV.

3.0 SELECTION OF SITES

Suitable site selection was an important element in this project. Several factors had to be considered before a site was accepted for data collection. A roadway with a steady flow of vehicles (medium volume) had to be selected, and it had to possess suitable characteristics for RSD and MOBILIZER© equipments.

3.1 Site Selection Criteria

According to Jack *et al.* (1995) the key criteria in site selection are:

(1) High average daily traffic volume: 5000 to 10,000

(2) Minimal perturbation of traffic flow

(3) Slight road grade

(4) Avoid locations causing hard accelerations

(5) Avoid sites causing no load decelerations

High average daily traffic volume increases the number of data collected. Minimal perturbation of traffic flow avoids biasing motorists' behavior. As reported in Jack *et al.* (1995), bursts of CO and HC can be emitted by vehicles that undergo variations in throttle position. A slight grade provides a consistent load on the vehicles. Hard acceleration and no load decelerations are reported to increase emissions substantially (Jack *et al.*, 1995).

The operating characteristics of the RSD equipment requires that its use be limited roadways with a maximum width of 13 meters. It is not possible for the RSD to measure emissions from two vehicles passing side-by-side in front of the equipment. For these reasons the RSD is typically used at roadways with one lane of traffic. Candidate sites for remote sensing must have adequate shoulder area for the location of the RSD equipment.

The traffic surveillance equipment required a hill or an overpass allowing a reasonable downward view field where individual vehicles were clearly visible over at least a 200 foot segment of roadway.

The constraints regarding both the RSD and video-traffic detector equipment significantly constrained the selection of sites. A number of possible sites were identified in discussions by the project team, including locations in Wake, Durham, and Orange counties. Project team members visited approximately one dozen candidate sites and selected two for data collection.

3.2 Selected Sites

A total of two sites were used during the course of the study. The first site is located at the junction of State Highway NC 86 and Interstate I-40 in the Chapel Hill/Durham area. The study interval was a straight section of the on-ramp from NC 86 onto I-40 eastbound. The ramp pavement is of concrete, and is in good condition with a

downgrade in the middle of the study interval of -5.9 percent. The entire ramp length is over 500 feet including the merge onto I-40. The study interval included only 200 feet approximately midway down the ramp. The entrance to the ramp is either a sharp right or sharp left turn at a signalized intersection. During the presence of the emissions study zone, a work zone sign was located at the entrance to the ramp. Once on the ramp, drivers have approximately 100 feet before entering the 200-foot marked study zone. The site layout with locations for the traffic camera, emission detection equipment and marked study zone are shown in Figure 3.1.

The second site is located at the junction of US 64 and US 1. The study interval was a slightly curved section of the on-ramp from US 64 eastbound onto US 1 northbound. The second site is different from the first in several ways. First, the US 64 site is an upgrade on-ramp, approximately 2 percent. Second, the second site is composed of a new asphalt surface. Finally, drivers are able to enter the measurement section at highway speeds despite the presence of an upstream signal. The entire ramp length is over 1000 feet, and the study zone used approximately 200 feet along the portion preceding another left-merge on-ramp. Approximately 100 feet remained after the study zone before the influence of the left-merge on-ramp. During the presence of the emission study zone, a work zone sign was located 200 feet upstream. This site map is given in Figure 3.2.

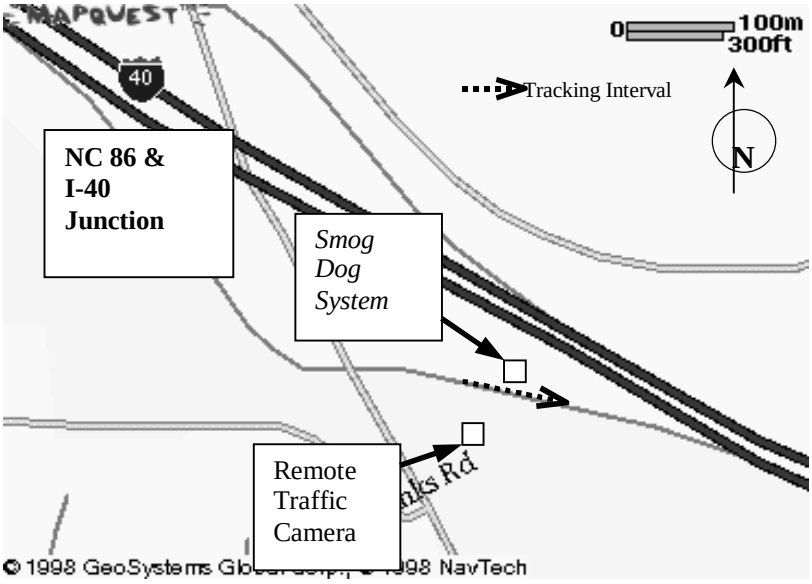


Figure 3.1. Map of Site 1.

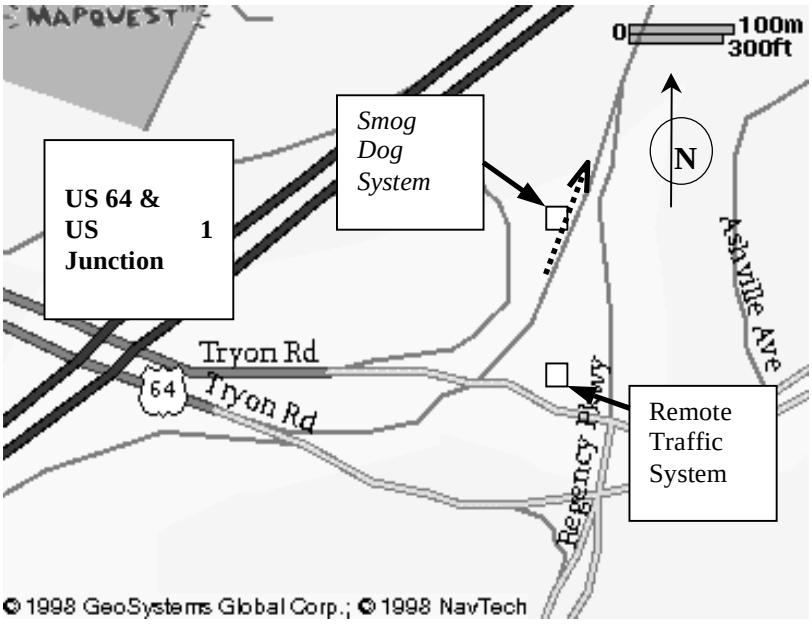


Figure 3.2. Map of Site 2.

4.0 DATA REDUCTION AND EXPLORATORY ANALYSIS

In this chapter, data post-processing and exploratory analysis activities are summarized. First, data post-processing procedures for database development are presented. This section includes information on the steps taken to identify the valid emissions data and the procedures employed to combine emissions and traffic data. Then, a summary of the measurements is presented. The rest of the chapter is devoted to an exploratory analysis of the data, which features the use of summary statistics, comparison of empirical distribution functions, and statistical multi-comparisons of the means of subsets of data. The purpose of the exploratory analysis is to identify possible trends or relationships within the overall data set.

4.1 Data Post-Processing and Database Development

Emissions measurements were obtained over 10 days of data collection at two sites. These sites are the junction of State Highway NC-86 and Interstate I-40 which is referred to as Site 1, and the junction of US-64 and US-1, which is referred to as Site 2. Table 4.1 summarizes the dates and location of data collection activities with the number of data collected. A total of 11,830 triggered measurement attempts were recorded. Of these 11,830 attempts, 7,056 were found to be valid data after post-processing. Table 4.2 summarizes the number of data points left after each post-processing step for Site 1, Site 2, and the combination of both.

Table 4.1. Summary of Total Emissions Data Collection

Date	Site	Total Number of Triggered Measurement Attempts
April 6	Site 1: NC86 - I-40 Junction	467
April 7	Site 1: NC86 - I-40 Junction	1,169
April 23	Site 1: NC86 - I-40 Junction	798
April 24	Site 1: NC86 - I-40 Junction	683
May 11	Site 2: US64 - US1 Junction	1,196
May 12	Site 2: US64 - US1 Junction	1,730
May 26	Site 2: US64 - US1 Junction	1,532
May 27	Site 2: US64 - US1 Junction	1,689
June 8	Site 2: US64 - US1 Junction	1,190
June 9	Site 2: US64 - US1 Junction	1,376

Table 4.2. Summary of Data After Post-Processing

Description	Site 1		Site 2		Combined	
	Number of Data	Percent	Number of Data	Percent	Number of Data	Percent
Total Triggered Measurement Attempts	3,117	100	8,713	100	11,830	100
Invalid Emissions Data	2,743	88	7,014	81	9,757	82
Negative Emissions Data	2,467	79	5,799	67	8,266	69
Truck and Trailer Data	2,161	69	5,010	58	7,171	61
PIV Data	2,126	68	4,930	57	7,056	60

Details of the post-processing procedure will be given in Section 4.1.2. As shown in Table 4.2 about 12 percent of the collected emissions data were invalid as reported by RSD for Site 1. For Site 2, 19 percent of the total collected emissions data were invalid. For the combined data set, a total of 2,073 data points were discarded

because they were invalid. Negative measurements of emissions were also discarded from the data sets. As described in Section 2.1.5. About 9 percent of the emissions estimates were negative for Site 1 and 14 percent for Site 2. The percent of the negative emissions estimates for the combined data is about 13 percent.

Data collected from trucks and trailers were eliminated from the data set. About 10 percent of the total emissions data were from trucks and trailers at Site 1 and 9 percent at Site 2. Measurements made of the calibration gas, referred to as "Puff-in-Vehicle" (PIV) mode measurements, were discarded from the vehicle data set. PIV measurements were made periodically in order to verify the calibration of the system. The PIV measurements were about 1 percent of the total emissions data for Site 1, Site 2 and the combined data set. After post-processing, about 32 percent of the data were discarded for Site 1 and 43 percent for Site 2. The combined data set has 7,056 data points that are considered valid measurements of emissions of light duty vehicles. This is 60 percent of the number of attempted measurements.

Valid emissions data were combined with speed-acceleration, license plate and MOBILIZER© data sets to form a database to be used in development of an empirical traffic-based emissions model. Table 4.3 summarizes the data available after each data matching step. Details of the matching procedure will be given in Section 4.1.2.

Table 4.3. Summary of Data Available For Regression Analysis

Description	Site 1		Site 2		Combined	
	Number of Data	Percent	Number of Data	Percent	Number of Data	Percent
Valid Emissions Data	2,126	100	4,930	100	7,056	100
After Match to Speed Data	1,730	81	4,652	94	6,382	90
After Match to License Plate	1,552	73	3,165	64	4,717	67
After Match to MOBILIZER© Data	1,816	85	3,841	78	5,657	80

The amount of emissions data available after matching to speed data is rather high. As seen in Table 4.3 above, for Site 1, 81 percent of the data were left after matching with speed data. For Site 2, 94 percent of the data were left. For the combined data set, 90 percent of the data were left. The third line of Table 4.3 gives the results of matching the valid emissions data to license plate data, independent of whether a particular data point was also matched to speed data. The last line of Table 4.3 gives the results of matching the valid emissions data to MOBILIZER© data, independent whether a particular data point was also matched to speed or license plate data. It is clear from the table that the process of matching valid emissions data to license plate data is the most restrictive, with only 67 percent of valid emissions measurements having a simultaneous valid license plate number.

In the next section, the first step of post-processing, which involves calculation of grams per gallon emission rates from the measured molar ratios of CO/CO_2 and HC/CO_2, will be explained.

4.1.1 Calculation of Grams-per-Gallon Emission Rate

The first step in post-processing is the calculation of grams per gallon emission rates for CO and HC from the molar CO/CO_2 and HC/CO_2 ratios that are reported by the RSD. In our study, we used a simplified combustion model developed by Frey and Eichenberger (1997). A modified version of this model was used in Section 2.1.6.3 to calculate emission factors associated with the reported detection limit of the instrument. This combustion model is intended to represent the conversion of fuel and air to the main products of combustion by using chemical mass balance principles. In actual vehicle engines, processes occur which lead to products of incomplete combustion. These processes are due to heterogeneity in the fuel, air, and exhaust product mixture in the cylinders, differences in temperature and pressure throughout the power stroke of the engine, and the performance of air pollution control devices. The typical products of incomplete combustion are carbon monoxide, unburned fuel, and intermediate hydrocarbon species associated with imperfect oxidation of the fuel. In addition, nitrogen oxides are formed from the nitrogen and oxygen in the inlet air (Flagan and Seinfeld, 1988).

Remote sensing data provides information regarding the molar ratios of CO to CO_2 and of HC to CO_2 as observed in the tailpipe exhaust gases. Since CO_2, CO, and HC are the products of the combustion which contain carbon, it is possible to develop a mass balance regarding the amount of carbon contained in the fuel and the amount of carbon contained in these three products of combustion. Since the carbon atoms must be conserved, the observed CO to CO_2 and HC to CO_2 molar ratios from the remote sensing

data enable the calculation of stoichiometric coefficients for the amount of CO_2, CO, and HC formed per mole of fuel combusted.

Frey and Eichenberger (1997) made some assumptions before developing the combustion model. These were: (1) complete consumption of the fuel was assumed; (2) the oxidation of nitrogen from the combustion inlet air was ignored; (3) the carbon released from the fuel is emitted as either CO_2, CO, or unburned hydrocarbon equivalent to propene (C_3H_6); (4) the ratio of hydrogen to carbon in the fuel was specified based upon typical fuel compositions for gasoline; and (5) only enough oxygen was assumed to be consumed to convert the fuel to CO_2, CO, HC, and water vapor. For simplicity, the combustion calculations were based upon stoichiometric amounts of air. The assumptions regarding the air-to-fuel ratio do not affect the ratios of CO to CO_2 and HC to CO_2. Furthermore, gasoline engine vehicles typically operate near stoichiometric conditions.

The starting point for the combustion calculation is to develop an equivalent molecular formula for the fuel. Frey and Eichenberger (1997) assumed that gasoline has a composition of 86 weight percent of carbon and 14 weight percent of hydrogen based upon the composition of gasoline as given by Avallone and Baumeister (1979) and Bosch (1986), neglecting trace species in the fuel. In Table 4.4 the development of an equivalent molecular formula is given for this composition of fuel. For the basis of the calculation, 100 pounds of fuel are assumed. The pounds of each component per 100 pounds of fuel are divided by their respective molecular weights, to obtain the lbmoles

of each component per 100 pounds of fuel. Then the lbmoles of each component per 100 pounds of fuel are divided by the lbmoles of carbon per 100 pounds of fuel. The result is the lbmoles of each component per lbmole of carbon in the fuel. For this calculation the resulting equivalent molecular formula is $CH_{1.95}$.

Table 4.4. Calculation of the Equivalent Molecular Formula of a Typical Gasoline

Component	lb per 100 lb of fuel	Molecular Weight	lb mole per 100 lb of fuel	lb mole per lb mole C
Carbon	86	12	7.17	1.00
Hydrogen	14	1	14.00	1.95

In general, it is assumed that the molecular weight of the fuel is given by the equivalent formula CH_y, where:

$$y = \left(\frac{wt - \%H}{wt - \%C}\right)\left(\frac{MW_C}{MW_H}\right) \tag{4-1}$$

where MW_C is the molecular weight of carbon, and MW_H is the molecular weight of hydrogen.

The fuel is combusted in air, which contains a mixture of approximately 21 volume percent of oxygen and 79 volume percent of nitrogen. The products of combustion are assumed to be CO, H_2O, C_3H_6, CO_2, and N_2. The mass balance for combustion, neglecting excess oxygen, is given by:

$$CH_y + m[0.21O_2 + 0.79N_2] \mapsto aCO + bH_2O + cC_3H_6 + dCO_2 + eN_2 \tag{4-2}$$

where the variables a, b, c, d, e, and m are unknown stoichiometric coefficients defined as follows:

m = moles of “air” consumed per equivalent mole of fuel consumed

a = moles CO formed per equivalent mole of fuel consumed

b = moles of H_2O formed per equivalent mole of fuel consumed

c = moles of C_3H_6 formed per equivalent mole of fuel consumed

d = moles of CO_2 formed per equivalent mole of fuel consumed

e = moles of N_2 in the product per equivalent mole of fuel consumed

Mass balance equations for each elemental species can be written as follows:

Table 4.5. Calculation of stoichiometric coefficients on the basis of consumption of one equivalent mole of fuel

Element	Moles of Elements in Reactants	=	Moles of Elements in Products
Carbon (C)	1	=	a + 3c + d
Hydrogen (H)	y	=	2b + 6c
Oxygen (O)	0.42 m	=	a + b + 2d
Nitrogen (N)	1.58 m	=	2e

There are six unknowns and four equations in Table 4.5. However, from the RSD measurements additional information is known regarding the molar ratios of CO to CO_2 and HC to CO_2, which allows creation of two more equations. These ratios are on a volume basis, which is also a molar basis for an ideal gas. Therefore, additional equations can be introduced as:

$$R_{CO} = \left(\frac{CO}{CO_2} \right) = \frac{a}{d} \tag{4-3}$$

$$R_{HC} = \left(\frac{HC}{CO_2} \right) = \frac{c}{d} \tag{4-4}$$

From these equations, the stoichiometric coefficients for CO and C_3H_6 can be expressed in terms of the stoichiometric coefficient for CO_2 and the ratios of CO to CO_2 and HC to CO_2:

$$a=R_{CO}\,d \tag{4-5}$$

$$c=R_{HC}\,d \tag{4-6}$$

Using Equations (4-5), (4-6), and the first equation in Table 4.5, *d* can be calculated as:

$$d = \frac{1}{R_{CO} + 3R_{HC} + 1} \tag{4-7}$$

Once the value of *d* is known, Equations (4-3) and (4-4) are used to solve for the moles of CO and C_3H_6, respectively, formed per mole of fuel combusted. As an example, assume that we obtained RSD measurements of R_{CO} = 0.19 and R_{HC} = 0.006. The estimated moles of CO_2 produced per mole of fuel would be:

$$d = \frac{1}{0.19 + 3(0.006) + 1} = 0.83 \frac{\text{lbmole } CO_2}{\text{lbmole fuel}} \tag{4-8}$$

Based upon *d*, the molar amounts of CO and C_3H_6 emitted per mole of fuel combusted can be estimated as:

$$a = (0.19)(0.83) = 0.16 \frac{\text{lbmole CO}}{\text{lbmole fuel}} \tag{4-9}$$

$$c = (0.006)(0.083) = 0.0049 \frac{\text{lbmole } C_3H_6}{\text{lbmole fuel}} \tag{4-10}$$

The next step is to calculate an emission factor on a grams per gallon basis. For this purpose, it is necessary to estimate the density of the fuel. For gasoline, Frey and

Eichenberger (1997) assumed that density is 0.742 g/cm^3 based upon Bosch (1986) and Avallone and Baumeister (1979). The equivalent molecular formula is used to estimate an equivalent molecular weight.

$$MW_{fuel} = 12.011\frac{\text{g C}}{\text{gmole C}}\left(1\frac{\text{gmole C}}{\text{gmole fuel}}\right)+1.0079\frac{\text{g H}}{\text{gmole H}}\left(\text{y}\frac{\text{gmole H}}{\text{gmole fuel}}\right) \quad (4\text{-}11)$$

For y = 1.95, the equivalent molecular weight is 13.976 grams of fuel per gmole of fuel. The estimated equivalent molar amount of fuel per gallon of fuel for gasoline is given by:

$$\rho' = \frac{\rho}{MWfuel}\left(\frac{3785\,\text{cm}^3}{gallon}\right) \quad (4\text{-}12)$$

where ρ‘ is the density in gmol/gallon, ρ is the density of the fuel in g/cm^3 and 3,785 cm^3/gallon is a standard conversion factor from gallons to cm^3. For gasoline, the estimated density is:

$$\rho'=\left(\frac{0.742g}{cm^3}\right)\left(\frac{gmol}{13.976g}\right)\left(\frac{3785\,\text{cm}^3}{gallon}\right)= 201\frac{gmol}{gallon} \quad (4\text{-}13)$$

The emission factors in grams of pollutant emitted per gallon of fuel consumed are given by:

$$EF'_{CO}= \rho' a MW_{CO} \quad (4\text{-}14)$$

$$EF'_{HC} = \rho' \text{ c } MW_{HC} \quad (4\text{-}15)$$

For our example, the emission factors for CO and HC are:

$$EF'_{CO} = \left(201\frac{\text{gmol}}{\text{gallon}}\right)\left(0.16\frac{\text{gmole CO}}{\text{gmole}}\right)\left(28\frac{\text{g CO}}{\text{gmole CO}}\right)= 900\frac{\text{g CO}}{\text{gallon}} \quad (4\text{-}16)$$

$$EF'_{HC} = \left(201\frac{\text{gmol}}{\text{gallon}}\right)\left(0.0049\frac{\text{gmole HC}}{\text{gmole}}\right)\left(42\frac{\text{g HC}}{\text{gmole HC}}\right) = 41\frac{\text{g HC}}{\text{gallon}} \qquad (4\text{-}17)$$

Emission factors on a grams of pollutant per vehicle-mile travel basis may be estimated based upon assumptions regarding the vehicle fuel economy:

$$EF_{CO} = \frac{EF'_{CO}}{MPG} \qquad (4\text{-}18)$$

$$EF_{HC} = \frac{EF'_{HC}}{MPG} \qquad (4\text{-}19)$$

For example, if the fuel economy of a car is 20 miles per gallon, then the CO emission factor would be estimated as:

$$\text{EF}_{\text{CO}} = \frac{900\frac{\text{g CO}}{\text{gallon}}}{20\frac{\text{miles}}{\text{gallon}}} = 45\frac{\text{g CO}}{\text{mile}} \qquad (4\text{-}20)$$

Similarly, the example emission factor for hydrocarbons is estimated to be 2.1 grams per mile.

Emission factor estimation equations are entered into an Excel spreadsheet. For each emissions measurement, the stoichiometric coefficient *d* is estimated using Equation (4-7) based upon the molar ratios of CO/CO_2 and HC/CO_2. Then, Equations (4-5) and (4-6) are used to calculate *a* and *c*. Equations (4-14) and (4-15) are applied to estimate pollutant concentration on a grams per gallon basis.

4.1.2 Quality Assurance of the Database

The second step in database formation is to screen the data for invalid measurements. The RSD reports 9.99 and 99 for invalid data points, as discussed in Section 2.1.3. In calculating the grams/gallon estimates, both the molar CO/CO_2 and HC/CO_2 ratios should be known. Therefore, we excluded from the database all data for any vehicles for which either the CO/CO_2 and/or HC/CO_2 ratios were reported as invalid by the RSD.

The RSD reports some measurements that have negative values for either the molar CO/CO_2 ratio and/or HC/CO_2 ratio. There are various reasons as to why these might be reported as negative values. As described in Section 2.1.4, the molar ratios are estimated based on a linear regression of data collected and analyzed by the RSD. The linear regression is based upon the estimated percent of CO in the path of the IR beam versus the estimated percent of CO_2 at different times during the sampling interval. Because the regressions are based on a finite and relatively small data set (e.g., n=70) and because there is usually scatter in the data, it is possible to obtain some results in which a slope (molar ratio) is calculated to be negative but is not statistically significantly different than zero or a small positive value. Hence, a negative ratio may in fact reflect a positive or zero emission rate. In some cases, however, the negative ratios are significantly different from zero. These cases may be heavily influenced by a few outliers in the approximately 70 data points used in the regression, or may simply represent an invalid attempt at a measurement due to irregularities in the ability to capture measurements within the actual exhaust plume of the vehicle.

It is not possible to determine, simply by inspection, which of the negative ratios might reflect valid measurements that are not significantly different from a small positive value or zero versus those negative ratios which are the result of an invalid attempt at a measurement. The on-board computer of RSD does not report standard errors or t-ratios for each slope. It would be possible, in principle, to separately analyze each individual vehicle emission file to recalculate each regression for the purpose of estimating the statistical significance of negative values. However, at this time, the effort required for such an analysis is beyond the scope of this work. Therefore, to avoid introducing into the data set any invalid measurement attempts, all negative ratios were excluded from the data set for the time being. It is recommended that the negative ratios be re-evaluated at a later time to determine which ones are not significantly different from zero; these could be retained in the database in the future.

To evaluate the implications of excluding negative ratios from the data base in cases where some may not be significantly different than zero, some sensitivity analyses on the database were performed. In the sensitivity analyses, the 7,056 valid data points, as given at the bottom of Table 4.2, were augmented with 1,491 negative data points that were assigned values of zero or a nominal detection limit of the RSD. The effect of including or excluding the adjusted negative data with respect to the mean emission factor estimated for each site was calculated. Three alternatives were considered. In the first, the mean emissions estimates were calculated based only on the valid non-negative emissions measurements, as shown in the first row of Table 4.6. In the second case, the 1,491 measurements with negative values were assigned a value of zero and combined with the 7,056 valid measurements. The mean for this case shown as the second row of

values in Table 4.6. In the third case, the negative values were assigned a nominal detection limit value, as described in Section 2.1.5, of 8 g/gal for CO and 2 g/gal for HC, and combined with the 7,056 valid measurements. The results for this case are shown in the third row of data in Table 4.6.

The comparison in Table 4.6 suggests that the mean emission rate may be over-estimated by as much as approximately 20 percent for CO and HC if all negative values repeated by the RSD are not statistically different from zero. However, since not all of the negative emission ratios are valid, the actual bias on the mean values associated with discarding all negative emissions values is likely to be less than is indicated here.

Table 4.6. Results of Sensitivity Analysis Regarding Effect on Mean Emissions Estimate of Approaches For Dealing with Emissions Ratios Departed as Negative Values by the RSD[a]

Description	Site 1		Site 2		Combined	
	CO (g/gal)	HC (g/gal)	CO (g/gal)	HC (g/gal)	CO (g/gal)	HC (g/gal)
Discarding Negative Values (n = 7056)	366	48	340	46	348	47
Assigning Negative Values to Zero (n = 8547)	324	43	273	37	287	39
Assigning Negative Values to Detection Limit (n = 8547)	325	43	274	37	289	39

[a]HC emissions are on a propane-equivalent basis

We analyzed the negative emissions measurements to determine whether the fleet characteristics for these measurements are different than the fleet characteristics associated with the positive emissions measurements. Vehicle speed, vehicle acceleration, vehicle model year data were available for the negative emissions data set.

We do not have information regarding the vehicle type and platoon positions since we did not match these data with the MOBILIZER© data set. Table 4.7 summarizes the vehicle activity data for negative emissions measurements for each site.

Table 4.7. Summary of Activity Data for Negative Emissions Measurements

Description	Site 1	Site 2
Speed		
Number of Data Points	146	1180
Average Speed (mph)	35	48
Std. Dev. of Speed (mph)	4.9	5.5
Acceleration		
Number of Data Points	146	1180
Average Acceleration (mph/sec)	4	0.4
Std. Dev. of Acceleration (mph/sec)	2.8	2.2
Vehicle Model Year		
Number of Data Points	126	560
% 1980&Earlier	1.1	0.3
% 1981-1985	3.4	2.5
% 1986-1988	7.9	9.1
% 1989-1990	12.5	3.5
% 1991-1992	14.3	18.5
% 1993-1994	20.4	16.3
% 1995-1996	20.0	27.0
% 1997-1999	20.4	22.9

The average speed for negative emissions measurements is approximately 35 mph for Site 1 and 48 mph for Site 2. The average speed for positive emissions data for Site 1 is approximately 36 mph and 48 mph for Site 2 as reported in Tables 4.8 and 4.9 in Section 4.2. The average speed values for negative emissions data are very close to the average speed values for positive emissions data for both sites.

The average acceleration for negative emissions data for Site 1 is 4 mph/sec and 0.4 mph/sec for Site 2. The average acceleration values for positive emissions measurements are 5.1 mph/sec for Site 1 and 0.2 for Site 2, as reported in Tables 4.8 and 4.9. Although not identical, those average accelerations are of similar magnitude when comparing positive and negative measurements at the same site.

For Site 1 approximately 75 percent of the vehicles having negative emissions measurements are of model years 1991 or newer. For Site 2 approximately 85 percent of the vehicles having negative emissions measurements are 1991 or newer models. For vehicles which positive emission measurements were obtained, approximately 70 percent of the measurements were of model years 1991 or newer, as reported in Tables 4.8 and 4.9 in Section 4.2.

The comparison of fleet activity data for vehicles having negative emissions measurements and for vehicles having positive emissions measurements indicates that there is not much difference in the fleet characteristics for the two cases. Average speed and acceleration values are similar. There is a slight difference in the distribution of model years of the vehicles in the two cases, with a relatively larger share of newer vehicles in the set of negative emissions estimates. This is expected, since emissions from newer vehicles are generally lower than for older vehicles as is discussed later in section 4.4.4. Overall, however, we conclude that discarding the set of negative emissions data does not have a substantial effect on the overall distribution of vehicles in the final data set.

A key step in database develop was to verify the license plate numbers in the database. The license plate numbers recorded by the on-board computer were compared with the video tape of passing vehicles to ascertain any errors by manual inspection. Furthermore, in cases where the on-board computer failed to record a license plate number, the video tape was reviewed and license plates numbers were entered manually when they could be read from the videotape. The resulting data set of license plate numbers was matched at NC DENR to a NC Department of Motor Vehicle (DMV) database to obtain individual vehicle information including vehicle type, model year, and vehicle identification number (VIN).

Additional information on vehicle specifications can be obtained through the VIN number. The VIN can be decoded to obtain information on engine displacement, cylinder configuration, fuel induction type, evaporative emission controls, exhaust emission controls, gross vehicle weight rating, and manufacturer of the vehicle. We sent our VIN data to the Department of Civil and Environmental Engineering at the Georgia Institute of Technology, where the VIN data were decoded with software developed by RADIAN International. We would like to thank Dr. Randall Guensler and William Bachman for their effort in supplying the vehicle information based upon the VIN numbers from our dataset.

Traffic variables were obtained from the video tracking system, MOBILIZER©. The variables include vehicle type, vehicle speed, and two registration time stamps for each vehicle tracked. Using speed and time stamp data, time and distance spacing

between consecutive vehicles (i.e. time headway or distance headway values) and platoon positions were calculated. Headway is referred to as the distance or time difference between a vehicle and the one in front of it. Tailway is referred to as the time or distance difference between a vehicle and the vehicle behind it. A platoon is a group of vehicles that are closely spaced and traveling together. Platoon order includes the first vehicle in the platoon, vehicles within the platoon, and the last vehicle in the platoon. Vehicles with headway and tailway higher than a cut-off point were judged to be free flowing. In the literature, the cut-off point for being in platoon is given as a range between 3 and 5 seconds (TRB, 1994). In our study, a headway of 3 seconds was used as a cut-off point for platoon positions. Detailed information regarding the estimation method is given in Dalton (1999).

A vehicle with a headway of greater than 3 seconds and tailway of less than or equal to 3 seconds was judged to be the first vehicle in a platoon. A vehicle with a headway of less than or equal to 3 seconds and a tailway of less than or equal to 3 seconds was judged to be within a platoon. Finally, vehicles having time headway less than or equal to 3 seconds and tailway greater than 3 seconds were judged to be last in a platoon.

Time stamps used in the RSD and MOBILIZER© differed for some data sets due to differences in the equipments' clocks settings. For some data sets, there was as much as a difference of 5 minutes between the RSD and MOBILIZER© clocks. In order to match emission and traffic data, we first adjusted the time recorded for data points by

watching the video footage from both the MOBILIZER© and RSD cameras to find the difference in time recorded for several vehicles. This time difference was used to adjust either of the data sets so that both would have the same values of time for the same vehicles. For some of the data sets, video footage from the RSD was not available. In that case, vehicles identified from license plate data by vehicle type and model year were searched and identified in MOBILIZER© video footage. After some tedious work, the time records in both of the data sets were adjusted. The time stamps for individual vehicles from the emission and traffic information files were then matched for time-of-day in an Excel spreadsheet macro. The spreadsheet macro, "TimeMatch," is given in Appendix E. The matched data were then combined with the vehicle information data set obtained from VIN data using Microsoft Access. The VIN data were added to the emissions data base based upon matching license plate numbers in the RSD and VIN data sets.

After matching, our dataset had two sets of vehicle type information: one from license plate data and the other from MOBILIZER© analysis. Thus, it was necessary to develop an approach by which to decide, based on one or both data sources, the vehicle type to assign to each vehicle. When available, vehicle type information obtained from DMV records was used, supplemented by the use of MOBILIZER© vehicle type identification for vehicles for which the license plate was not available. Vehicles were classified as one of two types. The vehicle type categorization obtained from MOBILIZER© is based upon Federal Highway Administration (FHWA) definitions. Type 2 represents small vehicles including compact cars, coupes, and station wagons,

whereas Type 3 represents light duty trucks such as pickup trucks, sport utility vehicles, and minivans. MOBILIZER© classifies vehicle type based on vehicle length.

The body type of vehicles obtained from license plates were used to assign a vehicle type to each vehicle. Two-door sedans, four-door sedans, coupe convertibles, compact cars were assigned vehicle Type 2 whereas vans, stationwagons, sports utility vehicles (SUVs), and pickup trucks were assigned vehicle Type 3. It is possible that there is some misclassification of vehicles, since sometimes a minivan or sport utility vehicle is recorded in the DMV database as a stationwagon or may appear to MOBILIZER© to be of the same length as a Type 2 vehicle.

In the next step, all measurements pertaining to large trucks or to vehicles towing trailers were removed from the database. This was necessary because many large trucks have an exhaust point above the vehicle, and hence it would be impossible to obtain an accurate measurement of vehicle exhaust. The presence of a trailer interferes with measurement of exhaust near the tailpipe. In order to remove trucks and trailers from the data base, vehicle type information was used. First, information regarding vehicle body type obtained from VIN data was used to screen all the trucks and trailers. Information obtained from VIN data generally provided more detail regarding the body type and series of the vehicle compared to data obtained from the DMV data base. For example, a vehicle which is identified as truck in the DMV data base can be identified by model, such as a F-100 Light Truck, in the VIN data base. For vehicles for which VIN information was not available, we used MOBILIZER© data to aid in screening out

large trucks. MOBILIZER© identifies trucks and trailers as vehicle Types 4 or 5. Under some conditions, such as sudden sun light changes, MOBILIZER© might detect two vehicles driving closely together as one long vehicle. However, a long vehicle such as a truck can not be mistaken for two separate vehicles. Therefore, MOBILIZER© may misclassify some smaller vehicles as Type 4 or 5, but does not misclassify larger trucks as smaller Type 2 or Type 3 vehicles. Using MOBILIZER© vehicle classifications as a screen for large trucks will therefore result in no errors of comission, in terms of inadvertently retaining unwanted large trucks in the database, but could result in some errors of omission, in terms of inadvertently discarding some vehicles that might actually be of Type 2 or Type 3. The percentage of vehicles identified as large trucks or vehicles towing trailers were 8 percent of the total triggered measurement attempts, or approximately 1,095 data points.

On-site calibration measurements taken and Puff-in Vehicle mode measurements (PIV) were also screened from the database since they are not emissions measurements for vehicles. For PIV measurements, a calibration gas having a CO/CO_2 molar ratio of 1.0 and a HC/CO_2 molar ratio of 0.133 was used. If the calibration results were within plus or minus 10 percent of these values, the previous calibration is accepted for continued use in data collection. If not, another calibration is made. The calibration and PIV measurements of the calibration gas result in measurement of emission rates, but not speed, acceleration, or license plate, since no vehicle is present at the time that these measurements are made. In order to screen out the PIV measurements from the dataset, measurements which do not have license plate information and speed data were

identified. Then the CO/CO_2 and HC/CO_2 ratios of these data points were checked to find the ones which are in the plus or minus 10 percent range of the calibration gas values. These data points were identified as PIV measurements. However, not all PIV measurements were in the plus or minus 10 percent range of calibration gas value. There were some PIV measurements which were out of the plus or minus ten percent range. Since a calibration is made after each of this type of PIV measurements, we identified these PIV measurements by checking the last measurements immediately preceding a calibration. All of the PIV measurements identified by both approaches were discarded from the data base. During this process, it is possible that some actual vehicle measurements, without license plate information, might have been discarded which could cause an error of omission. However, the probability of having such high measurements, for both CO and HC for one vehicle is very low. Specifically, a measurement of 1.0 for the CO/CO_2 ratio and of 0.133 HC/CO_2 ratio is equal to 2,300 g/gal CO and 490 g/gal HC. Each of these values are at approximately the 95^{th} percentile of CO and HC measurements.

The final step in database development is to correct, as needed, the speed and acceleration measurements reported by the RSD. MOBILIZER© also reports a point estimate of vehicle speed. However, the RSD speed is more representative since it is measured at the time and location of the emissions measurement, whereas the MOBILIZER© speed is inferred from vehicle time stamps over a larger distance and time interval.

In the speed data reported by the RSD, there were some errors due to computer data storage. When the speed is recorded by on-board computer, one digit of the measurement might be lost, e.g., 3 instead of 30 mph. These erroneous values were found by manually checking the data and then were removed from the dataset. Acceleration is also reported by the RSD. However, some internal calculation errors and unit conversion problems were encountered. We calculated acceleration by a basic physics equation, given in equation 4-21 below, using two speed measurements and the 6 feet distance between the pneumatic tubes.

$$a = \frac{\left(v^2 - u^2\right)\text{mi}^2}{\text{hour}^2} \times \frac{1}{\left(2 \times d\right)\text{feet}} \times \frac{1\,\text{foot}}{1.89 \times 10^{-4}\,\text{mi}} \times \frac{1\,\text{hour}}{3600\,\text{seconds}} \tag{4-21}$$

where u and v are the speeds measured at the first and second pneumatic tube respectively, and d is the distance between the pneumatic tubes, 6 feet.

For example, for a vehicle having two speed measurements of 30 and 31.5 mph, the acceleration is calculated as follows:

$$a = \frac{\left(32^2 - 31.5^2\right)\text{mi}^2}{\text{hour}^2} \times \frac{1}{2 \times 6\text{feet}} \times \frac{1\text{foot}}{1.89 \times 10^{-4}\,\text{mi}} \times \frac{1\text{hour}}{3600\text{seconds}} \tag{4-22}$$

$$a = \frac{3.89\text{mi}}{\text{hour} \times \text{second}} \tag{4-23}$$

This step concluded our database development. Figure 4.1 summarizes the post-processing procedure.

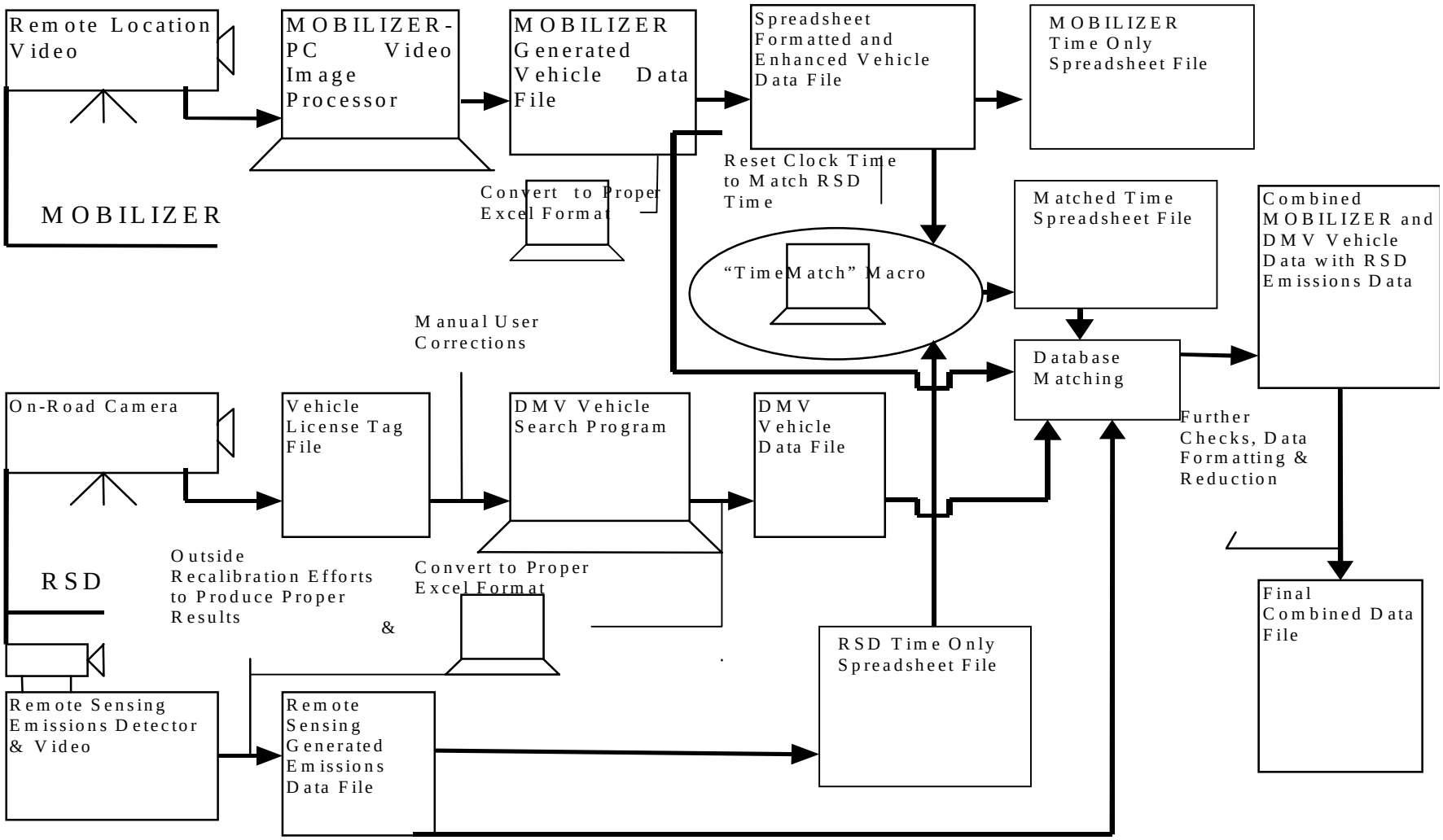


Figure 4.1. Data Post-Processing Chart.

The data fields in the final data set included: date and time of data collection; CO and HC emissions estimates (g/gallon); vehicle speed (mph); vehicle acceleration (mph/s); vehicle type (one of two categories, approximately equal to cars and light duty trucks); vehicle position (e.g., free flow versus position within a platoon); vehicle model year; and VIN information (e.g., body style, engine size, cylinder configuration, type of fuel delivery system, aspiration, type of exhaust control systems, information on air injection reactor, information on evaporative emission control, information on exhaust gas recirculation, engine type, information on closed loop combustion control, positive crankcase ventilation system, thermostatic air cleaner, country of manufacture, vehicle type, gross vehicle weight rating). A sample integrated traffic and emissions data file is given in Appendix D.

4.2 Summary of Measurements

After the post-processing procedure was completed, 7,056 valid emissions data points were obtained for all ten days of data collection. A summary of the emissions and activity data for each day of measurement at Site 1 is given in Table 4.8. Table 4.9 gives the summary of data collection activities at Site 2.

Table 4.8. Summary of Data Collected at Site 1

Description	*Date*				*All Days*
	April 6	April 7	April 23	April 24	
Time of Collection	15^{00} - 18^{00}	6^{55} - 12^{25}	13^{30} - 18^{00}	7^{20} – 11^{00}	
Weather Conditions					
Daytime High Temperature(^{0}F)	68	78	64	73	
Daytime Low Temperature(^{0}F)	31	37	48	44	
Emissions					
Number of Data Points	334	753	543	496	2,126
Mean CO (g/gal)	430	366	360	329	366
95% CI on Mean (g/gal)	358 - 502	325 - 408	313 - 408	280 - 378	341 - 391
Mean HC (g/gal)[a]	56	46	58	34	48
95% CI on Mean (g/gal)	42 - 70	38 - 53	48 - 67	27 - 42	43 - 52
Speed					
Number of Data Points	312	404	527	487	1,730
Average Speed (mph)	36.2	35.7	35.7	36.9	36
Std. Dev. Of Speed (mph)	4.5	4.1	4.2	4.8	4.4
Acceleration					
Number of Data Points	312	404	527	487	1,730
Average Acceleration (mph/sec)	4.6	4.0	4.2	5.3	5.1
Std. Dev. of Acceleration (mph/sec)	2.2	1.9	2.1	3.6	2.9
Vehicle Type					
Number of Data Points	297	719	509	491	2,016
% 2 (cars)	78.5	69.5	69.9	74.7	72.2
% 3 (e.g., minivans, SUVs)	21.5	30.5	30.1	25.3	27.8
Vehicle Position					
Number of Data Points	250	646	440	480	1,816
% Free	63.6	53.9	63.2	47.9	55.9
% First in a Platoon	26.0	25.9	27.7	32.3	28.0
% Within a Platoon	6.8	12.1	5.7	14.2	10.4
% Last in a Platoon	3.6	8.2	3.4	5.6	5.7
Vehicle Model Year					
Number of Data Points	189	586	421	356	1,552
% 1980&Earlier	2.1	2.4	2.1	2.2	2.3
% 1981-1985	7.4	7.8	7.1	5.9	7.2
% 1986-1988	9.5	13.6	14.0	11.5	12.8
% 1989-1990	13.2	11.4	9.0	11.0	10.9
% 1991-1992	15.9	14.0	15.0	19.7	15.8
% 1993-1994	16.4	15.7	19.0	18.3	17.3
% 1995-1996	20.1	20.1	22.6	18.0	20.3
% 1997-1999	15.3	14.8	11.2	13.5	13.6

[a] HC emissions data are on a propane-equivalent basis

Table 4.9. Summary of Data Collected at Site 2

Description	*Date*						*All Days*
	May 11	May 12	May 26	May 27	June 8	June 9	
Time of Collection	15^{00} - 18^{00}	11^{20} - 16^{20}	12^{20} - 17^{45}	9^{20} - 15^{40}	13^{40} - 18^{00}	9^{20} - 15^{00}	
Weather Conditions							
Daytime High Temperature(^{0}F)	67	64	86	85	79	74	
Daytime Low Temperature(^{0}F)	57	53	65	66	47	60	
Emissions							
Number of Data Points	617	815	889	863	849	897	4,930
Mean CO (g/gal)	468	412	381	357	210	252	340
95% CI on Mean (g/gal)	413-523	367-456	344-419	318-397	186-234	223-282	324-356
Mean HC (g/gal)[a]	66	64	61	35	29	30	46
95% CI on Mean (g/gal)	55 - 77	55 - 74	53 - 69	29 - 40	24 - 33	26 - 35	43 - 49
Speed							
Number of Data Points	597	781	877	787	835	775	4,652
Average Speed (mph)	50.4	47.1	47.1	46.6	47.9	46.5	47.5
Std. Dev. of Speed (mph)	5.9	6.7	5.3	5.2	5.4	4.9	5.7
Acceleration							
Number of Data Points	597	780	877	786	834	775	4,649
Average Acceleration (mph/sec)	0.2	0.6	-0.4	0.4	0.2	0.1	0.2
Std. Dev. of Acceleration (mph/sec)	2.5	2.0	2.2	1.5	1.9	1.6	2.0
Vehicle Type							
Number of Data Points	363	717	822	810	799	855	4,366
% 2 (cars)	65.8	83.3	73.5	66.4	75.0	62.2	71.2
% 3 (e.g., minivans)	34.2	16.7	26.5	33.6	25.0	37.8	28.8
Vehicle Position							
Number of Data Points	132	673	742	747	715	832	3,841
% Free	23.5	16.5	17.5	17.8	15.2	45.2	23.2
% First in a Platoon	31.8	36.3	35.3	34.8	30.9	38.5	35.1
% Within a Platoon	33.3	34.5	39.4	38.4	47.1	13.8	34.0
% Last in a Platoon	11.4	12.8	7.8	9.0	6.7	2.5	7.7
Vehicle Model Year							
Number of Data Points	303	374	547	648	600	693	3,165
% 1980&Earlier	1.0	2.9	3.3	2.0	3.3	1.7	2.4
% 1981-1985	7.9	5.9	4.8	4.2	6.2	4.3	5.2
% 1986-1988	12.2	9.9	10.1	11.1	12.7	11.1	11.2
% 1989-1990	9.9	13.6	9.9	9.3	10.7	9.5	10.3
% 1991-1992	11.9	14.4	13.2	11.3	11.7	11.8	12.2
% 1993-1994	20.1	18.7	19.4	20.4	16.7	20.3	19.3
% 1995-1996	20.1	17.9	19.9	21.0	22.2	20.9	20.6
% 1997-1999	16.8	16.6	19.6	20.8	16.7	20.2	18.8

[a] HC emissions data are on a propane-equivalent basis

The data in Tables 4.8 and 4.9 are divided into several categories. Within each category, the number of valid data points is given pertaining to all vehicles for which valid emissions data were obtained. Speed and acceleration values were obtained for about 90 percent of the valid emissions estimates as given in Table 4.8. Some data were lost because of the errors in the speed measuring device. On some days, such as April 7, May 27, and June 9, there were problems in setting up the speed equipment and a relatively smaller proportion of speed and acceleration measurements were obtained compared to the number of valid RSD CO/CO_2 and HC/CO_2 measurements. For example, on April 7, 404 speed measurements were collected, which constitute approximately 54 percent of the valid emissions data collected on that day.

The average speed is approximately 36 mph for data collected at Site 1 and 47.5 mph for Site 2. Average acceleration is approximately 5.1 mph/sec for Site 1 and 0.2 mph/sec for Site 2. The differences in average speed and acceleration between the two sites are attributed to differences in site geometries, traffic patterns, and the location of the study zone. At Site 1, measurements were taken at the middle of a down-grade ramp at a location where vehicles were accelerating to high speed. In contrast, at Site 2 drivers had ample time to reach highway speeds on an upgrade ramp before entering the study zone. Traffic patterns were different at Site 1, which had more free-flowing vehicles, compared to Site 2, where platoons were quite common. The variability in both speed and acceleration is indicated by the standard deviations in the Table 4.8 and 4.9.

The vehicle type categorization given in Tables 4.8 and 4.9 is based upon Federal Highway Administration (FHWA) definitions. Platoon position estimates are

obtained from the vehicle headway produced by the MOBILIZER© system as discussed in Section 4.1.2. The percentage of vehicles in different platoon positions are given in the tables. For Site 1, approximately 45 percent of the vehicles for which valid emissions measurements were obtained were associated with a platoon. For Site 2, vehicles in platoon comprised approximately 75 percent of the headway measurements having also valid emissions estimates. The difference in platooning is mainly due to different traffic patterns at each site as previously discussed.

The percentage of vehicles observed in each model year category is also given in Tables 4.8 and 4.9. The number of data points available classifying vehicle model years are low compared to the other categories. This is because model year is predicated upon correct reading of a license plate, and the DMV database accessible to the study was for North Carolina only. Hence, data were not obtained for out-of-state plates and could not be obtained, even manually, if the license plate camera failed to capture a legible image of the plate. The ratio of valid in-state license plate records to total valid emissions measurements ranges from 46 to 78 percent for different days. The percent of the vehicles in each model year category varies for data collected at different days, but nearly three quarters of the vehicles in each data set by day were of 1991 or more recent model years.

Daily temperatures reported in Tables 4.8 and 4.9 were taken from a local television station (WRAL, 1999). Daytime high temperatures range from 64 ^{0}F to 86 ^{0}F whereas daytime low temperatures range from 31 ^{0}F to 66 ^{0}F. Daytime low temperature

are usually collected at around 3 am in the morning. These temperature readings do not represent the lowest temperatures during our data collection activities but are given here as an extra information.

4.3 Empirical Distributions of Emissions Estimates

In this section, examples of data collected at each site are presented to illustrate the variability in observed data. For this purpose, the April 6 and June 8 data sets were selected. The probability distributions showing the inter-vehicle variabilities and mean estimates for these data sets are given here. Probability distributions for CO and HC emissions estimates of other days are given in Appendix F.

On April 6, data were collected from 3:00 PM to 6:00 PM at Site 1. The average CO emissions were 430 grams/gallon for 334 cars and light duty trucks. The 95 percent confidence interval for this mean value ranges from 360 grams/gallon to 500 grams/gallon, or a range of approximately plus or minus 16 percent. The variability in individual vehicle emissions estimates is illustrated in Figure 4.2. Most of the emissions estimates are within a range of two orders-of-magnitude (e.g., ranging from 14 to 2450 grams/gallon over a 95 percent probability range). The maximum value in the data set is 3,680 g CO/gallon which is well below the maximum possible value of 5,630 g CO/gallon, which would occur if all carbon in the fuel was emitted as CO. The data set of CO emissions estimates is positively skewed on a linear scale, resulting in the mean

value occurring at approximately the 77th percentile of the data values. Approximately 90 percent of the values were below 1500 grams/gallon.

The average HC emissions estimate was 56 grams/gallon, with a 95 percent confidence interval on the mean ranging from 42 grams/gallon to 70 grams/gallon (plus or minus 25 percent). The variability in individual estimates are within a range of two orders-of-magnitude (e.g., approximately 2 to 510 grams/gallon over a 95 percent probability range) as indicated in Figure 4.3. The maximum value in the data set is 817 g HC/gallon, which is well below a possible approximate upper bound of 2,800 g HC/gallon if all carbon in the fuel were to converted to propane-equivalent HC in the exhaust. The distribution is positively skewed on a linear scale and the mean is at approximately the 88th percentile. Approximately 90 percent of the values were below 94 grams/gallon. The HC emissions estimates are on a propane equivalent basis.

On June 8, 849 valid emissions data points were collected from 1:40 PM to 6:00 PM at Site 2. The average CO emissions were 210 grams/gallon. The 95 percent confidence interval for the average CO emissions range from 186 grams/gallon to 334 grams/gallon, or a range of approximately plus or minus 11 percent. The variability in individual vehicle emissions estimates is illustrated in Figure 4.4. Approximately 95 percent of the emission estimates are within a range of 18 to 1140 grams/gallon. The data set of CO emissions estimates is positively skewed, resulting in the mean value occurring at approximately 78th percentile of the data value.

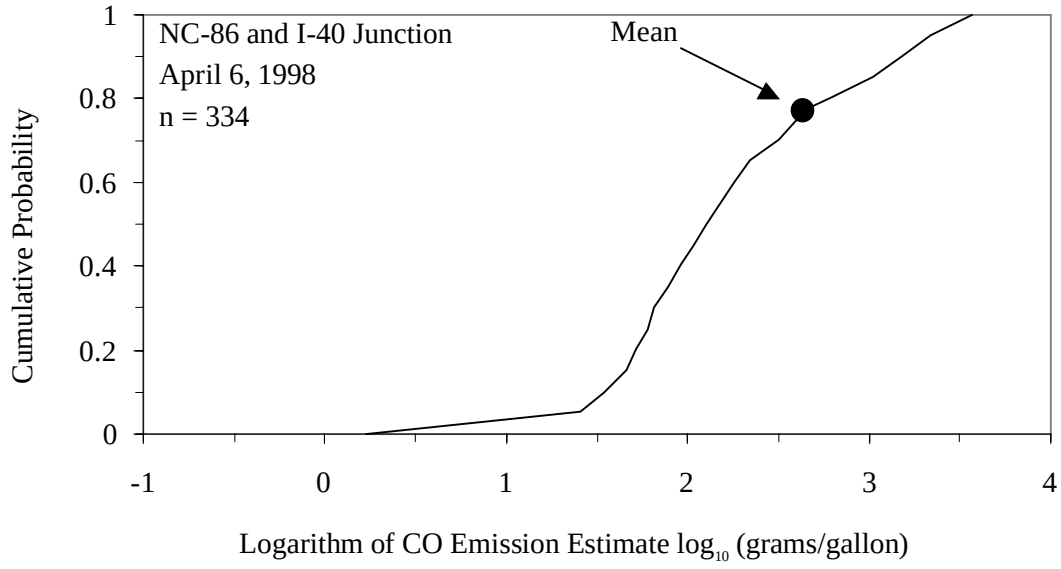


Figure 4.2. CO inter-vehicle variability and mean estimate for data collected on April 6.

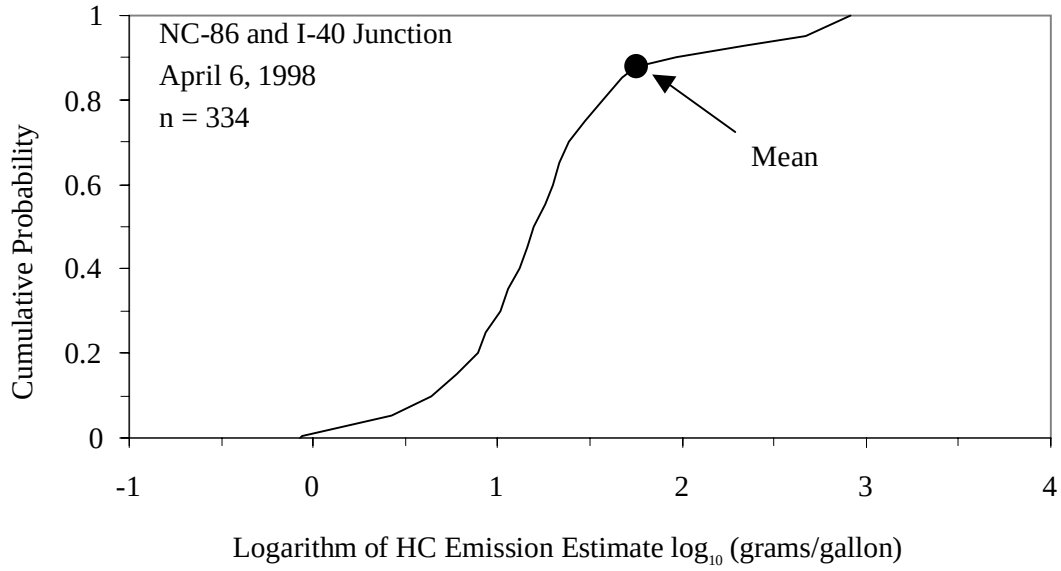


Figure 4.3. Propane-equivalent HC inter-vehicle variability and mean estimate for data collected on April 6.

The average HC emissions estimate for June 8 data was 29 grams/gallon with a 95 percent confidence interval on the mean ranging from 24 grams/gallon to 33 grams/gallon (plus or minus 16 percent). The variability in individual estimates are within a range of two orders-of-magnitude (e.g., approximately 3 to 113 grams/gallon over a 95 percent probability range) as indicated in Figure 4.5. The distribution is positively skewed and the mean is at approximately 81st percentile. Approximately 90 percent of the values were below 44 grams/gallon.

The empirical distributions of estimated CO and HC emissions at these two sites on two specific days illustrate the large range of inter-vehicle variability in emissions. In all cases for all days of data collection, the range of inter-vehicle variability is approximately two orders-of-magnitude or more. The 95 percent confidence interval for the mean depends upon the sample size and the variability in the data, and ranges from approximately plus or minus 10 percent to plus or minus 25 percent among these four example data sets. The confidence interval for the mean is an indication of uncertainty in the mean due to random sampling error. The mean values tend to be heavily influenced by a relatively small number of large estimated values, which is why the mean values occur at high percentiles of the data sets. These characteristics are also observed in the other daily data sets reported in Appendix F.

The following section focus on the difference in mean emissions and inter-vehicle variabilities among data sets collected on different days and at different sites.

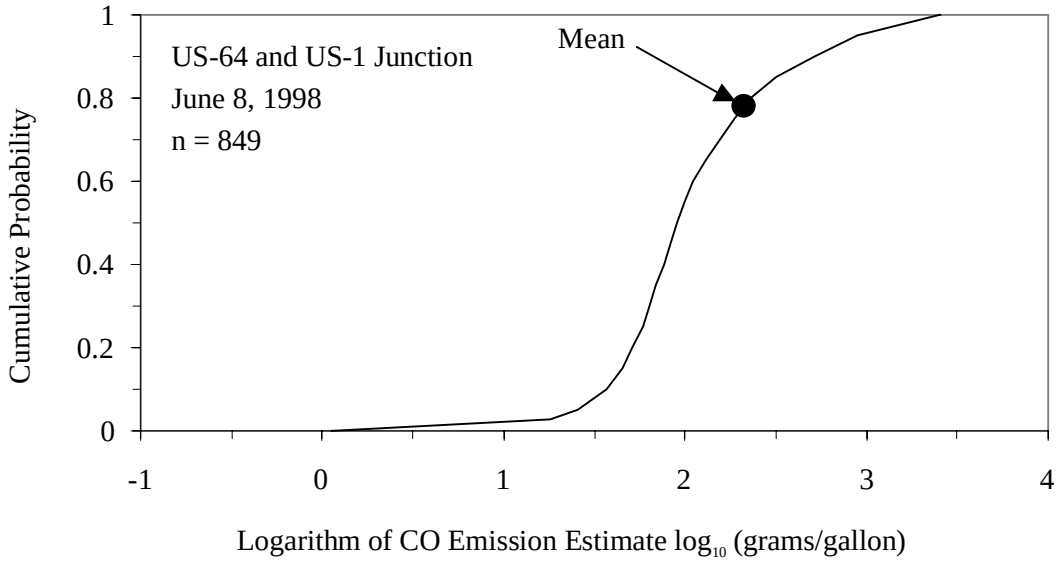


Figure 4.4. CO inter-vehicle variability and mean estimate for data collected on June 8.

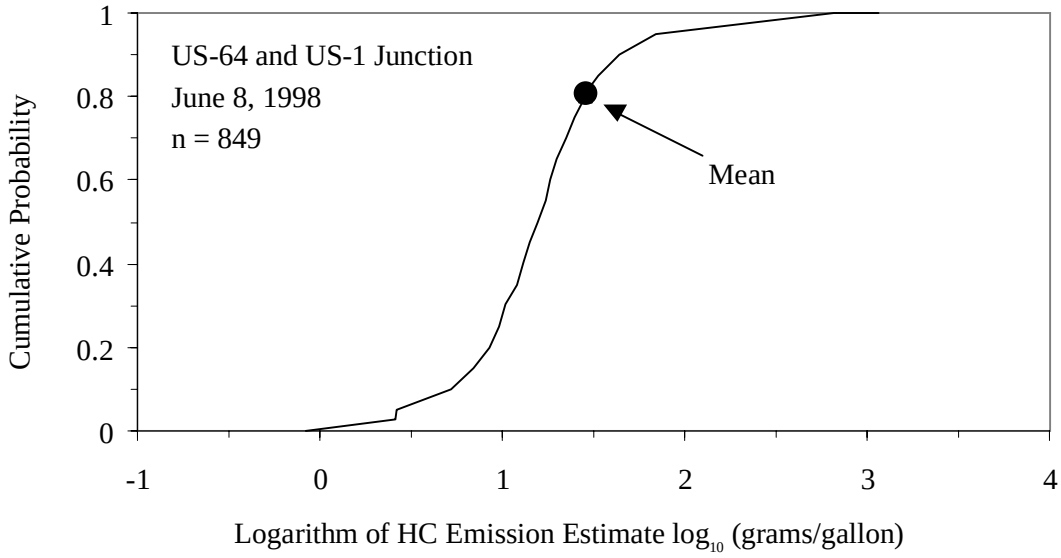


Figure 4.5. Propane-equivalent HC inter-vehicle variability and mean estimate for data collected on June 8.

4.4 Analysis of Emissions Data

In this section, emissions data are analyzed to identify possible differences in vehicle emissions estimates for data collected at different sites and on different days. Multiple observations for individual vehicles are evaluated to gain insight regarding intra-vehicle variability of emissions for comparison with observations regarding inter-vehicle variability. Finally, multi-comparison analysis of vehicle model year and vehicle type is given to determine whether there are any statistically significant differences in emissions by model year or vehicle type.

4.4.1 Analysis of Emissions Estimates: Site-to-Site Differences

In the previous section, a summary of data collected at different sites and on different days was presented. There are variations in the mean estimates and inter-vehicle variability of pollutants among some of the daily data sets. This section is devoted to further analysis of the overall dataset to evaluate whether there are statistically significant differences in emissions estimates with respect to site.

The cumulative probability distributions of emissions estimates for each site were plotted to compare emissions between the two sites in Figures 4.6 and 4.7 for CO and HC emissions, respectively. The figures indicate that the inter-vehicle variability in emissions is qualitatively similar for both sites. Specifically, the 95 percent probability range of values and the general shapes of the CDFs for CO at Sites 1 and 2 are similar, although there does appear to be some difference in the central portions of the

distributions. The two distributions for HC are nearly identical above the 40th percentile, with apparently only slight differences at the lower percentiles. Furthermore, there are similarities in the mean values.

In order to find out if there is a statistically significant difference between the two probability distributions, goodness-of-fit tests could be conducted. In our study we took the simpler approach of qualitatively comparing the empirical CDFs and of quantitatively evaluating the differences in mean values between the two sites. Specifically, a two-tailed t-test was used to test the hypothesis that two distributions have the same mean values. The calculated t-statistic for the difference in mean CO emissions values is 1.73, whereas it is 0.44 for the difference in mean emission values for HC. The critical t-value at a significance level of 0.05 is 1.96. Since the calculated t-statistics for CO and HC are less than the critical value at a significance level of 0.05, we cannot reject the hypothesis that means are equal. The results of the t-test are summarized in Tables 4.10 and 4.11 for comparison of the mean CO and HC values respectively.

Our overall conclusion is that, for practical purposes, there are not significant differences in the mean value or distribution of values at the two sites.

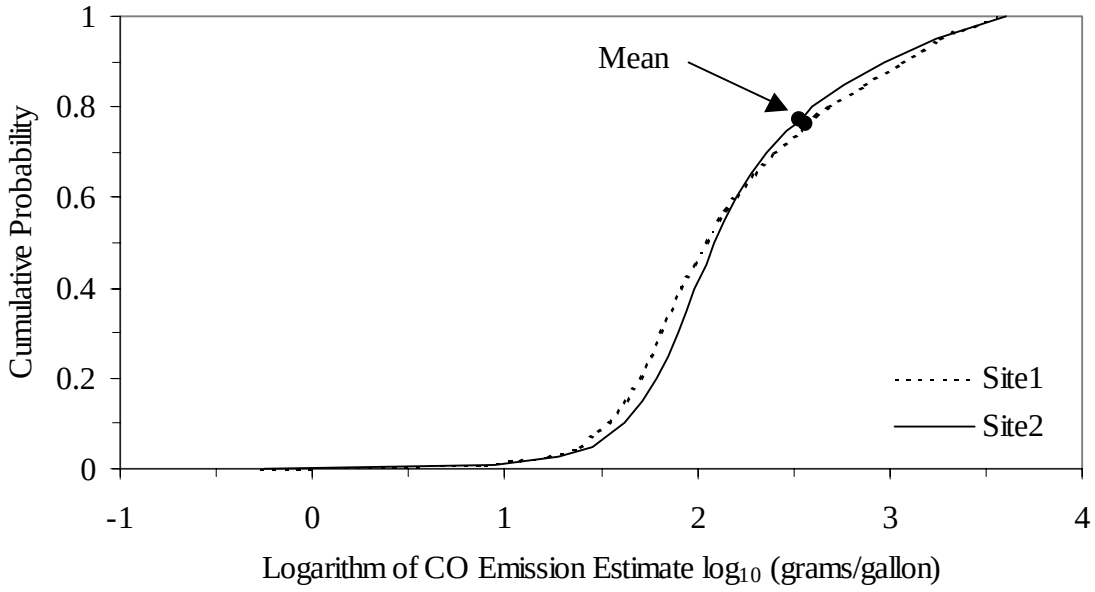


Figure 4.6. CO inter-vehicle variability and mean estimates for data collected at Site 1 and Site 2.

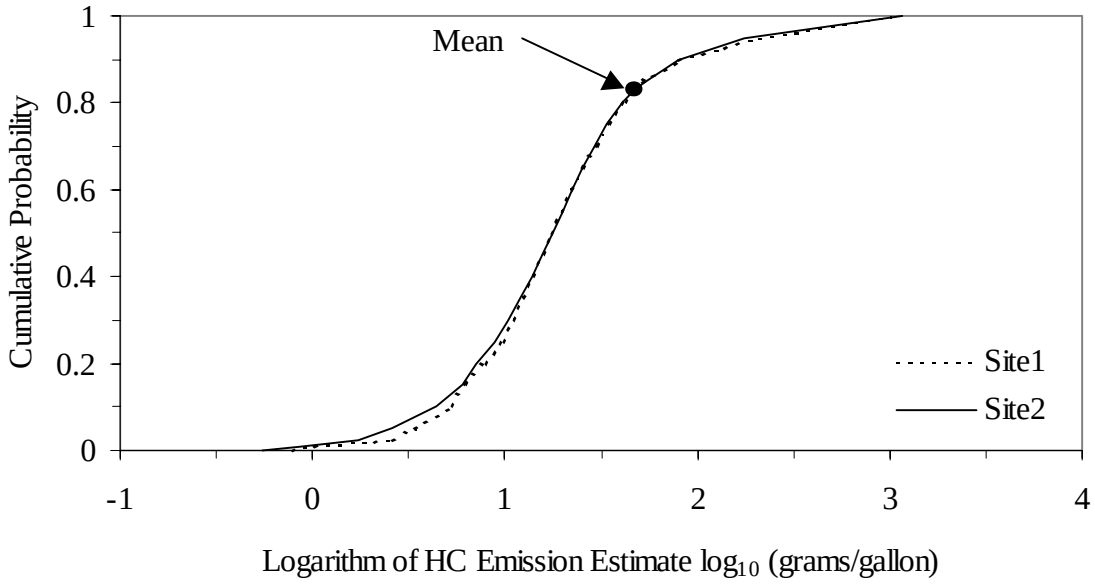


Figure 4.7. Propane-equivalent HC inter-vehicle variability and mean estimates for data collected at Site 1 and Site 2.

Table 4.10. Results of t-Test for Mean CO Emission Estimates of Site 1 and 2

	Site 1	*Site 2*
Mean	366.0	340.0
Standard Deviation	583.5	562.8
Observations	2,127	4,931
Hypothesized Mean Difference	0	
Degrees of Freedom	3,903	
t Statistic	1.73	
t Critical two-tail at α=0.05	1.96	

Table 4.11. Results of t-Test for Mean HC Emission Estimates of Site 1 and 2

	Site 1	*Site 2*
Mean	47.7	46.4
Standard Deviation	106.9	108.3
Observations	2,127	4,931
Hypothesized Mean Difference	0	
Degrees of Freedom	4,082	
t Statistic	0.44	
t Critical two-tail at α=0.05	1.96	

4.4.2 Analysis of Emissions Estimates: Day-to-Day Differences

On-road vehicle data were collected over the course of 10 days of field work between April and June in 1998. On the first four days of data collection, measurements were taken at Site 1 and the remaining data were collected at Site 2. In this section we seek to compare the conditions at each site and for each day of data collection, and to identify whether there are significant differences in emissions from one day to another.

The time period of data collection were not the same for all the days and varied between 7:00 am and 6:00 PM. The data collection times for each measurement day are

given in Tables 4.8 and 4.9. In order to facilitate a comparison, the data collection times are also plotted in Figure 4.8.

Data collection efforts took place in the afternoon for eight of the measurement days. On two days, April 7 and April 24, data were collected in the morning. On May 27 and June 9, data were collected from 9:30 AM until 3:00 PM and thus included both morning and afternoons. Afternoon rush hours were represented primarily by five of the data sets.

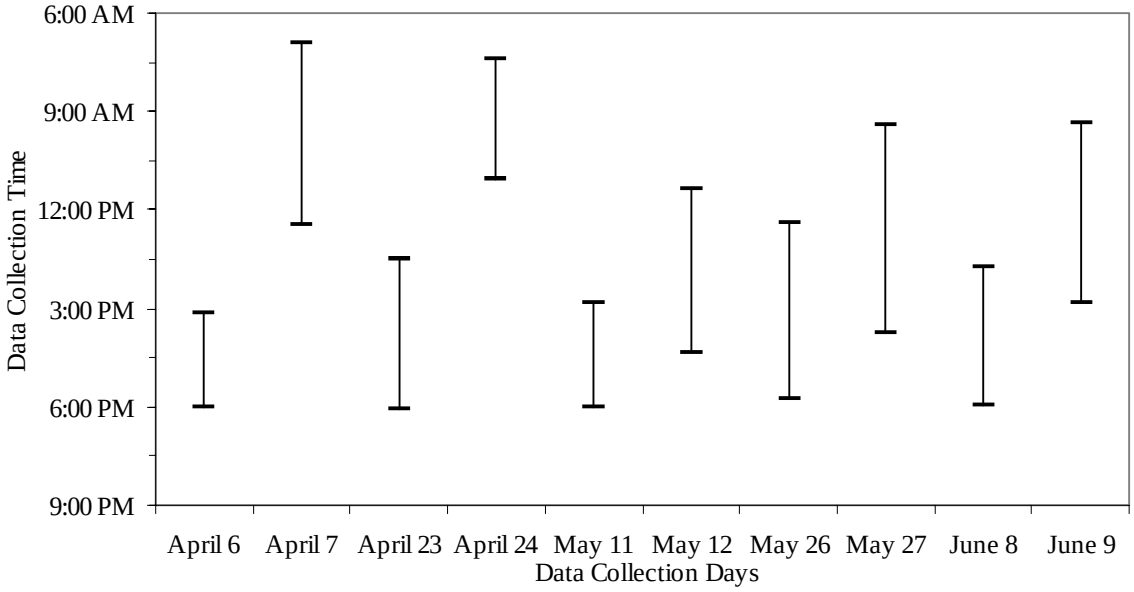


Figure 4.8. Data collection times by day.

4.4.2.1 Carbon Monoxide Data

In order to visually observe the differences in the inter-vehicle variability and the mean values of emissions for different days, the empirical CDFs of CO emissions estimates for data collected at different days were plotted in Figures 4.9 and 4.10. Figure 4.9 presents the data collected at Site 1 and Figure 4.10 illustrates the data collected at Site 2.

The variability in the individual vehicle CO emissions estimates is approximately within a range of three orders-of-magnitude for each of the data sets collected on different days. The 95 percent probability range varies for each of the distributions but typically is in a range approximately from 20 to 2,320 grams/gallon for data collected at Site 1. The probability distributions of data sets collected at Site 1 are similar to each other as shown in Figure 4.9. At the lower tail there are slight differences between the April 7 and April 23 data sets.

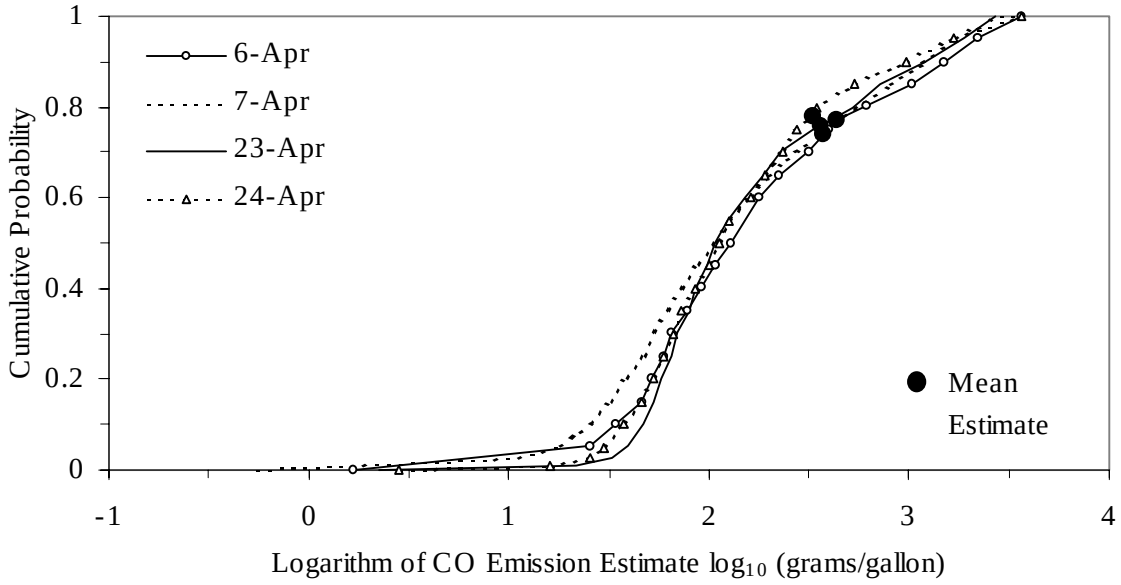


Figure 4.9. CO inter-vehicle variability and mean estimates for data collected at different days at Site 1.

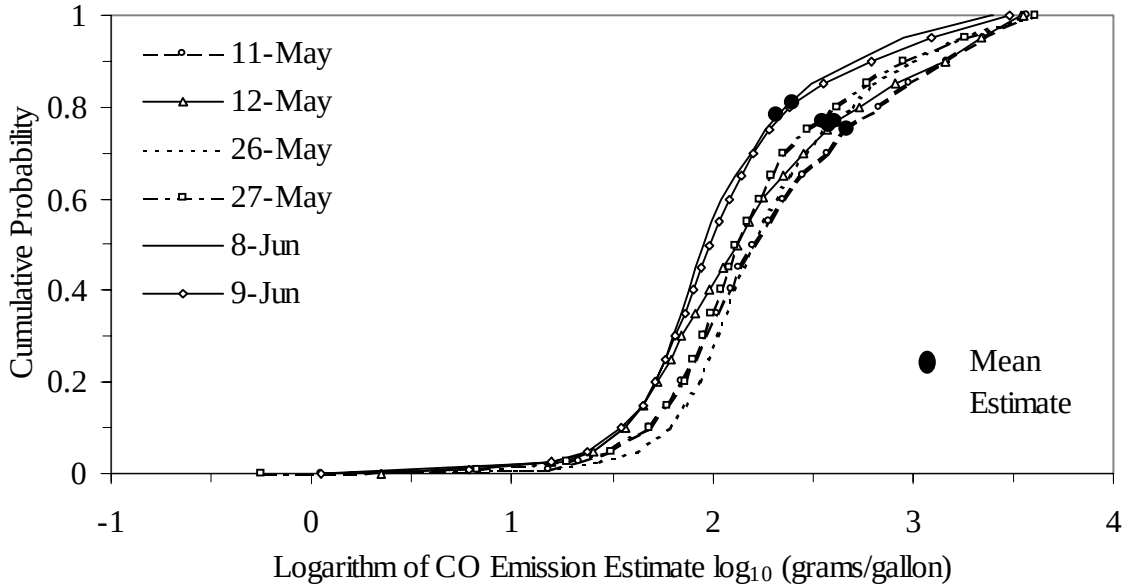


Figure 4.10. CO inter-vehicle variability and mean estimates for data collected at different days at Site 2.

The 95 percent probability range for CO emissions data collected at Site 2 ranges approximately from 17 to 2,270 grams/gallon. The June 8 and June 9 data sets seem to be slightly different from all others. These two distributions have lower values of emissions estimates at high cumulative probabilities as illustrated in Figure 4.10. For example the 90^{th} percentile for the June 8 and June 9 data sets is approximately 550 g/gal whereas it is in the range of 1,000 to 1,800 g/gallon for the other four data sets at that site. Mean estimates for June 8 and June 9 are smaller than the mean estimates of the other daily data sets at Site 2. However, the mean estimates are point estimates and uncertainty is associated with them. In order to make a comparison between mean estimates, the 95 percent confidence intervals on the mean are calculated and plotted in Figure 4.11.

In Figure 4.11, it appears that, although there seems to be a downward trend in the mean values for the four April dates, the confidence intervals of all four means overlap. In order to determine whether there is a statistically significant difference among these four means, we used a multi-comparison method.

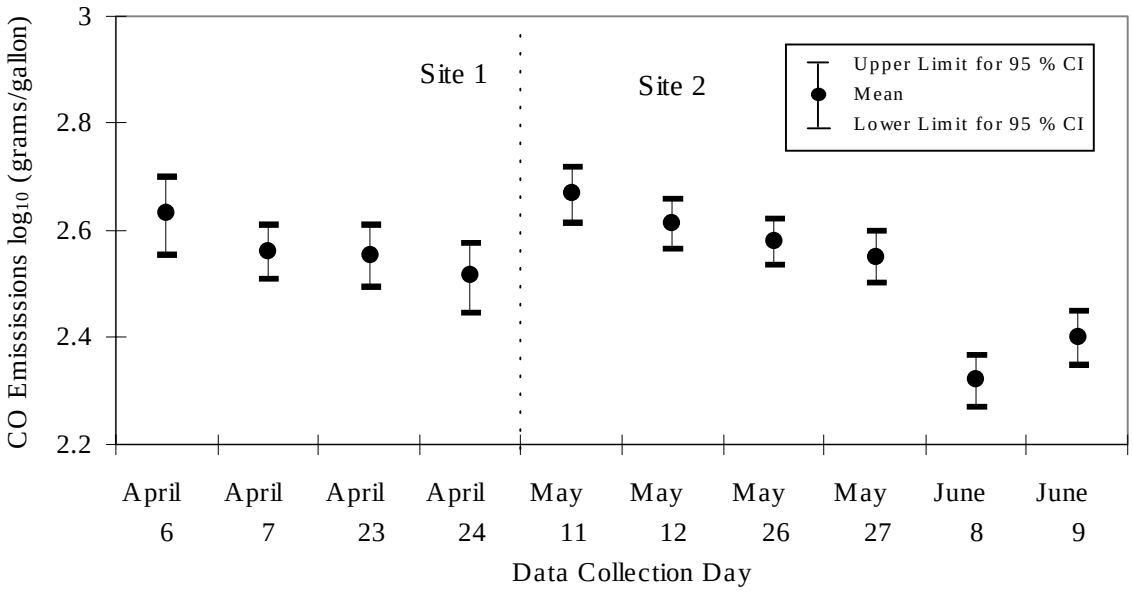


Figure 4.11. Mean CO estimates with 95 percent confidence intervals for data collected on different days.

Multi-comparison is a statistically useful method for determining the differences among different factor levels of the explanatory variables when there are more than two levels. For example, in this case the explanatory variable is the date of data collection, and there are 10 different days or levels. When factor levels are two we can use a t-test or F-test to compare the factor levels. However, as the number of factor levels increase the comparison of all possible pairs becomes challenging, due to combinatorics. For the case of k means, there will be $r = \frac{k(k-1)}{2}$ possible paired comparisons. For example if there are 10 levels of an explanatory variable, such as days of data collection, there are $r = 45$ possible paired comparisons. Assuming independent comparisons, the family-wise error rate (i.e. the probability of false rejection of at least one of the hypotheses) is given by $1-(1-\alpha)^r$, where α is the selected probability of type I error for a specific

comparison. This measure of family-wise type I error can be quite large. For example, even if there are only 45 comparisons, for comparing 10 means, and $\alpha = 0.05$, the family-wise error rate is $1-(0.95)^{45} \approx 0.90$. Thus, there is a 90 percent probability that at least one hypothesis is falsely rejected based on random sampling error.

As the number of factor levels increase, the family-wise error rate increases. In order to prevent this problem, several statistical approaches have been developed such as Tukey's method, Scheffe's method, and Bonferroni's method (Walpole and Myers, 1993). In our study we used S-PLUS™ 4.0, which has the capability to choose the best method for multi-comparison. This approach for choosing the best method is defined by Mathsoft (1997) based on finding the smallest critical point among all the available methods. The critical point is a statistic calculated to test the hypothesis for that problem. The critical point is the t-statistic, referred to as $t_{critical}$, for non-simulation-based methods such Tukey and Bonferroni, whereas an alternative simulation-based method generates a near-exact statistic via Monte Carlo simulation. By choosing the smallest critical point, narrower confidence limits can be obtained. As pointed out by Neter *et al.* (1996), the method which gives the narrowest confidence limit is preferred over the other methods. Such methods are said to be "efficient" in a statistical sense. The multi-comparisons of CO mean estimates are given in Table 4.12 .

The results of the 45 multicomparisons for mean CO emissions among all ten days of data collection are given in Table 4.12. The second column in Tables 4.12 gives the estimate of the difference between the means of the pairs of data sets that are being

compared and the third column gives the standard error of that estimate. The fourth and fifth columns give the lower and upper bound of the 95 percent confidence interval of the difference in means for the estimate. These values are calculated as:

$$(\text{estimate}) \pm (\text{critical point}) \times (\text{standard error of estimate}) \quad (4\text{-}24)$$

The critical point is calculated based upon the multi-comparison method chosen. For this case S-Plus™ estimated that the best method for estimating the critical point is the simulation-based method. This method is a Monte Carlo simulation that is used to generate a near-exact critical point.

The last column indicates whether the two means being compared are significantly different. Two means are significantly different when the confidence interval for the difference between them do not include zero. As seen in the table, there is no statistically significant differences among the means of the daily data sets collected at Site 1.

Some of the data sets collected at Site 2 have mean emissions estimates significantly different from each other. For example, mean CO emissions estimates of the June 8 and June 9 data sets are significantly different from all other mean emission estimates of daily data sets collected at Site 2. Furthermore, the June 8 and 9 data set means are not statistically different from each other. The May 11 and May 27 data sets have mean emissions estimates significantly different from each other. Other than these comparisons all of the other comparison within Site 2 are not statistically significantly different.

Table 4.12. Summary of Multi-Comparison of Mean CO Emissions Estimates for Different Dates

Date Compared	Difference in Mean Values (g/gallon)	Standard Error of Difference (g/gallon)	Lower Bound for 95% CI (g/gallon)	Lower Bound for 95% CI (g/gallon)	Significant Differences
April 6-April 7	63.7	37.1	-53.5	181	
April 6-April 23	69.6	39.3	-54.4	194	
April 6-April 24	101	40	-24.9	227	
April 6-May 11	-38.2	38.4	-159	82.9	
April 6-May 12	18.5	36.7	-97.3	134	
April 6-May 26	48.5	36.2	-65.9	163	
April 6-May 27	72.7	36.4	-42.2	188	
April 6-June 08	220	36.5	105	335	✔
April 6-June 09	178	36.2	63.7	292	✔
April 7-April 23	5.84	31.8	-94.5	106	
April 7-April 24	37.5	32.6	-65.6	141	
April 7-May 11	-102	30.7	-199	-5.16	✔
April 7-May 12	-45.3	28.5	-135	44.9	
April 7-May 26	-15.2	28	-104	73.1	
April 7-May 27	8.9	28.2	-80	97.8	
April 7-June 08	156	28.3	67.1	246	✔
April 7-June 09	114	27.9	26.1	202	✔
April 23-April 24	31.7	35.1	-79	142	
April 23-May 11	-108	33.2	-213	-2.91	✔
April 23-May 12	-51.1	31.3	-150	47.7	
April 23-May 26	-21	30.7	-118	76.1	
April 23-May 27	3.06	30.9	-94.6	101	
April 23-June 08	150	31	52.5	248	✔
April 23-June 09	108	30.7	11.4	205	✔
April 24-May 11	-140	34	-247	-32	✔
April 24-May 12	-82.8	32.1	-184	18.7	
April 24-May 26	-52.7	31.6	-153	47.2	
April 24-May 27	-28.6	31.8	-129	71.8	
April 24-June 08	119	31.9	18	220	✔
April 24-June 09	76.7	31.6	-23.1	176	
May 11-May 12	56.7	30.1	-38.4	152	
May 11-May 26	86.8	29.6	-6.64	180	
May 11-May 27	111	29.8	16.9	205	✔
May 11-June 08	258	29.9	164	353	✔
May 11-June 09	216	29.5	123	309	✔
May 12-May 26	30.1	27.4	-56.4	117	
May 12-May 27	54.2	27.6	-32.9	141	
May 12-June 08	202	27.7	114	289	✔
May 12-June 09	159	27.3	73.2	246	✔
May 26-May 27	24.1	27	-61.1	109	
May 26-June 08	172	27.1	86	257	✔
May 26-June 09	129	26.7	45	214	✔
May 27-June 08	147	27.3	61.2	234	✔
May 27-June 09	105	26.9	20.3	190	✔
June 08-June 09	-42.1	27	-127	43.3	

Some comparisons of data sets from both sites, such as May 11 versus April 7, 23, and 24 have mean emissions estimates significantly different from each other. When comparing the daily measurement between the two sites, there are many cases in which the daily means are not statistically different. For example, the mean value on April 6 at Site 1 is not significantly different from the mean values at Site 2 on all four of the May dates. In addition, the means of the other three April dates are not significantly different from the means of the last day of the three May dates. In contrast, the June 8 and 9 dates are both significantly different from the first three April dates, and the May 11 date is significantly different from the last three April dates.

Box-whisker plots were prepared for the ten daily data sets in order to gain insight regarding extreme values and the median estimates. Figure 4.12 gives the Box-whisker plots for each data set which were prepared using S-Plus™ 4.0.

The Box-whisker plot indicates the median (the white line inside the box), 95 percent confidence bounds for the median (dark gray area inside the box), the lower and upper quartiles of the data set (25^{th} and 75^{th} percentiles which is shown by the lower and upper ends of the box), and extreme values (top and bottom lines). Any observation which is more than 1.5 times the Interquantile Range (IQR = 75^{th} percentile – 25^{th} percentile) above the 75^{th} percentile or below the 25^{th} percentile is identified as extreme value.

The median estimates with 95 percent confidence interval are in the range of 100 to 200 g/gallon as shown in Figure 4.12. The plot suggests that the difference in median values are not as large as difference in mean estimates.

The difference in the conclusions drawn from mean and median analysis can be explained by the nature of these statistics. As explained by Cullen and Frey (1999), the median can be more robust measure of central tendency than the mean in some cases, since it is not sensitive to the shape of the tail of a distribution or as sensitive to the occurrence of outliers as is the mean.

The June 8 and June 9 data sets have the lowest mean emissions estimates although they have higher number of extreme values in the upper tail compared to other data sets. This might suggest that the probability distributions for these two data sets are significantly different than those of other data sets.

Possible reasons for differences between probability distributions might be: differences in temperature during the time of data collection; differences in the time of data collection; and differences in fleet characteristics such as speed, acceleration, headway, percent of vehicles in different vehicle type categories, percent of vehicles in different model year categories, and percent of vehicles in different platoon positions. We evaluate each of these possibilities to identify possible differences.

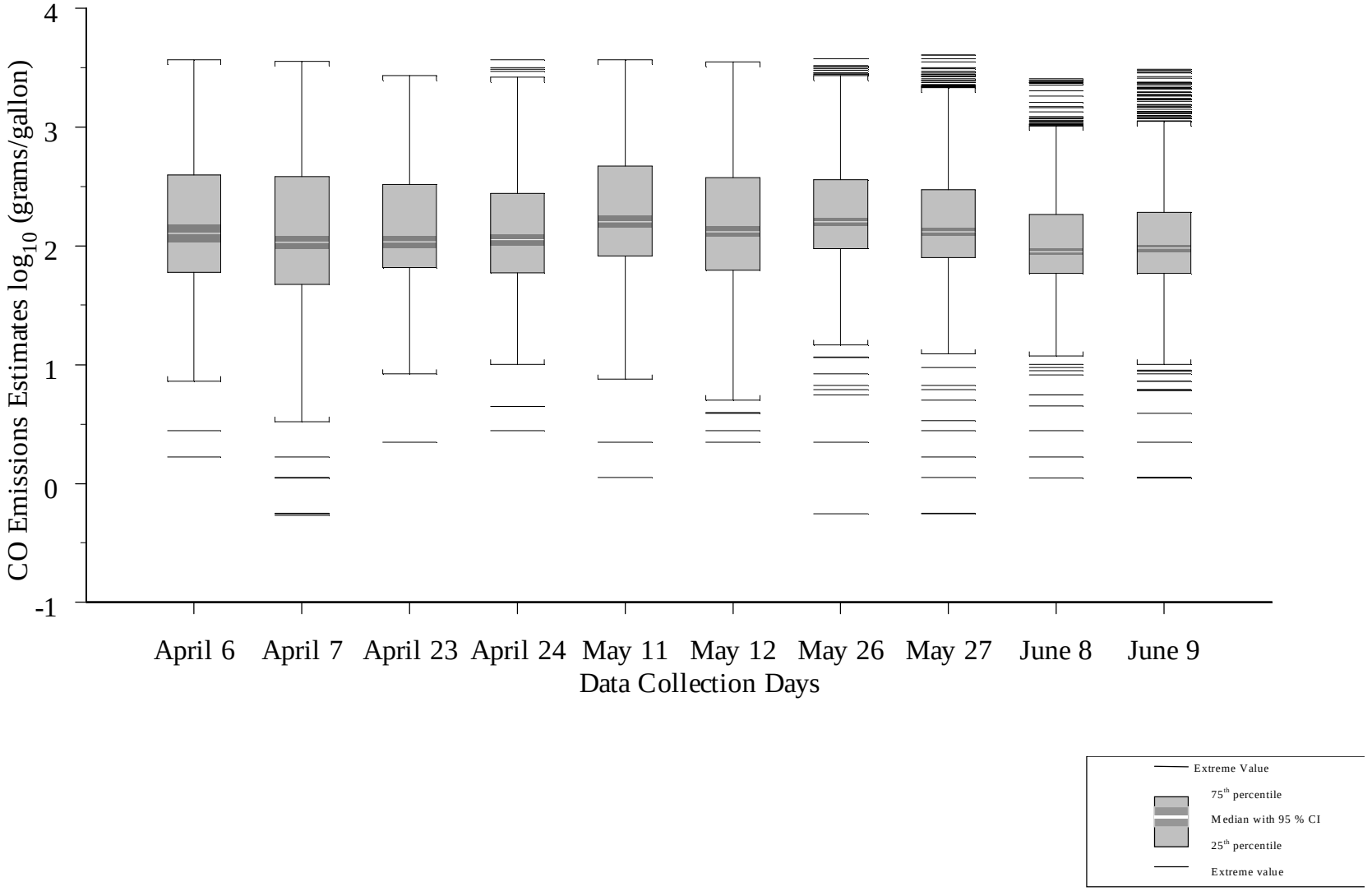


Figure 4.12. Box-Whisker plots for CO emissions estimates.

As shown in Figure 4.8, the times of data collection for these two days are not substantially different than the time of data collection for other data sets. Also these two data sets, which have similar probability distributions, were collected in different time period.

The daily temperatures, both daytime high and low, are reported in Tables 4.8 and 4.9. The values for June 8 and June 9 are in the range of the temperatures that were recorded for the other days of data collection. Differences in the activity data were also investigated. It is found that mean speed values for these two data sets are in the range of mean speed values for data collected on the other four days at Site 2. Mean speed values for these two days are significantly higher than mean values of data collected at Site 1. These differences are due to differences in sites such as site geometry and traffic patterns. Similarly, acceleration values of these two data sets are in the range of mean acceleration values estimated for data sets collected on different days at Site 2.

The percentage of vehicles in different vehicle type categories was also investigated. As given in Tables 4.8 and 4.9, for each data set the distribution of vehicle types vary slightly but is in a range of 62 to 83 percent for vehicle Type 2 and 17 to 38 percent for vehicle Type 3 for all days. There are no substantial differences of vehicle type percentages for June 8 and June 9 data sets, with most vehicles being of Type 2 as for other data collection days.

Based upon the investigation of the distribution of vehicle positions it is found that, for data sets collected at Site 2, free-flowing vehicles constitute 15 to 20 percent of

the whole data except for the June 9 data set, which had a free flowing vehicles percent of 45. Although the probability distributions of emissions for June 8 and June 9 are similar, the distribution of vehicle positions is quite different. Approximately 55 percent of the vehicles are in platoon for the June 9 data set, whereas platooning vehicles constitute 85 percent of the vehicles for the June 8 data set. For data sets collected at Site 1, free-flowing vehicles comprise 48 to 64 percent of the observed vehicles on a daily basis, which is similar to vehicle position data for the June 9 data set although there are differences in emissions distributions and mean estimates. These findings suggest that differences in the distribution of vehicle positions may not be helpful in explaining the differences in emissions estimates for different data sets.

About 70 percent of the vehicles observed at Site 2 were of 1991 or newer model years. There are slight differences in each model year category for each daily data set. However, the distribution of vehicle model years for June 8 and 9 is very similar to the overall average distribution for all days at Site 2.

Based upon the points discussed above we reached the following observations:

i) There are slight differences in mean emissions estimates of data sets for different days of data collection such as April 24, May 11, June 8, and June 9. In particular, the mean estimates for June 8 and June 9 are different than the mean estimates of the rest of the daily data sets.

ii) The differences in median values for different daily data sets are less pronounced than are the differences in mean estimates for different days of data collection. This suggests that differences in mean estimates might

be influenced by differences in the occurrence of extreme values observed in each data set.

iii) There are slight differences in the empirical CDFs of emissions estimates of data collected on some days. In particular, the June 8 and June 9 data sets seem to be different than the rest of the daily data sets. Time of data collection, daily temperatures, and activity data including average speed, average acceleration, percent of vehicles in different vehicle type categories, percent of vehicles in different model year categories, and percent of vehicles in different platoon positions were investigated. It seems unlikely that any of these fleet characteristics are sufficiently different to be solely responsible for differences in emissions estimates. However their combined effect might be important. It is possible that there are systematic errors in these two data sets because of some equipment errors. However, no difference in data collection procedure or errors were recorded for data collection on these two days. The data collection procedures on June 8 and June 9, including calibration, verification of the calibration, and recalibration as needed, were the same as for other days.

As a result of the considerations above we conclude that different CO emissions estimates of data collected on June 8 and June 9 versus other days is not likely to be explained on the basis of the observed traffic variables. Some undetectable error in the equipment and/or combination of the effects of activity data and data collection times with differences in temperature might be responsible.

Based upon the findings, it was decided that there is no specific basis for separating the data according to differences in mean emissions estimates and then analyzing them individually. The comparison of median estimates indicated that the median estimates for June 8 and June 9 are similar to median estimates for data sets collected at Site 1. Therefore, for further analysis, CO emissions estimates data from all daily data sets were combined together.

4.4.2.2 Hydrocarbon Data

The probability distributions of HC emissions estimates for each day of data collection are plotted in Figures 4.13 and 4.14 to observe potential differences in the inter-vehicle variability and mean estimates. Figure 4.13 presents the data collected at Site 1 and Figure 4.14 gives the data collected at Site 2.

The inter-vehicle variabilities are approximately within a range of three orders-of magnitude for all of the data collected at different days. The 95 percent probability range varies for each of the distributions but they are generally in the range of approximately 2 to 450 g/gallon for Site 1. The probability distributions of the daily data sets seem to be similar except for the April 23 data set. The data for April 23 has a different distribution than the other data sets collected at Site 1. The lower tail and center for this data set is shifted toward higher values of emissions for the same percentiles when compared with the other distributions of data sets collected at Site 1.

For Site 2, the 95 percent probability range varies for each distribution but they are generally in the range of approximately 3 to 480 g/gallon. There is no clear distinction between distributions of data sets collected at Site 2. However, the June 8, June 9 and May 27 data sets seem to be different at the higher tail of the distribution compared to the rest of the data sets collected at Site 2. In particular, the mean estimates for June 8, June 9, and May 27 seem to be lower than the mean estimates of the other data sets collected at Site 2. In order to account for the uncertainty due to random sampling error associated with the mean estimates, 95 percent confidence intervals are plotted in Figure 4.15.

According to Figure 4.15, it appears that mean estimates of the April 23 and April 24 data sets are different since the 95 percent confidence intervals do not overlap. For Site 2, the mean estimates of the May 27, June 8, and June 9 data sets are similar to each other, but different than the mean estimates of the May 11, May 12, and May 26 data sets. In order to determine if there is really a statistically significant difference among the mean estimates of the data sets we used the multi-comparison method as described in Section 4.4.2.1. In Table 4.13, the results of multi-comparisons for HC mean emissions estimates are given.

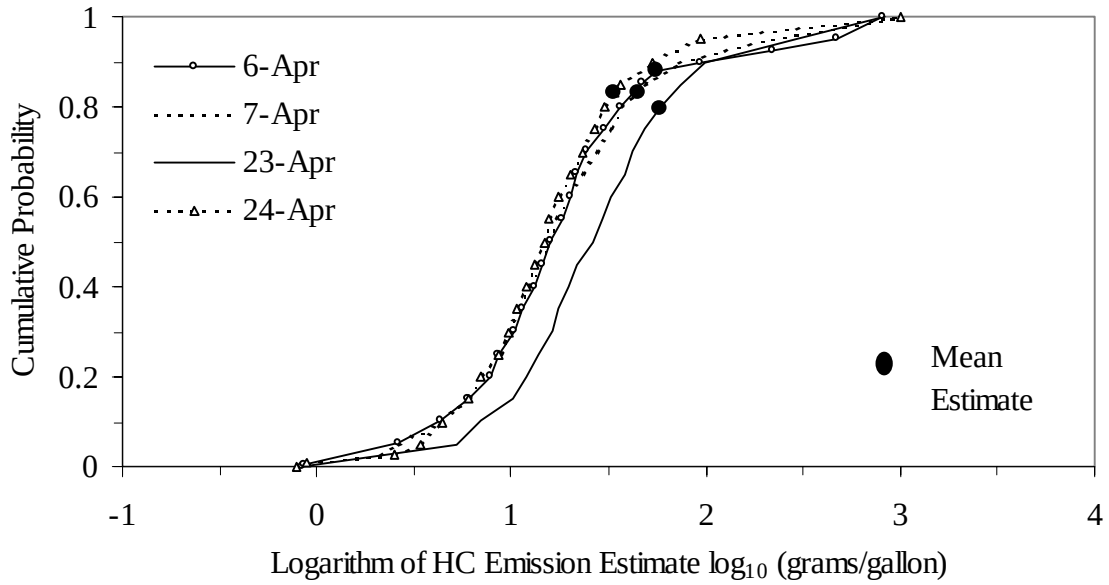


Figure 4.13. Propane-equivalent HC inter-vehicle variability and mean estimates for data collected at different days at Site 1.

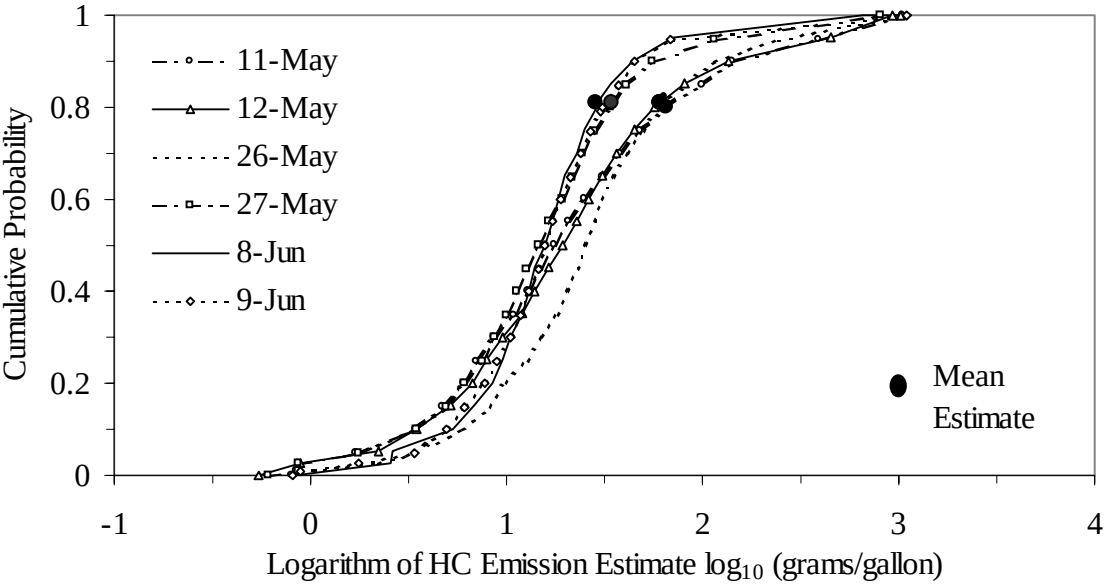


Figure 4.14. Propane-equivalent HC inter-vehicle variability and mean estimates for data collected at different days at Site 2.

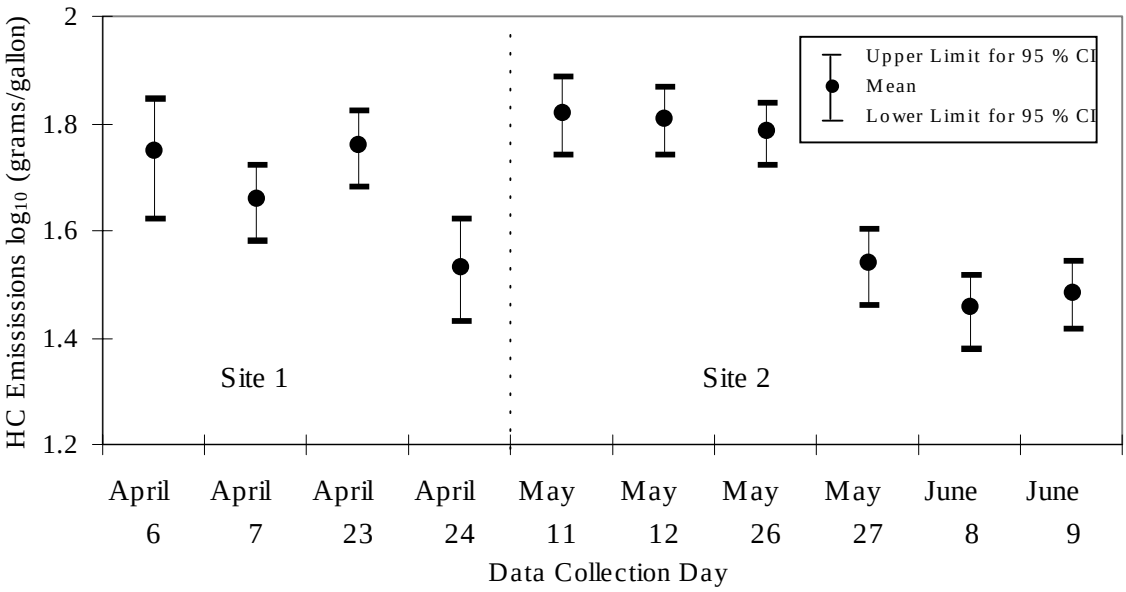


Figure 4.15. Mean propane-equivalent HC estimates with 95 percent confidence intervals for data collected on different days.

For Site 1 there is a statistical difference in mean emissions estimates of the April 23 and April 24 data sets. Comparison of other data sets at this site do not show a statistically significant difference in the mean emissions estimates of HC. For Site 2 there seems to be two different groups of data sets. The May 11, May 12, and May 26 data sets do not have statistically significant differences among them. Similarly, the May 27, June 8 and June 9 data sets also do not have statistically significant differences in mean HC emissions estimates. However, any pairwise comparison between data sets from these two groups have statistically significant different emissions estimates, such as between June 8 and May 11.

Table 4.13. Summary of Multi-Comparison of Mean HC Emissions Estimates for Different Dates

Date Compared	Difference in Mean Values (g/gallon)	Standard Error of Difference (g/gallon)	Lower Bound for 95% CI (g/gallon)	Lower Bound for 95% CI (g/gallon)	Significant Differences
April 6-April 7	10.6	7.04	-11.6	32.8	
April 6-April 23	-1.33	7.44	-24.8	22.1	
April 6-April 24	22.1	7.57	-1.83	45.9	
April 6-May 11	-9.65	7.27	-32.6	13.3	
April 6-May 12	-8.07	6.95	-30	13.9	
April 6-May 26	-4.85	6.87	-26.5	16.8	
April 6-May 27	21.7	6.9	-0.0824	43.4	
April 6-June 08	27.5	6.91	5.67	49.2	✔
April 6-June 09	25.7	6.86	4.12	47.4	✔
April 7-April 23	-11.9	6.02	-30.9	7.08	
April 7-April 24	11.5	6.19	-8.03	31	
April 7-May 11	-20.2	5.81	-38.5	-1.91	✔
April 7-May 12	-18.6	5.41	-35.7	-1.59	✔
April 7-May 26	-15.4	5.3	-32.1	1.29	
April 7-May 27	11.1	5.34	-5.74	27.9	
April 7-June 08	16.9	5.36	-0.00656	33.8	
April 7-June 09	15.2	5.29	-1.5	31.8	
April 23-April 24	23.4	6.65	2.43	44.3	✔
April 23-May 11	-8.32	6.3	-28.2	11.5	
April 23-May 12	-6.73	5.93	-25.4	12	
April 23-May 26	-3.51	5.83	-21.9	14.9	
April 23-May 27	23	5.86	4.51	41.5	✔
April 23-June 08	28.8	5.88	10.3	47.3	✔
April 23-June 09	27.1	5.82	8.74	45.4	✔
April 24-May 11	-31.7	6.45	-52	-11.4	✔
April 24-May 12	-30.1	6.09	-49.3	-10.9	✔
April 24-May 26	-26.9	6	-45.8	-7.99	✔
April 24-May 27	-0.395	6.03	-19.4	18.6	
April 24-June 08	5.4	6.05	-13.7	24.5	
April 24-June 09	3.7	5.99	-15.2	22.6	
May 11-May 12	1.59	5.71	-16.4	19.6	
May 11-May 26	4.81	5.61	-12.9	22.5	
May 11-May 27	31.3	5.64	13.5	49.1	✔
May 11-June 08	37.1	5.66	19.3	55	✔
May 11-June 09	35.4	5.6	17.8	53	✔
May 12-May 26	3.22	5.19	-13.1	19.6	
May 12-May 27	29.7	5.23	13.2	46.2	✔
May 12-June 08	35.5	5.25	19	52.1	✔
May 12-June 09	33.8	5.18	17.5	50.1	✔
May 26-May 27	26.5	5.11	10.4	42.6	✔
May 26-June 08	32.3	5.14	16.1	48.5	✔
May 26-June 09	30.6	5.06	14.6	46.6	✔
May 27-June 08	5.8	5.17	-10.5	22.1	
May 27-June 09	4.09	5.1	-12	20.2	
June 08-June 09	-1.71	5.12	-17.9	14.4	

In order to analyze the difference in median estimates and extreme values, Box-whisker plots were prepared as shown in Figure 4.16. The median estimates for all of the data sets are approximately 20 g/gallon except for April 23 and May 26, which have median values of approximately 30 g/gallon. These results are different than the mean estimates analysis where we saw larger differences between the mean values of some data sets, such as between May 11 and June 8. As explained in the previous section, the differences in results obtained from comparison of mean estimates and median estimates are due to the fact that median estimates are typically a more robust measure of central tendency than mean estimates.

The comparison of emissions estimates for the HC data set is rather complex. The results obtained from comparison of cumulative probability distributions, mean estimates and median values are somewhat different. For example, the CDFs and median values of April 23 and May 26 seem to be different than the rest of the data sets at Site 1 and Site 2 respectively; however, the mean estimates with 95 percent confidence intervals for these data sets overlap with the confidence interval of the mean estimates of other data sets at the respective sites.

The combination of effects of activity data and/or differences in measurement times and temperature differences might be considered as a potential explanation for the differences in data sets. However, as discussed for CO data, the differences in these parameters do not seem to be substantial. As a result it is concluded that differences in daily HC emissions estimates and distributions is not likely to be explained with the data

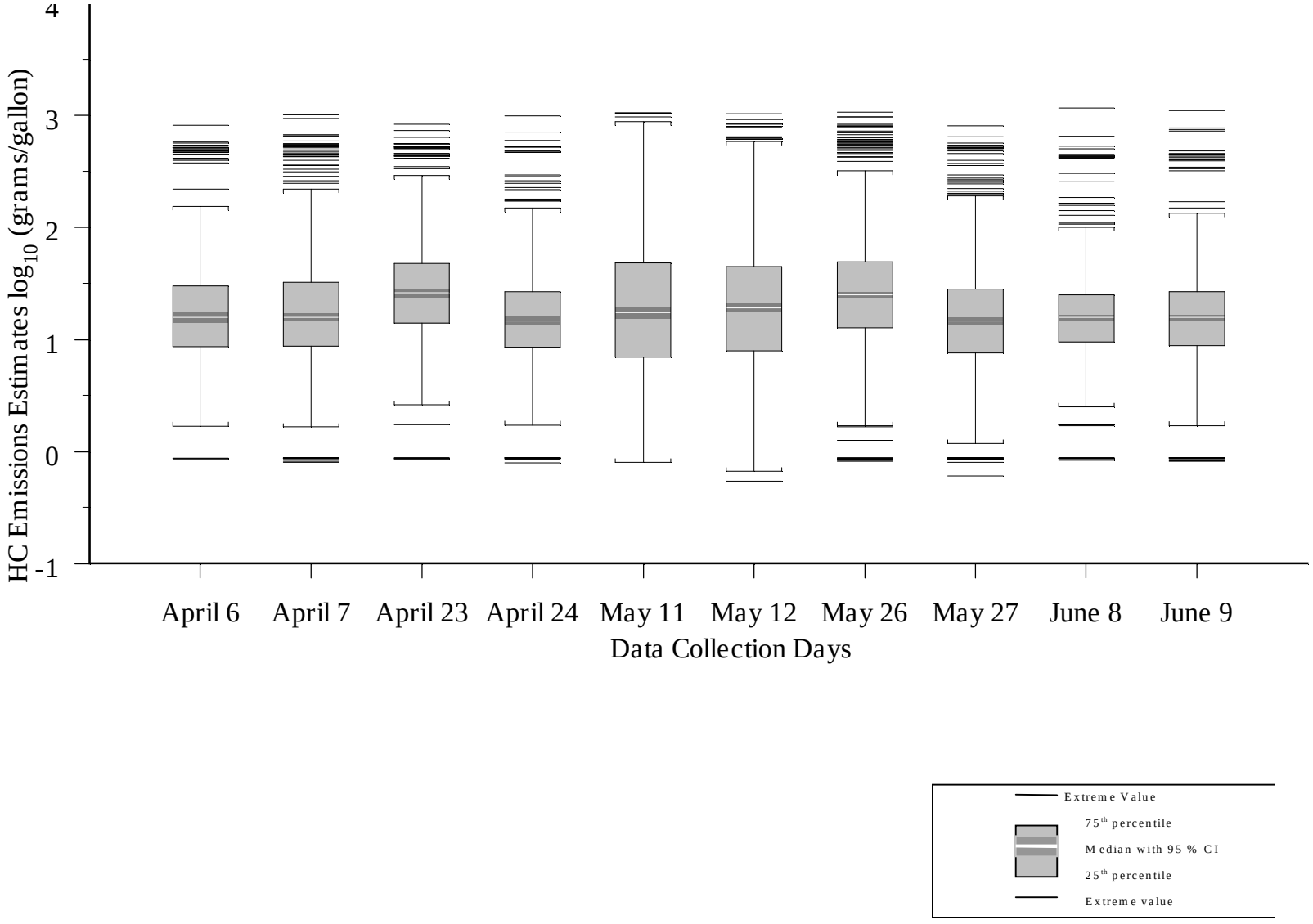


Figure 4.16. Box-whisker plots for HC emissions estimates.

collected. Based upon the findings, it was decided that there is no rationale to separate the data, such as by differences in mean emissions estimates, and then analyze them individually. Therefore, for further analysis, HC emissions data from all of the daily data sets were combined together.

4.4.3 Multiple Observations Analysis: Intra-Vehicle Variability in Emissions

Vehicles for which repeated measurements were obtained were identified for the purpose of quantifying intra-vehicle variability in emissions. The quantified intra-vehicle variability is then compared to overall variability. If intra-vehicle variability is high and comparable to inter-vehicle variability, then it could be difficult to explain observed inter-vehicle variability.

After forming the database, the vehicles which had multiple measurements were identified by using Microsoft Access. A total of 284 vehicles had more than one measurement. Out of this number, only 24 were measured three times, and one vehicle had four measurements. All others were measured twice.

A scatter plot of emission estimates of vehicles with three or four measurements is given to observe variability of the emissions estimates for the multiple measurements of each such individual vehicle. Figure 4.17 presents the scatter plot for CO data and Figure 4.18 gives the scatter plot for HC data. In each graph, different symbols were used for each vehicle to make the graphs easily readable.

As shown in Figure 4.17, CO emissions estimates can vary over a large range for the same vehicle. For example, the third vehicle has a CO emission estimate of about 103 grams/gallon for the first measurement, 194 grams/gallon for the second measurement and 1,613 grams/gallon for the third measurement. Of the 24 vehicles shown, 10 have emissions estimates that span more than one order-of-magnitude. In some cases, there is relatively little variability in the estimated emissions values. For example, for seven of the vehicles, the emissions estimates vary within one-half order-of-magnitude. These data values give insight regarding the intra-vehicle variability of CO emissions estimates.

Figure 4.18 illustrates the intra-vehicle variability for HC emission estimates. There are some vehicles which have large intra-vehicle variability, such as vehicle number fifteen. This vehicle has a measurement of 40 grams/gallon for the first observation. The second measurement is 20 grams/gallon and the third one is 436 grams/gallon. Similar to the case for CO, some vehicles have measurements spanning more than one order-of-magnitude while a few others have values within one-half order-of-magnitude.

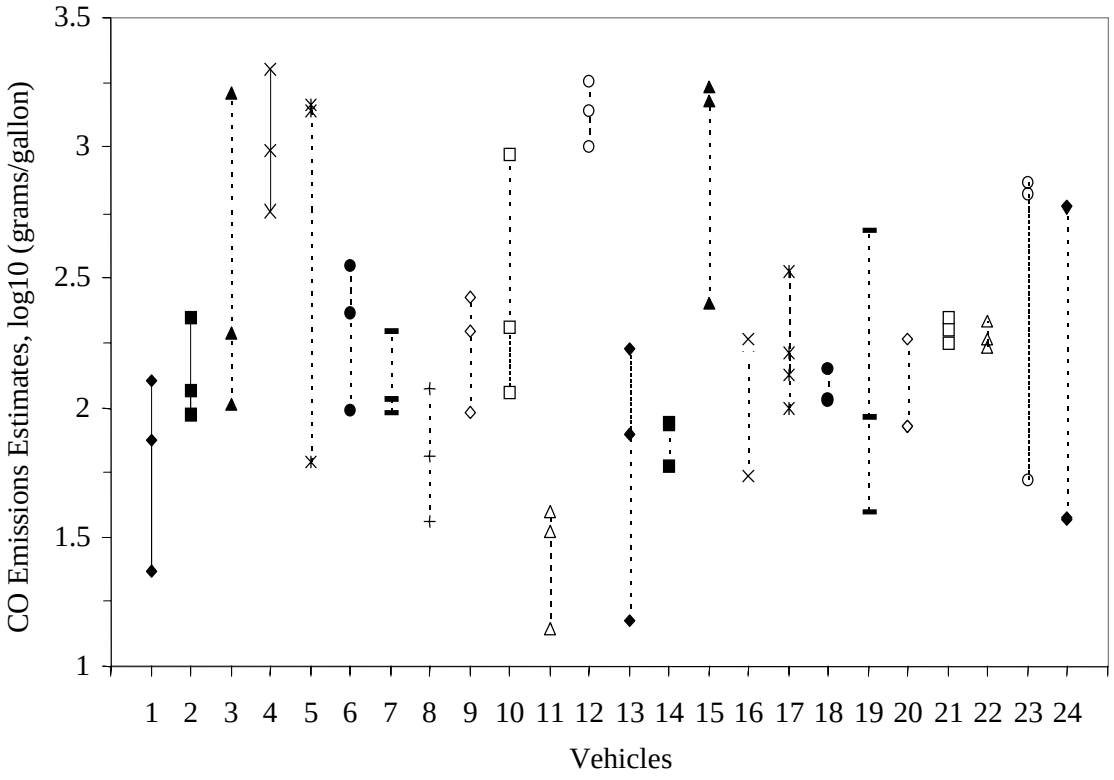


Figure 4.17. Scatter plot of CO emissions estimates for individual vehicles with 3 or 4 measurements.

Scatter plots for the vehicles having two measurements are given in Appendix F. CO emissions estimates for vehicles having two measurements range from 1 to 3,580 grams/gallon. For HC emissions estimates, the range is from 0.9 to 596 grams/gallon.

Every measurement, even for the same vehicle, was taken under different conditions, meaning that vehicle speed, acceleration, time of measurement and day of measurements might be different in each measurement. However, repeated values for a given vehicle were obtained at only one site for a given vehicle. There is no vehicle for which measurements were taken at both sites.

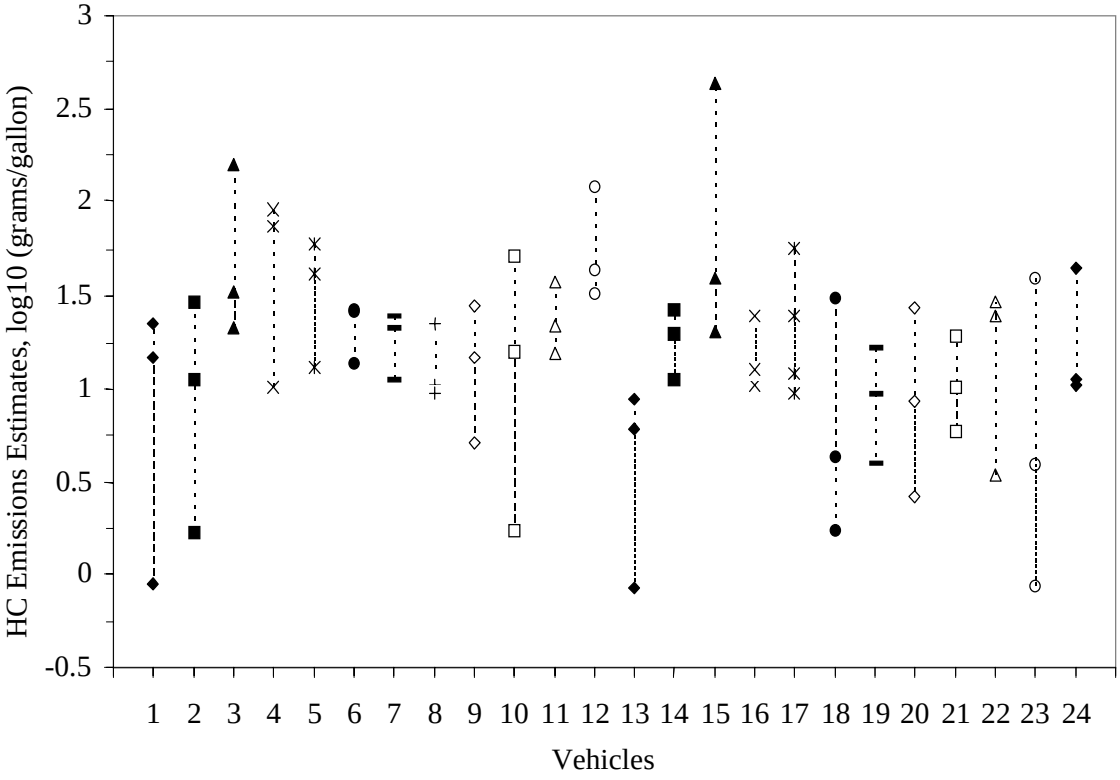


Figure 4.18. Scatter plot of HC emissions estimates for individual vehicles with 3 or 4 measurements.

In order to quantitatively estimate the intra-vehicle variability, we calculated standard deviations of emissions estimates for each of the 294 vehicles having multiple observations. In Figures 4.19 and 4.20, the cumulative probability distributions of individual vehicle standard deviations are plotted for CO and HC, respectively.

As shown in Figure 4.19 the individual vehicle CO emissions estimates have standard deviations from 0.6 to 750 g/gallon over a 95 percent probability range. The distribution is positively skewed and has an extreme value of standard deviation of 1,800 g/gallon. The standard deviation of emissions estimates for the complete data set,

including all vehicles at both sites, is 570 g/gallon, which corresponds to approximately the 90th percentile in Figure 4.19. This suggests that intra-individual variability might be as high as inter-vehicle variability but that for most vehicles it is less. It appears, for example, that 60 percent of the vehicles for which repeated measurements were available had standard deviations of less than 100 g/gallon, which is over five times lower than the standard deviations for inter-vehicle variability.

The standard deviations of individual vehicle HC emissions estimates range from 0.03 to 150 g/gallon on a probability range of 95 percent. The distribution is positively skewed, as shown in Figure 4.20, having an extreme value of approximately 400 g/gallon. The standard deviation of inter-vehicle variability of the overall emissions data set is 108 g/gallon, which corresponds to approximately the 95th percentile of the intra-vehicle variability standard deviations in Figure 4.20. The calculated intra-individual variability for HC is typically less than the inter-vehicle variability. For example, approximately 80 percent of the vehicles with multiple observations have an intra-vehicle variability standard deviation of less than 20 g/gallon, which is less than one-fifth the standard deviation for inter-vehicle variability.

Possible reasons for intra-individual variability might be differences in speed, acceleration, and weather conditions. These were not investigated in this work but could be analyzed in more detail in the future.

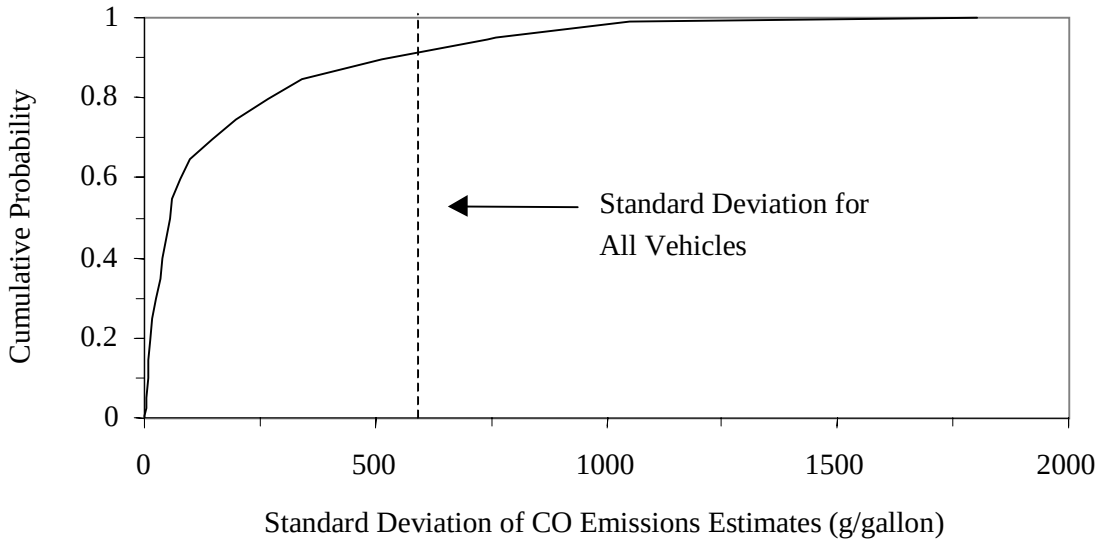


Figure 4.19. Distribution of the standard deviation of CO emissions estimates for vehicles for which 2, 3, or 4 repeated measurements were obtained.

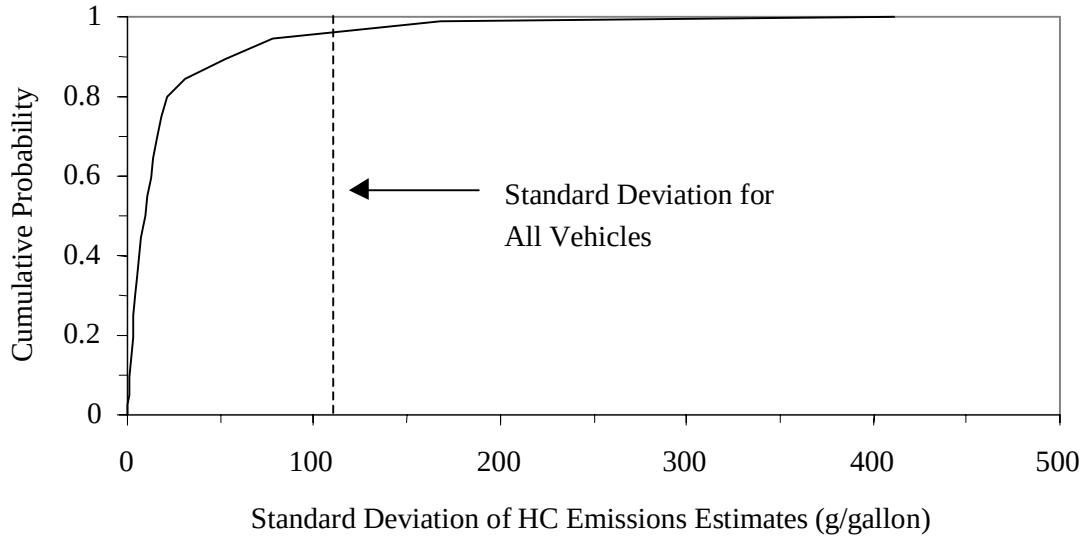


Figure 4.20. Distribution of the standard deviation of HC emissions estimates for vehicles for which 2, 3, or 4 repeated measurements were obtained.

4.4.4 Comparison of Model Years

In our study we did multi-comparison analysis for model years. As discussed in Section 4.1.2, measured vehicles were divided into eight different groups with respect to model years. These model year groups are: earlier than or equal to 1980; from 1981 to 1985; from 1986 to 1989; from 1990 to 1991; from 1993 to 1994; from 1995 to 1996; and from 1997 to 1999. These groups of model years were selected based upon differences in emission control technologies.

As summarized by Rendahl (1995), there are five different periods for vehicle emission control technology. Vehicles having model years before 1974 form the pre-catalyst period. Model years 1975 through 1980 were the beginning years for catalytic converter emission controls. Model years 1981 through 1987 were transition years. Newly designed vehicles were being produced that included catalytic converters. These vehicles also included some electronic engine process control as part of throttle body fuel injection. Model years 1988 through 1991 had no carbureted vehicles. These vehicles had multi-port fuel injection (MFI) techniques for air/fuel mixing. Vehicles which have model years later than 1992 have mass flow control of air/fuel mixture. These vehicles have also three-way catalytic converters. In addition to representing changes in emission control technology, model year also reflects the age of the vehicle, which is related to mileage accumulation and deterioration of the performance of the emission control system in time.

The multi-comparison results for model years are given in Tables 4.14 and 4.15 for CO and HC, respectively. Intervals that are statistically significantly different from zero are noted by a check mark. Mean estimates with 95 percent confidence intervals are given in Figure 4.21.

As seen in Table 4.14, the model year of the vehicles has an important effect on the CO emissions estimates. Out of the 28 pairwise comparisons, 24 of them show a significant difference. For example, the difference in mean emissions of vehicles of model years 1981 through 1985 and vehicles of model years 1993 through 1994 is 416 grams per gallon. Since the lower and upper 95 percent bounds do not include zero, these two sets of vehicles have mean CO emissions significantly different than each other. It is observed from Table 4.14 that, in general, when the model years of vehicles are close, the difference in mean CO emission estimates is not significant. There are three exceptions for this finding: comparison of model years 1981 through 1985 with model years 1986 through 1988; comparison of model years 1981 through 1985 with model years 1989 through 1990; and comparison of model years 1986 through 1988 with model years 1989 through 1990. Another finding is that, as the difference between factor levels increases the difference in mean estimates also increases. For example, the difference between vehicles having model years less than or equal to 1980 and model year 1997 to 1999 vehicles is 555 grams/gallon, with a lower bound of 424 grams/gallon and upper bound of 684 grams/gallon. However, the difference between the very old vehicles and vehicles that are closer to that model year range, such as vehicles having model years between 1986 to 1988, is not as large, with a mean difference of only 261

Table 4.14. Summary for Model Year Multi-Comparisons for Differences in CO Emissions Estimates

Model Years Compared	Difference in Mean Values (g/gallon)	Standard Error (g/gallon)	Lower[a] Bound (g/gallon)	Upper[a] Bound (g/gallon)	Significant Differences
"<=1980" -"81-85"	41.6	48.4	-103	186	
"<=1980" -"86-88"	261	44.8	127	395	✔
"<=1980" -"89-90"	352	45.2	217	487	✔
"<=1980" -"91-92"	426	44.3	294	559	✔
"<=1980" -"93-94"	458	43.4	328	587	✔
"<=1980" -"95-96"	531	43.1	402	659	✔
"<=1980" -"97-99"	554	43.6	424	684	✔
"81-85"-"86-88"	219	31.8	124	314	✔
"81-85"-"89-90"	311	32.4	214	408	✔
"81-85"-"91-92"	385	31.1	292	478	✔
"81-85"-"93-94"	416	29.8	327	505	✔
"81-85"-"95-96"	489	29.5	401	577	✔
"81-85"-"97-99"	512	30.1	422	602	✔
"86-88"-"89-90"	91.6	26.8	11.6	172	✔
"86-88"-"91-92"	166	25.2	90.4	241	✔
"86-88"-"93-94"	197	23.5	127	267	✔
"86-88"-"95-96"	270	23.1	201	339	✔
"86-88"-"97-99"	293	23.9	222	364	✔
"89-90"-"91-92"	74	26	-3.54	152	
"89-90"-"93-94"	105	24.3	32.6	178	✔
"89-90"-"95-96"	178	23.9	107	250	✔
"89-90"-"97-99"	202	24.7	128	275	✔
"91-92"-"93-94"	31.2	22.5	-36.2	99	
"91-92"-"95-96"	104	22.1	38.1	170	✔
"91-92"-"97-99"	128	23	58.9	196	✔
"93-94"-"95-96"	73	20.2	12.8	133	✔
"93-94"-"97-99"	96.3	21.1	33.3	159	✔
"95-96"-"97-99"	23.3	20.6	-38.3	85	

[a]The upper and lower bounds represent a 95 percent confidence interval on the difference in mean values

grams/gallon. In Figure 4.21 a clear trend in mean CO emission estimates are shown. As the model year gets newer the mean estimate decreases.

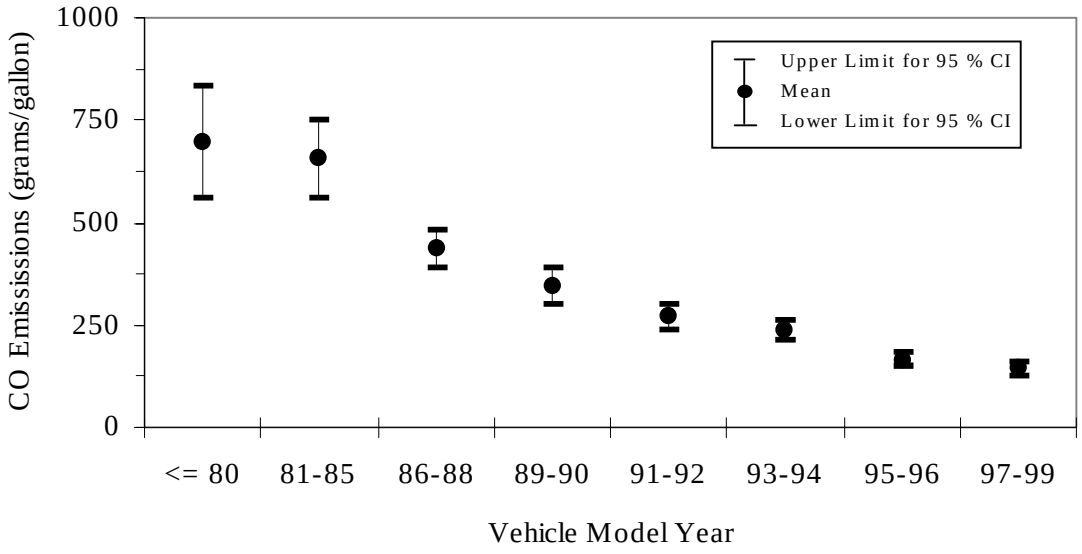


Figure 4.21. Mean CO estimates with 95 percent confidence intervals for different model years of vehicles.

Table 4.15 summarizes the results of the multi-comparisons for propane-equivalent HC data. Out of 28 pairwise comparisons, 15 of the pairs are significantly different. For very old vehicles, such as vehicles having model years earlier than 1980, the mean HC emissions are significantly different than the newer vehicles, such as vehicles having model years later than 1989. As the gap in model years increases, the difference in mean estimates also increases. For example, the mean difference between very old vehicles, with model years earlier than or equal to 1980, and vehicles with model years from 1986 to 1988, is 8.73 grams/gallon, which is not significantly different from a mean difference of zero, whereas the mean difference between very old cars and model year 1997 to 1999 vehicles is 27.6 grams/gallon. For vehicles which have model years later than 1989, the difference in mean emission estimates within the newer model

year categories are not significantly different. These results are also shown in Figure 4.22.

These findings show the importance of the model year of the vehicle as an explanation for differences in mean CO and HC emission estimates.

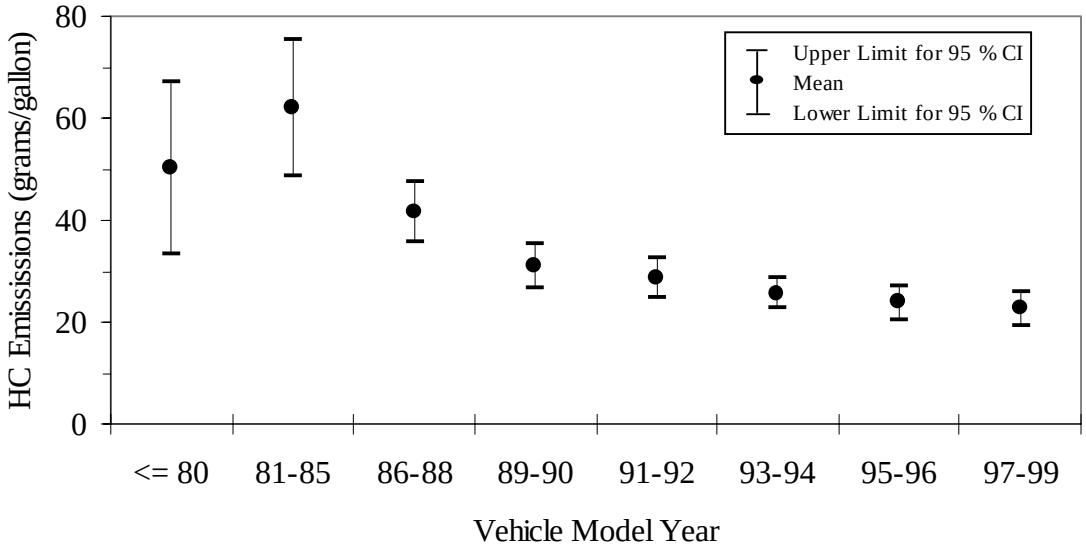


Figure 4.22. Propane-Equivalent Mean HC estimates with 95 percent confidence intervals for different model years of vehicles.

Table 4.15. Summary for Model Year Multi-Comparisons for Difference in Propane-Equivalent HC Emissions Estimates

Model Years Compared	Difference in Mean Values (g/gallon)	Standard Error (g/gallon)	Lower[a] Bound (g/gallon)	Upper[a] Bound (g/gallon)	Significant Differences
"<=1980" -"81-85"	-11.8	6.58	-31.5	7.94	
"<=1980" -"86-88"	8.73	6.09	-9.53	27	
"<=1980" -"89-90"	19.1	6.15	0.70	37.6	✔
"<=1980" -"91-92"	21.6	6.03	3.58	39.7	✔
"<=1980" -"93-94"	24.6	5.9	6.89	42.2	✔
"<=1980" -"95-96"	26.5	5.87	8.88	44.1	✔
"<=1980" -"97-99"	27.6	5.93	9.79	45.3	✔
"81-85"-"86-88"	20.5	4.33	7.54	33.5	✔
"81-85"-"89-90"	30.9	4.41	17.7	44.2	✔
"81-85"-"91-92"	33.4	4.24	20.7	46.1	✔
"81-85"-"93-94"	36.4	4.05	24.2	48.5	✔
"81-85"-"95-96"	38.3	4.01	26.2	50.3	✔
"81-85"-"97-99"	39.3	4.1	27.1	51.6	✔
"86-88"-"89-90"	10.4	3.64	-0.50	21.3	
"86-88"-"91-92"	12.9	3.43	2.65	23.2	✔
"86-88"-"93-94"	15.8	3.19	6.26	25.4	✔
"86-88"-"95-96"	17.7	3.14	8.33	27.1	✔
"86-88"-"97-99"	18.8	3.25	9.09	28.6	✔
"89-90"-"91-92"	2.5	3.53	-8.08	13.1	
"89-90"-"93-94"	5.42	3.31	-4.49	15.3	
"89-90"-"95-96"	7.32	3.25	-2.42	17.1	
"89-90"-"97-99"	8.42	3.36	-1.65	18.5	
"91-92"-"93-94"	2.92	3.07	-6.27	12.1	
"91-92"-"95-96"	4.82	3.01	-4.19	13.8	
"91-92"-"97-99"	5.91	3.12	-3.45	15.3	
"93-94"-"95-96"	1.9	2.74	-6.32	10.1	
"93-94"-"97-99"	2.99	2.87	-5.6	11.6	
"95-96"-"97-99"	1.09	2.81	-7.31	9.5	

[a]The upper and lower bounds represent a 95 percent confidence interval on the difference in mean values

4.4.5 Comparison of Vehicle Types

The effect of vehicle type on emissions is evaluated by using the multi-comparison method. In our study, we label the vehicles as Type 2 or Type 3 based upon Federal Highway Administration (FHWA) definitions. Table 4.16 gives the multi-comparison results for vehicle type for CO and HC. S-Plus™ using Fisher's LSD method to calculate the critical point for this calculation. Intervals that are statistically significantly different from zero are noted by a check mark.

Table 4.16. Summary for Vehicle Type Multi-Comparisons for Emissions Estimates

	Comparison	Difference in Mean Values (g/gallon)	Standard Error (g/gallon)	Lower[a] Bound (g/gallon)	Upper[a] Bound (g/gallon)	Significant Differences
CO	2-3	32.9	13.8	5.81	60.1	✔
HC	2-3	7.11	2.32	2.56	11.7	✔

[a]The upper and lower bounds represent a 95 percent confidence interval on the difference in mean values

In Figures 4.23 the mean CO emissions from Type 2 and Type 3 vehicles are plotted. Based upon the multi-comparison result we can say that there is a statistically significant difference between the mean estimates of Type 2 and Type 3 vehicles. Figure 4.24 illustrates the mean HC emissions measured for Type 2 and Type 3 vehicles. As suggested by the multi-comparison results, there is a significant difference among the mean HC estimates of vehicle types on grams per gallon basis. For both CO and HC the mean emissions estimate of Type 2 vehicles are higher than for Type 3 vehicles. This result is counter-intuitive, since generally it is assumed that pickup trucks, SUVs and other light duty trucks emit more than cars. However, there is a difference in fuel efficiency between cars and the larger Type 3 vehicles. In order to determine whether

fuel economy affects the comparison, we calculated grams/mile emission estimates for CO and HC by using typical numbers for fuel efficiency. In our calculations, we used a number of 21 miles/gallon for Type 3 vehicles and 28 miles/gallon for Type 2 vehicles. These numbers are the mean values of fuel efficiency for vehicles having model years from 1980 to 1997 (DOT,1999). The multi-comparison of grams/mile mean emissions estimates of CO and HC are given in Table 4.17.

Table 4.17. Summary for Vehicle Type Multi-Comparisons for Emissions Estimates

	Comparison	Difference in Mean Values (g/mile)	Standard Error (g/mile)	Lower[a] Bound (g/mile)	Upper[a] Bound (g/mile)	Significant Differences
CO	2-3	-0.085	0.531	-1.12	0.956	
HC	2-3	0.120	0.088	-0.053	0.293	

[a]The upper and lower bounds represent a 95 percent confidence interval on the difference in mean values

There is not a statistically significant difference between Type 2 and Type 3 vehicles when grams/mile emissions estimates are compared. This finding can be observed in Figures 4.25 and 4.26, which illustrate mean CO and HC emissions, estimates respectively, for Type 2 and Type 3 vehicles., The finding suggests that fuel efficiency makes a difference when comparing mean emissions estimates of different vehicle types.

Grams per gallon emissions estimates are statistically different for different vehicle types but not grams per mile emissions estimates.

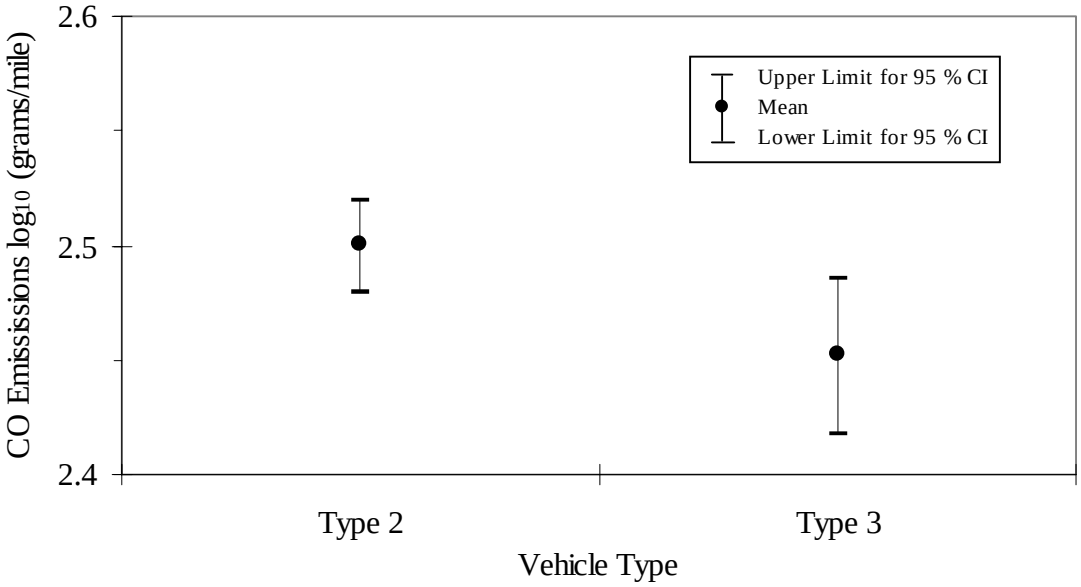


Figure 4.23. Mean grams per gallon CO estimates with 95 percent confidence intervals for different vehicle types.

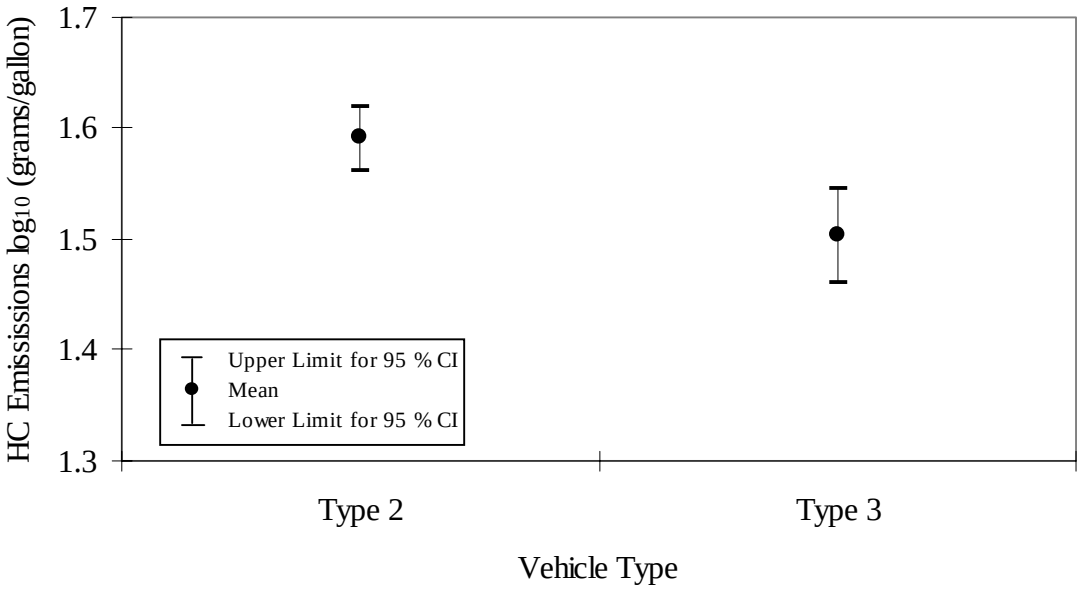


Figure 4.24. Mean grams per gallon Propane-Equivalent HC estimates with 95 percent confidence intervals for different vehicle types.

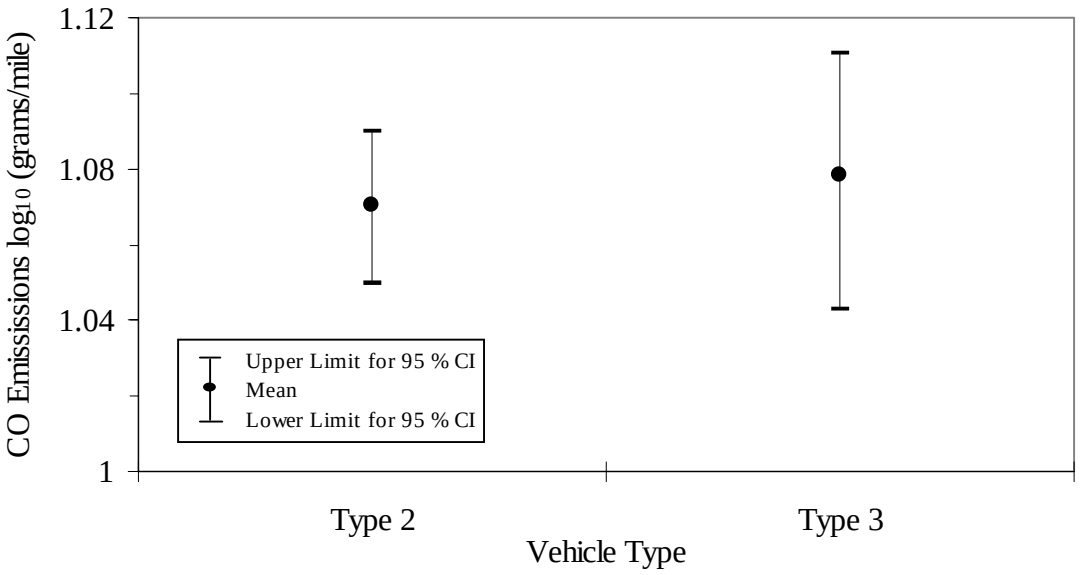


Figure 4.25. Mean grams per mile CO estimates with 95 percent confidence intervals for different vehicle types.

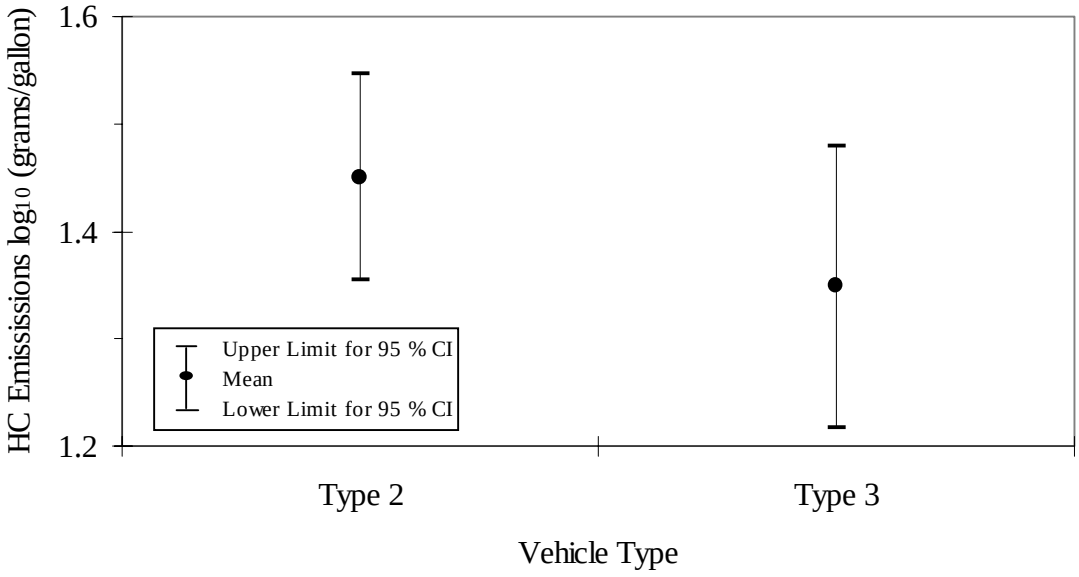


Figure 4.26. Mean grams per mile Propane-Equivalent HC estimates with 95 percent confidence intervals for different vehicle types.

5.0 MODEL DEVELOPMENT FOR VEHICLE EMISSION ESTIMATION

One of the goals of this study is to develop an empirical traffic-based model which will estimate vehicle emissions based upon measurable traffic and vehicle parameters. This section focuses on the statistical methods used for model development. Brief information on Hierarchical Tree-Based Regression (HTBR) and Ordinary Least Squares (OLS) regression will be presented. The following section will explain the methodology used to combine these two methods. Then, the results obtained from the analysis will be given. A sensitivity analysis of the models is also presented.

5.1 Statistical Modeling Methodology

In developing a statistical model, scatter plots are used to explore the data and capture trends and information explaining the underlying processes. They are essential if no *a priori* information on the process being modeled is available or if such information is inadequate. As a primary step we plotted scatter plots of CO and HC with respect to possible explanatory variables. These include: vehicle speed; vehicle acceleration; time headway and vehicle specific power.

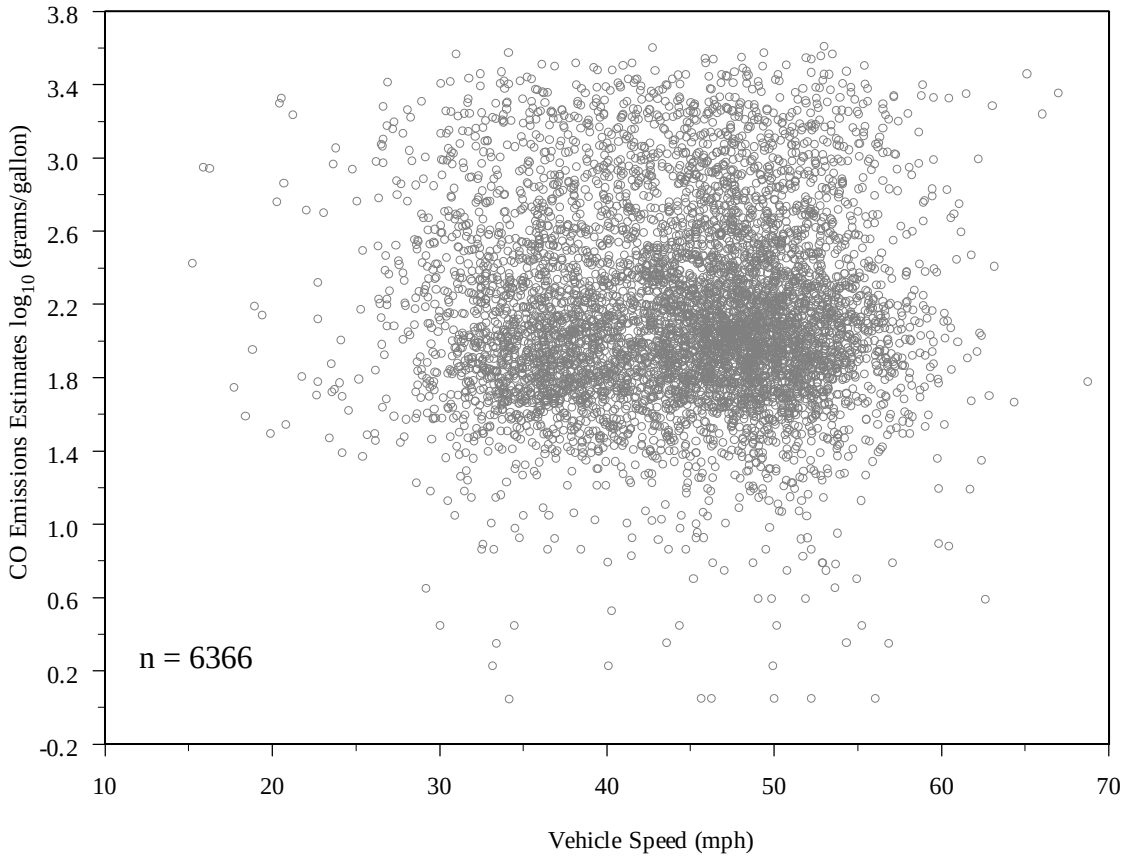


Figure 5.1. Scatter Plot of CO Emissions Estimates with respect to Vehicle Speed.

CO emissions estimates versus speed values are given in Figure 5.1 above. CO emissions estimates range from approximately 1 grams/gallon to about 3000 grams/gallon. Speed values range from approximately 15 mph to about 70 mph. There is a large block of data points in the middle of the plot, between 30 mph to 55 mph on the speed axis and between 25 grams/gallon to about 630 grams/gallon on the CO emissions axis. It is difficult to detect any trend in this scatter plot. The sample correlation coefficient between the logarithm of CO emissions estimates and speed values are calculated as –0.004. This correlation is very low, implying no statistical dependence between these two parameters.

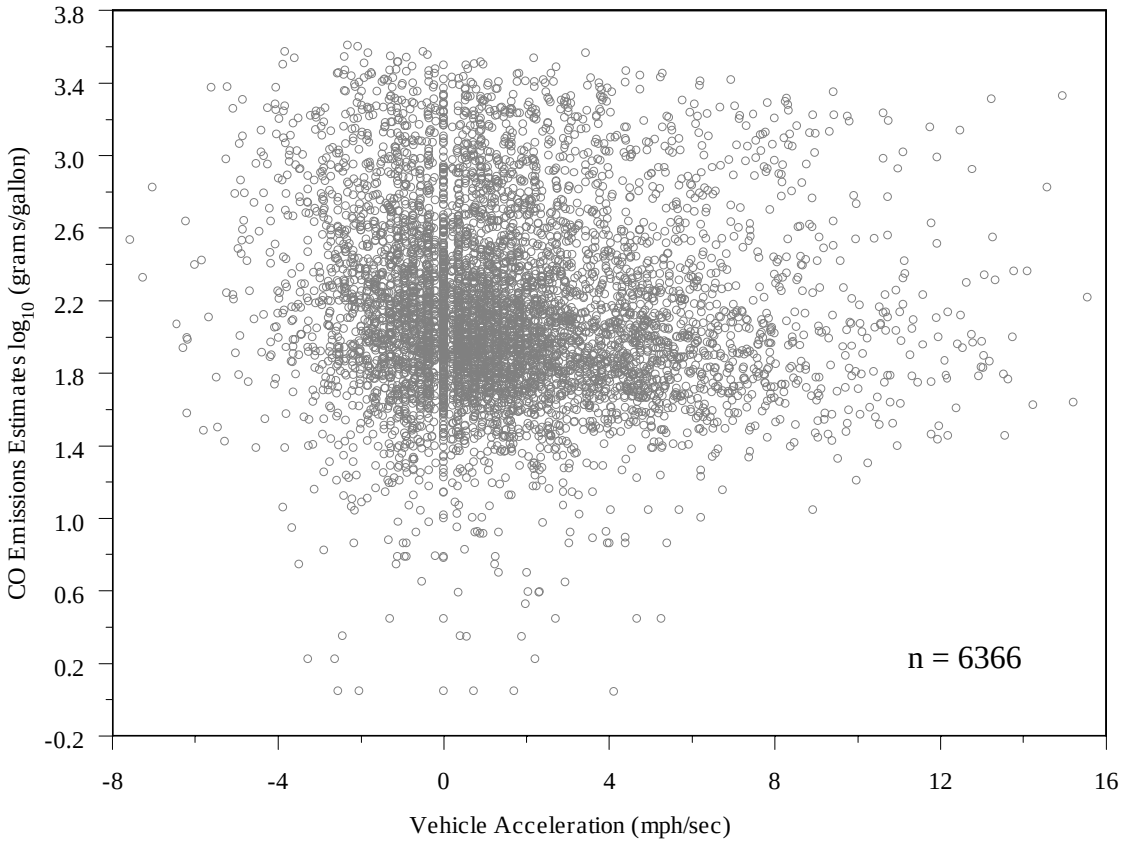


Figure 5.2. Scatter Plot of CO Emissions Estimates with respect to vehicle acceleration.

In Figure 5.2 CO emissions estimates are plotted against vehicle acceleration values. Approximately 88 percent of the vehicle acceleration values lie between –2 mph/sec to 5 mph/sec, but there are values lower than –2 mph/sec and higher than 5 mph/sec. Acceleration values as high as 15 mph/sec were obtained during the measurements. Deceleration values as low as -7 mph/sec were also obtained. There is a weak correlation among the logarithm of CO emissions estimates and vehicle acceleration values which is indicated by a sample correlation coefficient of –0.08 for these two variables.

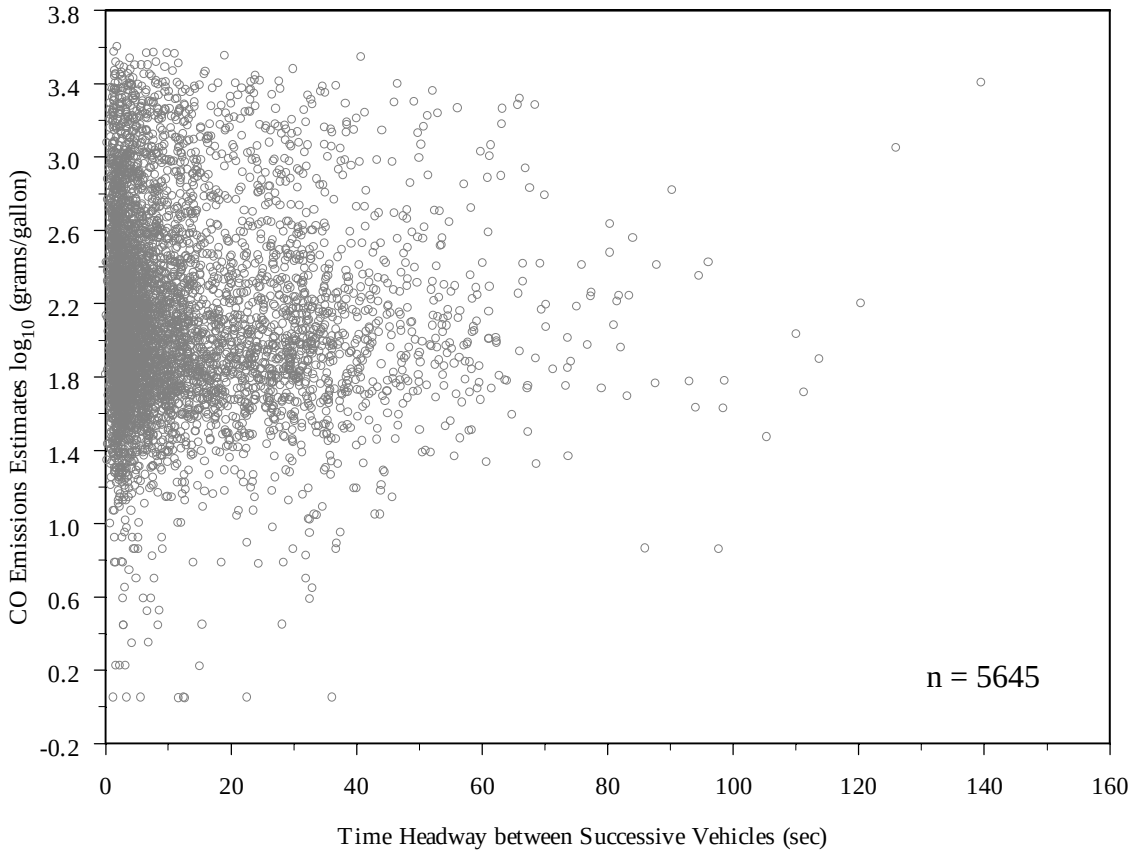


Figure 5.3. Scatter Plot of CO Emissions Estimates with respect to time headway between the successive vehicles.

CO emissions estimates were plotted against time headway between successive vehicles to see possible trends, as shown in Figure 5.3. Time headway values vary from very close to zero to approximately 140 seconds. Approximately 80 percent of time headway values are less than 20 seconds. It should be noted that approximately 65 percent of the vehicles have time headway values below 3 seconds were estimated to be in platoon, whereas vehicles having time headway values greater than 3 seconds are free-flowing. It seems that there is no clear trend between time headway values and CO emissions estimates. The sample correlation coefficient between CO emissions

estimates and time headway values is calculated as –0.03,which is very low and indicates a lack of linear dependence between the two values.

Scatter plots were also plotted for HC emissions estimates against vehicle speed, vehicle acceleration, and time headway values, and are given in Appendix F. Similar results were obtained. The sample correlation coefficients between the logarithm of HC emission estimates and vehicle speed, vehicle acceleration, and time headway values are estimated as –0.03, -0.1, and 0.01 respectively.

We also analyzed the effect of Vehicle Specific Power (VSP) on emissions estimates. This parameter represents the effect of engine load on emissions. An approach to calculate the instantaneous power of an on-road vehicle has been proposed by Jimenez *et al*. (1998). They defined specific power using the following equation:

$$\text{Specific Power} = v\left[a(1+e_i)+g\times grade+g\times C_R\right]+\frac{1}{2}\rho_A\frac{C_D.A}{m}(v+v_w)^2 v \qquad (5\text{-}1)$$

where:

m	=	vehicle mass
v	=	vehicle speed
a	=	vehicle acceleration
e_i	=	rotational mass factor which is gear dependent (0.1)
grade	=	vertical rise/slope length
g	=	acceleration of gravity

C_R	=	coefficient of rolling resistance (typically assumed to be 0.015)
C_D	=	drag coefficient (typically assumed to be 0.2 for sedan and 0.6 for van)
A	=	frontal area of the vehicle
ρ_A	=	ambient air density
v_W	=	headwind into the vehicles

The values given in parenthesis are typical values for those parameters obtained from Bosch (1986). A typical value of 0.0005 was used for $\frac{C_D.A}{m}$. Using these values, the equation for a flat surface and no wind condition is:

$$SP = 0.22va + 0.0657v + 0.000027v^3 \qquad (5\text{-}2)$$

Where v is in mph and a is in mph/s, and calculated SP is kilowatts per metric ton (McClintock, 1999).

By using Equation 5-2 vehicle specific power values were calculated and plotted with CO and HC emission estimates. Figures 5.4 and 5.5 gives scatter plots of CO emission estimates against vehicle specific power for Site 1. Figure 5.5 presents HC emission estimates versus vehicle specific power values calculated for Site 1.

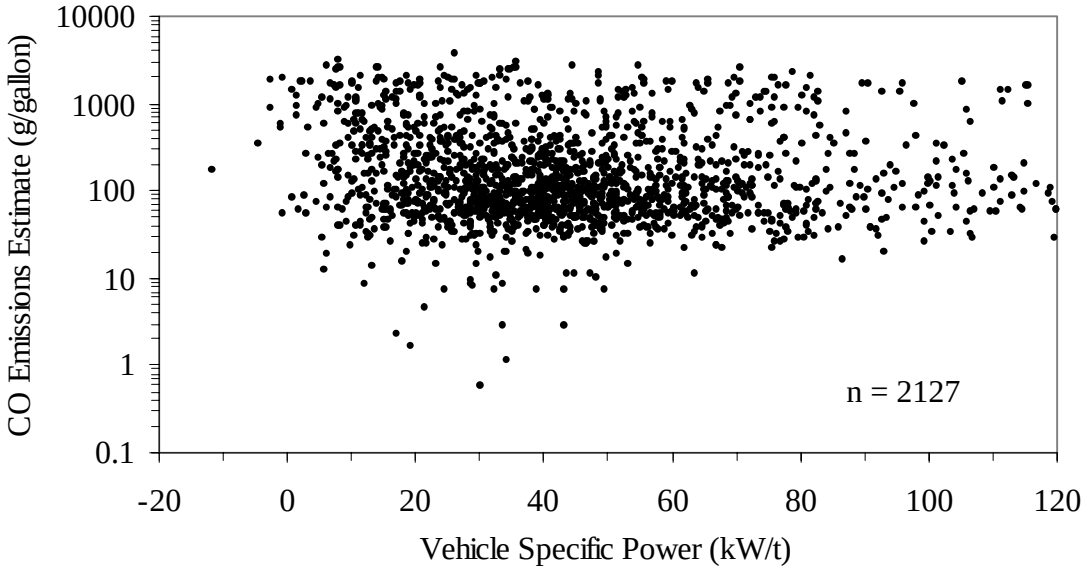


Figure 5.4. Scatter Plot of CO Emission Estimates with respect to Vehicle Specific Power Values.

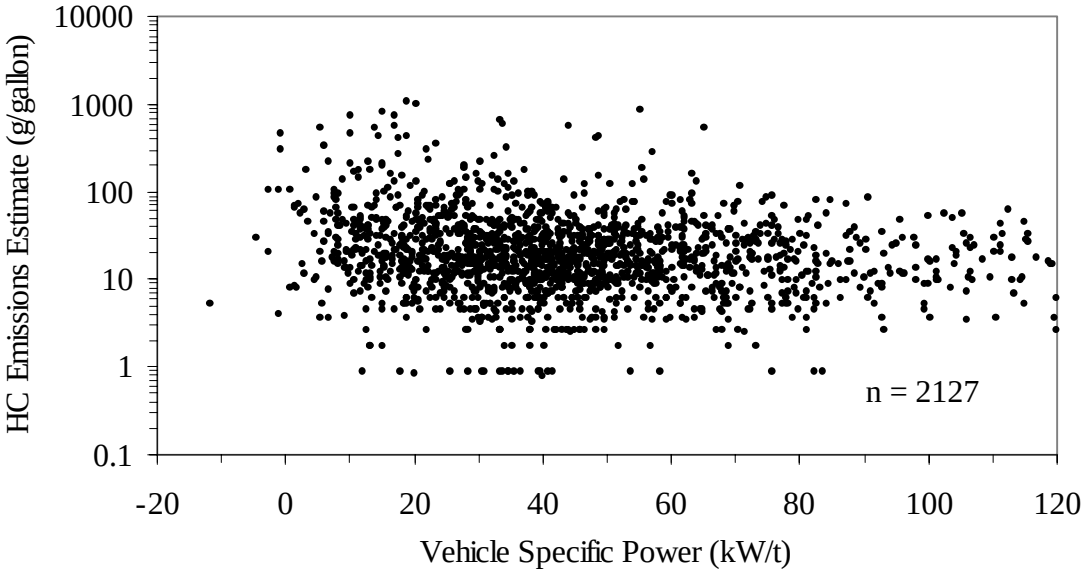


Figure 5.5. Scatter Plot of HC Emission Estimates with respect to Vehicle Specific Power Values.

As shown in Figures 5.4 and 5.5, there is no clear indication of correlation between VSP and emission estimates. The sample correlation coefficient between VSP and emissions estimates is –0.04 and –0.11 for CO and HC, respectively. Similar results were obtained for Site 2. The scatter plots are given in Appendix F. The sample correlation coefficients for Site 2 were -0.08 for CO and –0.07 for HC.

Statistical methods, including regression trees and least squares regression analysis, were utilized to find the relationship between emissions estimates and measured traffic and vehicle parameters. The following sections give a brief introduction to the methodology of the statistical techniques and explain the approach used in this study.

5.1.1 Hierarchical Tree-Based Regression

Hierarchical Tree-Based Regression (HTBR) can be interpreted as a forward step-wise variable selection method, similar to forward stepwise regression. Another name used for this method is Classification and Regression Trees (CARTs). The method is based upon iteratively asking and answering the following questions: (1) which variable of all of the variables 'offered' in the model should be selected to produce the maximum reduction in variability of the response?; and (2) which value of the selected variable (discrete or continuous) results in the maximum reduction in variability of the response? The method uses numerical search procedures to answer these questions. The

HTBR terminology is similar to that of a real tree; there are branches, branch splits or internal nodes, and leaves or terminal nodes (Washington *et al.*, 1997).

In order to explain the method in mathematical terms the definitions presented by Washington *et al.* (1997) will be used. The first step is to define the deviance at a node. A node represents a data set containing L observations. The deviance, D_a, can be estimated as follows:

$$D_a = \sum_{l=1}^{L} \left(y_{l,a} - \bar{x}_a\right)^2 \qquad (5\text{-}3)$$

where,

D_a = total deviance at node a, or the sum of squared error (SSE) at the node

$y_{l,a}$ = l^{th} observation of dependent variable y at node a

$\bar{x}_a$ = estimated mean of L observations in node a

For each of k variables, the algorithm seeks to split the domain of a variable, X_i, (where i has a value from 1 to k) into two half-ranges at node a, resulting in two branches and corresponding nodes *b* and *c*, each containing M and N of the original L observations (M + N = L) of the variable X_i. The reduction in deviance function is then defined as follows:

$$\Delta_{(allX)} = D_a - D_b - D_c \qquad (5\text{-}4)$$

where:

$\Delta_{(allX)}$ = the total deviance reduction function evaluated over the domain of all X_i's (i.e. for k number of X variables)

$$D_b = \sum_{m=1}^{M} \left(y_{m,b} - \bar{x}_b\right)^2$$

$$D_c = \sum_{n=1}^{N} \left(y_{n,c} - \bar{x}_c\right)^2$$

D_b = total deviance at node b

D_c = total deviance at node c

$y_{m,b}$ = m^{th} observation of dependent variable y in node b

$y_{n,c}$ = n^{th} observation of dependent variable y in node c

$\bar{x}_b$ = estimated mean of M observations in node b

$\bar{x}_c$ = estimated mean of N observations in node c.

The method seeks to find X_k and its optimum split at a specific value of X_k, $X_{k(i)}$, so that the reduction in deviance is maximized.

The maximum reduction occurs at a specific value $X_{k(i)}$, of the independent variable X_k. When the data are split at this $X_{k(i)}$, the remaining samples have a smaller variance than the original data set. Numerical methods are used to maximize (Equation

5-4) by varying the selection of which variable to use at a basis for a split and what value to use at the split point.

The iterative partitioning process is continued at each node until one of the following conditions is met: (1) the node of a tree has met minimum population criteria which is the minimum node size at which the last split is performed; or (2) minimum deviance criteria at a node have been met. Some software, such as S-Plus™, allows the user to select either criteria.

In order to select an appropriate HTBR model, a general procedure is to estimate a "full" tree model, by the procedure just described, that includes all of the variables of interest, and then 'trim' the tree with one of two methods. The trade-offs involved in tree model selections are analogous to statistical models in general. A statistical model with numerous number of exploratory variables may be difficult to maintain. In contrast, models with a limited number of exploratory variables are easier to work with and understand. Finally, the presence of many highly intercorrelated explanatory variables might increase the problem of round off errors. When the number of X variables is small, round off effects, especially when calculating model parameters, can be controlled by carrying a sufficient number of digits in intermediate calculations (e.g., most computer regression programs use double-precision, 16 digits). Still, with a large number of X variables, serious round off effects can arise despite the use of many digits in intermediate calculations (Neter *et al.*, 1996).

The two methods used to trim trees are called ‘pruning’ and ‘shrinking’. Each method trims the ‘full’ tree model into a smaller and more manageable tree size.

Pruning a tree successively snips off the least important splits. Shrinking a tree reduces the number of effective nodes by shrinking the fitted value of each node towards its parent node when the difference in fitted values is relatively small. A shrinking parameter, s, can be chosen from a value of 0 to 1 (s = 0 represents the root tree, and s = 1 represents the full tree). The detailed information on how to use these methods and mathematical formulation is given by Venables and Ripley (1997), and MathSoft (1997).

5.1.2 Ordinary Least Squares Regression

Ordinary Least Squares (OLS) regression is a common statistical technique for quantifying the relationship between a continuous dependent variable and one or more independent variables. The dependent variables may be either continuous or discrete.

This method has been used since the late 19th century by many analysts. Part of the reason that OLS is so popular is that it is easy to comprehend, it is incorporated into most statistical packages, and its statistical properties are well understood.

The basic OLS regression equation for a single variable regression can be written as follows;

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_i \times X_i + \varepsilon_i \qquad (5\text{-}4)$$

where;

$\hat{Y}_i$	=	value of the response variable in the i^{th} trial
$\hat{\beta}_0, \hat{\beta}_i$	=	estimators of regression parameters
X_i	=	value of the predictor variable in the i^{th} trial
ε_i	=	random error term, generally required to be normally distributed with a mean of zero and a variance of σ^2.

The parameters of the OLS regression equation, $\hat{\beta}_0$ and $\hat{\beta}_i$, are found by the method of least squares which requires that the sum of squares of errors would be minimized.

In order to fit a linear regression there are key assumptions that should be made. These include:

(1) X and Y values should be randomly selected

(2) The error terms are normally distributed

(3) The error terms have a constant variance

(4) The error terms are independent

(5) The error terms are normally distributed

If the above assumptions are violated the regression equation might yield biased results. A detailed discussion and presentation on OLS methods is not provided here because of the widespread knowledge of OLS techniques. For further information Neter *et al.* (1996) can be consulted.

5.1.3 Combination of HTBR and OLS

In developing a statistical model, the task is to specify a model that represents the underlying processes to the greatest extent possible. Given a dependent variable Y and a set of independent variables, X_1, X_2, X_3, …, X_p, the analyst must choose the most effective independent variables, and specify a suitable relationship between them such that it is most likely to be the same as the 'real' relationships that generated the observed data.

In OLS methods, as summarized by Washington (1999), there are three general categories of model mis-specifications: omitting relevant variables from the model, including irrelevant variables in the model, and incorrectly specifying the functional form of dependent variables with respect to the independent variable in the model. Omitting relevant variables cause the estimates of the standard errors and intercept to become biased if variables orthogonal with the variables with included variables. If variables correlated (not orthogonal) with included variables are omitted from the regression, then all estimates (intercept, slope coefficients, and standard errors) are biased. Including irrelevant variables does not cause a bias in the parameter estimates

but the parameter estimates become statistically inefficient. Finally, incorrectly specified variables result in biased estimates of parameters.

In order to solve the parameter selection problem, generally all of the parameters that are thought to be relevant are included in the model initially. With the help of step-wise regression techniques, the best independent variables are chosen by sequentially removing the least important variable from the model until some criteria is not, such as a minimum significance level for the parameters of all included variables. There might be some problems using the step-wise approaches. These methods can be misleading or may not give the best results when number of variables are large since they are 'mechanical' procedures that are driven primarily by the number of variables contained in the data (Washington,1999). As noted by Neter *et al.* (1996), automatic search procedures such as stepwise regression can sometimes err by identifying a poor regression model as "best". In addition, the identification of a single regression model may hide the fact that several other regression models may also be "good".

HTBR methods do not use priori information on the number of variables and their relationship with the dependent variable. These models use their data mining properties to identify the relevant variables. Complex relationships contained in the data can be captured via the process of stratification of the data inherent in the approach.

As suggested by Washington (1999), HTBR lacks some desirable properties of OLS procedures, specifically, properties of statistical parameters such as available statistical tests which might be used to test the differences in HTBR model formulations.

Without estimated parameters and their related properties, it is difficult to determine whether patterns identified in the data are likely to be explained by long-term stable patterns showing the real relationships in the data, or whether they are just noise reflecting spurious relationships by random fluctuations in the sample. In other words, tree is not as explicit a measure of statistical significance of HTBR methods as tree is for OLS methods.

In our study, we combined HTBR and OLS regression methods to be able to use strengths of both of the methods. We stratified the data into smaller data sets by using regression tree, and then OLS regression was done to capture relationships within the data strata.

First, a 'full' tree is grown for the data. For this purpose, all the variables available are included in the model. Then, using the pruning techniques, the least important splits are snipped off from the tree. In order to decide on the number of terminal nodes of the final tree, a reduction in deviance versus node size graph is used. This graph is plotted easily for the tree using S-Plus™ commands. An example plot of such a graph is given and is more fully explained in Figure 4.19 in the next section.

After trimming the tree, the next step is to fit OLS regression equations at the terminal nodes of the tree. The variables used for OLS are the measurable traffic parameters which include vehicle speed, vehicle acceleration, and time headway between successive vehicles. Possible interaction terms and some higher terms of vehicle speed, vehicle acceleration, and time headway are also included in the

regression equations. In order to choose the relevant parameters step-wise regression techniques are used for the data at every terminal node of the tree.

In OLS regressions when the predictor variables are correlated among themselves, intercorrelation or multicollinearity among them is said to exist. Multi-collinearity causes the parameter estimates to become unstable. One way to prevent multi-collinearity is the centering of the independent variables. Centering involves taking the difference between each observation and the mean of all observations for the variable. After obtaining the regression equations based upon the centered variables, regression coefficients for the model with the original variables can be found by simple transformation (Neter, 1996). In our study we used centering method to prevent multi-collinearity effect.

The model which includes regression tree and OLS regressions at the terminal nodes of the tree is called as “vehicle-based” model. In order to use this model individual vehicle speed, vehicle acceleration , and time headway values should be know apart from the parameters needed in the regression tree, such as vehicle model year. If traffic parameters for individual vehicles are not available, one can use regression trees to estimate the mean emissions at the terminal nodes of the tree. This approach is referred to as “flow-based” model. Both “vehicle-based” and “flow-based” models estimate emissions on grams/gallon basis. Fuel economy data obtained from the literature can then be used to estimate grams/mile emissions.

In the following section, model development and model results will be presented.

5.2 Statistical Model Development

In this section, model development processes will be explained. First, a nomenclature section will be given. Then, the models developed for different case studies will be addressed together with the results. Finally, sensitivity analysis for the model parameters will be presented.

5.2.1 Nomenclature

This section presents the nomenclature used throughout the text.

s	= Vehicle speed (mph)
a	= Vehicle acceleration (mph/sec)
h	= Time headway between successive vehicles (sec)
Site	= The site at which measurements were taken, 1 for the NC-86 Site, 2 for the US-64 Site
Type	= Vehicle Type based upon FHWA categories (i.e. 2 = Type 2 and 3 = Type 3 vehicles)
Model Year	= Model year group of the vehicle (i.e. 1980 and earlier; between 1981 and 1985; between 1986 and 1988; between 1989 and 1990; between 1991 and 1992; between 1993 and 1994; between 1995 and 1996; and between 1997 and 1999)
IND	= Information on fuel induction system of the vehicle (i.e. 1 = fuel injected, -1 = carbureted, and 0 = diesel)

AIR = Parameter giving information on whether vehicle has air injection reactor or not (0 if no, 1 if yes)

ECT = Exhaust control equipment type (i.e. 0 = no catalyst, 1 = oxidation catalyst, -1 = three-way catalyst)

CLL = Parameter giving information whether vehicle has closed loop combustion control or not (0 if no, 1 if yes)

TAC = Parameter giving information whether vehicle has a thermostatic air cleaner or not (0 if no, 1 if yes)

Body = Vehicle body type (i.e. 2S = two-door sedan, 4S = four-door sedan, SW = station wagon, V = van, CN = convertible, CP = coupe, MP = multi-purpose vehicle, TK= pickup truck)

CNTRY = The country where the vehicle was manufactured (e.g., USA, Japan, Germany)

CYL = Cylinder configuration of the vehicle (i.e. V6, V8, L6, L3 etc.)

ASP = Parameter giving information on the aspiration system of the vehicle (i.e. –1 = normal, 0 = turbocharged, 1 = supercharged)

EVP = Parameter giving information whether vehicle has evaporative emissions control equipment or not (0 if no, 1 if yes)

EGR = Parameter giving information whether the vehicle has exhaust gas recirculation system or not (0 if no, 1 if yes)

PCV = Parameter giving information whether the vehicle has positive crankcase ventilation system or not (0 if no, 1 if yes)

DISP = Engine displacement in liters (e.g., 1600, 2000, 2500 etc.)

5.2.2 Resulting Models

Emission estimation models have been developed for four different cases. These cases were based upon different data collection strategies. The only difference between these case studies are the parameters used to develop the regression trees. Before explaining the physical meaning of the different case studies, the parameters used in each of them are given in Table 5.1.

Table 5.1. Parameters Available in Different Case Studies

Parameters	*Case 1*	*Case 2*	*Case 3*	*Case 4*
s	✓	✓	✓	✓
a	✓	✓	✓	
h	✓	✓	✓	✓
Site	✓	✓	✓	✓
Type	✓	✓	✓	✓
Year	✓	✓		
IND	✓			
AIR	✓			
ECT	✓			
CLL	✓			
TAC	✓			
Body	✓			
CNTRY	✓			
CYL	✓			
ASP	✓			
EVP	✓			
EGR	✓			
PCV	✓			

The four cases developed in our study are based upon different assumptions about what type of traffic detector is used for traffic data collection. The first case represents the most extensive collection of data, using advanced video-based traffic detectors, hypothetically directly linked to DMV and VIN data bases for real time

decoding of license plates. The second case assumes that no VIN data are available. The last two cases represent use of conventional magnetic induction loop detectors.

Case 1 represents the conditions where all traffic and vehicle parameters can be collected. For this purpose, a license plate reader is needed to capture the images of the license plates of passing vehicles during data collection. It is assumed that there would be a direct link between license plate reader and DMV database. After obtaining vehicle specific information, VIN number, which is available in DMV database, can be decoded to obtain more information on the vehicle. Traffic parameters would have to be measured with instruments which have the capability of measuring vehicle speed, acceleration, and time headway between successive vehicles. Pneumatic tubes, used in this study, can be used to measure speed and acceleration. Time headway values can be obtained from an area-wide traffic detector such as MOBILIZER ©.

The parameters discussed for Case 1 are similar to parameters used in Case 2. The only difference between these two cases is that in Case 2, detailed vehicle information obtained from VIN number are not available.

Case 3 represents a situation which only a few traffic parameters can be collected. Conventional speed measurement devices such as conventional magnetic induction loop detectors. Vehicle type information can be obtained from area-wide traffic-detectors such as MOBILIZER©.

Case 4 is similar to Case 3 except vehicle acceleration is not included in the Case 4. This case study simulates the condition where traffic parameters are collected using a Loop detector. Loop detectors can measure speed, and give approximate information on vehicle type but can not measure acceleration. For time headway measurements MOBILIZER© can be used.

For each of the case studies, the model development begins with formation of a regression tree using the parameters specific for that case, as indicated in Table 5.1.

The number of data points used for Cases 1 and 2 is 3,262 whereas it is 6,382 for Cases 3 and 4. The difference is due to the fact that for valid emissions when license plates data set were used the available data after matching reduce to 3,262 from 6,382. Errors in the license plate reader, and limitations in matching license plates to North Carolina's DMV database caused these difference in the number of data values. In order to make a comparison among all the case studies, Cases 3 and 4 were modeled using both their original data, having 6,382 data points, and the smaller data set, having 3,262data points. Cases 1 and 2 are developed based upon the smaller data set.

The first step in model building is to form the 'full' tree for all of different case studies. Then the trees were analyzed in order to see which parameters were chosen for the model and how many times they were selected. The results of these analysis are given in Tables 5.2 and 5.3 for CO and HC respectively.

Table 5.2. Summary of the analysis of regression trees for CO

Case	*Total Number of Splits*	*Percent of Splits for Speed*	*Percent of Splits for Acceleration*	*Percent of Splits for Other Variables*
Case 1	182	21	22	Body(16), Year(11)
Case 2	211	34	31	Year(13), Type(11)
Case 3	264 (282)	40	28	Headway(18), Type(14)
Case 4	118 (80)	51	-	Headway(39), Type(10)

Table 5.3. Summary of the analysis of regression trees for HC

Case	*Total Number of Splits*	*Percent of Splits for Speed*	*Percent of Splits for Acceleration*	*Percent of Splits for Other Variables*
Case 1	168	-	-	Year(18), Body(18), Country(18), CYL(15),
Case 2	57	-	-	Year(53), Site(33), Type(14)
Case 3	189 (180)	37	34	Headway(18), Type(11)
Case 4	89 (72)	55	-	Headway(32), Type(9)

The second columns in Tables 5.2 and 5.3 give the total number of splits for the full tree. The values in the parenthesis for Cases 3 and 4 are the total number of splits for the model developed using the smaller data set. The total number of splits vary for each case. In S-Plus™, there are two methods to stop the tree growth. One is to define the minimum number of data points available at the terminal node. This means that whenever a node has data points equal or smaller than the minimum number of data points, the splitting procedure will stop at that node. In our study we set the minimum number of data points to 20 data points. Another method for stopping the tree growth is the node homogeneity principle. This approach is just based upon the idea that splitting will continue till the deviance at the node will be equal to the defined value. This

number is given as percentage of the deviance of the starting node. A default value of 0.01 is given in S-Plus™, indicating that splitting procedure will continue till the given reduction in deviance at the terminal node is 0.01 of the deviance at the starting node. Depending on the node homogeneity principle and the minimum number of data points in the node, different number of number of splits, or terminal nodes, were obtained.

In the third and the fourth columns of Tables 5.2 and 5.3, the percent of splits for speed and acceleration are given. This information is obtained to see how important is the vehicle speed and acceleration in reducing deviance. These numbers might be an indication of the variability that speed and acceleration can explain as part of linear regression. This is true if and only if the relation between speed, acceleration and emissions estimates are linearly related.

The percent of splits caused by speed range from 21 to 51 for CO. The less the number of variables used in the model, the more is the relative importance of speed. A similar trend occurs for acceleration, where the percent of splits that are on this variable range from 22 to 31. In Case 4 acceleration was not used to develop the model.

The percent of splits for speed and acceleration is not given for HC emissions for Cases 1 and 2 in Table 5.3. The reason is that speed and acceleration were not used in tree formation in these case studies. Initially, these parameters were used in regression trees, but it was found out that first two splits were with respect to vehicle speed and vehicle acceleration parameters. Since OLS regressions fitted at the terminal nodes, generally after the second split, were fitted by using vehicle speed, acceleration, and

headway, second time use of these parameters were prevented by not including them in regression tree development.

In the last columns of Tables 5.2 and 5.3, other variables which were used in the regression trees and the percent of splits based upon these variables are given. Vehicle model year and body type are the most important parameters in Cases 1 and 2 for CO and HC. For Cases 3 and 4, the relative importance of the other parameters, including headway and vehicle type, increased substantially, since the number of other possible explanatory variables decreased.

After full tree development and analysis, pruning techniques were applied to get the final form of the tree. Based on the deviance versus node size graphs, the lowest number of nodes giving a comparatively largest reduction in deviance were selected. A sample of a deviance versus node size graph for CO Case 1 is given in Figure 5.6.

The deviance for a given node size can be read directly from the deviance-node size graph. Then this value is divided to the initial deviance value in order to get the percent reduction in deviance for that node size. For example, for node size 20 the reduction in deviance is about 16 percent.

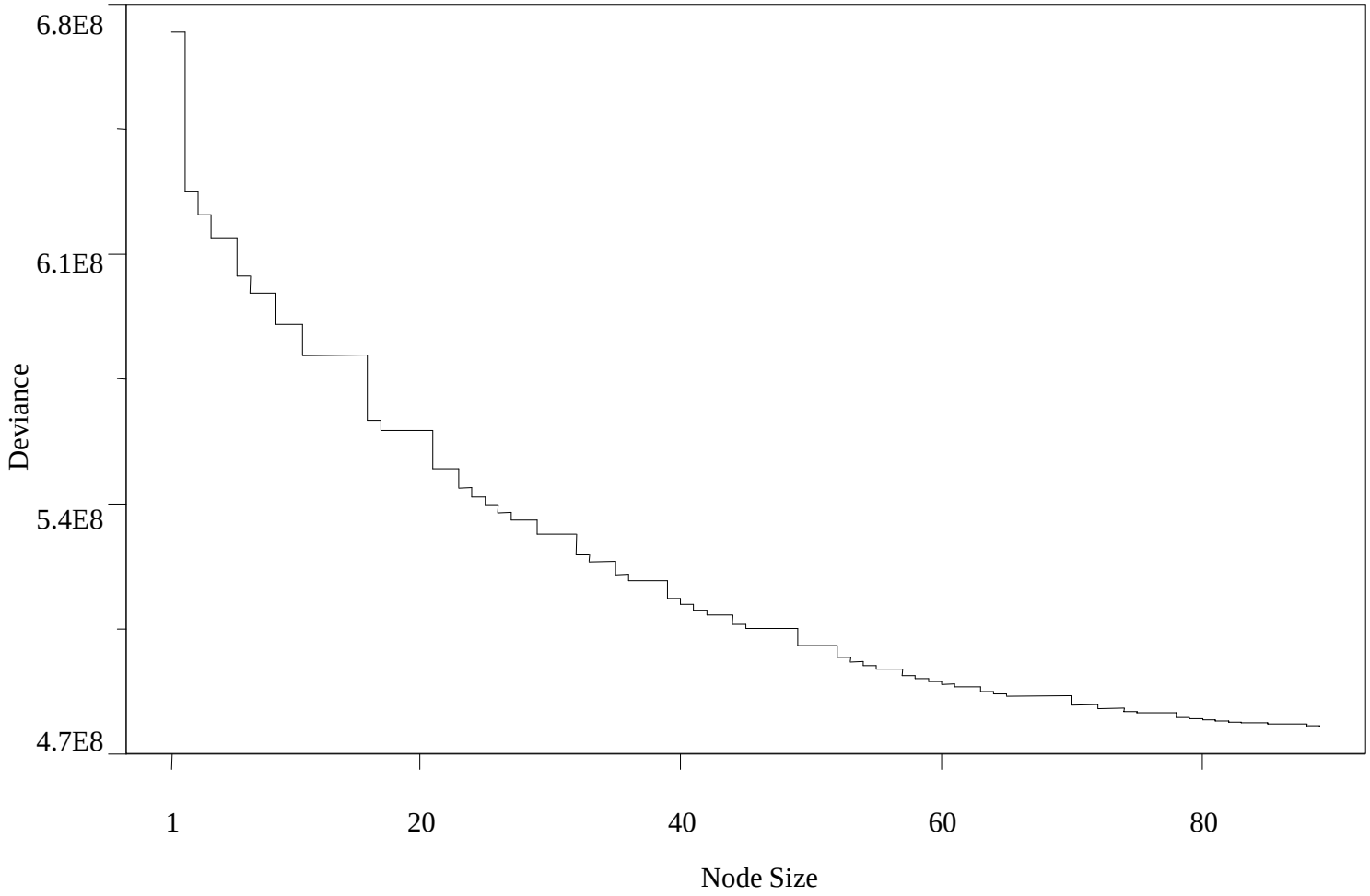


Figure 5.6. Reduction in deviance with respect to node size for CO Case 1.

After deciding on the final form of the tree, linear regressions using traffic parameters were fit to the data at the terminal nodes of the tree. As discussed before, step-wise regression techniques were used in S-Plus™ to choose the parameters. Tables 5.4 and 5.5 provide a summary of the results of the model development activities.

Table 5.4. Summary of the results of the statistical emissions estimate models for CO

Case	*Number of Data Points*	*Total Deviance (10^8)*	*Maximum Reduction in Deviance*	*Reduction in deviance for 20 Nodes*	*Reduction in deviance before regression*	*Reduction in deviance after regression*
Case 1	3262	6.72	29 %	16 %	7 %	12 %
Case 2	3262	6.72	21 %	14 %	7 %	10.6 %
Case 3	6382 (3262)	1.59 (6.72)	6 % (8.9 %)	3 % (4.5 %)	1 % (0.5 %)	1.5 %
Case 4	6382 (3262)	15.9 (6.72)	2 % (1.5 %)	2 % (1.3 %)	0.3 % (0.3 %)	0.3 %

Table 5.5. Summary of the results of the statistical emissions estimate models for HC

Case	*Number of Data Points*	*Total Deviance (10^8)*	*Maximum Reduction in Deviance*	*Reduction in deviance for 20 Nodes*	*Reduction in deviance before regression*	*Reduction in deviance after regression*
Case 1	3262	9.61	15 %	14 %	5 %	9.7 %
Case 2	3262	9.61	4 %	4 %	2 %	8.1 %
Case 3	6382 (3262)	45 (9.61)	6.7 % (10.5 %)	4 % (7.5 %)	1.6 % (2 %)	2.0 %
Case 4	6382 (3262)	45 (9.61)	2 % (1.6 %)	2 % (1.3 %)	0.5 % (0.4 %)	0.6 %

In Table 5.4 and 5.5 the second columns give the number of data used in the model. The third column gives the total deviance at the root node, starting node, of the regression trees. The total deviance is lower when the data sets are smaller, as easily can be seen for Case 3. The value inside the parenthesis is for the smaller data set. The total deviance is a property of the original data sets and for the same initial value for a given pollutant for all case studies using that data set.

Maximum reduction in deviance values are given in the fourth column. These values show how much reduction in deviance was accomplished by using the full tree. For CO data maximum reduction in deviance occurs for Case 1, which has a reduction in deviance of 29 percent. The reduction in deviance decreases from Case 1 to Case 4, because the number of parameters included in the model decreases and relevant parameters were omitted in the latter case studies. When the smaller data set was used for Case 3, the maximum reduction increased approximately 48 percent for CO data since the data are more homogeneous in smaller data set than in the larger one. However, for Case 4, for CO data, maximum reduction decreased for the small data set, from 2 to 1.3 percent. One possible explanation might be that the parameters used in the model were not powerful and decreasing the sample size and making them more homogenous did not effect the maximum reduction in deviance.

Maximum reduction in deviance for HC data follows the same trend as CO data do, having a reduction 15 percent for Case 1, and decreasing to 2 percent for Case 4, for the large data set. Reduction in deviance for Case 4 is 1.6 percent for the smaller data set. It is shown by the Table 5.5 that maximum reduction for Case 2 is lower than maximum reduction in Case 3. This is due to the fact that speed and acceleration parameters were not included in regression tree development for Case 2. The result suggest that vehicle speed and acceleration are important variables that should be considered in the model.

In the fifth column the reduction in deviance for 20 nodes are given. The aim of using 20 nodes here is to compare the regression trees for the same number of nodes.

For CO data, maximum reduction in deviance for 20 nodes occurs for Case 1 having a value of 16 percent. This values is lower for the other cases. The reduction in deviance for 20 nodes for Case 4 is reported as 1.5 percent. Similar, trend is observed for HC data, except Case 3 has higher reduction in deviance than Case 2, since vehicle speed and acceleration parameters were not included in Case 2.

Reduction in deviance before linear regression is given in the sixth columns. These values show the reduction in deviance in the final 'trimmed' tree before fitting the linear regressions. There is a substantial decrease in the deviance reduction when compared to deviance reduction for the full tree. The final tree is chosen based upon the fact that it would include the most important splits, causing a major reduction in deviance, but not include the noise in the regression tree. Our approach in developing this combined model was to stratify the data into smaller data sets by using regression tree, and then fit OLS regression to capture the relationships within the data strata. For this purpose we do not need the full tree. The final trees chosen for the model included at most 5 terminal nodes.

After deciding on the final tree, linear regressions were fitted to the data at the terminal nodes of the tree. Step-wise linear regression was used to select the most relevant parameters among several of them. These parameters include; speed; acceleration; $speed^2$; $acceleration^2$; speed×acceleration; $speed^2$×acceleration; speed×$acceleration^2$; $speed^2$×$acceleration^2$;headway; and headway×speed. These parameters were selected because analysis of full trees, given in Tables 5.2 and 5.3,

showed that vehicle speed and acceleration were among the most important parameters. The multiplicative terms of speed, acceleration and their second order terms are chosen to capture any interaction effect between speed and acceleration. In a study conducted by Washington *et al.* (1997) it was found that high emitting vehicles are sensitive to speed×acceleration, which is defined as kinetic energy, whereas normal emitting vehicles are sensitive to speed^2×acceleration, defined as the change in power. The term headway×speed was included to observe any interactions between speed and headway values.

After running the stepwise regression in S-Plus™, most relevant parameters were identified. Then regressions were fitted to the data set using these parameters. Resulting linear regression equations are given in Figures 5.7 to 5.10 for every regression tree model.

The last columns in Tables 5.4 and 5.5 give reduction in deviance after regressions were fitted. In order to calculate the deviance left after linear regressions were fitted to the data, error sums of squares (SSE) were used. As discussed before deviance is exactly the same as SSE. So, SSE for the regression equation indicates the deviance left at that node. A similar approach was used by Washington (1999) to calculate residual error for a combined model of HTBR and OLS regression.

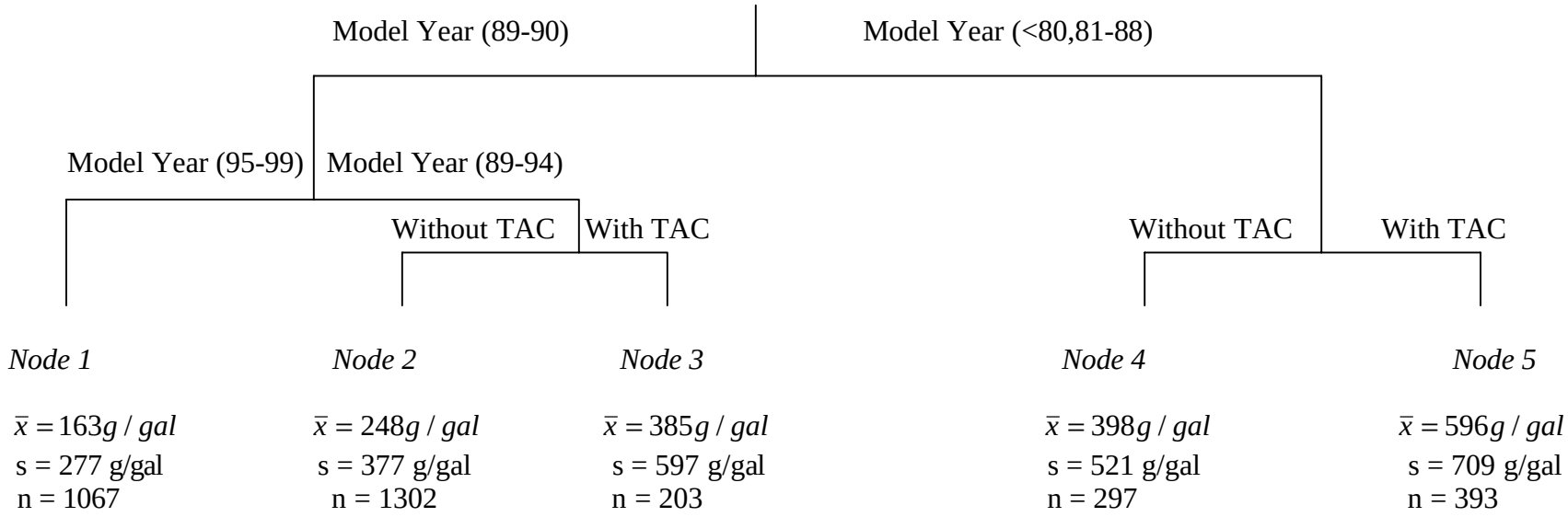
The sum of the deviances in all nodes divided by the total deviance before linear regressions give the percent reduction in deviance due to linear regressions. The

sum of this value with the reduction in deviance due to regression tree will give the total reduction in deviance due to combined model.

As shown in Table 5.4, for Case 1 the combined model reduce the deviance 12 percent. For Case 2, the reduction is about 11 percent. As the relevant parameters are omitted from the model the reduction in deviance decreases substantially. About 1.5 percent reduction occurs for Case 3 and 0.3 for Case 4.

The reduction in deviance for combined model Case 1 for HC is about 10 percent. For Case 2, this value is 8 percent. Case 3 has reduction of 2 and Case 4 has 0.6 percent.

The percent reduction in deviance after linear regressions are not reported in Tables 5.4 and 5.5 for cases 3 and 4 for the small dataset. The reason is that percent reduction in deviance for large data set after regressions were so low that it was decided not to develop models for these two cases. It was decided that the capability of cases 3 and 4 are limited and would not help to estimate emissions. The final model includes the first two cases for CO and HC which are given in Figure 5.7 through Figure 5.10 below. The regression trees with the linear regressions fitted at the terminal nodes are presented.



Linear Regressions Fitted at the Terminal Nodes:

Node 1; $EF_{CO11} = \exp\left(4.95 - 7.80\times10^{-3}\times s - 2.61\times10^{-1}\times a + 4.60\times10^{-3}\times s\times a + 6.30\times10^{-3}\times a^2\right)$ $R^2 = 0.03$

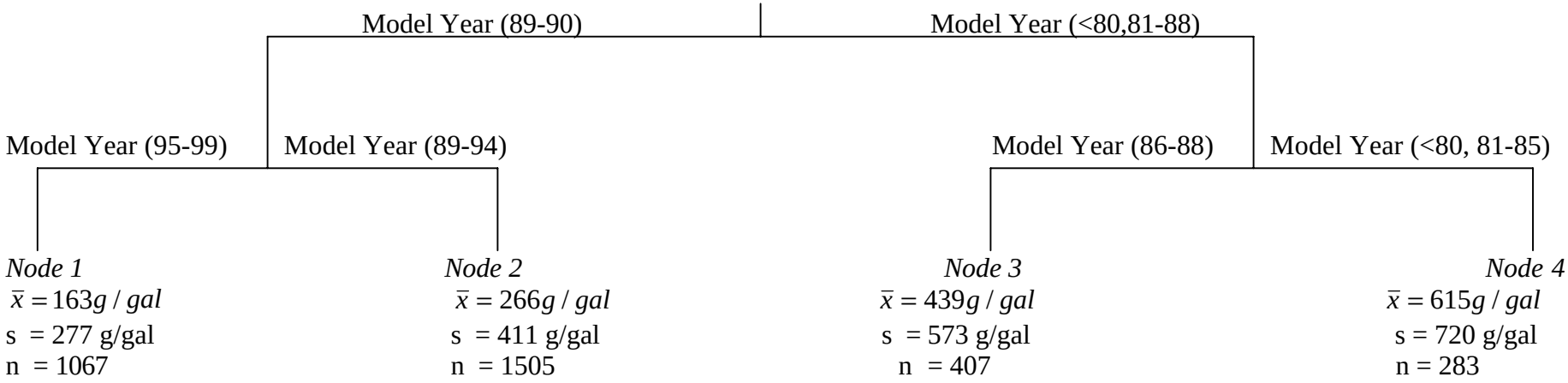
Node 2; $EF_{CO12} = \exp\left(6.16 - 2.78\times10^{-2}\times s - 1.01\times10^{-1}\times a + 2.00\times10^{-3}\times s\times a - 4.38\times10^{-2}\times a^2 + 1.10\times10^{-3}\times s\times a^2\right)$ $R^2 = 0.03$

Node 3; $EF_{CO13} = \exp\left(6.80 - 3.03\times10^{-2}\times s - 1.26\times a + 2.58\times10^{-2}\times s\times a + 1.83\times10^{-1}\times a^2 - 4.20\times10^{-3}\times s\times a^2\right)$ $R^2 = 0.11$

Node 4; $EF_{CO14} = \exp\left(6.51 - 2.58\times10^{-2}\times s - 1.71\times10^{-1}\times a + 1.69\times10^{-2}\times a^2\right)$ $R^2 = 0.06$

Node 5; $EF_{CO15} = \exp\left(6.38 - 1.57\times10^{-2}\times s - 5.54\times10^{-2}\times h + 1.30\times10^{-3}\times s\times h\right)$ $R^2 = 0.02$

Figure 5.7. Emissions Estimation Model for CO Case 1.



Linear Regressions Fitted at the Terminal Nodes:

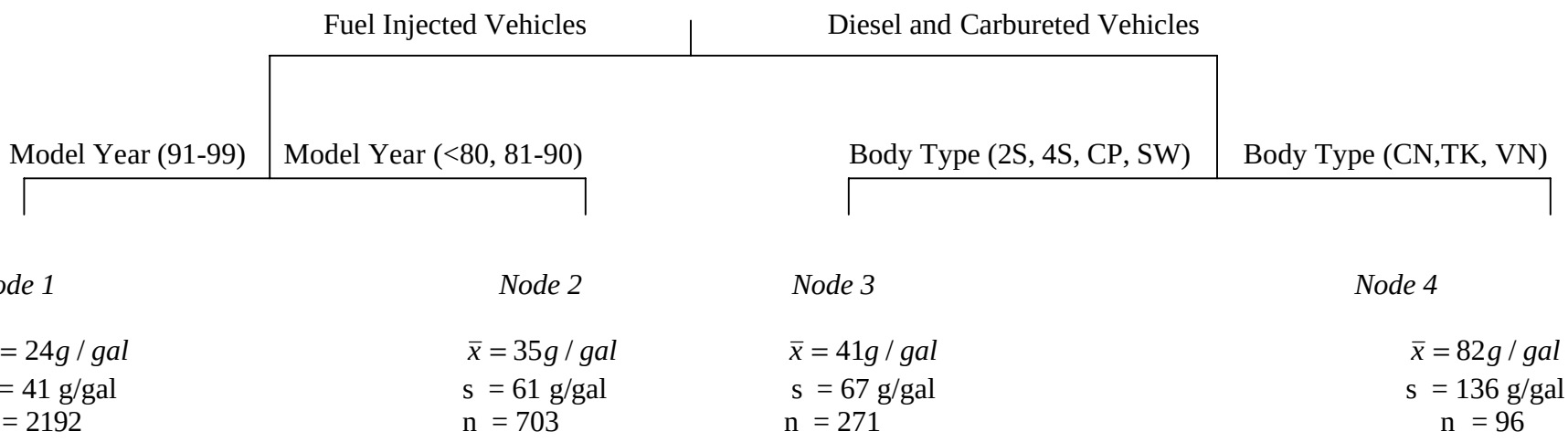
Node 1; $EF_{CO11} = \exp\left(4.95 - 7.80\times10^{-3}\times s - 2.61\times10^{-1}\times a + 4.60\times10^{-3}\times s\times a + 6.30\times10^{-3}\times a^2\right)$ $R^2 = 0.03$

Node 2; $EF_{CO22} = \exp\left(6.25 - 2.78\times10^{-2}\times s - 3.27\times10^{-1}\times a + 6.20\times10^{-3}\times s\times a - 5.00\times10^{-3}\times a^2\right)$ $R^2 = 0.02$

Node 3; $EF_{CO23} = \exp\left(10.0 - 2.01\times10^{-1}\times s - 5.75\times10^{-1}\times a + 1.10\times10^{-2}\times s\times a + 2.10\times10^{-3}\times s^2 + 3.10\times10^{-2}\times a^2 - 1.30\times10^{-3}\times a^3\right)$ $R^2 = 0.04$

Node 4; $EF_{CO24} = \exp\left(5.82 - 6.63\times10^{-2}\times a\right)$ $R^2 = 0.02$

Figure 5.8. Emissions Estimation Model for CO Case 2.



Linear Regressions Fitted at the Terminal Nodes:

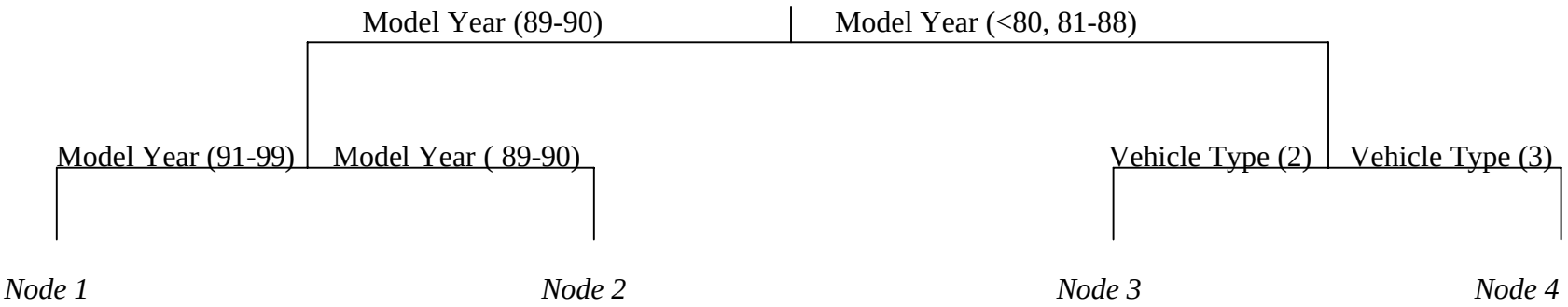
Node 1; $EF_{HC11} = \exp\left(3.18 - 9.60\times10^{-3}\times s - 5.74\times10^{-2}\times a + 2.70\times10^{-3}\times a^2\right)$ $R^2 = 0.02$

Node 2; $EF_{HC12} = \exp\left(3.93 - 2.06\times10^{-2}\times s - 1.01\times10^{-1}\times a + 7.60\times10^{-3}\times a^2\right)$ $R^2 = 0.04$

Node 3; $EF_{HC13} = \exp\left(3.70 - 1.11\times10^{-2}\times s + 3.34\times10^{-1}\times a - 7.80\times10^{-3}\times s\times a - 9.45\times10^{-2}\times a^2 - 2.20\times10^{-3}\times s\times a^2\right)$ $R^2 = 0.04$

Node 4; $EF_{HC14} = \exp\left(5.55 - 4.56\times10^{-2}\times s + 5.93\times10^{-1}\times a - 1.43\times10^{-2}\times s\times a - 2.12\times10^{-2}\times a^2 + 5.10\times10^{-3}\times s\times a^2\right)$ $R^2 = 0.11$

Figure 5.9. Emissions Estimation Model for HC Case 1.



Linear Regressions Fitted at the Terminal Nodes:

Node 1; $EF_{HC21} = \exp\left(3.16 - 1.01\times10^{-2}\times s - 5.75\times10^{-2}\times a + 3.50\times10^{-3}\times h + 2.50\times10^{-3}\times a^2\right)$ $R^2 = 0.02$

Node 2; $EF_{HC22} = \exp\left(3.75 - 1.76\times10^{-2}\times s - 5.51\times10^{-2}\times a\right)$ $R^2 = 0.03$

Node 3; $EF_{HC23} = \exp\left(3.88 - 1.66\times10^{-2}\times s + 2.05\times10^{-1}\times a - 6.50\times10^{-3}\times s\times a - 6.15\times10^{-2}\times a^2 + 1.60\times10^{-3}\times s\times a^2\right)$ $R^2 = 0.04$

Node 4; $EF_{HC24} = \exp\left(5.02 - 3.86\times10^{-2}\times s + 2.86\times10^{-1}\times a - 6.60\times10^{-3}\times s\times a - 1.16\times10^{-1}\times a^2 + 2.70\times10^{-3}\times s\times a^2\right)$ $R^2 = 0.06$

Figure 5.10. Emissions Estimation Model for HC Case 2.

After developing the regression equations, a sample calculation was estimated to compare the results of the "flow-based" and "vehicle-based" models. It was found that when the regression result is back transformed to estimate grams/gallon from natural logarithm scale, results were biased. This problem of back-transformation is not uncommon in environmental field, where log transformations are used frequently, as reported by Gilbert (1987). Our regression model was in the form of:

$$\ln CO = l\text{inear model} + \varepsilon \tag{5-5}$$

where ε is the residual error term for the linear regression equation which has a mean of zero and variance of σ^2. Zero mean and constant variance for the residual terms are the key assumptions for the OLS regression to be valid. However, when we back transform the equation to calculate CO in terms of grams/gallon from natural logarithm scale, the residual term is also transformed as shown in Equation 5-6 below:

$$CO = \exp(\text{linear model}) \times \exp(\varepsilon) \tag{5-6}$$

If we call exp(ε) as ε', then the new residual termε' is biased and no longer have zero mean but have mean of exp(0), equal to 1, and a different variance σ'^2. This violates the basic assumption of OLS regression. In order to solve this problem several methods have been proposed including modifying the regression methodology so that when it is back-transformed the results would not be biased (Heien, 1968; Bradu and Mundlak, 1970) Another approach to solve this problem is by developing a correction factor for the regression equation. To do that the individual residual terms were back transformed from natural logarithm scale to grams/gallon and their average was calculated. The average value of the transformed residuals is the correction factor for that regression equation for calculating average emissions estimate. The average of emissions estimates from linear

regression should be multiplied by that correction factor. Table 5.6 gives the correction factors and unbiased average emissions estimated by using correction factors. For comparison purposes actual average values are also given.

Table 5.6. Correction Factor for Regression Equations

Model	Mean of Residuals ln(g/gal)	Average of Actual Data (g/gal)	Estimated Mean (g/gal)	Residual Mean (g/gal)	Corrected Estimated Mean (g/gal)
EF_{CO11}	0.00	162.97	95.35	1.71	163.05
EF_{CO12}	0.00	247.63	136.73	1.80	246.11
EF_{CO13}	0.00	381.43	208.86	1.84	384.30
EF_{CO14}	0.00	400.34	213.93	1.89	404.33
EF_{CO15}	0.00	596.69	296.07	2.02	598.06
EF_{CO21}	0.00	162.97	95.35	1.71	163.05
EF_{CO22}	0.00	266.27	143.46	1.85	265.40
EF_{CO23}	0.00	439.24	226.55	1.95	441.77
EF_{CO24}	0.00	615.14	311.80	1.99	620.48
EF_{HC11}	0.00	24.42	14.88	1.61	24.01
EF_{HC12}	0.00	35.06	19.74	1.75	34.59
EF_{HC13}	0.00	40.85	24.73	1.67	41.32
EF_{HC14}	0.00	82.29	40.83	1.96	79.93
EF_{HC21}	0.00	24.99	14.92	1.65	24.58
EF_{HC22}	0.00	31.08	18.08	1.71	30.97
EF_{HC23}	0.00	39.66	23.26	1.71	39.77
EF_{HC24}	0.00	54.89	28.85	1.83	52.83

The second column in Table 5.6 gives the mean of residuals for natural logarithm regression equation. As expected all the values are zero. In the third column average of the actual data used to develop the regression model are given. In the fourth column the result of the regression equation, after back transformed to grams/gallon values, are given. As noticed, there are huge differences between the actual average emission value

and the one predicted by the regression equation. For example, for model EF_{CO23}, the observed mean CO value is 439.24 grams/gallon whereas the predicted value from the regression equation 226.55 grams/gallon. Predicted value is approximately half of the observed value.

In the fifth column the correction factors developed for each regression equation is given. The last column gives the corrected estimated mean for the regression equations, predicted emission value from regression equation times the correction factor. For example for EF_{CO23} the estimated correction factor is 1.95. The corrected mean CO emission is :

$$EFCO_{(corrected)\,i,i} = EFCO_{(predicted)\,i,i} \times \text{Correction Factor}_{i,i} \tag{5-7}$$

$$EFCO_{(predicted)2,3} = 226.55\ \frac{\text{grams CO}}{\text{gallon}} \tag{5-8}$$

$$\text{Correction Factor}_{2,3} = 1.95 \tag{5-9}$$

$$EFCO_{(corrected)\,i,i} = 226.55\,\frac{\text{grams CO}}{\text{gallon}} \times 1.95 = 441.77\,\frac{\text{grams CO}}{\text{gallon}} \tag{5-10}$$

The corrected mean emission values and the observed values are very close to each other. The differences are less than 1 percent. In order to summarize the information for the regression equation Table 5.7 is given below:

Table 5.7. Summary of the regression equations

Model	Number of data	Standard Error	R^2	Correction Factor
EF_{CO11}	1065	0.948	0.01	1.71
EF_{CO12}	1302	1.036	0.03	1.80
EF_{CO13}	203	1.089	0.11	1.84
EF_{CO14}	294	1.161	0.06	1.89
EF_{CO15}	394	1.266	0.02	2.02
EF_{CO21}	1065	0.948	0.01	1.71
EF_{CO22}	1504	1.058	0.02	1.85
EF_{CO23}	407	1.201	0.04	1.95
EF_{CO24}	282	1.252	0.02	1.99
EF_{HC11}	2188	0.975	0.02	1.61
EF_{HC12}	700	1.019	0.04	1.75
EF_{HC13}	272	0.990	0.03	1.67
EF_{HC14}	95	1.181	0.11	1.96
EF_{HC21}	2081	0.983	0.02	1.65
EF_{HC22}	339	0.999	0.03	1.71
EF_{HC23}	423	1.009	0.04	1.71
EF_{HC24}	244	1.090	0.06	1.83

The second column in Table 5.6 gives the number of data used to develop the regression equation. These numbers are slightly lower than the number of data at the terminal nodes of the trees because for some of the data points speed, or acceleration or headway data are not available. Third column gives the standard error for linear regressions which were calculated by S-Plus™ 4.0.

In the fourth column coefficient of multiple determination, R^2 values were reported for the regression equations R^2 is a measure of the proportionate reduction of total variation in dependent variable with the use of independent variables. For our regression equations R^2 values are quite low ranging from 2 to 11 percent. As stated by

Neter et. al., (1996) one should not depend only on R^2 values to decide whether the regression is a good model or not. Our R^2 values are low but all the parameters used in these regression equations have p-values less than 0.05 indicating that at a significance level of 0.05, we reject the hypothesis that parameter is equal to zero. As a result linear regression equations are significant although they can explain small portion of the variability in emissions estimates. It should be also reminded that in our model some portion of the variability is also explained with the regression tree. Total variability explained, or reduction in deviance which has similar meaning, for the combined models are given in Tables 5.4 and 5.5.

The models presented in our study can be used in the following manner.

(1) Select the regression tree to be used both for CO and HC depending on the purpose and type of information available.

(2) Use "vehicle-based" or "flow-based" model depending on availability of vehicle specific speed, acceleration, and headway information.

(3) If "vehicle-based" model is selected, for each vehicle corresponding regression equation should be selected. This can be done by putting all these models in an Excel spreadsheet and write a function to choose the correct path in the regression tree based upon vehicle parameters; Model Year; TAC; IND; Body; and Type. Then regression equation at the end node of the tree should be used to calculate the emissions estimates in grams/gallon. Emissions estimates for all vehicles are then averaged to calculate the average vehicle fleet emission estimate. The result should be corrected by multiplying with the correction factors given in Table 5.6. Then the average value, in grams/mile,

can be estimated by multiplying grams/gallon value with the average fuel efficiency value.

(4) If “flow-based” model is selected, the percentage of the vehicles for different levels of the tree parameters should be estimated same. Then same approach that discussed in step 3 should be used to select the right node for each percentage of the vehicle fleet. Then the average emission estimate given at that node should be used as an average emission estimate for particular percentage of the fleet. After estimating the average emission value for all the fleet an overall weighted average should be calculated. This value gives the average emission estimate in grams/gallon for the whole fleet. An average fuel efficiency can be used to convert this value to grams/mi of average emission for the fleet.

In the next section, sensitivity analysis will be presented for the linear regression parameters for each case study.

5.2.3 Sensitivity Analysis for Model Parameters

Sensitivity analysis is used to measure the potential importance of inputs on the model outputs. Several methods can be used to accomplish this objective. Some of these techniques are; examining coefficients of variation; scatter plots; correlation coefficients; multivariate regression; contribution to variance; and probabilistic methods. Brief

information on these techniques and additional references are given in Cullen and Frey (1999).

In our study we used scatter plot technique to visually observe the effect of input parameters on the model output. For this purpose emissions estimates were calculated for the range of model parameters using the linear regressions. The range of input parameters were selected based upon the input data which were used to develop the regression equations. This was important to prevent extrapolations. There is also the possibility of hidden extrapolation, which occurs when the input values which are not used to develop the model, even though all of the values fall within the range, are used for prediction. A thorough analysis was made to prevent hidden extrapolation and the actual range of model parameters were identified. Then the scatter plots of emissions estimates versus model parameters were prepared.

A total of 18 scatter plots prepared for sensitivity analysis. Some of the scatter plots are presented and discussed in this section and the rest of the plots are given in Appendix F.

Generally emissions estimates were plotted against the vehicle speed values at different acceleration levels since these are the only parameters for many of the models. For the models, such as model EF_{CO15}, having time headway as an input variable, emissions estimates were plotted against speed at two different time headway values, 2 and 6 seconds. These time headway values were chosen to reflect the vehicle being in platoon or free-flowing, respectively.

For the models having vehicle speed, acceleration, and time headway as input parameters, such model EF_{HC21}, two graphs were prepared for time headway values where emissions estimates were plotted against vehicle speed at different acceleration levels. Finally, for the models having only acceleration as the input parameter, emissions estimates were plotted against the acceleration values.

In Figure 5.11, scatter plot of CO emissions estimates versus vehicle speed values were given at different acceleration values for the model EF_{CO13}.

CO emissions estimates range from about 100 to 1,000 grams/gallon. Speed values vary for each acceleration level but roughly in the range of 30 to 60 mph. The highest emissions estimate of 1,000 grams/gallon occurs for deceleration of 2 mph/sec at a speed value of 35 mph and decreases linearly to about 100 grams/gallon at a speed value of 55 mph. As the acceleration value increases the emissions estimates decreases for lower speeds. For higher speeds we see that highest emissions occur for acceleration value of 2 mph/sec. For acceleration of 5 mph/sec the emissions estimates are stable at approximately 150 grams/gallon.

Similarly scatter plot of emissions estimates versus vehicle speed values were prepared at two different time headway values for the model EF_{CO15} as shown in Figure 5.12.

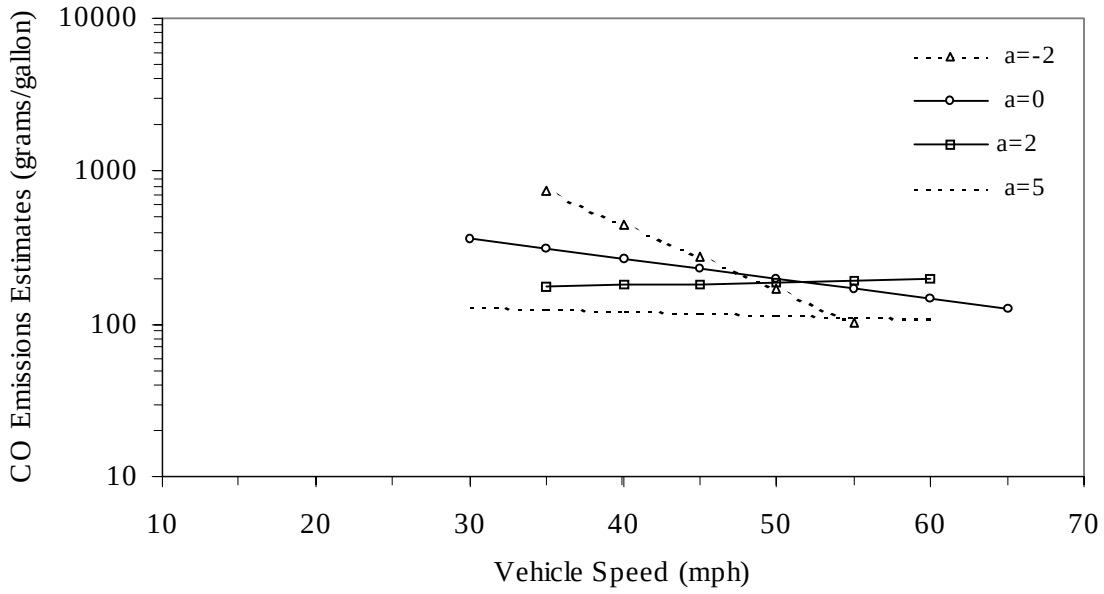


Figure 5.11. Sensitivity analysis for the model EF_{CO13}.

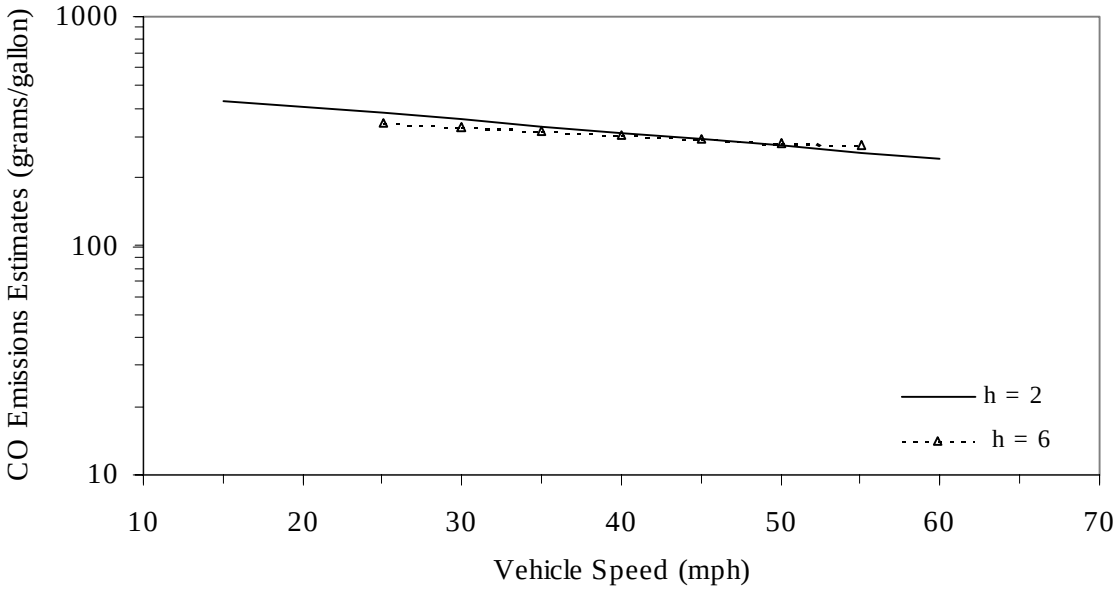


Figure 5.12. Sensitivity analysis for the model EF_{CO15}.

CO emissions estimates vary from 250 to 450 grams/gallon. As vehicle speed increases emissions estimates also increase. Emissions estimates for different time headway values are nearly the same as seen in Figure 5.12. This plot suggest that compared to vehicle speed, time headway does not have an important effect on the model output. Similar results were obtained for the model EF_{HC21} where emissions estimates were plotted against speed at different acceleration levels for two different time headway values.

Emissions estimates for CO were plotted for the model EF_{CO23} in Figure 5.13. Similar to other plots highest emissions estimates occur for high deceleration values and as acceleration increases, or deceleration decreases, emissions estimates increase for a given vehicle speed. There are some differences in this graph. First one is that the plots are not linear but more close to a curve which is due to the fact that this model includes second order speed and acceleration terms. Another difference is that highest emissions estimates occur at the highest acceleration value, 5 mph/sec, for high speed values. Similar results were obtained for EF_{CO12}, where emissions estimates linearly increase for an acceleration value of 5 mph/sec.

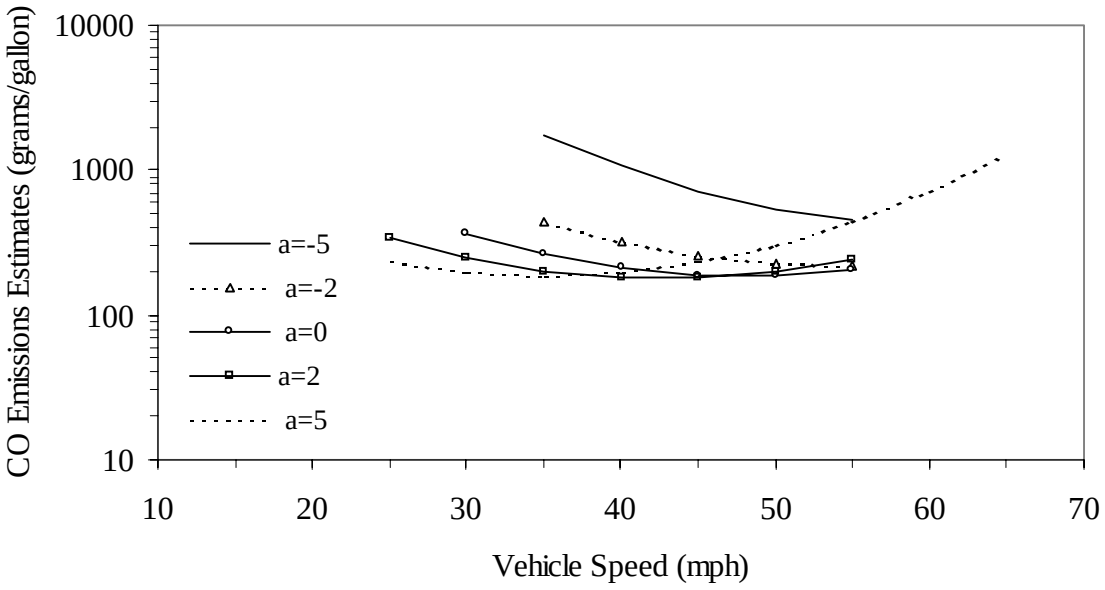


Figure 5.13. Sensitivity analysis for the model EF_{CO23}.

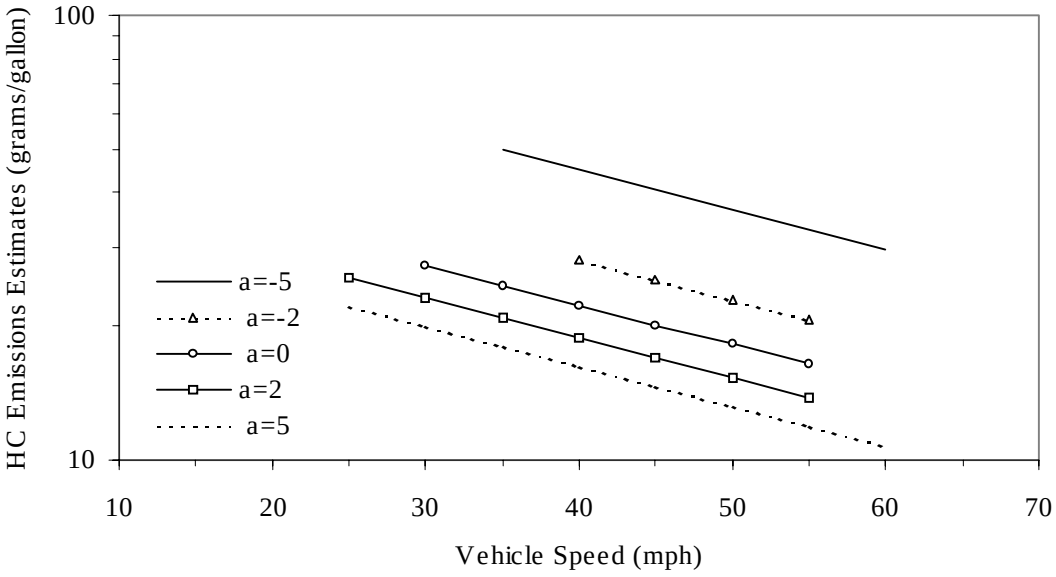


Figure 5.14. Sensitivity analysis for the model EF_{HC12}.

In Figure 5.14 above, HC emissions estimates for the model EF_{HC12} were plotted against speed values for different acceleration values. Maximum emissions estimates occur at high acceleration values for low speeds. As acceleration decreases from 2 mph/sec to 0 emissions estimates also decrease. However, for deceleration of 2 mph/sec emissions estimates were higher . Emissions estimates for this acceleration level increase as speed increases.

A different trend occurs in scatter plot of emissions estimates versus speed values for the model EF_{HC22}. The plot is given in figure 5.15. The lines showing the emissions estimates for different acceleration values are parallel to each other . Maximum emissions estimates occur for an acceleration level of -2 mph/sec and decrease as acceleration increases. Emissions estimates decrease as speed increases for all acceleration levels.

The sensitivity analysis based upon the scatter plots discussed in this section suggests that input parameters vehicle speed acceleration has an observable effect on the emissions estimates whereas outputs are not very sensitive to time headway parameter. Another important finding is that relationship between output and input variables is not constant and changes for each linear regression model.

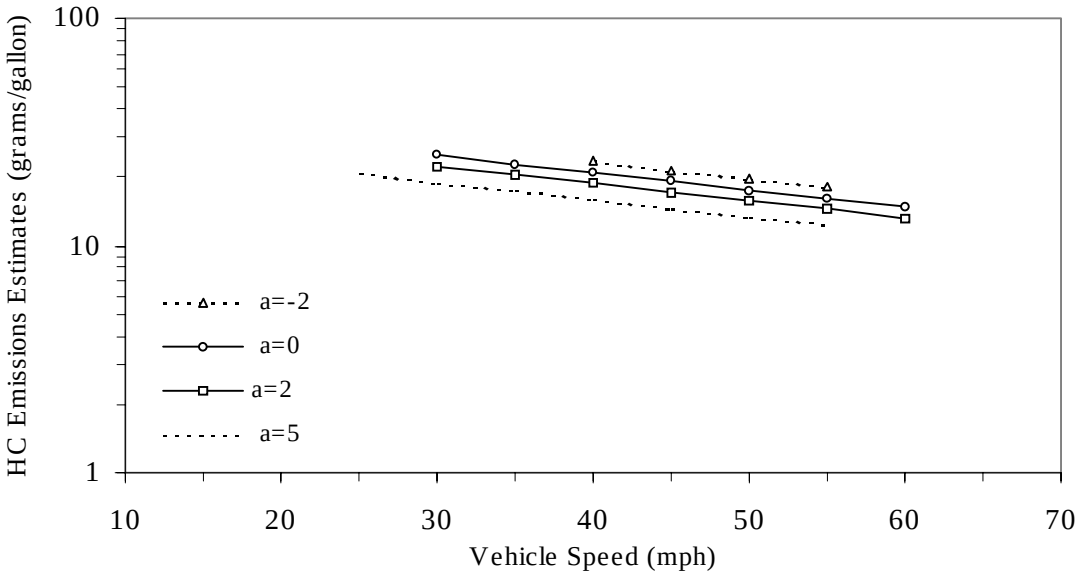


Figure 5.15. Sensitivity analysis for the model EF_{HC22}.

5.2.4 Comparison of Model Results with MOBILE5a

We compared our model results with MOBILE5a results. We can not compare "vehicle-based" model with MOBILE5a since MOBILE5a estimates vehicle emissions for average speed values, information on individual vehicle speed are not used. For this reason, a comparison between "flow-based" model and MOBILE5a has been done.

The first step is the input preparation for the MOBILE5a model. Input variables for MOBILE5a are: average speed; ambient temperature; operating mode fractions; vehicle miles traveled mix; annual mileage accumulation rate and registration distributions by age; basic exhaust emission rates; I/M programs; air-conditioning usage, extra loading, trailer towing, and humidity corrections; tampering rate; anti-tampering

program (ATP); refueling emission; local area parameters such as fuel volatility, oxygenate content; and trip length distribution (USEPA, 1994).

For our calculations we did not enter all these input parameters. Parameters other than quantifiable for our model are kept as default values in MOBILE5a model. The input parameters we entered to the MOBILE5a model are: average speed; ambient temperature; registration distribution by age.

Average speeds were calculated for each regression node, but they were very close to each other and the average speed of the all data set. For example, average speed of data points at model CO14 is 43.13 mph whereas the average speed of all the data points is 44.4 mph. Since the difference between the average speed is so small the average speed of the whole data set, 44.4 mph, was entered to MOBILE5a model. For ambient temperature, average values of observed daytime highest and lowest temperatures were used. Maximum temperature was 74 ^{0}F, and minimum temperature was 51 ^{0}F. There were daily differences in ambient temperature but since the data at the terminal node of regression tree are a combination of all the data collected this assumption does not seem to be bad.

Registration distribution of the vehicles by age were identified for all the regression models and then entered into the MOBILE5a model.

For scenario section, several other variables are needed: region (e.g., low altitude region was selected); calendar year of evaluation (e.g., 1999); average speed (e.g., 44.4

mph); average ambient air temperature (e.g., 60 ^{0}F); operating mode fractions (i.e. percentage of vehicle operation under cold-start, hot-start, and hot-stabilized conditions); and month of evaluation (e.g., July). For operating mode fractions default values were used since no measurements were made regarding those parameters. An example input file from MOBILE5a is given in Appendix C.

The results of MOBILE5a runs are reported with the results obtained from our "flow-based" model in Tables 5.8 and 5.9.

Table 5.8. Summary of the comparison between MOBILE5a and "Flow-Based" Model for CO emissions

Node	MOBILE5a Results (grams/gallon)	"Flow-Based" Model Results (grams/gallon)
CO_{11}	80	163
CO_{12}	183	248
CO_{13}	207	385
CO_{14}	653	398
CO_{15}	668	596
CO_{21}	80	163
CO_{22}	187	266
CO_{23}	477	439
CO_{24}	1021	615

As shown in Table 5.8, mean CO emissions estimated by our model, "flow-based" model, are close to the values predicted by MOBILE5a model. For some of the data sets, such as CO_{11}, the mean emissions predicted by "flow-based" model is twice the values predicted by MOBILE5a. However, for other data sets, such as CO_{24}, the mean emissions predicted by MOBILE5a is 1.6 times higher than the one predicted by "flow-based" model.

Table 5.9. Summary of the comparison between MOBILE5a and "Flow-Based" Model for HC emissions

Node	MOBILE5a Results (grams/gallon)	"Flow-Based" Model Results (grams/gallon)
HC_{11}	NE[a]	24
HC_{12}	NE	35
HC_{13}	NE	41
HC_{14}	NE	82
HC_{21}	10	25
HC_{22}	35	32
HC_{23}	55	39
HC_{24}	62	55

[a]NE: Not Estimated

Table 5.9 summarizes the mean HC emissions results estimated by MOBILE5a and "flow-based" model. The first four data points could not be estimated from MOBILE5a model since these data could not be represented in MOBILE5a input. The first regression tree for HC data was first split according to the vehicle fuel induction systems. It is not possible to represent this kind of information in MOBILE5a.

Results obtained, both by MOBILE5a and the "flow-based" model, for the second HC regression tree showed that mean HC emissions predicted are similar. However, it should be pointed out that HC emissions measured by RSD are based upon propane and does not account for all types of HC emitted from vehicle exhaust. The amount of bias is not well known but might be as much as 50 percent or more lower than the actual total HC emissions. This suggests that values predicted by "flow-based" model are approximately twice the values predicted by MOBILE5a.

Possible reason for differences in mean emissions might be due to the fact that MOBILE5a model is based upon testing vehicles under laboratory conditions and does not consider on-road emissions of the vehicles. However, this approach has been criticized since depending on the engine load, emissions might be very different. As stated by USEPA (1993b), it was found that HC and CO emissions increased by almost 20-100 times during fuel enrichment conditions, which might be due to high accelerations, high speeds, positive road grades, air conditioning operation or any combination of any of these items.

6.0 CONCLUSIONS AND RECOMMENDATIONS

As noted in the introduction section, our main objective is to answer the following questions:

(1) What are the mean CO and HC emission factors for an average vehicle fleet in the Research Triangle area of North Carolina?
(2) What are the traffic and vehicle characteristics of those vehicle fleets?
(3) Do site and day of data collection have an affect on mean emission estimates?
(4) Are there any variables that can explain variability in vehicle emissions?
(5) Can we develop a statistical model which can predict vehicle emissions based upon measurable traffic variables and vehicle parameters?

Throughout the study, data were collected and analyzed to answer these questions.

Emissions and traffic data were collected over 10 days of field work at two sites: the junction of State Highway NC-86 and Interstate I-40, which is referred to as Site 1, and the junction of US-64 and US-1, which is referred to as Site 2. A total of 11,830 triggered emissions measurement attempts were recorded.

6.1 Results of Data Collection and Database Development Activities

Preliminary analysis of the emissions data showed that there were negative CO/CO_2 and HC/CO_2 molar ratios reported by the RSD. Detailed investigation of RSD data processing steps indicate that the regression method used by the on-board computer to calculate CO/CO_2 and HC/CO_2 molar ratios is based on a "floating-intercept" approach in which the intercept of the linear regression is allowed to vary. With this method, there is a probability that negative CO/CO_2 and HC/CO_2 molar ratios might be estimated. A different regression method was considered in which the intercept was forced to be zero for each regression analysis. However, with the "zero-intercept" method, systematic drift occurred when background readings were taken manually. Furthermore, the zero-intercept method requires that the intercept be accurately known to be equal to zero. This is not possible because the RSD has a detection limit which prevents estimating the zero value accurately. For these reasons, it was decided to use the floating-intercept approach which is employed by RSD on-board computer. Our conclusion is that the statistical methods employed by the RSD are reasonable.

The "floating-intercept" method for calculating CO/CO_2 and HC/CO_2 molar ratios can produce negative values of these ratios. Some of the results obtained in this fashion may not be statistically significantly different than zero or a small positive value. In other cases, the results might be statistically significantly different from zero. We conclude that not all negative ratios represent invalid measurements. However, it is not possible to determine, simply by inspection, which of the negative ratios might reflect valid measurements that are not significantly different from a small positive value or zero

versus those negative ratios which are the result of an invalid attempt at a measurement. In particular, the on-board computer of RSD does not report standard errors or t-ratios for each slope. Therefore, to avoid introducing into the data set any invalid measurement attempts, all negative ratios were excluded from the data set for the time being. It is recommended that the negative ratios be re-evaluated at a later time to determine which ones are not significantly different from zero; these could be retained in the database in the future.

To evaluate the implications of excluding negative ratios from the data base in cases where some may be not significantly different than zero, some sensitivity analyses on the data base were performed. Three alternatives were considered: (1) calculating the mean emissions estimates based only on the valid non-negative emissions measurements; (2) calculating the mean emissions for dataset where negative values were assigned a value of zero and combined with the non-negative measurements; and (3) calculating the mean emissions estimate for dataset where negative values were assigned a nominal detection limit value, 8 g/gal for CO and 2 g/gal for HC, and combined with the non-negative measurements. The comparison of the mean estimates from three approaches suggested that the mean emission rate may be over-estimated by as much as approximately 20 percent for CO and HC. However, since not all of the negative emission ratios represent valid measurements, the actual bias on the mean values associated with discarding all negative emissions values is likely to be less than is indicated here. We conclude that the amount of possible bias in our average emission estimates is likely to be within the range of the precision of our mean estimates and, hence does not significantly influence our results.

We analyzed the fleet characteristics of vehicles having negative emissions measurements to determine whether the fleet characteristics for these are different than for the vehicles having positive emissions measurements. The comparison indicated that mean speed and acceleration values are similar in both data sets. There appear to be slightly more newer model year vehicles in the negative measurements data set. Overall however, we conclude that discarding the set of negative emissions data does not have a substantial effect on distribution of vehicles in the final data set.

We verified the RSD equipment by the "Puff-in-Vehicle" (PIV) mode measurements during the data collection. The average PIV CO/CO_2 ratio was found to be 1.009, and the average for the PIV HC/CO_2 ratio was calculated as 0.132. These averages are nearly identical to the true ratios of the calibration gas, and in fact are not statistically different from the true values. Therefore, we conclude that the system is accurate. The PIV calculations offer a means by which the precision of measurements may be ascertained since they are based upon repeated reading of a gas of known composition. The 95 percent probability range for PIV CO/CO_2 measurements were in the range of 0.88 to 1.13, approximately within plus or minus 12 percent of the mean value. For the HC/CO_2 PIV measurements, the 95 percent of the values were within a range of 0.119 to 0.144, approximately plus or minus 10 percent of the mean value. We conclude that the observed precisions of the measurements of the calibration gas are approximately plus or minus 10 percent.

After post-processing 7,056 valid emissions data were left. Most of the data that were removed from the data set were associated with measurements reported either

invalid or as negative by the RSD. We conclude that when deciding on how many data points to collect in future studies, it should be assumed that only about 60 percent of measurement attempts will result in valid emissions estimate for on-road light-duty vehicles for sites such as the ones for which we collected data.

6.2 Exploratory Analysis of Database and Results

After database development, a summary of the results were prepared. Then exploratory analysis of the data was conducted to identify possible trends or relationships within the overall data set. Key findings from these analysis are given as follows:

- The average CO emissions for Site 1 is calculated as 366 grams/gallon. For Site 2 the average CO emissions is calculated as 340 grams/gallon. The average HC emissions estimate for Site 1 is 48 grams/gallon and 46 grams/gallon for Site 2. We compared our mean emission estimate results with the results obtained from MOBILE5a model. The range for the mean CO emissions estimate were estimated from 300g/gal to 408 grams/gallon for light duty gasoline vehicles. For light duty gasoline trucks the range is approximately from 248 to 443 grams/gallon. The average HC emissions estimate was calculated to be in the range from 23 to 32 grams/gallon for light duty gasoline vehicles. For light duty gasoline trucks the range is from 20 to 37 grams/gallon. Based upon these findings we conclude that our mean CO estimates are in the range that is estimated by MOBILE5a. The average

HC emissions, on the other hand, are higher than the values predicted MOBILE5a.

- Fleet characteristics were also determined for vehicles observed at each site. Average speed for Site 1 is 36 mph whereas 47.5 mph for Site 2. The average acceleration for Site 1 is estimated as 5.1 mph/sec and 0.2 mph/sec for Site 2. The differences in average speed and acceleration between the two sites are attributed to differences in site geometries, traffic patterns, and the location of the study zone. The percent of the vehicles in each model year category varies for data collected at different days, but approximately three quarters of the vehicles in each data set by day were of 1991 or more recent model years. Approximately, 45 percent of the vehicles at Site were associated with platoon whereas approximately 75 percent of the vehicles were in platoon for Site 2.
- The site effects on emissions were investigated by comparing the data collected at two sites both for CO and HC emissions estimates. Visual comparison of CDFs of emissions data collected at different sites showed that the inter-vehicle variability in emissions is qualitatively similar. A two-tailed t-test were conducted on the mean estimates of the data collected at different sites indicate that there is no statistically significant difference between mean emissions estimates of CO and HC for data collected at Site 1 and Site 2. We conclude that in order to get an average emission estimate for Research Triangle area data collection at two sites seem to be enough. However, to ensure this finding data should be collected at sites which have different

traffic and vehicle fleet properties than the ones for which we have taken data.

- Empirical distributions of CO and HC emissions data collected on different days showed qualitative differences between some of the daily data sets. Multi-comparison tests were conducted to determine whether the difference in mean emissions of pairs are statistically different or not. The results indicate that for some CO data sets, such as June 8 and June 9 data sets, the mean emissions estimates are significantly different from the mean emissions estimates of other data sets. There are differences in median estimates but they are not pronounced as much as the differences in mean estimates. For HC there are also some differences in emissions estimates distributions and mean estimates. For example mean estimates of May 27, June 8, and June 9, data sets are similar to each other, but different than mean estimates of May 11, May 12. Vehicle fleet characteristics for different days, time of data collection, and difference in temperatures were investigated to explain the differences among some of the daily data sets. Our findings suggested that differences in the observed traffic variables along with time of data collection and temperature is not likely to explain the differences between data sets, both for CO and HC data sets. Based upon the findings it was concluded that there is no rationale to separate the data such as by differences in mean emissions, and then analyze them individually.
- Vehicles for which repeated measurements were obtained were identified for the purpose of intra-vehicle variability in emissions. Analysis showed that individual vehicle CO emissions estimates have standard deviations from 0.6

to 750 g/gallon over a 95 percent probability range. This suggests that intra-vehicle variability might be as high as inter-vehicle variability, 60 percent of the vehicles for which repeated measurements were available had standard deviations of less than 100 g/gallon, which is over five times lower than the standard deviations for inter-vehicle variability. Individual vehicle HC emissions estimates had standard deviations from 0.03 to 150 g/gallon on a probability range of 95 percent. It was found that 80 percent of the vehicles with multiple observations have an intra-vehicle variability, standard deviation of less than 20 g/gallon, which is less than one-fifth of the standard deviation for inter-vehicle variability. We conclude that for some vehicles intra-vehicle variability might be considerably high but in general compared to inter-vehicle variability it is relatively low.

- Multi-comparison analysis of vehicle model years was conducted to investigate the effect of vehicle model year on CO and HC emissions estimates. It was found that for very old vehicles, such as model years earlier than 1980, the mean CO and HC emissions are significantly different than the newer vehicles, model years later than 1989. As the gap in model years increases, the difference in mean estimates tends to increase. Out of 28 pairwise comparisons of the mean emissions estimates 24 of them are significantly different from each other for CO data. For HC 15 pairs of comparisons were significantly different from each other. We conclude that vehicle model year is an important variable to explain the variability in emissions estimates. This parameter should be used in a model which is used to predict the vehicle emissions.

- The effect of vehicle type on emissions estimates was also evaluated. It was found that, the mean estimates of Type 2 vehicles are significantly different from the mean emission estimates of Type 3 vehicles for grams per gallon estimates. For emissions estimates on grams per mile basis there was not a statistically significant difference in the mean estimates between these two types of vehicles. We conclude that vehicle type might be an important parameter to explain the variability in vehicle emissions, however, a detailed type category or vehicle body types might be more useful than aggregating them into two types.

6.3 Model Development for Vehicle Emissions Estimation

In our study an emission estimation model was developed by combining Hierarchical Tree-Based Regression (HTBR) and ordinary least squares regression (OLS) methods. The model is basically a regression tree, having ordinary regression equations fitted at the terminal nodes. This model let us use strengths of both of these methods: data mining aspects of regression trees and prediction capability and practicality aspects of OLS.

Four different regression trees, both for CO and HC, were developed in order to simulate different data collection scenarios. First and second model had extensive amount of input including detailed vehicle parameters and traffic variables. Third and fourth models were used using only traffic variables. Later it was found that the power of the

third and fourth models, which is expressed as reduction in deviance, are not strong enough to be used for prediction purposes. We conclude that only traffic variables are not enough to explain the variability in emissions. Later we discarded these models.

Our model consists of regression trees having linear regressions fitted at the terminal nodes. Actually, this model can be used for both “vehicle-based” and “flow-based” estimation. If the individual vehicle parameters such as speed, acceleration measurements are available emissions factors can be calculated through regression equations. If those measurements can not be obtained, the mean pollutant values given at the terminal nodes of the tree can be used to calculate the emissions probabilistically. The first approach is called “vehicle-based” and the second is “flow-based”. It should be noted that both of these models should be used to calculate average mean estimates for the fleets, although “vehicle based” model estimates individual vehicle emissions. By using the developed model one can estimate average emission rates for CO and HC for a selected fleet.

In order to compare the results, we run the MOBILE5a model for different nodes of the “flow-based” model. It is not possible to compare “vehicle-based” model with MOBILE5a since MOBILE5a uses average speed values for emission estimation. The same input variables were used in the MOBILE5a model as the one used in the “flow-based” model. For example, the percent of vehicles in each model year were entered to the MOBILE5a model. In general the results looked similar, being in the same order-of-magnitude. Some results of MOBILE5a were less than values given by our model, such as node 1 in the first CO model. However, there were some nodes in the same regression

tree where MOBILE5a estimated a value twice the one given by our model. Similar results were obtained for HC data. We conclude that our results were comparable to the results obtained using the model that is legally used in emission factor development.

6.4 Limitations

During the study some limitations has been recognized. It is necessary to identify these limitations. Some of the key factors to consider include:

The data were collected at two sites. Statistical analysis showed that the mean emissions estimates are not significantly different from each other. However, it would have been better to collect data at more sites so that effect of different sites on emissions might have been quantified better.

Because the RSD does not report standard errors of the calculated CO/CO_2 and HC/CO_2 molar ratios we can not determine whether negative values are statistically significantly different than zero or not.

The RSD is calibrated using a cylinder gas that represents a high-emitting vehicle. This might affect the accuracy of the measurements for lower values of emissions.

6.5 Recommendations and Future Work

As discussed in Section 6.1 negative emissions values were discarded from the data base. As a future work negative ratios should be re-evaluated to determine which ones are not significantly different from zero. This could be done by automatically getting standard errors of the slopes which are reported in the individual emissions output files by a software, such as MATLAB. Then these numbers could be used to test if the CO/CO_2 and HC/CO_2 molar ratios are significantly different that zero or not. Values which are significantly different than zero might be eliminated from the data base since they are invalid data points. Other negative values could be retained in the database.

In this study data were collected at two sites in Research Triangle area of North Carolina. Statistical analysis showed that there is no statistically significant difference between the mean emissions estimates for the data collected at these two different sites. This suggests that site does not have an effect on the mean emission estimates. It is possible that these two sites may not represent the Research Triangle area properly. It is recommended that more data be collected in order to have representative data for the area of concern. In the course of analysis, multi-comparison tests were conducted. These tests did not give statistically significant differences for many of the explanatory variables. It is possible to see statistically significant differences when larger data sample sizes are used.

7.0 REFERENCES

An, F., M. Barth, G. Scora, and T. Younglove (1996), "Catalyst Cold-Start Characterization and Modeling," *Proceedings of the Sixth CRC On-Road Vehicle Emissions Workshop*, Coordinating Research Council, San Diego, CA, pp.2.1-2.21.

An, F., and M. Ross (1996), "A Simple Physical Model for High Power Enrichment Emissions," *J. Air and Waste Mgmt. Assoc.*, 46(3):216-223.

An, F., M. Barth, G. Scora, and J. Norbeck (1997) "Characterization and Modeling of Vehicle Lean-Burn HC Emissions," *Proceedings of the Seventh CRC On-Road Vehicle Emissions Workshop*, Coordinating Research Council, Atlanta, GA, pp. 6.97-6.114.

Ashbaugh, L., and D.R. Lawson (1991), "A Comparison of Emission from Mobile Sources Using Random Roadside Surveys Conducted," *Proceedings of the 84th Air and Waste Management Association*, Paper-91-180.58.

Asbaugh, L., D.R. Lawson, G.A. Bishop, *et. al.* (1992), "On-Road Remote Sensing of Carbon Monoxide and Hydrocarbon Emissions During Several Vehicle Operating Conditions," Presented at AWMA/EPA Conference on PM10 Standards and Nontraditional Particulate Source Controls, Phoenix, AZ, January.

Avallone E. A., and T. Baumeister (1979). *Mark's Standard Handbook for Mechanical Engineers*., McGraw Hill: New York. Pp. 7.14 -7.21

Barth, M., F. An, J. Norbeck, and M. Ross (1996), "Modal Emissions Modeling: A Physical Approach," *Transportation Research Board*, 1520:81-88.

Barth, M., and J. Norbeck (1997), "NCHRP Project 25-11: The Development of a Comprehensive Modal Emission Model," *Proceedings of the Seventh CRC On-Road Vehicle Emissions Workshop*, Coordinating Research Council, Atlanta, GA, pp. 6.53-6.71.

Bishop, G.A., J.R. Starkey, A. Ihlenfeldt, W.J. Williams, and D.H. Stedman (1989), "IR Long Path Photometry: A Remote Sensing Tool for Automotive Emissions," *Analytical Chemistry*, 61(10): 671A-677A.

Bishop, G., and Stedman, D.H. (1990), "On-Road Carbon Monoxide Emission Measurement Comparisons for the 1988-1989 Colorado Oxy-Fuels Program," *Environmental Science and Technology*, 24(6):843-847.

Bishop, G.A., D.H. Stedman, and L. Ashbaugh (1996), "Motor Vehicle Emissions Variability," *J. Air and Waste Mgmt. Assoc*., 46(7):667-678.

Bosch (1986), *Automotive Handbook*. Robert Bosch GmbH: Stuttgart.

Bradu, D., and Y. Mundlak (1970), " Estimation in Lognormal Linear Models," *Journal of the American Statistical Association*, 65(3):198-211.

Brady, J. E., and G.E. Humiston (1978). *General Chemistry-Principle and Structure*. John Wiley and Sons: New York.

Cadle, S. and R. D. Stephens (1994), "Remote Sensing of Vehicle Exhaust Emissions," *Environmental Science and Technology*, 28(6) pp 258A-64A.

Cicero-Fernandez, P., and J.R. Long (1994), "Instantaneous and Short Term Emission Rates Observed during Modal Acceleration Testing," *Proceedings of the Fourth CRC-APRAC On-Road Vehicle Emission Workshop*, San Diego, CA, pp. 3.1-3.25.

Cooper, C.A., and F.C. Alley (1994). *Air Pollution Control, A Design Approach.* 2nd Edition. Waveland Press Inc: Illinois, pp. 45-57.

CMS (1997), "MOBILIZER-PC© Version 2.2 Operator Manuals," Prepared by CMS Engineering Systems, Inc.

Cullen, A. C., and H. C. Frey (1999). *Probabilistic Techniques in Exposure Assessment*. Plenum Press: New York.

Curran, T., T. Fitz-Simons, W. Freas, J. Hemby, D. Mintz, S. Nizich, B. Parzygnat, and M. Wayland (1994), "National Air Quality and Emission Trends Report, 1993," Report No. EPA-45-R-94-026, U.S. Environmental Protection Agency, Research Triangle Park, N.C., October.

Dalton, R. (1999), "Traffic Impacts of a Temporary Mobile Emission (RSD) Work Zone," Master's Thesis, Department of Civil Engineering, North Carolina State University, Raleigh.

DOT (1999) "Average Fuel Efficiency of U.S. Passenger Cars, Light Trucks, and Light Duty Vehicles," http://www.bts.gov/btsprod/nts/chp4/tbl4x22.html, Located 18 May 1999.

Effa, R. C., and L. C. Larsen (1993), "Development of Real-World Driving Cycles for Estimating Facility-Specific Emissions from Light-Duty Vehicles," Presented at the Air and Waste Management Association Specialty Conference on the Emission Inventory: Perception and Reality, Pasadena, California, October 18-20.

FHWA (1992), "A Summary: Transportation Programs and Provisions of the Clean Air Act Amendments of 1990," Report No. FHWA-PD-92-023, U.S. Department of Transportation, October.

Flagan, R. C., and J. H. Seinfeld (1988). *Fundamentals of Air Pollution Engineering*. Prentice Hall: New York.

Frey, H.C. (1997), "Variability and Uncertainty in Highway Vehicle Emission Factors," *Proceedings, Emission Inventory: Planning for the Future*, Air and Waste Management Association, Pittsburgh, PA, October, pp. 208-219.

Frey, H.C., and D.A. Eichenberger (1997), "Remote Sensing of Mobile Source Air Pollutant Emissions: Variability and Uncertainty in On-Road Emissions Estimates of Carbon Monoxide and Hydrocarbons for School and Transit Buses," Report No. FHWY/NC/97-005, Prepared by North Carolina State University for the North Carolina Department of Transportation, Raleigh, NC.

Gautam, M., D. Gupta, S. Popuri, and D. Lyons (1997), "Speciation and Reactivities of Diesel Exhaust Emission," *Proceedings of the Seventh CRC On-Road Vehicle Emissions Workshop*, Coordinating Research Council, Atlanta, GA, pp.5.19-5.48.

Gilbert, R. O. (1987). *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold: New York.

Gorse, R.A. (1984), "On-Road Emission Rates of CO, NO_X and Gaseous HC," *Environmental Science and Technology*, 18(7):500-507.

Guensler, R. (1994), "Data Needs for Evolving Motor Vehicle Emission Modeling Approaches," *Proceedings of National Conference on Transportation Planning and Air Quality,* Danvers, MA, pp. 167-196.

Guensler, R., W. Bachman, S. Washington, M.O. Rodgers, and M.D. Meyer (1997), "Motor Vehicle Activity and Emissions Algorithms in the Georgia Tech GIS-Based Modal Emissions Model," *Proceedings of the Seventh CRC On-Road Vehicle Emissions Workshop*, Coordinating Research Council, Atlanta, GA, pp. 6.1-6.11.

Heien, D. M. (1968) "A Note on Log-Linear Regression," *Journal of the American Statistical Association*, 63(9):1034-1038.

Jack, M.D., T.P. Bahan, M. N. Gray, J. L. Hanson, T. L. Heidt, F. A. Huerta, D. R. Nelson, A. J. Paneral, C . M. Peterson, J. Peterson, M. Sullivan, G. C. Polchin, L. H. Rubin, C. B. Tacelli, W. C. Trautfield, R. O. Wageneck, G. A. Walter, J. D. Wills, J. F. Alves, B. A. Berger, J. Brown, J. A. Shelton, G. A. Smith, E. J. Palen, and N.W. Sorbo (1995*)*, "Remote and On-Board Instrumentation for Automotive Emissions Monitoring," Paper No: 951943, Society of Automotive Engineers, Warrendale, PA, pp. 1-15.

Jimenez, J. L, P. McClintock, G. J. McRae, D. D. Nelson, and M. S. Zahniser (1998), "Vehicle Specific Power: A Useful Parameter for Remote Sensing and Emission Studies," Presented at 9th CRC On-Road Vehicle Emissions Workshop, San Diego, April 21st .

Kelly, N. A., P. J. Groblicki (1993), "Real-World Emissions from a Modern Production Vehicle Driven in Los Angeles," *J. Air and Waste Mgmt. Assoc.,* 43(10):1351-1357.

Kini, M.D., and H. C. Frey (1997), "Probabilistic Modeling of Exhaust Emissions From Light Duty Gasoline Vehicles," Prepared by North Carolina State University for the Center for Transportation and the Environment, North Carolina State University, Raleigh, NC.

Kirchstetter, T.W., B.C. Singer, and D.A. Harley (1996), "Impact of Oxygenated Gasoline use on California Light-Duty Vehicle Emissions," *Environmental Science and Technology*, 30(2):661-670.

Lawson, D.R., P.J. Groblicki, D.H. Stedman, G.A. Bishop, and P.L. Guenther (1990), "Emissions from In-Use Motor Vehicles in Los-Angeles: A Pilot Study of Remote Sensing and the I/M Program," *J. Air and Waste Mgmt. Assoc*., 40(8):1096-1105.

Lawson, D.R. (1993), "Passing the Test: Human Behavior and California's Smog Check Program," *J. Air and Waste Mgmt. Assoc.*, 43(10):1567-1575.

LeBlanc, D. C., F. M. Saunders, M.D. Meyer, D. Michael, and G. Guensler (1995), "Driving pattern variability and impacts on vehicle carbon monoxide emissions " *Transportation Research Record*, 1472:45-52.

Lodge, J. P. (1991). *Methods of Air Sampling and Analysis*. Lewis Publishers: Michigan. 1991.

McClintock, P. (1999), "Remote Sensing Measurements of Real World High Exhaust Emitters," Report No. E-23-Interim Report, Prepared by Colorado Department of Public Health and Environment for Coordinating Research Council, , Atlanta, Georgia.

Mathsoft (1997). *S-Plus 4 Guide to Statistics*. Data Analysis Products Division, Mathsoft, Seattle, WA.

Mercer, A.P., and S.S. McArragher (1997), "Remote Sensing Measurement of On-Road Emissions in Five European Cities," *Proceedings of the Seventh CRC On-Road Vehicle Emissions Workshop*, Coordinating Research Council, Atlanta, GA, pp. 8.81-8.91.

Neter, J., M. H. Kutner, C. J. Nachtsheim, and W. Wasserman (1996). *Applied Linear Statistical Models*. Irwin: Chicago.

Nizich, S. V., T. C. McMullen, and D. C. Misenheimer (1994), "National Air Pollutant Emission Trends 1900-1993," Report No. EPA-454-R-94-027, Office of Air Quality Planning and Standards, Research Triangle Park, N.C., October, 314pp.

NRC (1992), *Rethinking the Ozone Problem in Urban and Regional Air Pollution*, National Research Council, National Academy Press: Washington, D.C.

Popp, P. J., G.A. Bishop, and D.H. Stedman (1997), "Development of a High-Speed Ultraviolet Spectrophotometer Capable of Real-Time NO and Aromatic Hydrocarbon

Detection in Vehicle Exhaust," *Proceedings of the Seventh CRC On-Road Vehicle Emissions Workshop,* Coordinating Research Council, Atlanta, GA, pp.4.3-4.12.

Rendahl, C. S. (1995) "Data Handling and Validation from Wisconsin's Remote Vehicle Emissions Sensing Studies," *Proceedings of the SPIE International Society of Optical Engineers*, Bellingham, WA, pp. 72-83.

Sargeant, K. A. (1994), "Transportation Conformity Final Rule: Summary of Provisions and Their Implications," Paper No. 94-MP13.02, *Proceedings, 87th Annual Meeting*, Air and Waste Management Association, Pittsburgh, PA.

SBRC (1994). *Remote Emissions Sensor RES-100 Smog Dog User's Manual*, Reference NO. 94-0327, Prepared by Hughes Santa Barbara Research Center for State of North Carolina Dept. of Environmental, Health, and Natural Resources, Raleigh, N.C., SBRC.

Shih R., and R.F. Sawyer (1996), "The Relation Between Throttle Positioning and Emissions," *Proceedings of the Sixth CRC On-Road Vehicle Emissions Workshop*, Coordinating Research Council, San Diego, CA, pp. 2.41-2.51.

Shih R., S. Fable, and R.F. Sawyer (1997), "Effects of Driving Behavior on Automobile Emissions," *Proceedings of the Seventh CRC On-Road Vehicle Emissions Workshop*, Coordinating Research Council, Atlanta, GA, pp. 6.115-6.124.

Singer, B.C., and D.A. Harley (1996), "A Fuel-Based Motor Vehicle Emission Inventory," *J. Air Waste Mgmt. Assoc.*, 46(6):581-593.

Singer, B. C., R. A. Harley, D. Littlejohn, J. Ho, and T. Vo (1998), " Scaling of Infrared Remote Sensor Hydrocarbon Measuerements for Motor Vehicle Emission Inventory Calculations," *Environmental Science and Technology*, 32(11):3241-3248.

Stedman, D.H. (1989), "Automobile Carbon Monoxide Emissions," *Environmental Science and Technology*, 23(2):147-149.

Stephens, R.D., and S.H. Caddle (1991), "Remote Sensing Measurements of Carbon Monoxide Emissions from On-Road Vehicles," *J. Air and Waste Mgmt. Assoc.*, 41(1):39-46.

Stephens, R.D., S.H. Cadle, and T.Z. Qian (1996a), "Analysis of Remote Sensing Errors of Comission and Omission Under FTP Conditions," *J. Air and Waste Mgmt. Assoc*., 46(7):510-516.

Stephens, R.D., P.A. Mulawa, M.T. Giles, K.G. Kennedy, P.J. Groblicki, and S.H. Cadle (1996b), "An Experimental Evaluation of Remote Sensing-Based HC Measurement: A Comparison to FID Measurements," *J. Air and Waste Mgmt. Assoc*., 46(2):148-158.

Stephens, R. D., M.T. Giles, and K. McAlinden (1997), "An Analysis of Michigan and California CO Remote Sensing Measurement," *J. Air and Waste Mgmt. Assoc.*, 47(5):601-607.

Stern, A. (1976), *Air Pollution, Volume III: Measuring, Monitoring, and Surveillance of Air Pollution*, Third Edition. Academic Press: London. pp. 764-767.

TRB (1994), *Highway Capacity Manual*, Special Report 209, Transportation Research Board, National Research Council: Washington, D.C.

TRB (1995), *Expanding Metropolitan Highways: Implications for Air Quality and Energy Use*, Special Report 245, Transportation Research Board, National Research Council: Washington, D.C.

USEPA (1993a), "Materials from the FTP Review Project Public Workshop," Office of Mobile Sources, Certification Division, US Environmental Protection Agency, Ann Arbor, MI.

USEPA (1993b), "Federal Test Procedure Review Project: Preliminary Technical Report," Report No. EPA/420/R-93-007, U.S. Environmental Protection Agency, Washington D.C., pp.16-20.

USEPA (1994), "MOBILE5a User's Guide," http://www.epa.gov/orcdizux/m5.htm, Located 18 May 1999.

USEPA (1995), "Air Quality Trends," Report No. EPA-454/F-95-003, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, N.C., pp. 3-10.

Vojtisek-Lom, Michal, and J.T. Cobb (1997), "Vehicle Mass Emissions Measurements Using a Portable 5-Gas Exhaust Analyzer and Engine Computer Data," Paper presented at Emission Inventory: Planning for the Future, October 29, 1997, Research Triangle Park, N.C., pp. 1-14.

Yu, Lei (1997), "Collection and Evaluation of Modal Traffic Data for Determination of Vehicle Emission Rates Under Certain Driving Conditions," Report No. TxDOT 1485-3F, Prepared by Texas Southern University for Texas Department of Transportation, Houston, TX.

Venables, W. N., B. D. Ripley (1997). *Modern Applied Statistics with S-Plus*. Springer: New York. 1997.

Walpole, R. E., R. H. Myers (1993). *Probability and Statistics for Engineers and Scientists*. Macmillan Publishing Company: New York. 1993.

Washington, S., J. Wolf, and R. Guensler (1997), "A Binary Recursive Partitioning Method for Modeling Hot-Stabilized Emissions from Motor Vehicles," Prepared by

School of Civil and Environmental Engineering, Georgia Institute of Technology for the 76th Annual Meeting of the Transportation Research Board, Atlanta, Georgia.

Washington, S. (1999), Personal contact with Alper Unal.

West, B. H., and R. N. McGill (1997), "Emissions and Fuel Consumption Predictions from Data-Based Modal Models," *Proceedings of the Seventh CRC On-Road Vehicle Emissions Workshop*, Coordinating Research Council, Atlanta, GA, pp. 6.33-6.51.

WRAL (1999), "RDU Weather Data," http://www.wral-tv.com/weather/data/, Located 18 May 1999.

Zhang, Y., D.H. Stedman, G.A. Bishop, P.L. Guenther, and S. P. Beaton (1995), "Worldwide On-Road Vehicle Exhaust Emissions Study by Remote Sensing," *Environmental Science and Technology*, 29(9): 2286-2294.

Zhang, Y., D.H. Stedman, G.A. Bishop, S.P. Beaton, P.L. Guenther, and L.F. McVey (1996a), "Enhancement of Remote Sensing for Mobile Source Nitric Oxide," *J. Air and Waste Mgmt. Assoc.*, 46(1): 25-29.

Zhang, Y., D.H. Stedman, G.A. Bishop, S.P. Beaton, and P.L. Guenther (1996b), "On-Road Evaluation of Inspection and Maintenance Effectiveness," *Environmental Science and Technology*, 30(5): 1445-1450.

Zub, R. W. (1981), "A Computer Program (VEHSIM) for Vehicle Fuel Economy and Performance Simulation", National Traffic Safety Administration, Department of Transportation, Washington D. C.

APPENDIX A: SAMPLE SINGLE EMISSION FILE

Vehicle File = c:/alper/Ascii/06-08/V060898A.034
Vehicle Number = 34
Coefficient File = puffcal.dat
Data Acq Time/Date = 13:56:15 / 06-08-1998
Intercept = 0.0831336 0.0986418
Slope = 0.0047919 0.0018652
Standard Deviation = 0.0008799 0.0003806

Coefficients:
Order of Fit for CO = 4
Order of Fit for CO_2 = 4
Order of Fit for HC = 4

No	CO	CO_2	HC	Coefficient
1	-1973.627685546875	-235.521133422852	-252.008987426758	Constant
2	91.283119201660	13.047317504883	11.550007820129	X
3	-1.555695056915	-0.251983582973	-0.196325734258	X^2
4	0.011631903239	0.002084425185	0.001473723794	X^3
5	-0.000032296444	-0.000006337780	-0.000004134658	X^4

Mean Peak-to-peak values..., 4 Channels Pre/Post

Pre Ref	Post Ref	Pre CO	Post CO	Pre CO2	Post CO2	Pre HC	Post HC
19756	0	22154	0	12954	0	16361	0

Peak-to-Peak Array 4 Channels Pre/Post...

No.	Pre Ref	Post Ref	Pre CO	Post CO	Pre CO2	Post CO2	Pre HC	Post HC
1	19769	19914	22159	22051	12944	11919	16367	16012
2	19759	19905	22160	22045	12945	11762	16358	16000
3	19758	19909	22159	22057	12938	11762	16359	16004
4	19756	19899	22161	22053	12940	11776	16355	15998
5	19740	19893	22142	22057	12937	11920	16345	15990
6	19754	19892	22148	22052	12943	11942	16358	15992
7	19761	19902	22161	22058	12948	12228	16356	15997
8	19756	19906	22159	22066	12945	12402	16356	16004
9	19759	19905	22151	22061	12949	12484	16369	15995
10	19758	19903	22146	22069	12949	12464	16364	15994
11	19774	19892	22163	22050	12960	12477	16378	15988
12	19769	19890	22171	22053	12961	12444	16370	15988

APPENDIX A: SAMPLE SINGLE EMISSION FILE

13	19763	19908	22164	22067	12957	12428	16364	16001
14	19767	19899	22167	22057	12963	12319	16367	15992
15	19754	19900	22150	22057	12948	12316	16360	15994
16	19767	19904	22158	22065	12953	12289	16367	15995
17	19750	19905	22144	22062	12950	12266	16354	15993
18	19770	19900	22162	22056	12964	12226	16371	15991
19	19759	19901	22157	22061	12953	12240	16356	15998
20	19754	19891	22151	22053	12957	12245	16355	15992
21	19764	19892	22155	22052	12961	12247	16366	15989
22	19757	19887	22145	22052	12952	12276	16359	15989
23	19768	19892	22154	22048	12963	12280	16375	15992
24	19754	19890	22148	22041	12961	12285	16362	15989
25	19740	19886	22144	22038	12961	12312	16354	15983
26	19733	19906	22141	22060	12964	12383	16355	16002
27	19705	19896	22120	22060	12947	12454	16323	15995
28	19717	19905	22123	22059	12953	12511	16334	16005
29	19799	19897	22217	22055	13007	12560	16411	15999
30	19828	19894	22206	22054	12837	12578	16423	16002
31	19806	19898	22183	22055	12822	12576	16408	16002
32	19970	19892	22536	22053	13306	12584	16997	16000
33	19972	19899	22539	22052	13308	12598	16998	16002
34	19977	19892	22535	22050	13308	12611	17002	15994
35	19624	19899	21761	22059	12909	12609	16702	15997
36	20028	19890	22680	22059	13388	12598	17211	15989
37	20009	19875	22671	22043	13378	12591	17197	15977
38	20010	19897	22674	22059	13383	12613	17206	15995
39	19766	19886	22474	22048	13086	12607	17024	15988
40	19764	19893	22468	22055	13084	12606	17025	15996
41	19760	19894	22469	22050	13085	12600	17017	15993
42	19746	19898	22461	22061	13078	12601	17004	16002
43	19748	19900	22458	22054	13083	12582	17010	15996
44	19737	19894	22454	22052	13077	12582	17006	15984
45	19739	19904	22453	22057	13080	12593	17009	16001
46	19759	19900	22474	22059	13093	12602	17021	15996
47	19750	19902	22466	22062	13080	12617	17004	16003
48	19759	19900	22474	22059	13090	12636	17019	16005
49	19757	19893	22474	22052	13092	12639	17013	15994
50	19747	19900	22471	22062	13088	12658	17008	16008
51	19749	19896	22461	22051	13081	12656	17008	16001
52	19749	19894	22455	22050	13074	12662	17006	16000

APPENDIX A: SAMPLE SINGLE EMISSION FILE

53	19758	19896	22461	22053	13081	12677	17015	16002
54	19751	19892	22465	22049	13084	12692	17012	16001
55	19752	19907	22459	22062	13081	12712	17015	16010
56	19762	19898	22475	22061	13090	12721	17019	16006
57	19741	19905	22459	22066	13083	12732	17008	16010
58	19743	19901	22467	22066	13085	12735	17002	16003
59	19733	19891	22456	22059	13077	12736	16992	16000
60	19744	19898	22468	22064	13090	12744	17004	16001
61	19741	19886	22465	22051	13080	12737	16997	15993
62	19738	19894	22462	22052	13084	12731	17003	15999
63	19740	19894	22453	22054	13086	12727	17010	15997
64	19736	19894	22454	22062	13089	12741	17003	15998
65	19749	19911	22450	22072	13090	12757	17013	16014
66	19742	19905	22453	22074	13095	12763	17007	16009
67	19744	19909	16616	22075	9187	12763	6634	16010
68	21333	19905	19156	22074	13133	12748	9659	16008
69	26618	19906	23549	22070	14612	12749	10367	16002
70	26609	19904	23539	22076	14608	12750	10360	16001

Post-Car Peak-to-Peak/Pre-Car Mean Ratio of 4 Channels...

No.	Reference	CO	CO2	HC
1	1.00800	0.99535	0.92010	0.97867
2	1.00754	0.99508	0.90798	0.97794
3	1.00774	0.99562	0.90798	0.97818
4	1.00724	0.99544	0.90906	0.97781
5	1.00693	0.99562	0.92018	0.97732
6	1.00688	0.99540	0.92188	0.97745
7	1.00739	0.99567	0.94396	0.97775
8	1.00759	0.99603	0.95739	0.97818
9	1.00754	0.99580	0.96372	0.97763
10	1.00744	0.99616	0.96217	0.97757
11	1.00688	0.99531	0.96318	0.97720
12	1.00678	0.99544	0.96063	0.97720
13	1.00769	0.99607	0.95939	0.97800
14	1.00724	0.99562	0.95098	0.97745
15	1.00729	0.99562	0.95075	0.97757
16	1.00749	0.99598	0.94866	0.97763
17	1.00754	0.99585	0.94689	0.97751
18	1.00729	0.99558	0.94380	0.97739
19	1.00734	0.99580	0.94488	0.97781
20	1.00683	0.99544	0.94527	0.97745

21	1.00688	0.99540	0.94542	0.97726
22	1.00663	0.99540	0.94766	0.97726
23	1.00688	0.99522	0.94797	0.97745
24	1.00678	0.99490	0.94836	0.97726
25	1.00658	0.99476	0.95044	0.97690
26	1.00759	0.99576	0.95592	0.97806
27	1.00709	0.99576	0.96140	0.97763
28	1.00754	0.99571	0.96580	0.97824
29	1.00714	0.99553	0.96958	0.97787
30	1.00699	0.99549	0.97097	0.97806
31	1.00719	0.99553	0.97082	0.97806
32	1.00688	0.99544	0.97144	0.97794
33	1.00724	0.99540	0.97252	0.97806
34	1.00688	0.99531	0.97352	0.97757
35	1.00724	0.99571	0.97337	0.97775
36	1.00678	0.99571	0.97252	0.97726
37	1.00602	0.99499	0.97198	0.97653
38	1.00714	0.99571	0.97368	0.97763
39	1.00658	0.99522	0.97321	0.97720
40	1.00693	0.99553	0.97314	0.97769
41	1.00699	0.99531	0.97267	0.97751
42	1.00719	0.99580	0.97275	0.97806
43	1.00729	0.99549	0.97128	0.97769
44	1.00699	0.99540	0.97128	0.97696
45	1.00749	0.99562	0.97213	0.97800
46	1.00729	0.99571	0.97283	0.97769
47	1.00739	0.99585	0.97398	0.97812
48	1.00729	0.99571	0.97545	0.97824
49	1.00693	0.99540	0.97568	0.97757
50	1.00729	0.99585	0.97715	0.97842
51	1.00709	0.99535	0.97700	0.97800
52	1.00699	0.99531	0.97746	0.97794
53	1.00709	0.99544	0.97862	0.97806
54	1.00688	0.99526	0.97977	0.97800
55	1.00764	0.99585	0.98132	0.97855
56	1.00719	0.99580	0.98201	0.97830
57	1.00754	0.99603	0.98286	0.97855
58	1.00734	0.99603	0.98309	0.97812
59	1.00683	0.99571	0.98317	0.97794
60	1.00719	0.99594	0.98379	0.97800
61	1.00658	0.99535	0.98325	0.97751
62	1.00699	0.99540	0.98279	0.97787

63	1.00699	0.99549	0.98248	0.97775
64	1.00699	0.99585	0.98356	0.97781
65	1.00785	0.99630	0.98479	0.97879
66	1.00754	0.99639	0.98526	0.97849
67	1.00774	0.99643	0.98526	0.97855
68	1.00754	0.99639	0.98410	0.97842
69	1.00759	0.99621	0.98417	0.97806
70	1.00749	0.99648	0.98425	0.97800

Normalized gas ratios for CO, CO2 and HC...

No.	CO	CO2	HC
1	0.98745	0.91280	0.97090
2	0.98763	0.90119	0.97061
3	0.98797	0.90100	0.97066
4	0.98829	0.90253	0.97079
5	0.98876	0.91384	0.97059
6	0.98859	0.91557	0.97076
7	0.98836	0.93703	0.97058
8	0.98852	0.95017	0.97081
9	0.98835	0.95650	0.97031
10	0.98881	0.95507	0.97035
11	0.98850	0.95659	0.97052
12	0.98873	0.95416	0.97062
13	0.98847	0.95207	0.97053
14	0.98847	0.94415	0.97042
15	0.98842	0.94387	0.97049
16	0.98858	0.94161	0.97036
17	0.98839	0.93980	0.97019
18	0.98837	0.93697	0.97031
19	0.98855	0.93800	0.97069
20	0.98869	0.93885	0.97081
21	0.98859	0.93896	0.97058
22	0.98884	0.94142	0.97083
23	0.98841	0.94149	0.97076
24	0.98820	0.94197	0.97068
25	0.98826	0.94423	0.97051
26	0.98825	0.94872	0.97069
27	0.98875	0.95464	0.97075
28	0.98826	0.95857	0.97092
29	0.98848	0.96271	0.97094
30	0.98858	0.96424	0.97127

31	0.98843	0.96389	0.97108
32	0.98864	0.96480	0.97125
33	0.98824	0.96553	0.97103
34	0.98850	0.96687	0.97089
35	0.98856	0.96637	0.97073
36	0.98900	0.96597	0.97068
37	0.98903	0.96616	0.97068
38	0.98866	0.96678	0.97070
39	0.98871	0.96685	0.97081
40	0.98868	0.96643	0.97096
41	0.98840	0.96593	0.97073
42	0.98870	0.96581	0.97108
43	0.98828	0.96425	0.97062
44	0.98849	0.96455	0.97018
45	0.98822	0.96490	0.97072
46	0.98851	0.96579	0.97062
47	0.98854	0.96684	0.97094
48	0.98851	0.96839	0.97116
49	0.98854	0.96896	0.97084
50	0.98864	0.97008	0.97134
51	0.98835	0.97012	0.97111
52	0.98840	0.97068	0.97115
53	0.98844	0.97173	0.97118
54	0.98846	0.97308	0.97131
55	0.98829	0.97387	0.97112
56	0.98870	0.97501	0.97132
57	0.98857	0.97551	0.97122
58	0.98877	0.97593	0.97099
59	0.98895	0.97650	0.97130
60	0.98883	0.97677	0.97102
61	0.98884	0.97682	0.97112
62	0.98849	0.97597	0.97109
63	0.98858	0.97566	0.97097
64	0.98894	0.97673	0.97103
65	0.98854	0.97713	0.97117
66	0.98893	0.97788	0.97116
67	0.98878	0.97768	0.97103
68	0.98893	0.97673	0.97110
69	0.98870	0.97676	0.97069
70	0.98907	0.97693	0.97072

APPENDIX A: SAMPLE SINGLE EMISSION FILE

Computed Concentrations of CO, CO2 and HC...

No.	CO	CO2	HC
1	0.09489	1.22062	0.09959
2	0.09346	1.37817	0.10054
3	0.09073	1.38067	0.10038
4	0.08817	1.35959	0.09998
5	0.08433	1.20680	0.10061
6	0.08574	1.18388	0.10005
7	0.08757	0.90642	0.10066
8	0.08629	0.73784	0.09990
9	0.08769	0.65551	0.10154
10	0.08400	0.67430	0.10142
11	0.08646	0.65435	0.10085
12	0.08458	0.68615	0.10053
13	0.08672	0.71330	0.10082
14	0.08673	0.81537	0.10117
15	0.08713	0.81893	0.10094
16	0.08585	0.84785	0.10138
17	0.08733	0.87100	0.10193
18	0.08749	0.90717	0.10153
19	0.08609	0.89406	0.10030
20	0.08498	0.88313	0.09989
21	0.08574	0.88177	0.10065
22	0.08374	0.85031	0.09985
23	0.08718	0.84941	0.10005
24	0.08891	0.84329	0.10033
25	0.08839	0.81434	0.10089
26	0.08845	0.75663	0.10030
27	0.08445	0.67991	0.10010
28	0.08841	0.62832	0.09954
29	0.08665	0.57338	0.09946
30	0.08581	0.55294	0.09837
31	0.08705	0.55760	0.09902
32	0.08538	0.54545	0.09845
33	0.08854	0.53555	0.09918
34	0.08646	0.51745	0.09965
35	0.08601	0.52415	0.10018
36	0.08241	0.52965	0.10033
37	0.08218	0.52705	0.10032
38	0.08521	0.51867	0.10026
39	0.08478	0.51766	0.09989
40	0.08505	0.52332	0.09941

APPENDIX A: SAMPLE SINGLE EMISSION FILE

41	0.08726	0.53020	0.10017
42	0.08489	0.53179	0.09902
43	0.08821	0.55273	0.10054
44	0.08654	0.54882	0.10197
45	0.08873	0.54399	0.10018
46	0.08641	0.53207	0.10054
47	0.08613	0.51781	0.09946
48	0.08641	0.49665	0.09874
49	0.08614	0.48883	0.09981
50	0.08533	0.47351	0.09814
51	0.08770	0.47293	0.09890
52	0.08726	0.46524	0.09877
53	0.08697	0.45067	0.09870
54	0.08682	0.43192	0.09825
55	0.08813	0.42073	0.09887
56	0.08489	0.40482	0.09822
57	0.08588	0.39775	0.09854
58	0.08429	0.39171	0.09930
59	0.08281	0.38365	0.09829
60	0.08381	0.37980	0.09922
61	0.08370	0.37906	0.09889
62	0.08654	0.39119	0.09897
63	0.08581	0.39554	0.09937
64	0.08293	0.38028	0.09917
65	0.08612	0.37470	0.09871
66	0.08300	0.36390	0.09874
67	0.08424	0.36671	0.09919
68	0.08300	0.38033	0.09894
69	0.08484	0.37994	0.10030
70	0.08188	0.37745	0.10018

Appendix B: Sample Vehicle Emissions Data File

Vehicle No	Date	Time	Number of Detector Channels	License Plate	CO (%)	CO2 (%)	HC (ppm)	CO/CO_2	HC/CO_2	Axle 1 Speed	Axle 2 Speed	Acceleration
123	4/6/98	16:03:38	4	FSB7601	0.36	14.80	692.00	0.0247	0.0047	38.83	39.07	1.75
124	4/6/98	16:04:01	4	JWV1697	0.12	14.97	144.00	0.0082	0.0010	34.75	35.23	2.81
125	4/6/98	16:04:15	4	NOPLATE	999	999	99999	0.0114	0.0014	32.99	33.57	1.92
126	4/6/98	16:04:32	4	JXL2211	0.12	14.97	217.00	0.0083	0.0015	32.88	33.42	3.11
127	4/6/98	16:05:04	4	DTK9355	0.14	14.96	251.00	0.0093	0.0017	38.07	38.64	3.64
128	4/6/98	16:05:14	4	LPV1100	2.24	13.46	539.00	0.1663	0.0040	42.72	43.49	5.12
129	4/6/98	16:05:19	4	NOPLATE	0.38	14.79	228.00	0.0255	0.0015	38.90	39.64	3.71
130	4/6/98	16:05:48	4	NOPLATE	999	999	99999	9.99	9.99	24.24	24.64	1.30
131	4/6/98	16:05:53	4	NOPLATE	0.99	14.40	99999.00	0.0684	0.0288	27.43	27.69	0.93
132	4/6/98	16:06:08	4	HPJ4682	0.39	14.78	297.00	0.0264	0.0020	33.61	34.04	2.47
133	4/6/98	16:06:19	4	NOPLATE	999	999	99999	1.9874	0.0171	99	99	999
134	4/6/98	16:06:27	4	NOPLATE	999	999	99999	0.5035	-0.0616	37.73	38.42	2.23
135	4/6/98	16:07:15	4	NOPLATE	0.54	14.67	368.00	0.0369	0.0025	39.35	39.89	3.80
136	4/6/98	16:07:29	4	HWX5189	0.36	14.80	606.00	0.0245	0.0041	36.47	37.13	3.94
137	4/6/98	16:08:07	4	NOPLATE	0.37	14.79	306.00	0.0252	0.0021	4.00	40.74	4.33
138	4/6/98	16:08:28	4	KXD5278	0.94	14.39	395.00	0.0650	0.0027	39.60	40.08	3.12
139	4/6/98	16:08:35	4	NOPLATE	0.15	14.95	264.00	0.0103	0.0018	37.42	3.00	3.47
140	4/6/98	16:08:44	4	NOPLATE	0.47	14.71	86.00	0.0322	0.0006	40.62	41.51	4.80
141	4/6/98	16:08:51	4	HYN2953	1.35	14.09	312.00	0.0956	0.0022	4.00	40.60	3.86
142	4/6/98	16:09:21	4	NOPLATE	0.55	14.68	99999	0.0372	0.0077	34.99	35.52	2.35
143	4/6/98	16:09:24	4	NOPLATE	0.41	14.76	169.00	0.0281	0.0011	35.02	35.74	4.55
144	4/6/98	16:09:40	4	JZB2079	0.51	14.69	419.00	0.0347	0.0029	30.32	30.83	2.29

APPENDIX C: SAMPLE MOBILE55A OUTPUT

0Scenario title. Minimum Temp: 51. (F) Maximum Temp: 74. (F)

Period 1 RVP: 9.0 Period 2 RVP: 8.7 Period 2 Start Yr: 1999

0Total HC emission factors include evaporative HC emission factors.

0Emission factors are as of Jan. 1st of the indicated calendar year.

0User supplied veh registration distributions.

0Cal. Year: 1999 I/M Program: No Ambient Temp: 68.1 / 68.1 / 68.1 (F) Region: Low

Anti-tam. Program: No Operating Mode: 20.6 / 27.3 / 20.6 Altitude: 500. Ft.

Reformulated Gas: No

0 Veh. Type:	LDGV	LDGT1	LDGT2	LDGT	HDGV	LDDV	LDDT	HDDV	MC	All Veh
Veh. Speeds:	44.4	44.4	44.4	44.4	44.4	44.4	44.4	44.4		
VMT Mix:	0.605	0.191	0.086	0.040	0.001	0.001	0.070	0.006		

0Composite Emission Factors (Gm/Mile)

Total HC:	1.11	1.34	1.52	1.40	1.49	0.31	0.40	1.12	2.49	1.212
Exhaust HC:	0.71	0.90	1.07	0.95	0.65	0.31	0.40	1.12	1.25	0.806
Evaporat HC:	0.13	0.13	0.13	0.13	0.39				0.99	0.133
Refuel L HC:	0.14	0.19	0.19	0.19	0.31					0.152
Runing L HC:	0.09	0.09	0.10	0.10	0.11					0.088
Rsting L HC:	0.04	0.03	0.03	0.03	0.03				0.25	0.033
Exhaust CO:	7.55	9.41	11.04	9.92	11.14	0.74	0.80	5.28	9.08	8.184
Exhaust NOX:	1.41	1.60	1.91	1.69	5.24	1.26	1.33	8.85	1.20	2.157

Emission factors are as of July 1st of the indicated calendar year.

0User supplied veh registration distributions.

APPENDIX D : SAMPLE INTEGRATED TRAFFIC AND EMISSIONS DATA FILE

Site	Date	Time	Speed	Acce.	EFCO (gpg)	EFHC (gpg)	Type	Year	PlatPos	RelPos	Headway	Disp(L)	IND	AIR	ECE	CLL	TAC	GVWR	Make	Series	Body	Manufac.	Country	CYL	ASP	EVP	EGR	PCV
1	24-Apr	7:43:32	33.09	6.21	10.10	2.65	2	1995	1	free	5.28	2.4	-1	1	-1	1	0		NISSAN	Sentra	4S	NISSAN	MEXICO	L4	1	1	1	1
1	24-Apr	7:43:47	44.03	6.49	21.84	2.64	2	1997	1	first	15.82	3.8	-1	0	-1	1	0		PONTIAC	Grand Prix SE	4S	GM	U.S.A.	V6	1	1	1	1
1	24-Apr	7:44:03	35.19	4.59	34.56	9.63	2		2	last	2.9																	
1	24-Apr	7:44:18	42.45	12.39	40.65	8.75	3	1993	1	free	4.01	1.6	-1	0	-1	1	0		TOYOTA	Corolla	SW	TOYOTA	JAPAN	L4	1	1	1	1
1	24-Apr	7:44:25	34.99	5.69	206.17	19.46	2		1	first	7.37																	
1	24-Apr	7:44:36	36.42	10.11	106.54	14.67	3	1996	1	free	9.01	3.0	-1	0	-1	1	0	1	MERCURY	Villager Wagon	VN	FORD	U.S.A.	V6	1	1	1	1
1	24-Apr	7:44:43	39.87	8.75	66.01	6.97	2	1995	3	within	1.73	2.2	-1	0	-1	1	0		HONDA	Accord or Passport	4S	HONDA	U.S.A.	L4	1	1	1	1
1	24-Apr	7:46:21	35.55	7.82	27.95	4.39	3	1995	1	free	12.79	4.0	-1	0	-1	1	0	1	FORD	Explorer XL 2dr	SW	FORD	U.S.A.	V6	1	1	1	1
1	24-Apr	7:47:20	36.97	8.22	49.59	2.63	3	1995	1	first	44.15	3.0	-1	0	-1	1	0		MERCURY	Sable 4Dr. Stn Wgn GS	SW	FORD	U.S.A.	V6	1	1	1	1
1	24-Apr	7:47:24	39.16	11.05	95.40	8.67	2	1990	3	within	2.37	2.2	-1	0	-1	1	0		HONDA	Accord or Passport	4S	HONDA	U.S.A.	L4	1	1	1	1
1	24-Apr	7:47:42	45.25	11.95	100.39	5.20	2		1	free	12.7																	
1	24-Apr	7:47:55	33.39	3.94	1986.53	149.20	2		1	free	7.63																	
1	24-Apr	7:48:08	36.26	9.97	540.08	7.97	2	1991	1	first	12.84	2.2	-1	0	-1	1	1		PLYMOUTH	Sundance RS; & America	2S	CHRYSLEI	U.S.A.	L4	1	1	1	1
1	24-Apr	7:48:19	37.61	7.70	39.05	4.38	2	1991	1	free	4.02	2.0	-1	0	-1	1	0		PONTIAC	Sunbird LE	CN	GM	U.S.A.	L4	1	1	1	1
1	24-Apr	7:48:29	36.97	6.50	31.26	7.02	2		1	free	4.16																	
1	24-Apr	7:48:45	40.12	11.02	137.32	9.46	2	1989	1	first	16.29	2.4	-1	1	-1	1	0		NISSAN	240SX	2S	NISSAN	JAPAN	L4	1	1	1	1
1	24-Apr	7:48:57	42.92	18.05	23.49	5.27	2	1994	1	free	9.03	1.6	-1	0	-1	1	0		NISSAN	Maxima	4S	NISSAN	JAPAN	L4	1	1	1	1
1	24-Apr	7:49:05	40.24	9.04	34.57	8.76	3	1994	1	first	4.19	3.8	-1	1	-1	1	0	1	CHRYSLER	Town & Country	SW	CHRYSLEI	U.S.A.	V6	1	1	1	1
1	24-Apr	7:49:10	40.06	13.52	62.17	6.98	2		4	last	1.4																	
1	24-Apr	7:49:34	39.60	9.59	261.97	20.92	2	1986	1	first	23.87	1.5	1	1	-1	1	0		HONDA	Civic	4S	HONDA	JAPAN	L4	1	1	1	1
1	24-Apr	7:49:41	34.11	9.13	106.13	11.23	3	1994	1	free	4.7	4.0	-1	0	-1	1	0	1	FORD			FORD	U.S.A.	V6	1	1	1	1
1	24-Apr	7:49:48	38.64	13.63	58.34	6.11	2	1993	2	within	4.89	1.9	-1	0	-1	1	0		FORD	;cort 4 Dr. Sedan Notchback	4S	FORD	MEXICO	L4	1	1	1	1
1	24-Apr	7:49:53	39.25	11.57	90.33	14.71	3	1996	1	first	3.3	5.7	-1	0	-1	1	0	0	CHEVROLET	Conventional Cab GMT400 4x	SW	GM	MEXICO	V8	1	1	1	1
1	24-Apr	7:49:59	41.59	12.77	838.33	18.67	2	1992	1	first	5.13	1.9	-1	0	-1	1	0		FORD	;cort 4 Dr. Sedan Hatchback	4S	FORD	U.S.A.	L4	1	1	1	1
1	24-Apr	7:50:15	36.49	7.20	50.10	5.25	2		1	first	13.64																	
1	24-Apr	7:50:19	37.98	10.06	118.34	15.50	2	1997	3	last	2.67	2.2	-1	0	-1	1	0		TOYOTA	Camry	4S	TOYOTA	U.S.A.	L4	1	1	1	1
1	24-Apr	7:50:45	40.22	13.27	354.28	12.37	2	1986			26.59	1.6	1	1	-1	1	1		TOYOTA	Corolla	4S	TOYOTA	JAPAN	L4	1	1	1	1
1	24-Apr	7:51:08	35.79	9.39	26.83	5.27	2	1990	1	first	13.79	1.6	-1	1	-1	1	0		NISSAN	Sentra	2S	NISSAN	U.S.A.	L4	1	1	1	1
1	24-Apr	7:51:19	36.33	9.81	120.76	8.63	2		1	first	3.38																	
1	24-Apr	7:51:27	38.50	11.35	124.52	8.62	2	1994	1	free	5.88	2.5	-1	0	-1	1	0		MAZDA	626/MX6	4S	MAZDA	U.S.A.	V6	1	1	0	1
1	24-Apr	7:51:54	41.91	14.24	42.29	10.49	2	1995	1	free	17.55	2.8	-1	0	-1	1	0		MERCEDES	C280	4S	MERCEDES	GERMANY	L6	1	1	0	1
1	24-Apr	7:52:18	33.86	7.20	139.82	1.72	2		1	first	17.53																	
1	24-Apr	7:52:40	36.41	9.01	44.59	4.38	2	1996	1	free	17.99	1.6	-1	0	-1	1	0		NISSAN	Sentra	4S	NISSAN	MEXICO	L4	1	1	1	1
1	24-Apr	7:53:12	39.62	11.78	29.01	9.64	3	1997	3	last	2.25	3.4	-1	0	-1	1	0	1	CHEVROLET	All Purpose Vehicle	VN	GM	U.S.A.	V6	1	1	1	1
1	24-Apr	7:53:21	38.55	9.05	27.93	6.15	2	1986	1	free	8.88	3.0	-1	0	-1	1	0		OLDSMOBILI	Calais Supreme	4S	GM	U.S.A.	V6	1	1	1	1
1	24-Apr	7:54:02	34.27	10.30	40.74	2.63	2	1994	1	free	16.06	3.0	-1	0	-1	1	0		MERCURY	Sable 4Dr. Sedan GS	4S	FORD	U.S.A.	V6	1	1	1	1
1	24-Apr	7:54:55	33.09	4.15	63.61	16.51	2	1988	4	last	2.21	1.2	1	1	-1	1	1		SUBARU	GL	2S	SUBARU	JAPAN	L3	1	1	1	1
1	24-Apr	7:55:06	37.94	10.43	36.25	7.01	2		1	first	4.77																	
1	24-Apr	7:55:33	35.61	8.63	277.86	13.38	2	1988	1	free	10.43	1.6	-1	1	-1	1	1		NISSAN	Sentra	2S	NISSAN	U.S.A.	L4	1	1	1	1
1	24-Apr	7:55:39	41.42	10.85	33.47	7.89	2		2	within	4.12																	
1	24-Apr	7:55:46	37.03	5.28	25.15	6.15	3	1993	1	free	4.97	3.0	-1	1	-1	1	1	1	MERCURY	Villager Wagon	SW	FORD	U.S.A.	V6	1	1	1	1
1	24-Apr	7:56:01	29.75	2.56	404.82	13.07	2		1	free	11.52																	
1	24-Apr	7:56:15	40.64	12.18	137.98	6.88	2	1991	1	free	7.53	2.2	-1	0	-1	1	0		HONDA	Accord or Passport	4S	HONDA	U.S.A.	L4	1	1	1	1
1	24-Apr	7:56:22	29.79	5.00	37.88	9.63	2	1992	1	free	6.51	3.1	-1	0	-1	1	0		OLDSMOBILI	Cutlass Supreme Convertible	CN	GM	U.S.A.	V6	1	1	1	1
1	24-Apr	7:56:34	32.99	5.45	181.85	8.53	2	1991	1	free	12.06	1.5	-1	0	-1	1	0		TOYOTA	Tercel	4S	TOYOTA	JAPAN	L4	1	1	1	1
1	24-Apr	7:57:20	36.88	7.92	45.65	7.00	2	1995	1	free	7.2	3.5	-1	0	-1	1	0		CHRYSLER	Concorde	4S	CHRYSLEI	CANADA	V6	1	1	1	1
1	24-Apr	7:57:30	37.30	10.72	128.42	5.17	2	1992	1	first	10.07	2.4	-1	1	-1	1	0		NISSAN	Stanza	4S	NISSAN	JAPAN	L4	1	1	1	1
1	24-Apr	7:57:39	35.74	8.48	136.95	6.03	3	1991	1	first	6.76	3.3	-1	1	-1	1	0	1	PLYMOUTH	Plymouth Voyager	SW	CHRYSLEI	U.S.A.	V6	1	1	1	1
1	24-Apr	7:58:13	31.74	5.88	29.55	11.39	2		1	first	20.95																	
1	24-Apr	7:58:18	36.18	7.50	66.54	7.84	3	1990	4	last	2.51	2.9	-1	0	-1	1	1	1	FORD			FORD	U.S.A.	V6	1	1	0	1
1	24-Apr	7:58:28	33.99	7.90	73.70	6.10	2		1	free	10.58																	
1	24-Apr	7:58:39	32.76	0.72	1601.28	18.81	3	1992	1	free	4.85	2.4	-1	1	-1	1	0		NISSAN			NISSAN	U.S.A.	L4	1	1	1	1

APPENDIX E: "TIMEMATCH" MACRO MODEL CODE FILE

```
Sub TimeMatch()
' TimeMatch Macro
' Macro recorded 7/6/98 by Russell Dalton


Dim MOBTime As Long
Dim RSDTime As Long
Dim NextRSDTime As Long
Dim Diff As Long
Dim NextDiff As Long
Dim ExpDiff As Long
Dim a As Long
Dim b As Long
Dim c As Long
a = 0
b = 0
c = 0
ExpDiff = 2

Sheets("MatchedData").Select
Columns("A:B").Select
Selection.ClearContents

Range("A1").Select
ActiveCell.Value = "MOBTime"

Range("B1").Select
ActiveCell.Value = "RSDTime"

Do Until ActiveCell.Value = "stop"

Sheets("423").Select

Range("A1").Select
ActiveCell.Offset(1 + a, 0).Select
MOBTime = ActiveCell.Value

Range("B1").Select
ActiveCell.Offset(1 + b, 0).Select
RSDTime = ActiveCell.Value

Diff = MOBTime - RSDTime
ActiveCell.Offset(1, 0).Select
```

APPENDIX E: "TIMEMATCH" MACRO MODEL CODE FILE

```
NextRSDTime = ActiveCell.Value
Appendix C (Continuation)

NextDiff = MOBTime - NextRSDTime

   If (Abs(Diff) <= Abs(ExpDiff)) And (Abs(Diff) < Abs(NextDiff)) Then
      Sheets("MatchedData").Select
      Range("A1").Select
      ActiveCell.Offset(1 + c, 0).Select
      ActiveCell.Value = MOBTime

      Range("B1").Select
      ActiveCell.Offset(1 + c, 0).Select
      ActiveCell.Value = RSDTime

      a = a + 1
      c = c + 1

      End If

   If (Abs(Diff) > Abs(ExpDiff)) And (Abs(Diff) < Abs(NextDiff)) Then
      a = a + 1

   If (Abs(Diff) <= Abs(ExpDiff)) And (Abs(Diff) > Abs(NextDiff)) Then
      a = a + 1
       b = b + 1

      End If

   Else
       b = b + 1

   End If

Range("A1").Select
ActiveCell.Offset(1 + a, 0).Select

Loop

End Sub
```

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

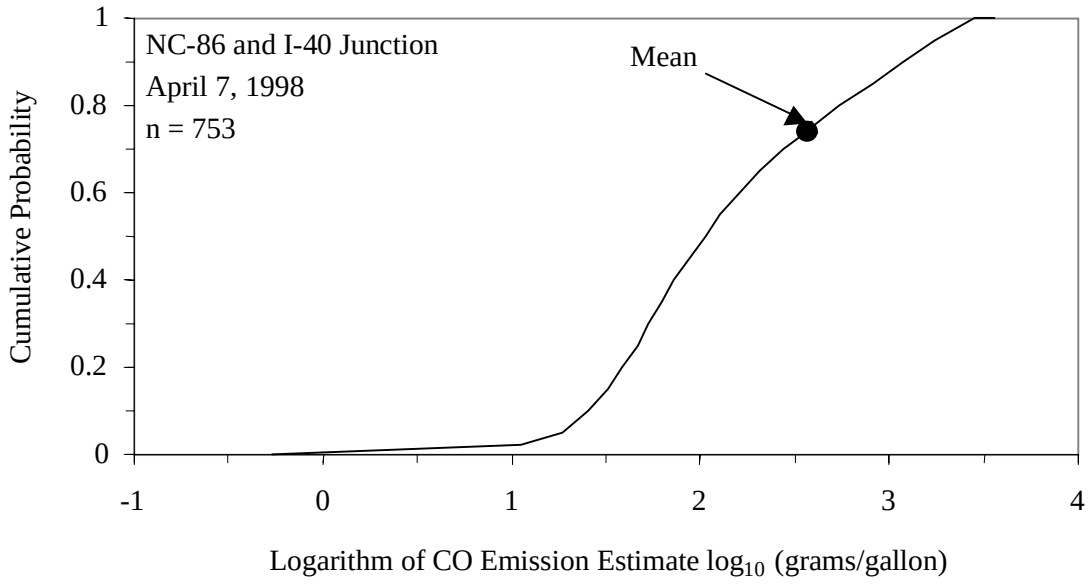


Figure F.1. CO Inter-individual variability and mean estimate for data collected on April 7.

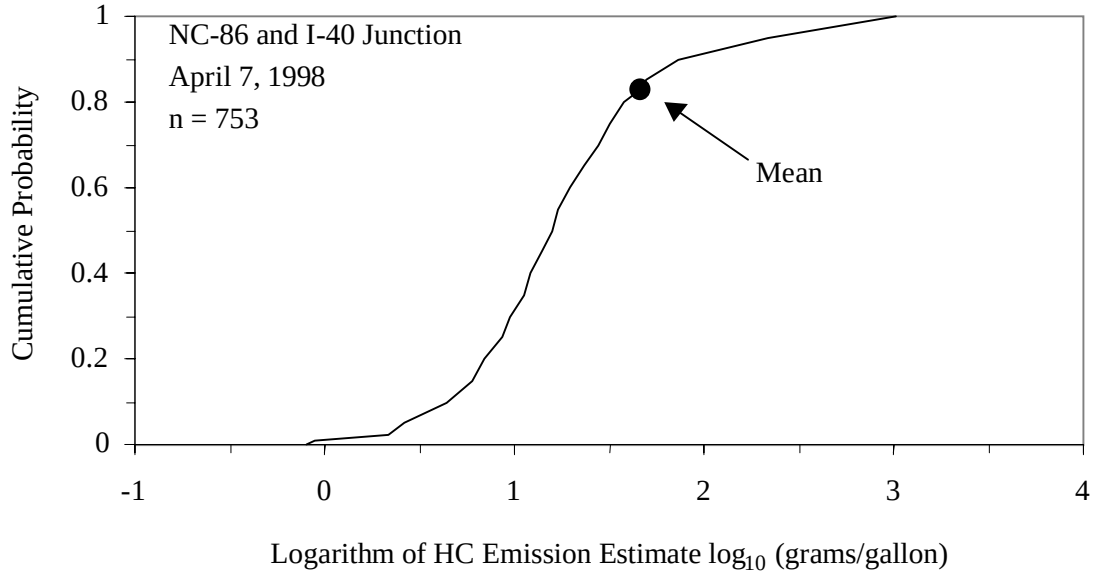


Figure F.2. HC Inter-individual variability and mean estimate for data collected on April 7.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

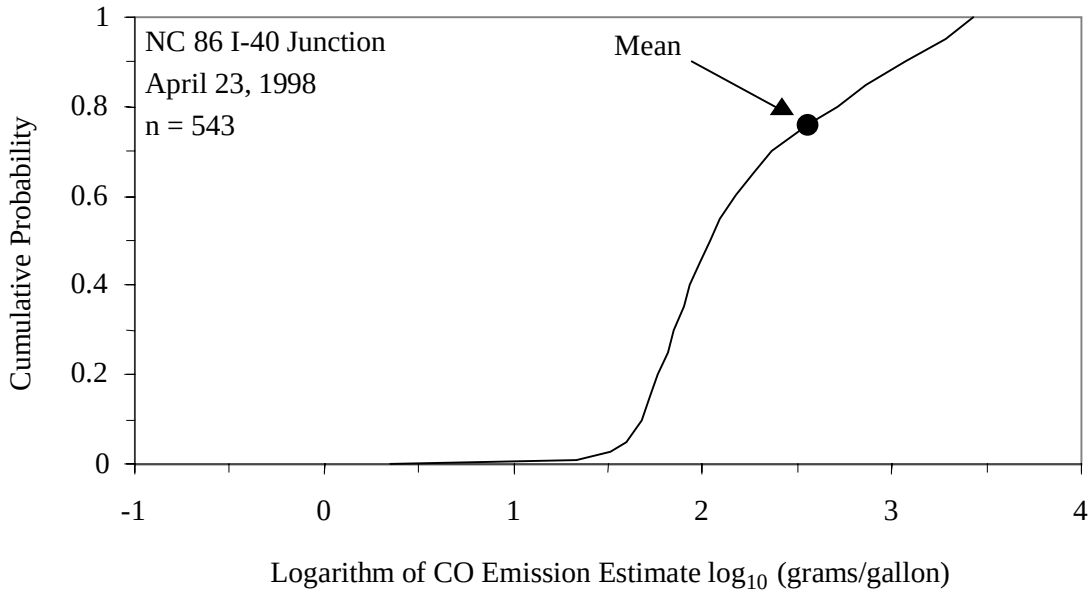


Figure F.3. CO Inter-individual variability and mean estimate for data collected on April 23.

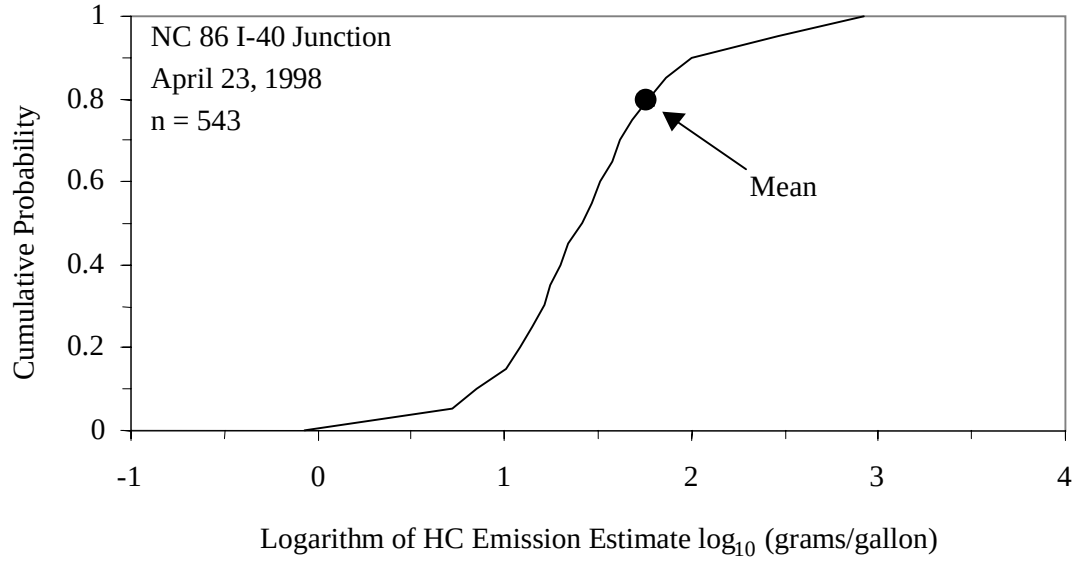


Figure F.4. HC Inter-individual variability and mean estimate for data collected on April 23.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

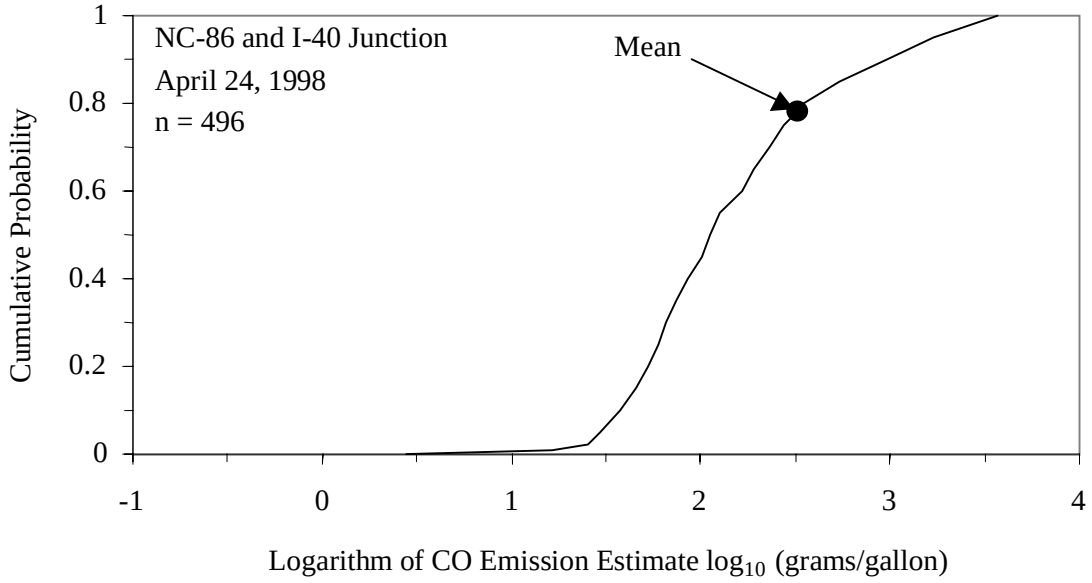


Figure F.5. CO Inter-individual variability and mean estimate for data collected on April 24.

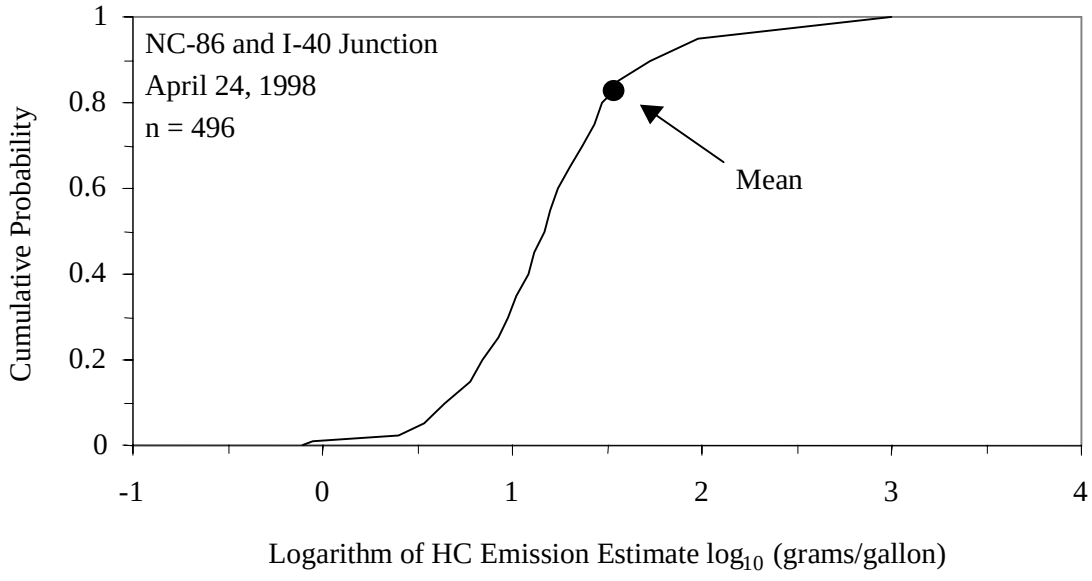


Figure F.6. HC Inter-individual variability and mean estimate for data collected on April 24.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

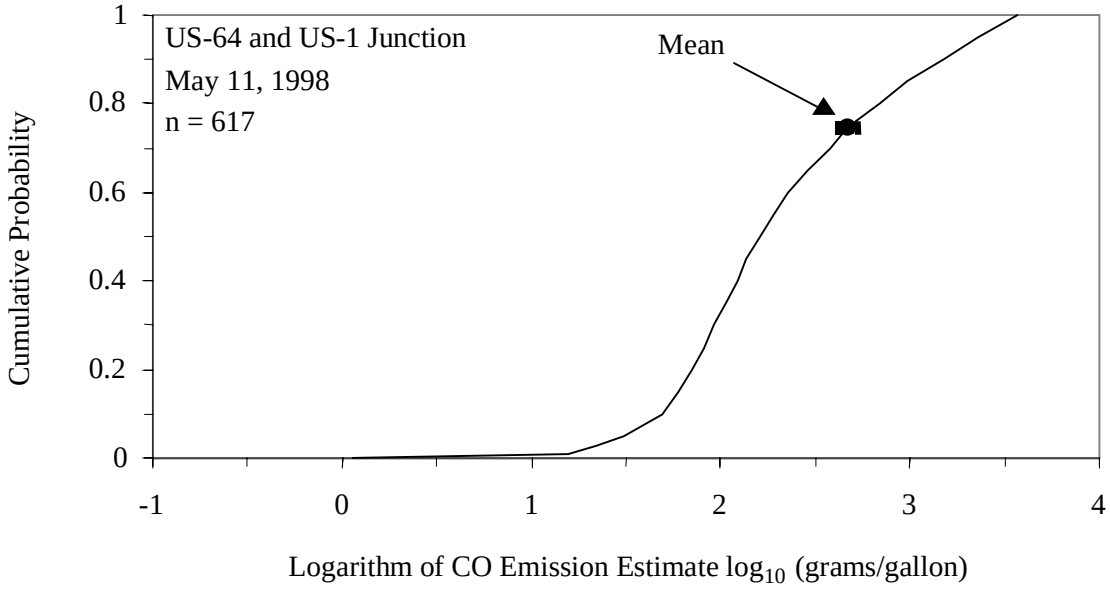


Figure F.7. CO Inter-individual variability and mean estimate for data collected on May 11.

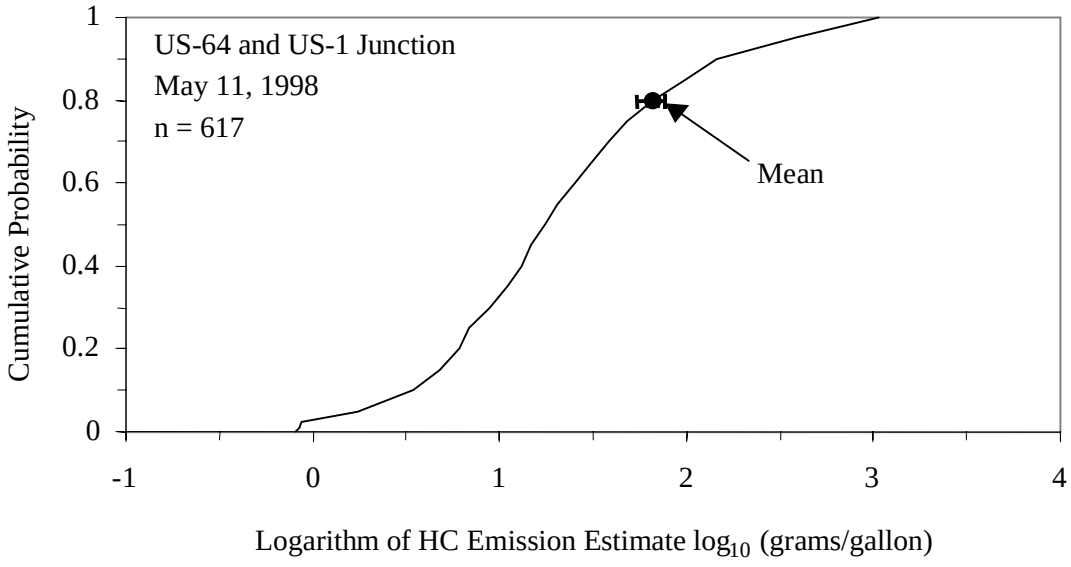


Figure F.8. HC Inter-individual variability and mean estimate for data collected on May 11.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

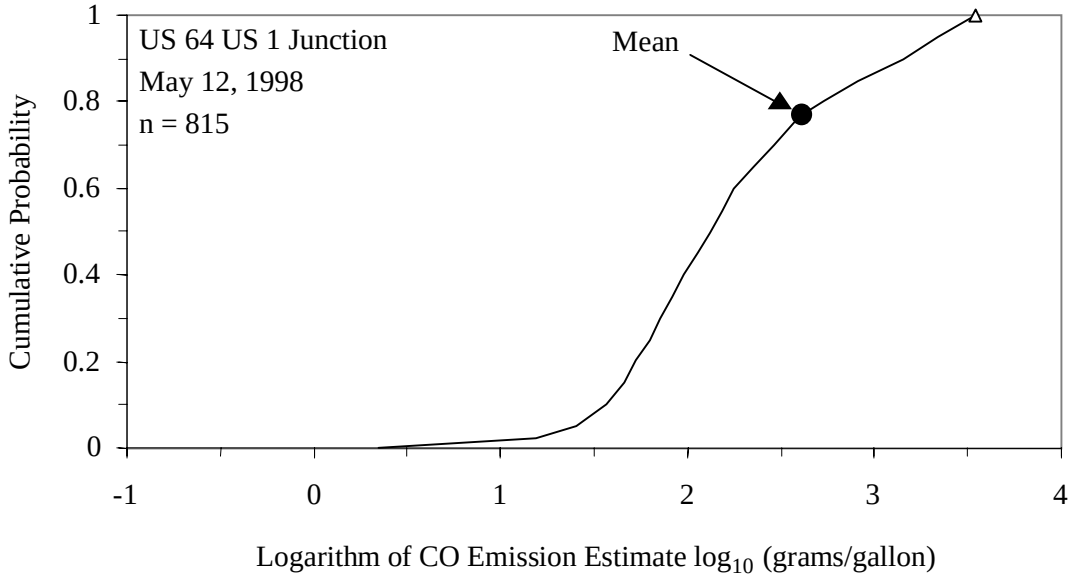


Figure F.9. CO Inter-individual variability and mean estimate for data collected on May 12.

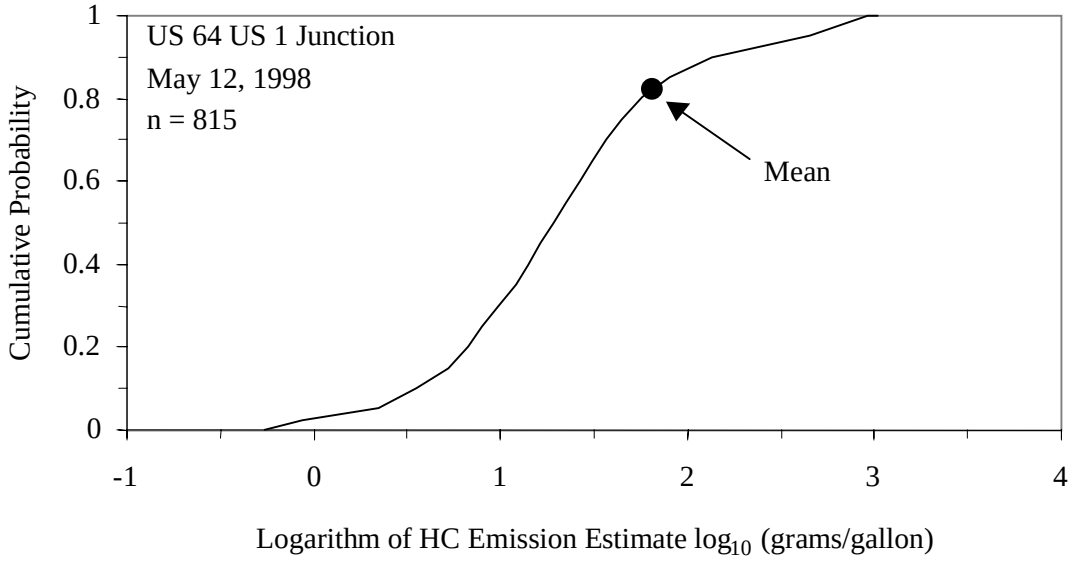


Figure F.10. HC Inter-individual variability and mean estimate for data collected on May 12.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

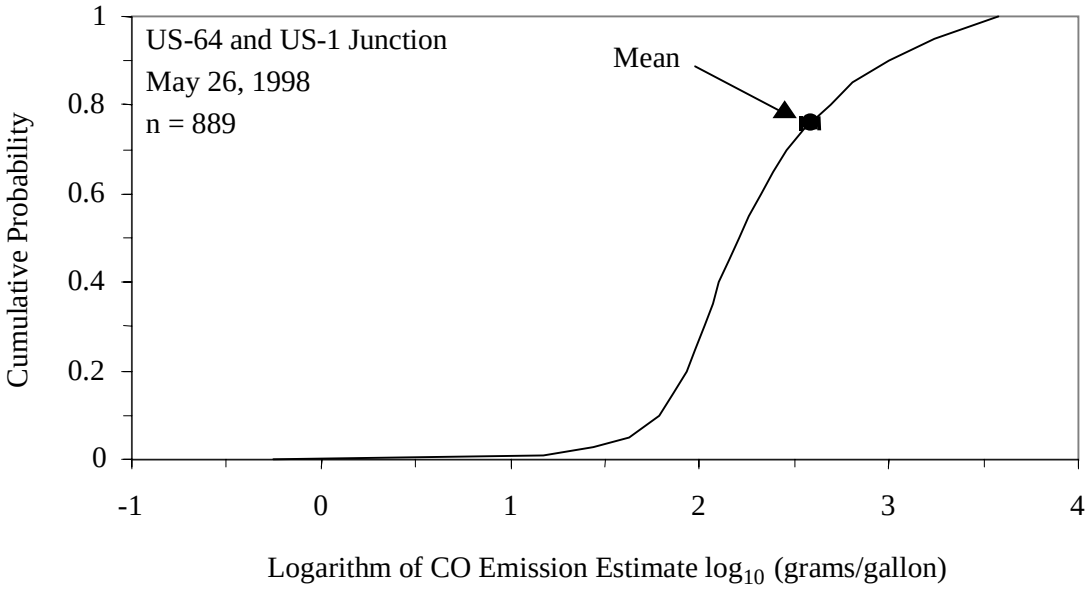


Figure F.11. CO Inter-individual variability and mean estimate for data collected on May 26.

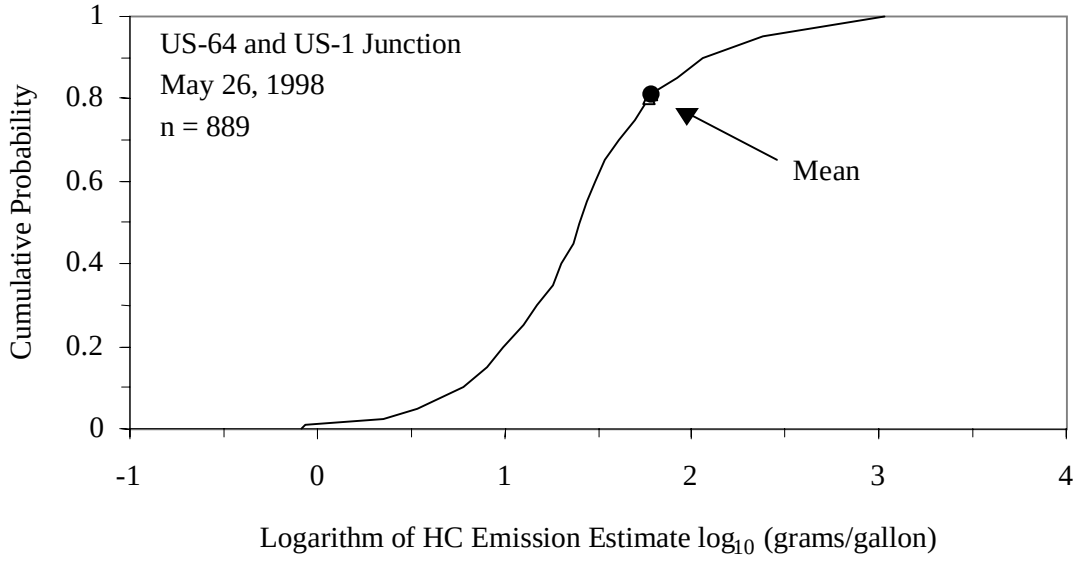


Figure F.12. HC Inter-individual variability and mean estimate for data collected on May 26.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

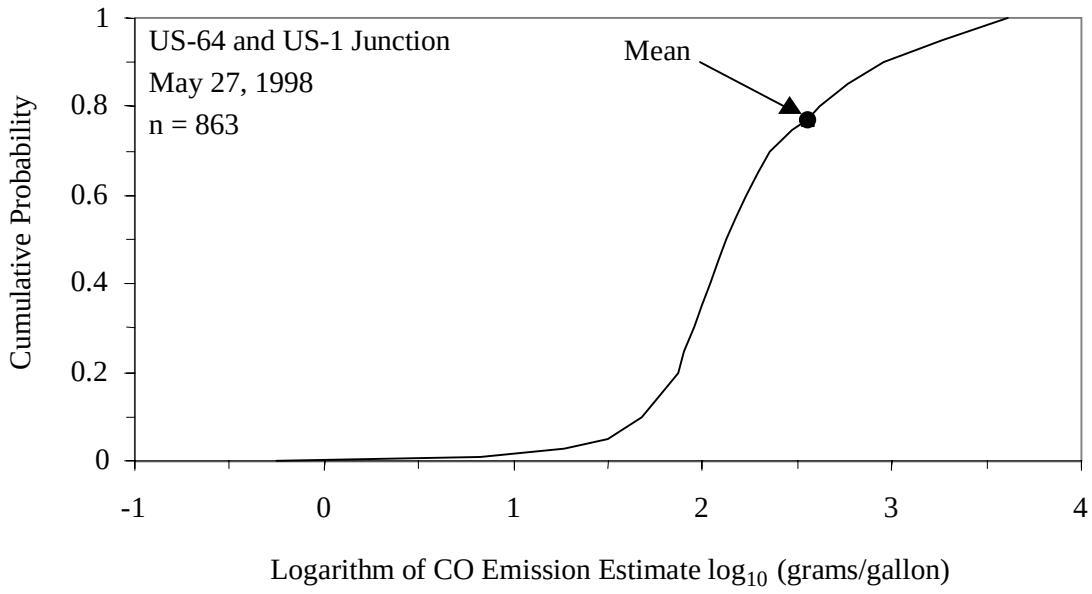


Figure F.13. CO Inter-individual variability and mean estimate for data collected on May 27.

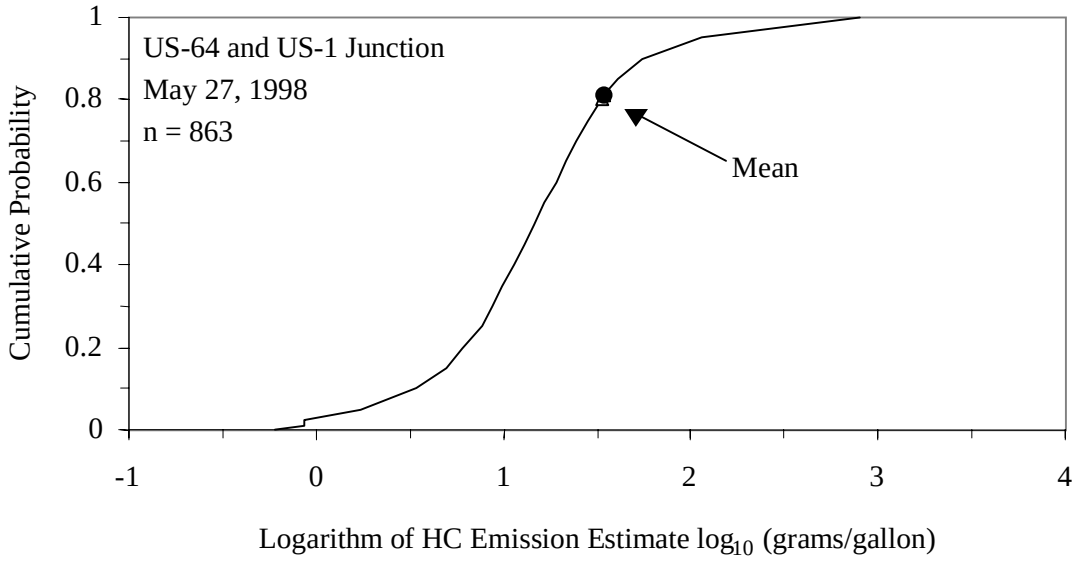


Figure F.14. HC Inter-individual variability and mean estimate for data collected on May 27.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

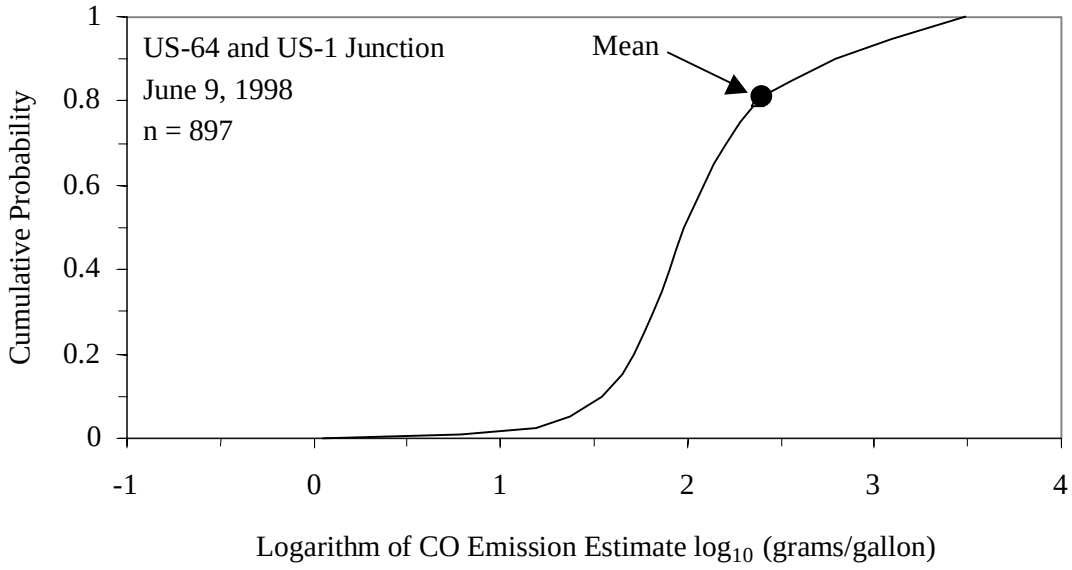


Figure F.15. CO Inter-individual variability and mean estimate for data collected on June 9.

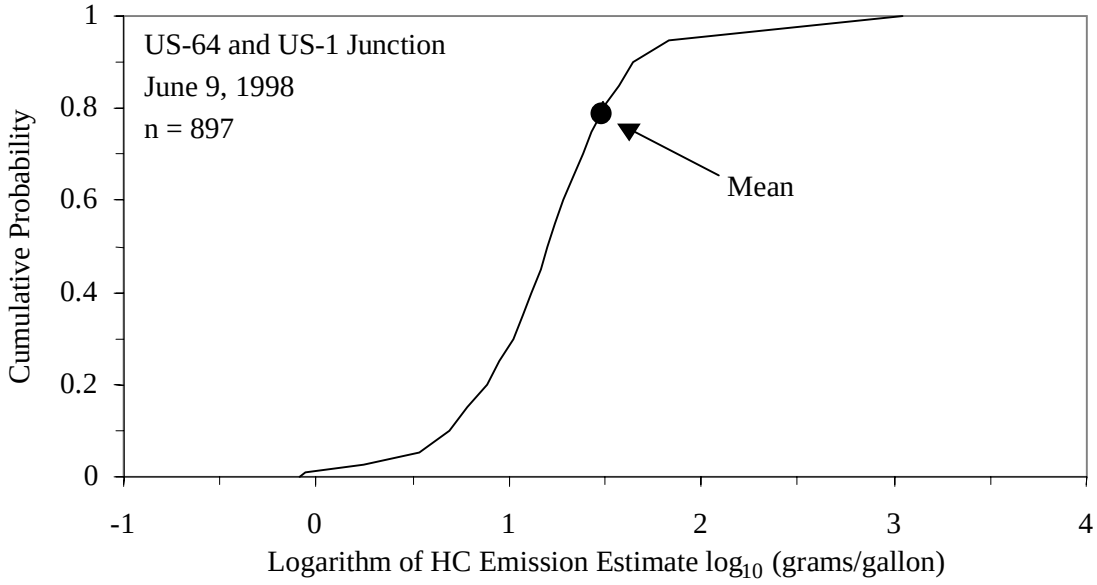


Figure F.16. HC Inter-individual variability and mean estimate for data collected on June 9.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

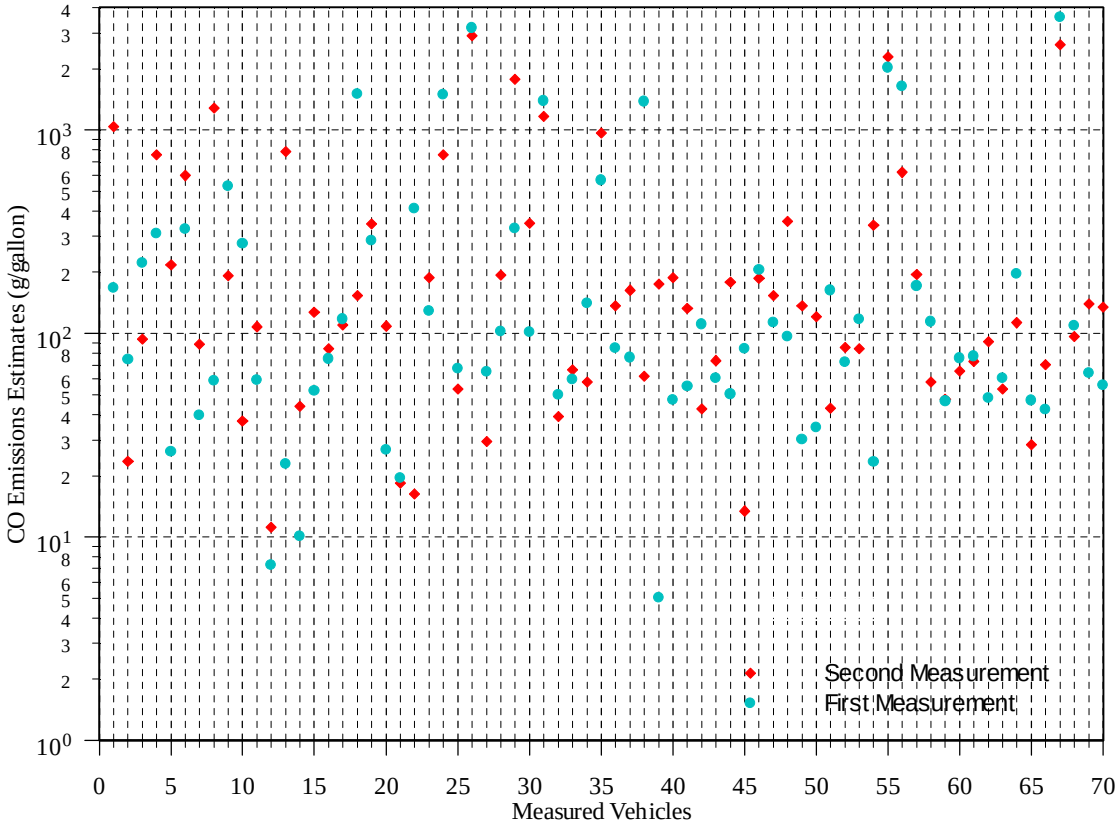


Figure F.17. Scatter plot of multiple CO measurements for first 70 data points.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

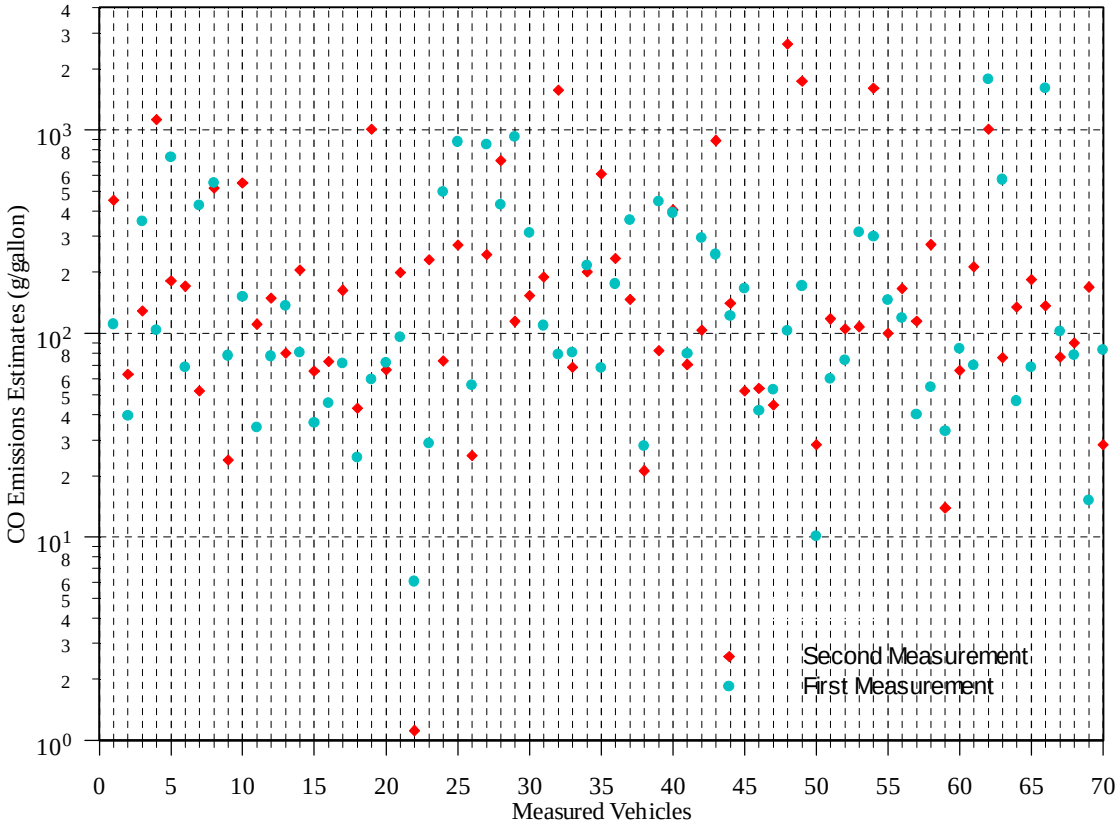


Figure F.18. Scatter plot of multiple CO measurements for data points from 70 to 140.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

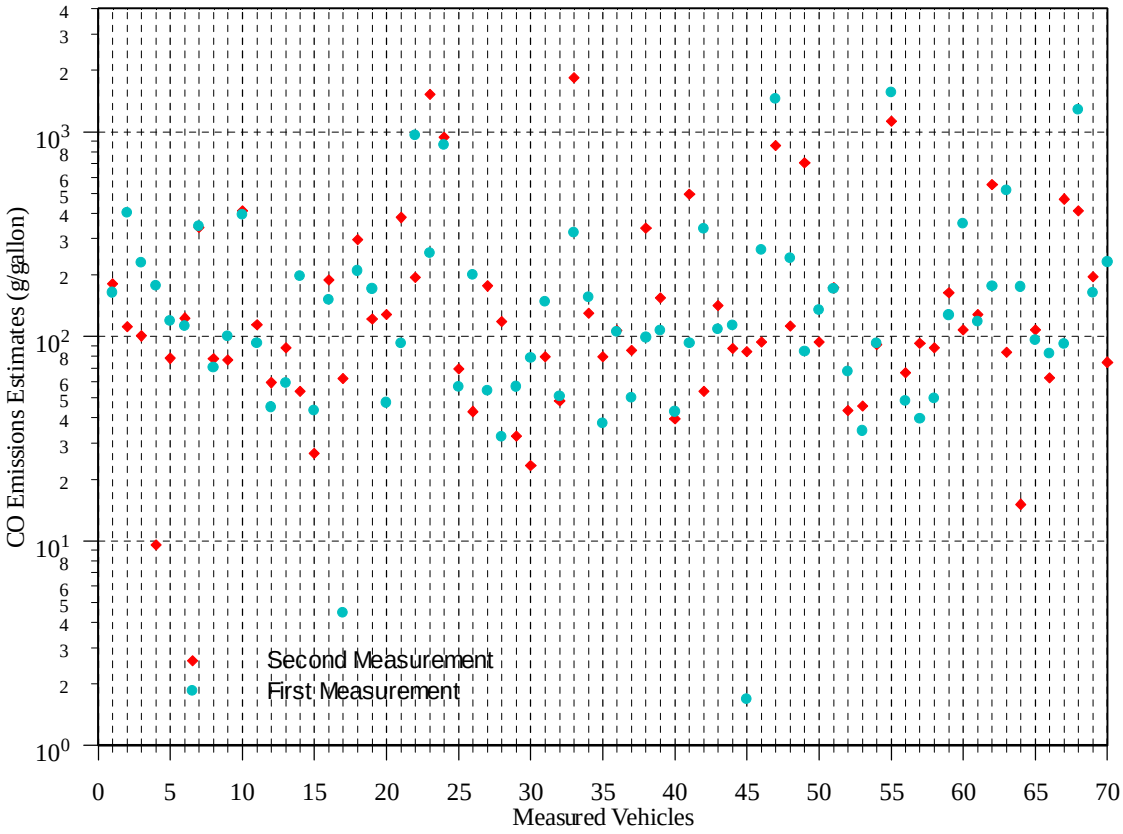


Figure F.19. Scatter plot of multiple CO measurements for the data points from 140 to 210.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

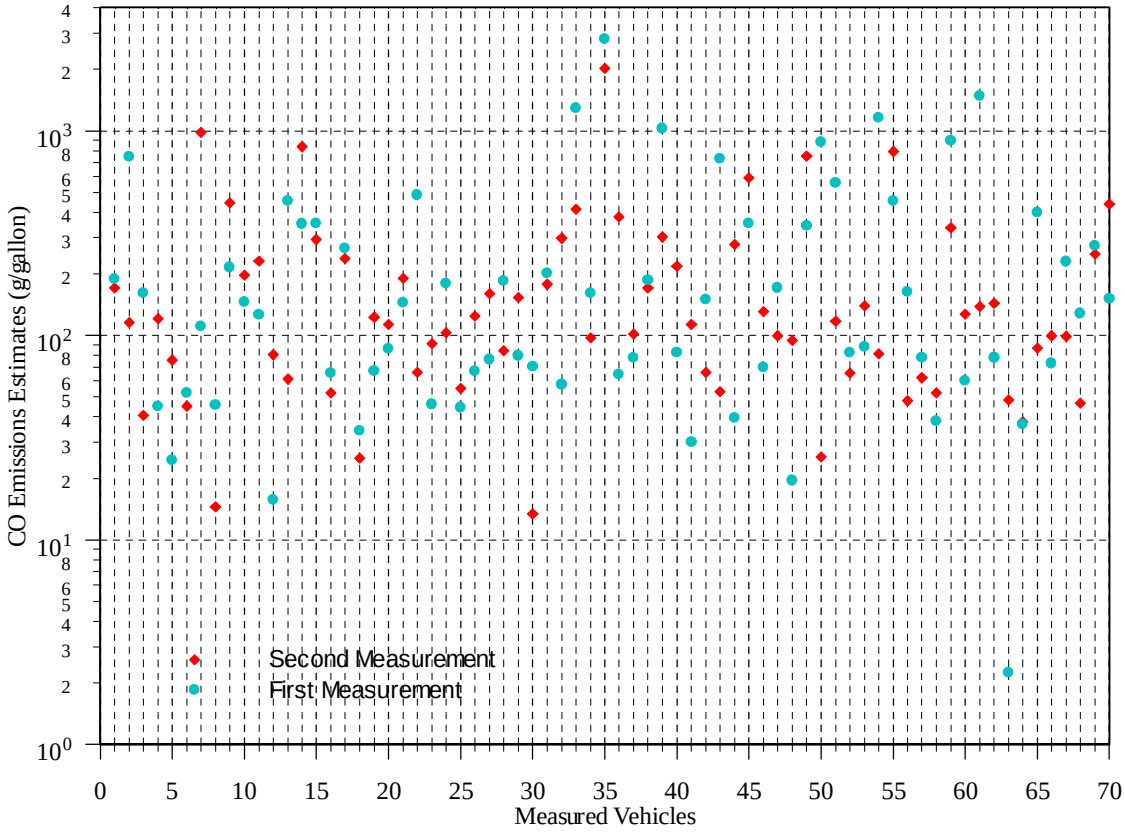


Figure F.20. Scatter plot of multiple CO measurements for the data points from 210 to 280.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

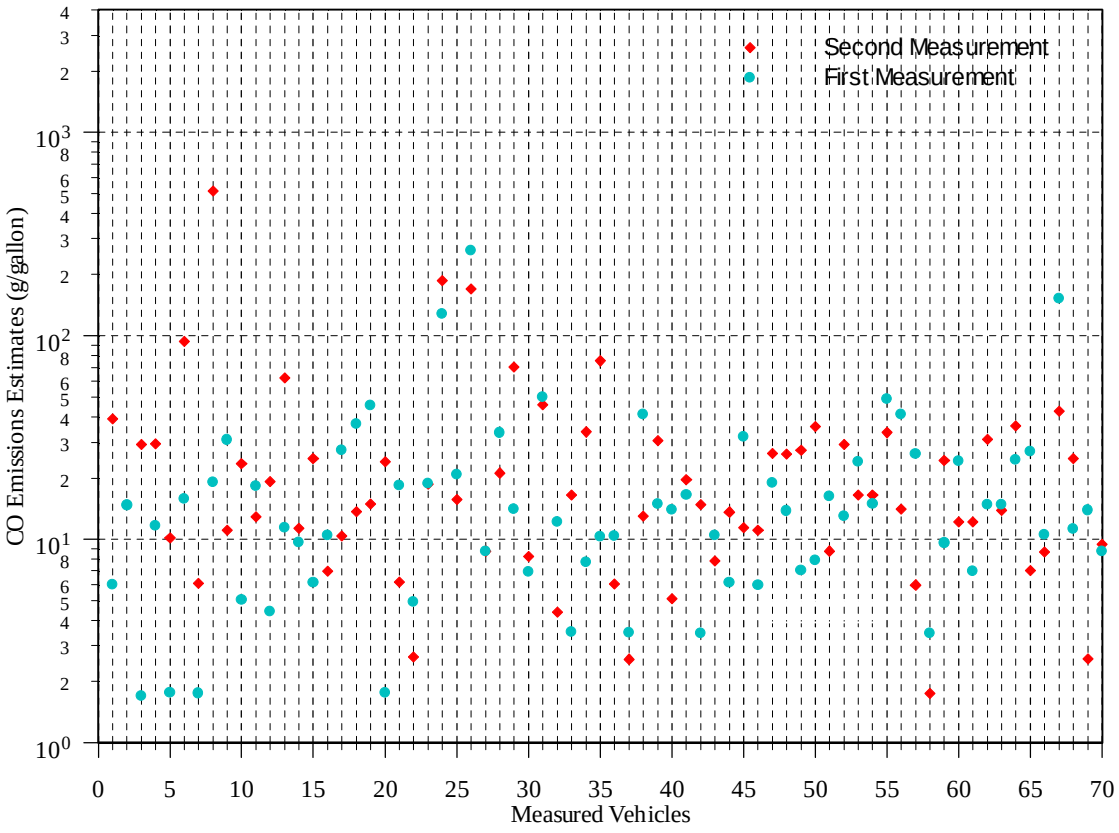


Figure F.21. Scatter plot of multiple HC measurements for first 70 data points.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

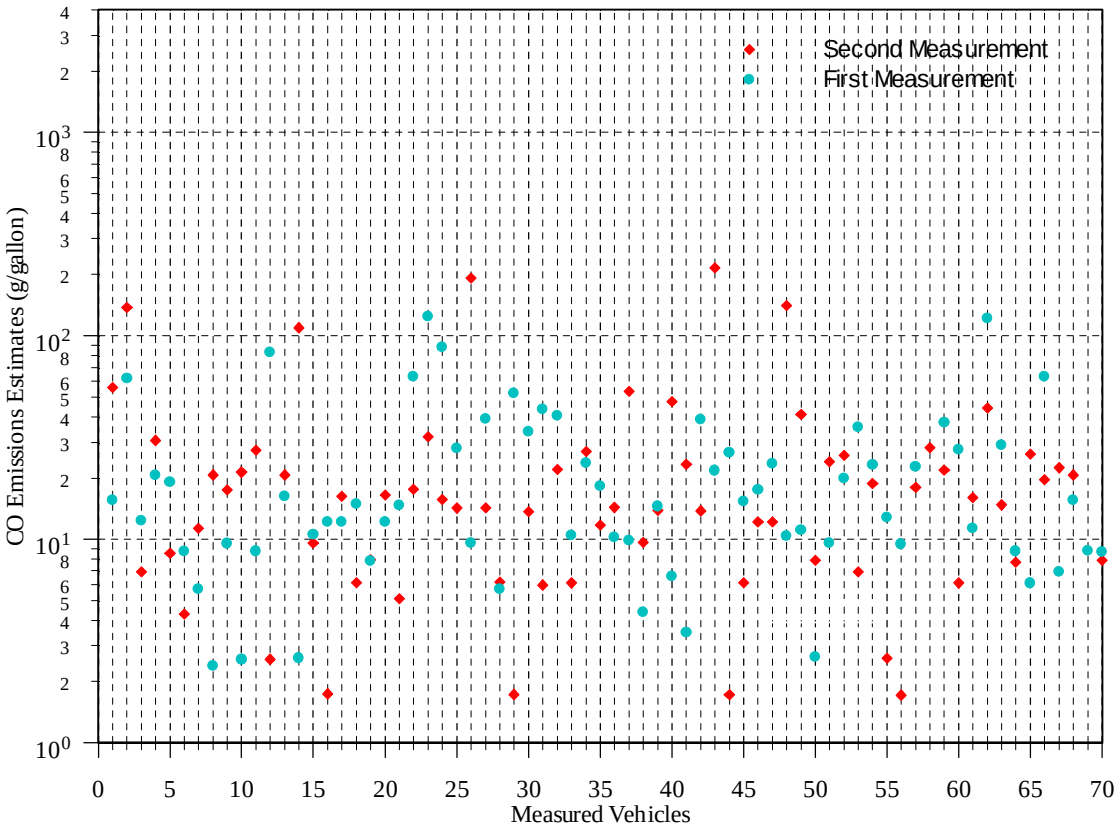


Figure F.22. Scatter plot of multiple HC measurements for data points from 70 to 140.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

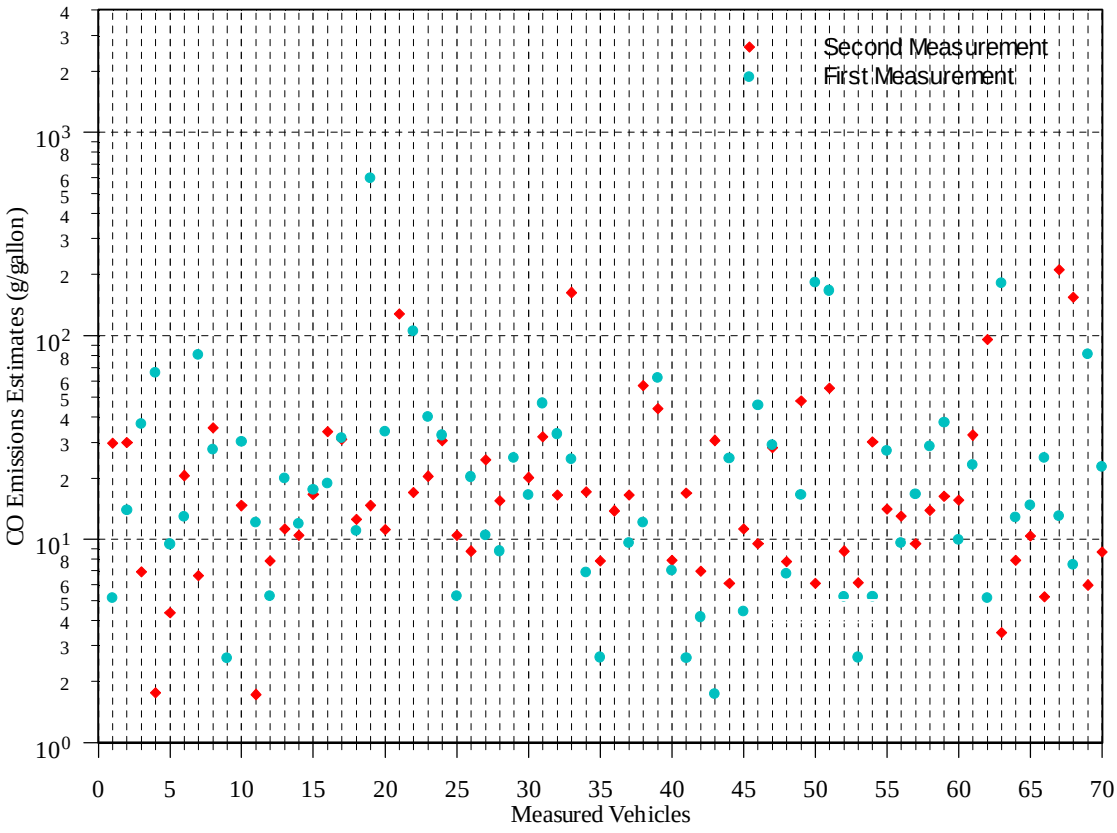


Figure F.23. Scatter plot of multiple HC measurements for data points from 140 to 210.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

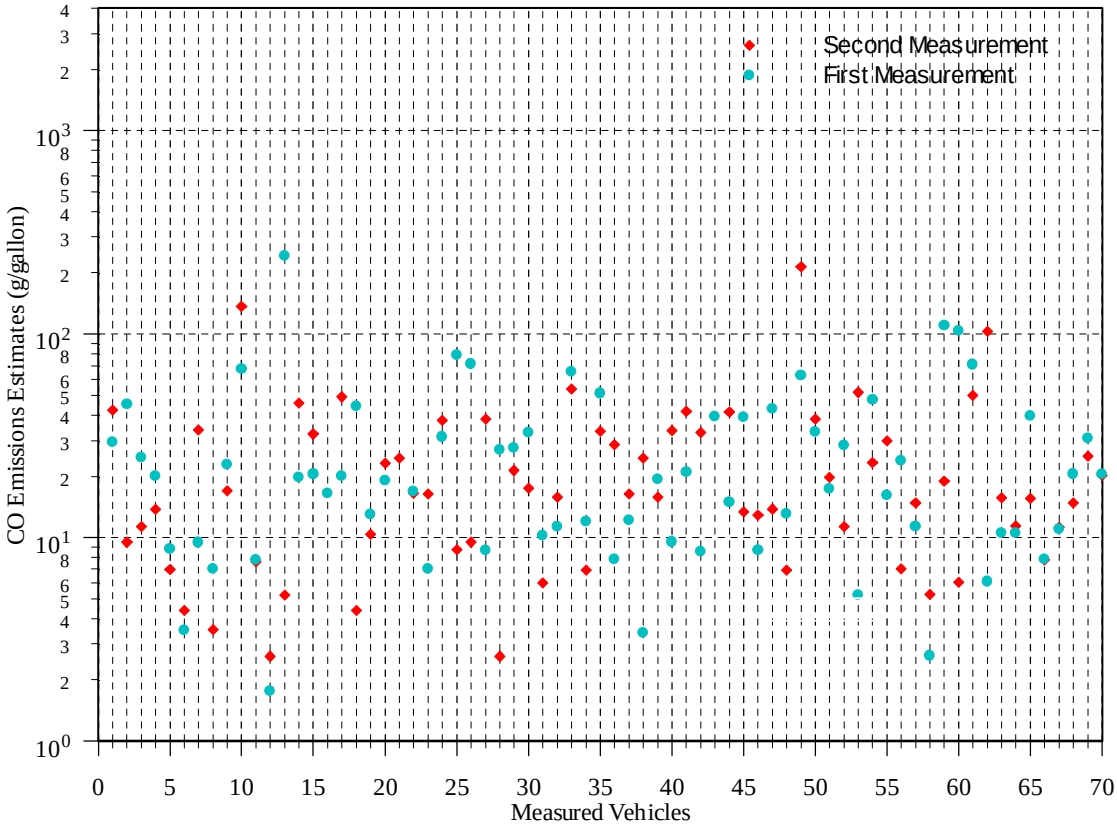


Figure F.24. Scatter plot of multiple HC measurements for data points from 210 to 280.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

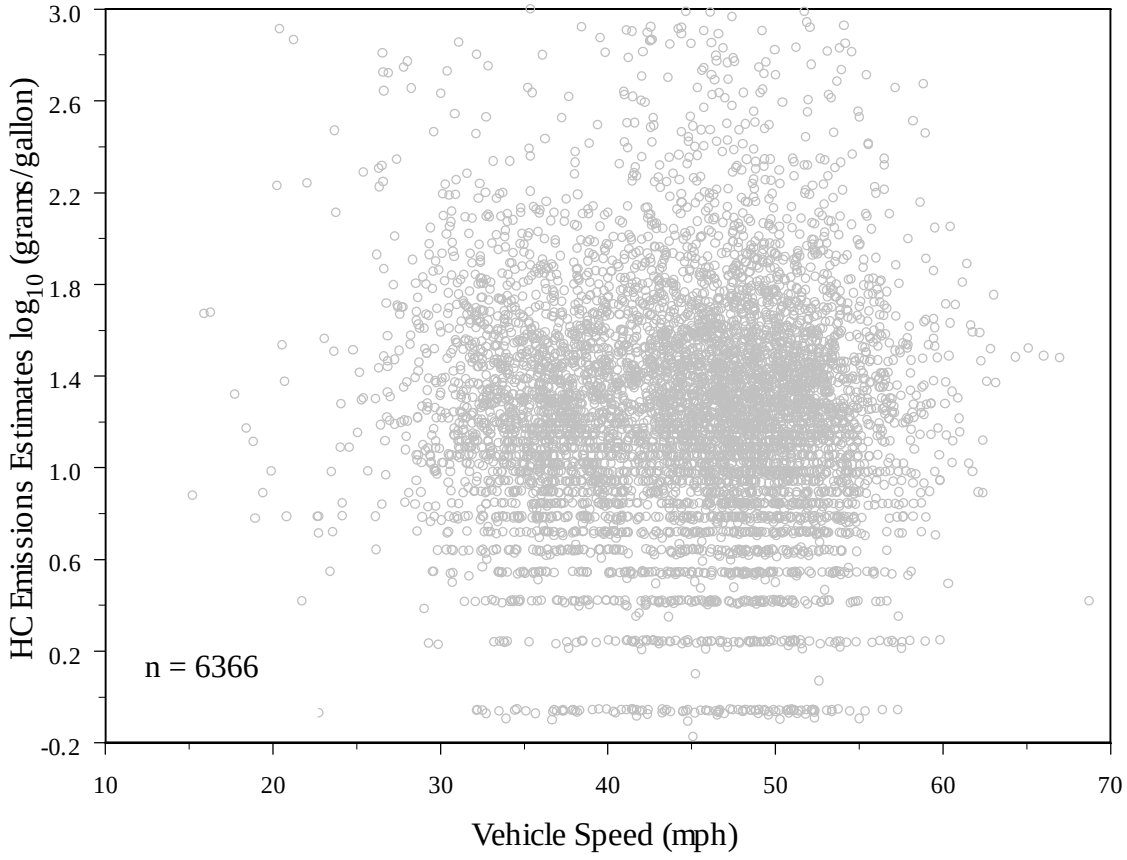


Figure F.25. Scatter plot of HC emissions estimates with respect to vehicle speed.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

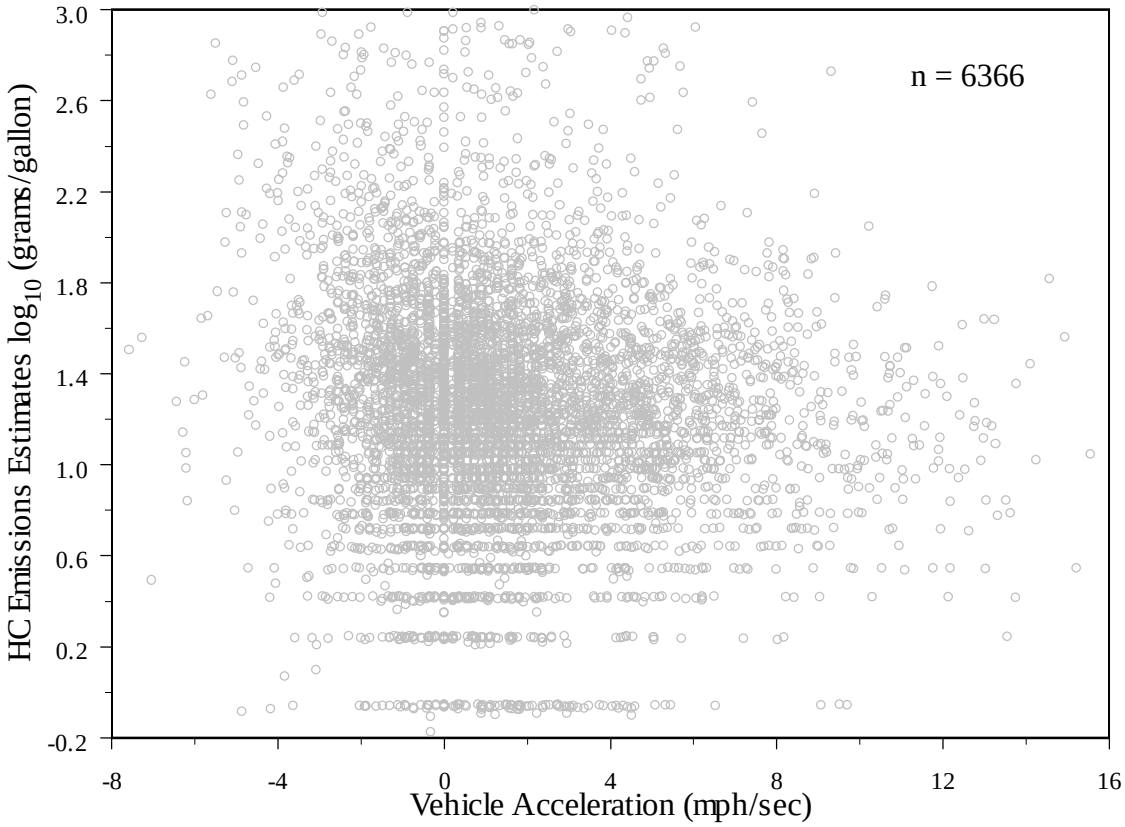


Figure F.26. Scatter plot of HC emissions estimates with respect to vehicle acceleration.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

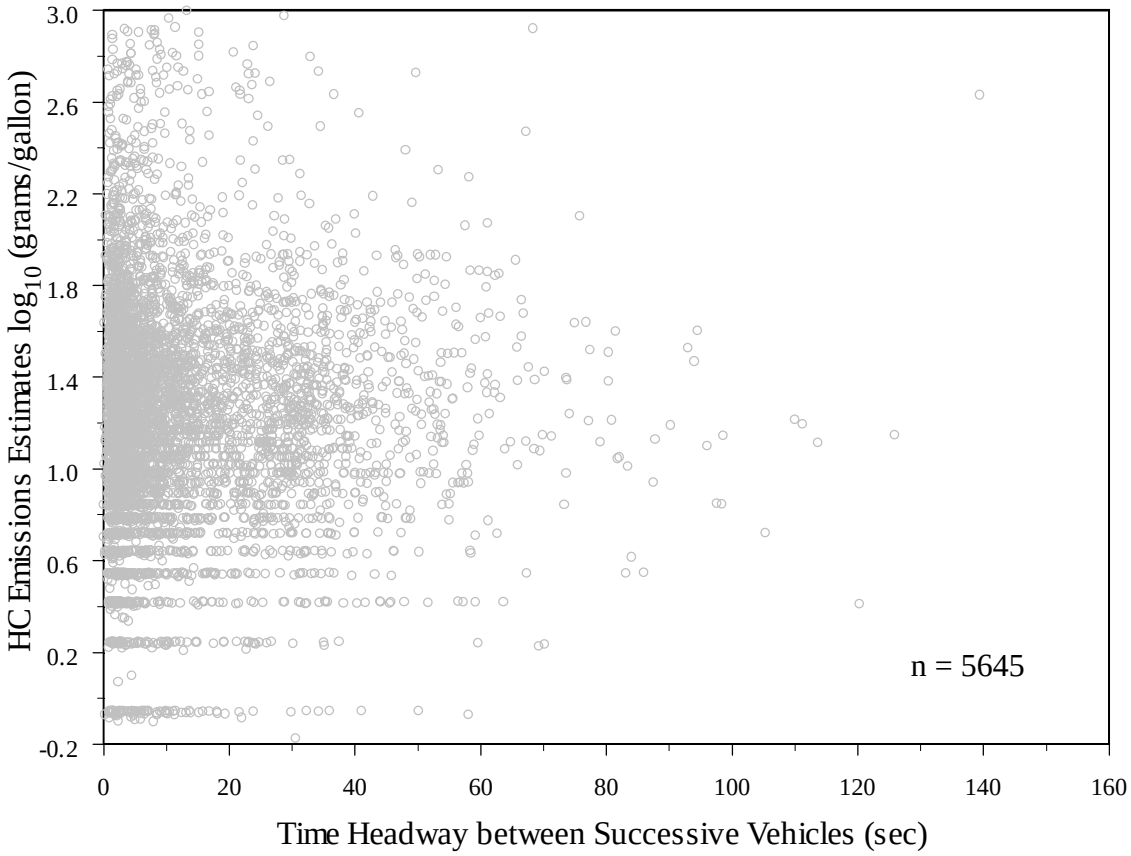


Figure F.27. Scatter plot of HC emissions estimates with respect to time headway between successive vehicles.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

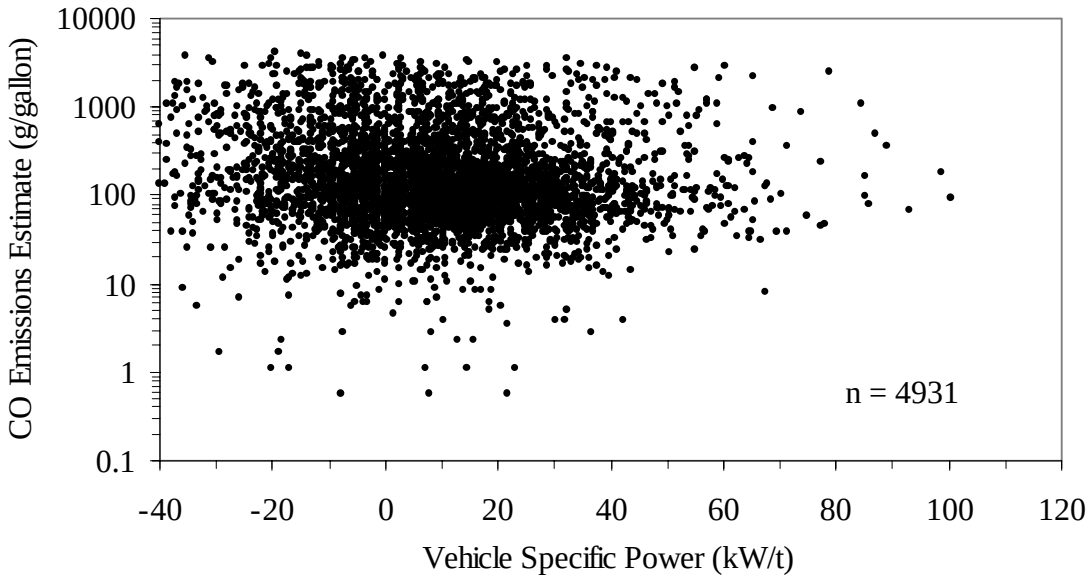


Figure F.28. Scatter plot of CO emissions estimates with respect to Vehicle Specific Power (VSP) values.

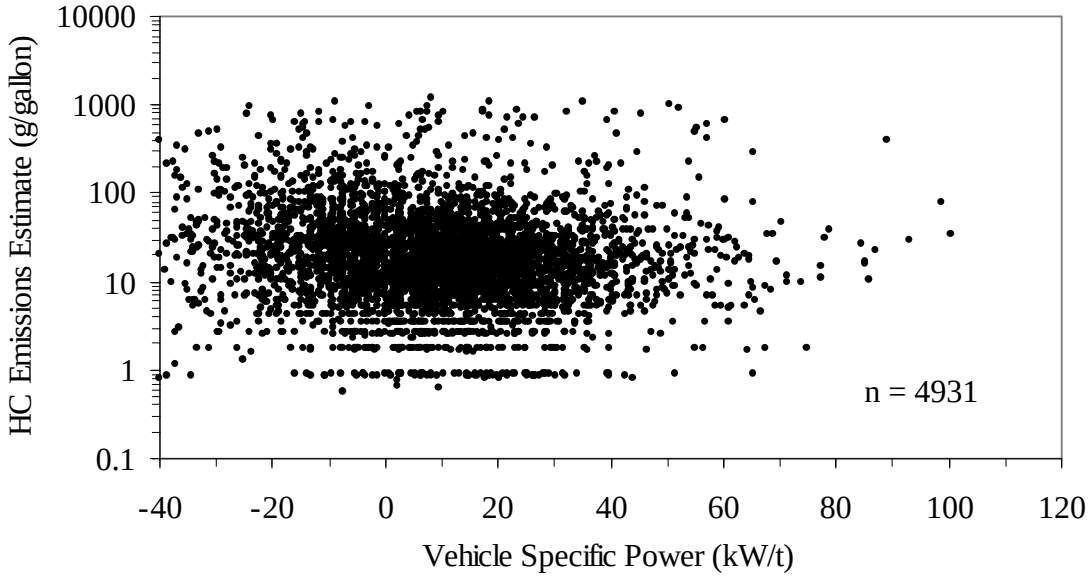


Figure F.29. Scatter plot of HC emissions estimates with respect to Vehicle Specific Power (VSP) values.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

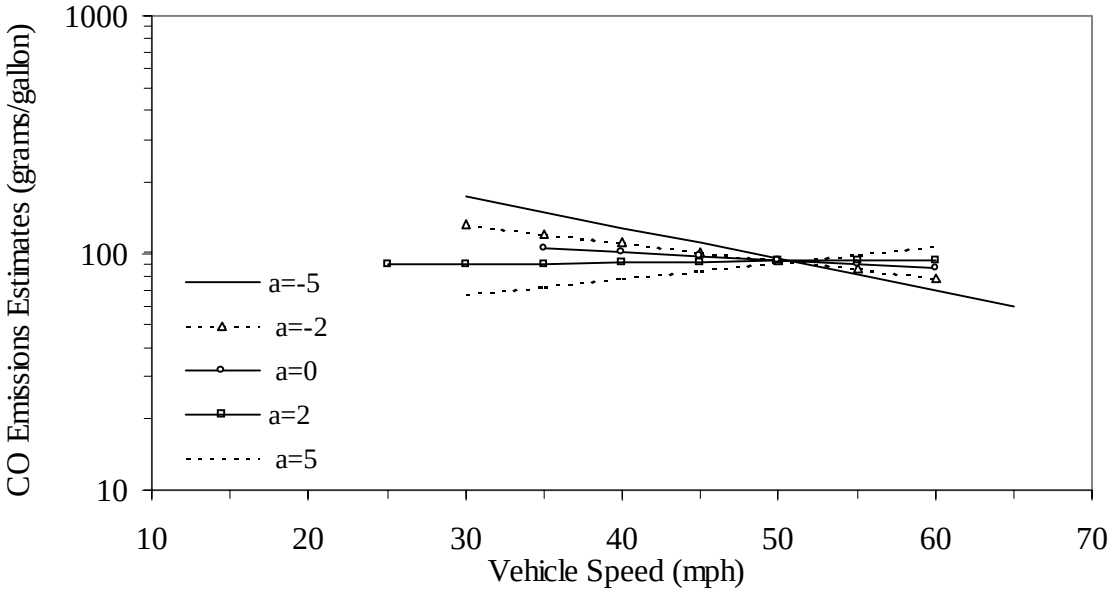


Figure F.30. Sensitivity analysis for the model EF_{CO11}.

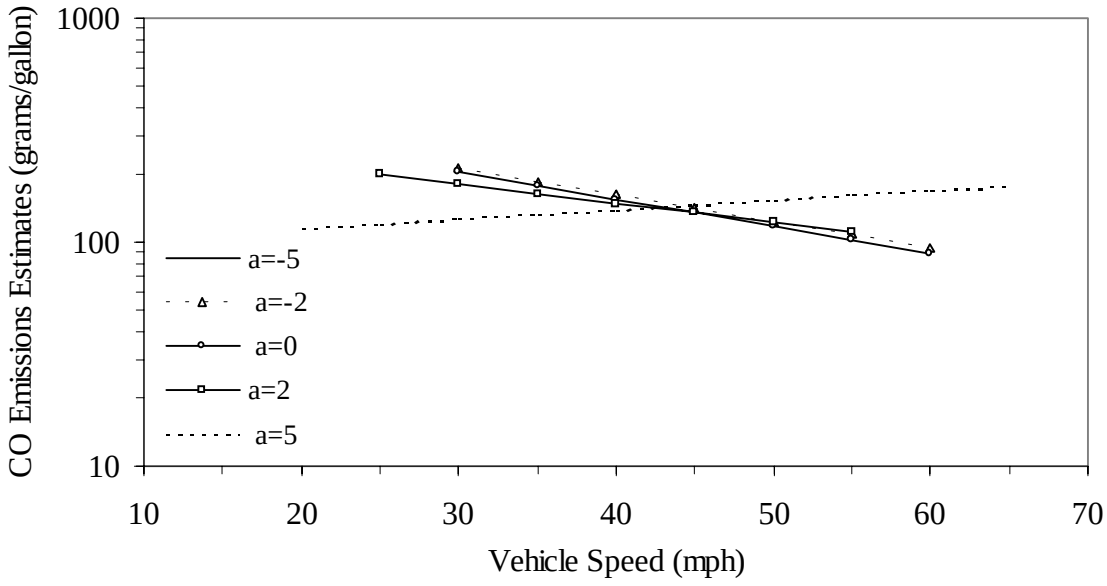


Figure F.31. Sensitivity analysis for the model EF_{CO12}.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

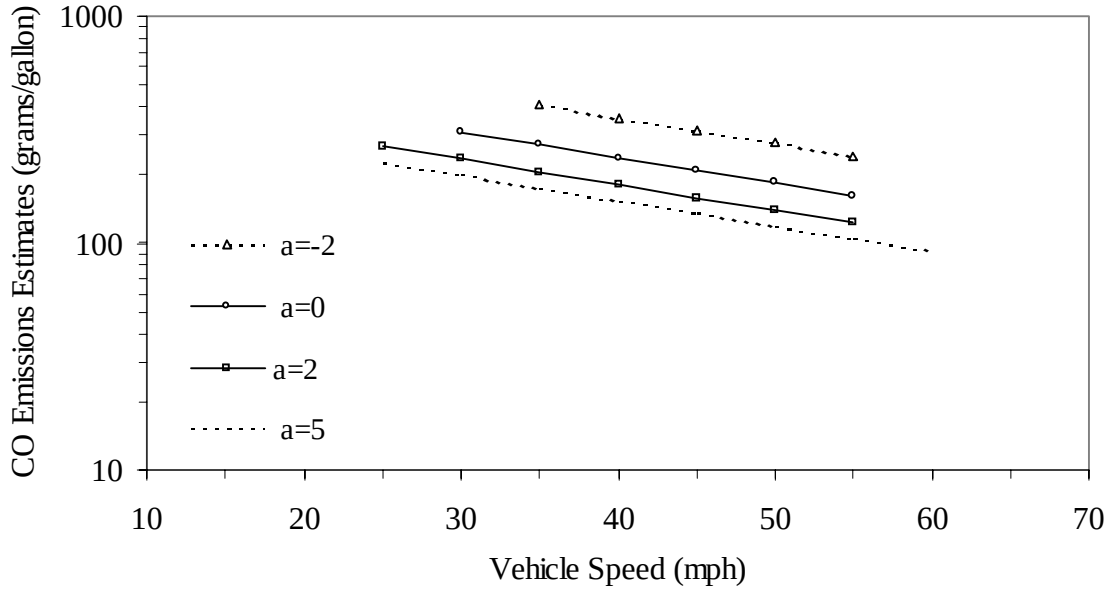


Figure F.32. Sensitivity analysis for the model EF_{CO14}.

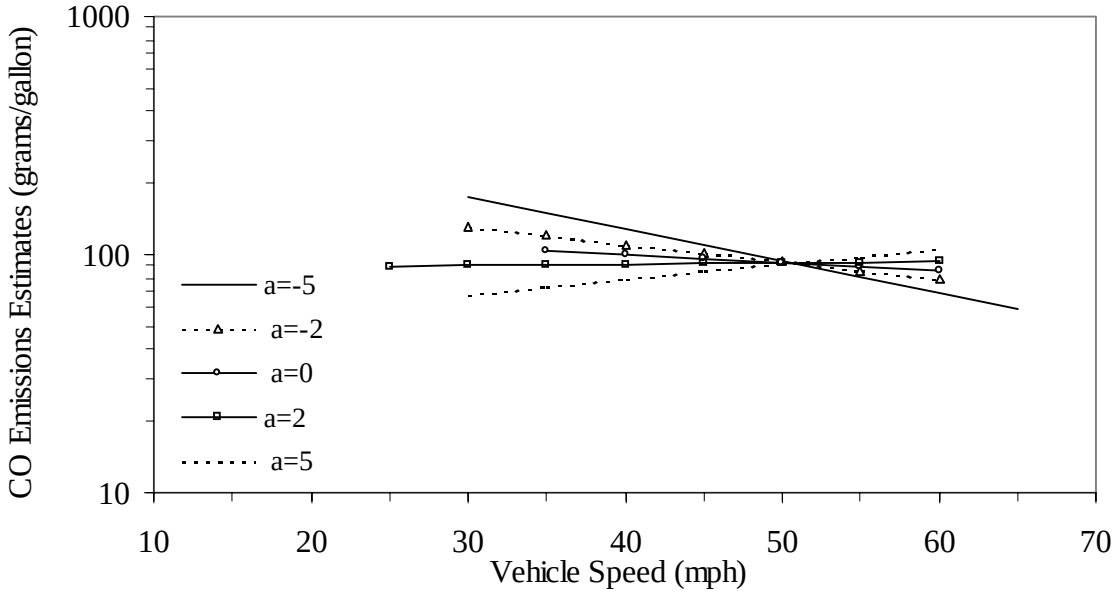


Figure F.33. Sensitivity analysis for the model EF_{CO21}.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

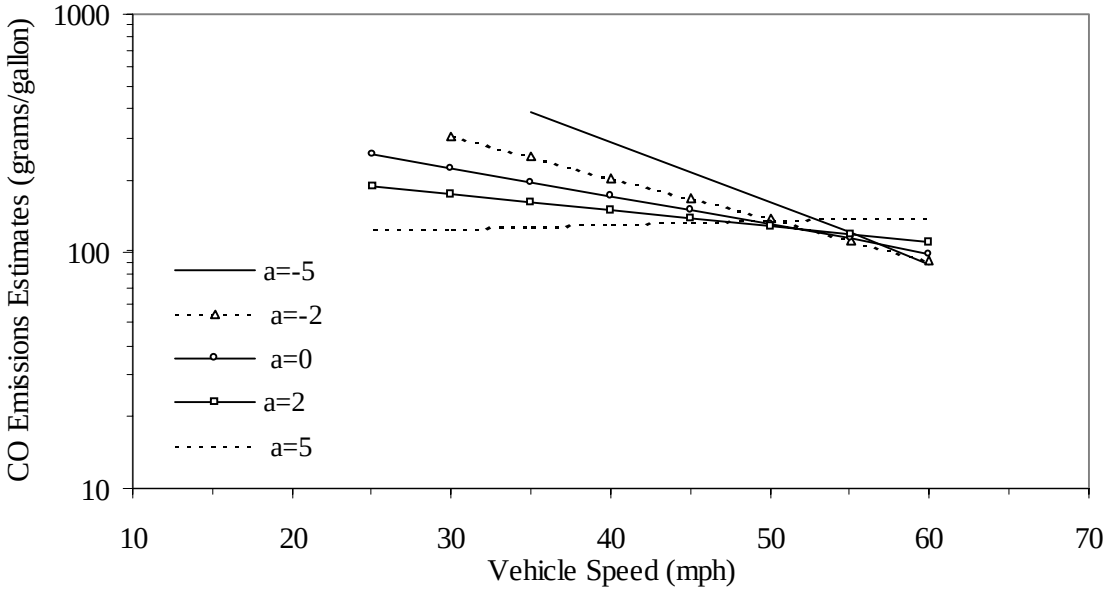


Figure F.34. Sensitivity analysis for the model EF_{CO22}.

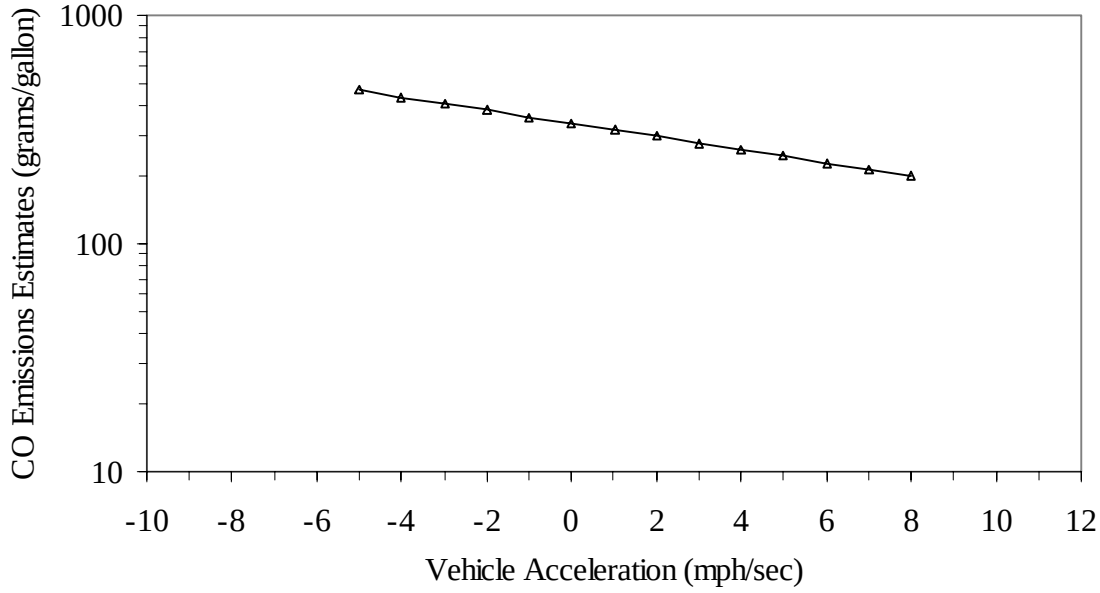


Figure F.35. Sensitivity analysis for the model EF_{CO24}.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

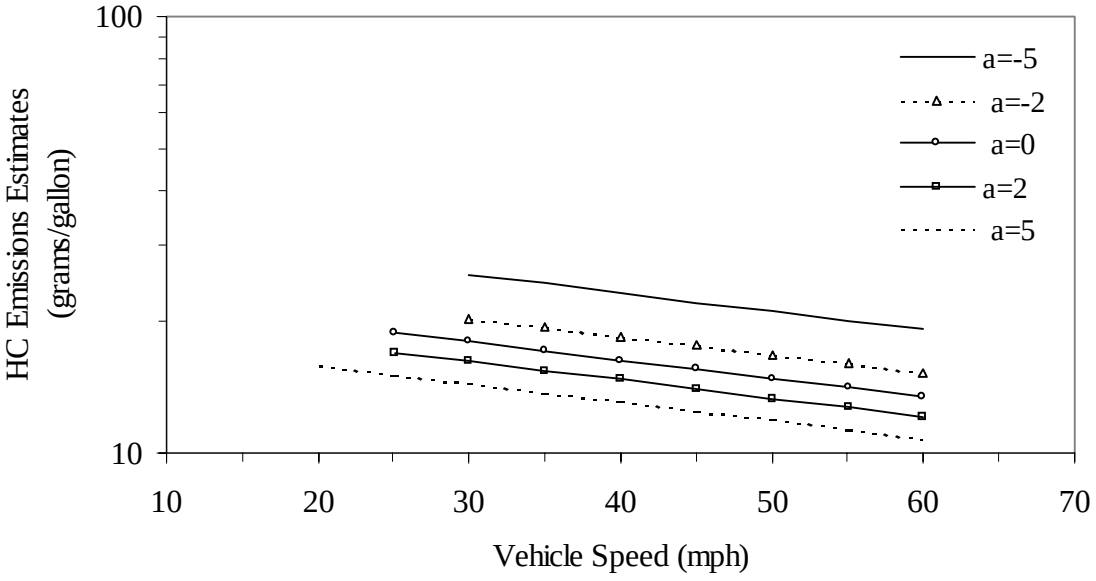


Figure F.36. Sensitivity analysis for the model EF_{HC11}.

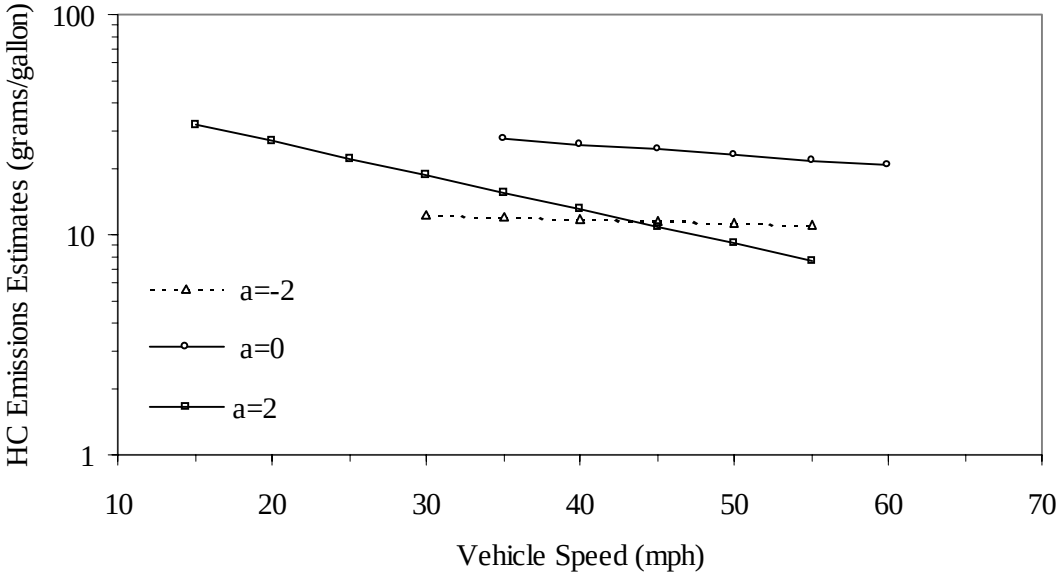


Figure F.37. Sensitivity analysis for the model EF_{HC13}.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

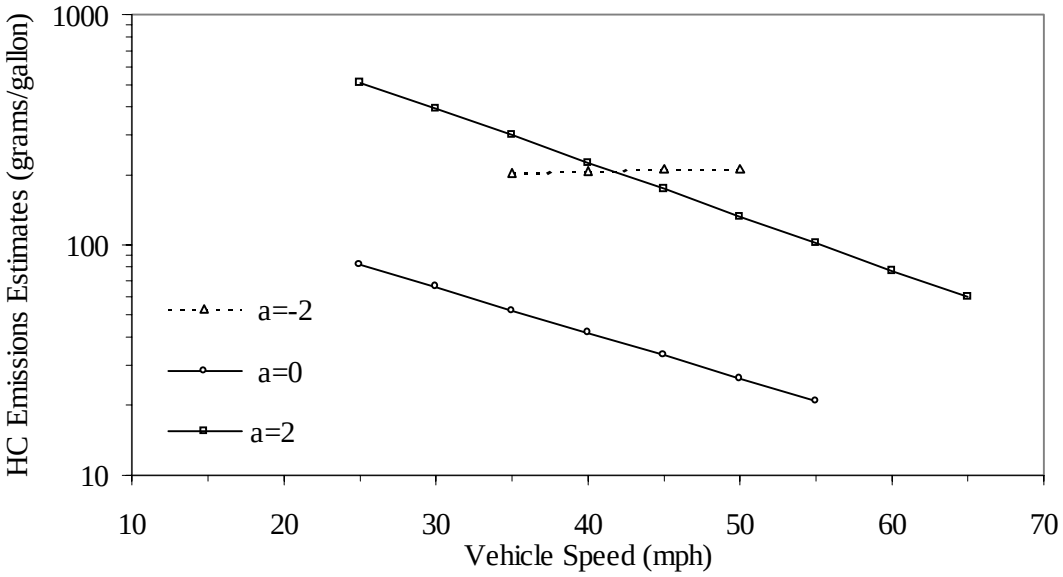


Figure F.38. Sensitivity analysis for the model EF_{HC14}.

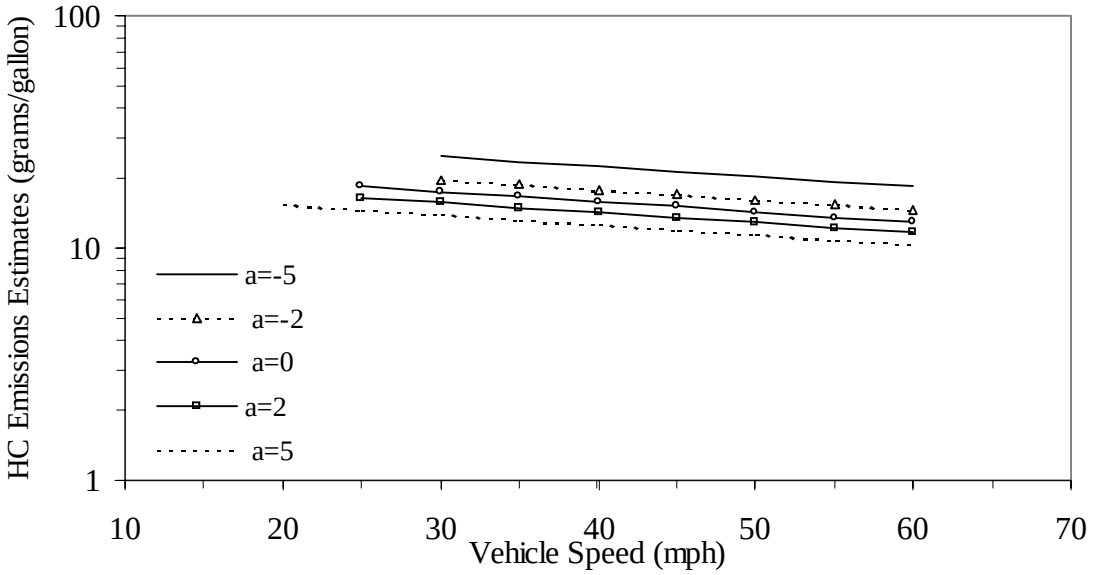


Figure F.39. Sensitivity analysis for the model EF_{HC21} for headway value of 2 seconds.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

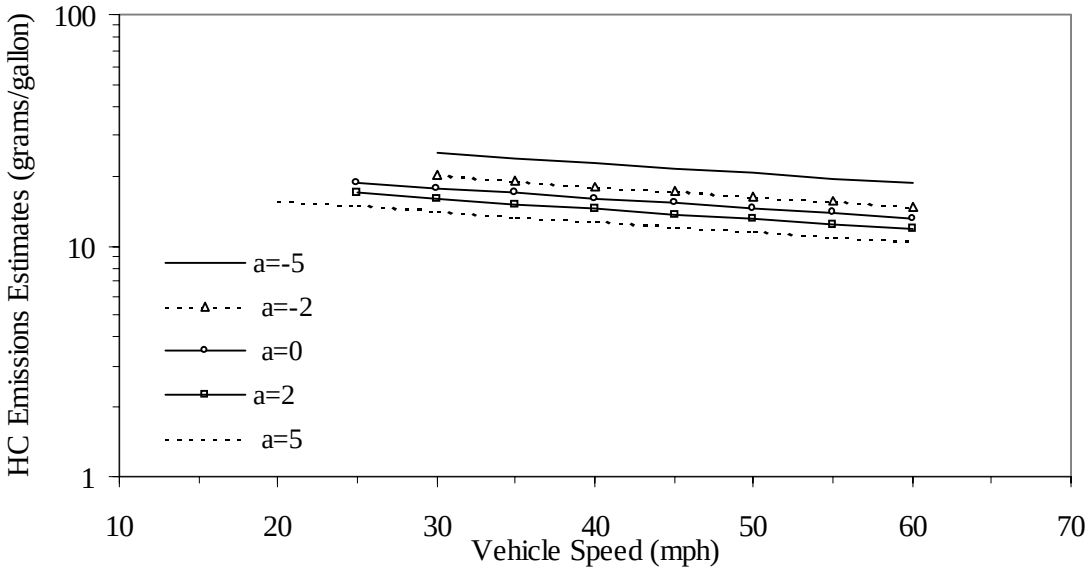


Figure F.40. Sensitivity analysis for the model EF_{HC21} for headway value of 6 seconds.

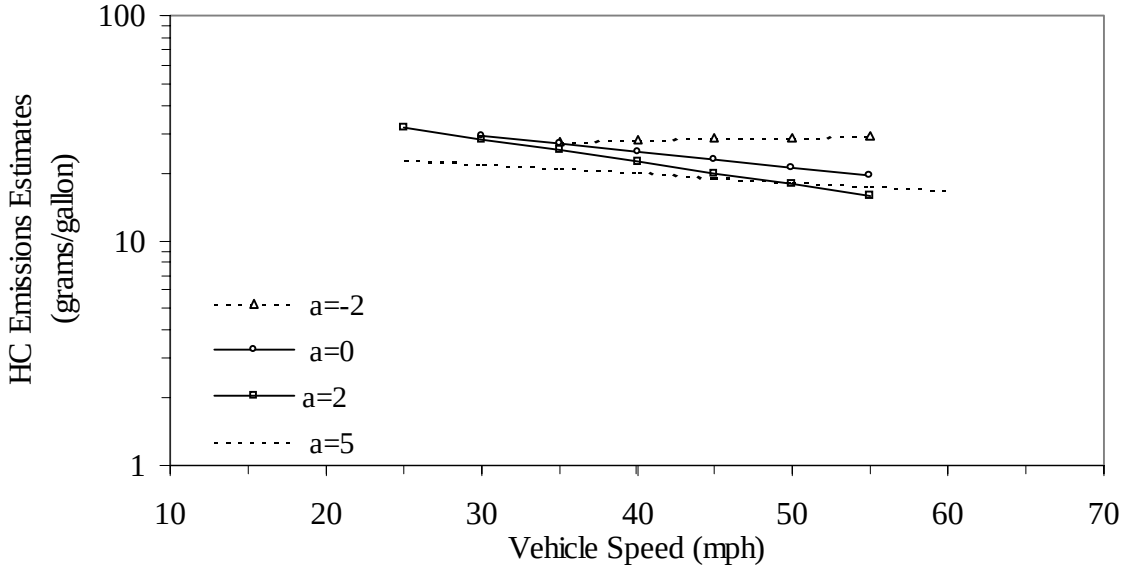


Figure F.41. Sensitivity analysis for the model EF_{HC23}.

APPENDIX F: SUPPLEMENTAL GRAPHS AND FIGURES

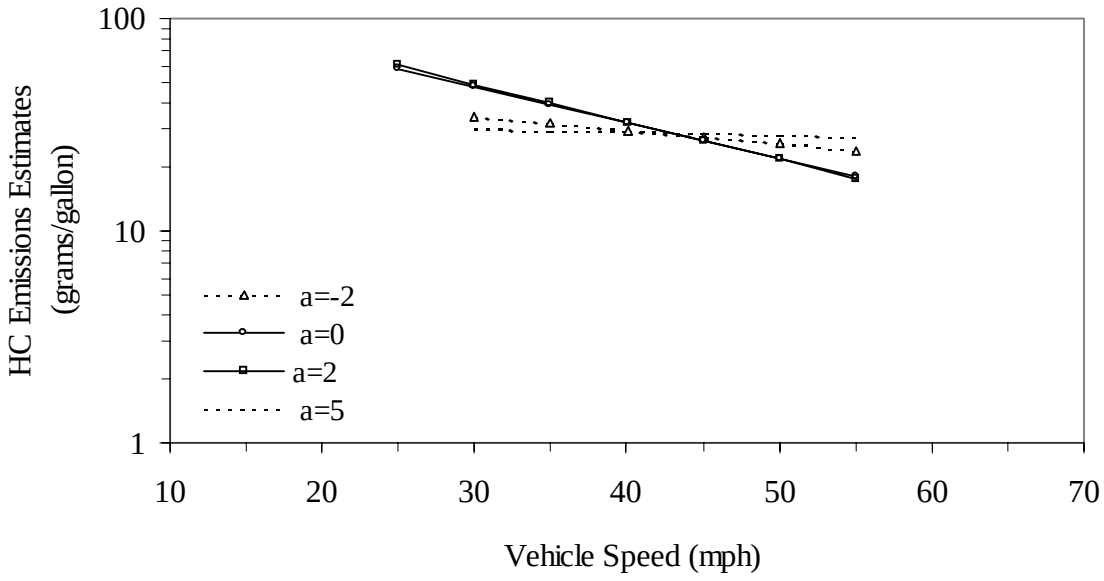


Figure F.42. Sensitivity analysis for the model EF_{HC24}.