VOLUME IV: CHAPTER 1

Preferred and Alternative Methods for Gathering and Locating Specific Emission Inventory Data

June 1996

Prepared by:

Post Office Box 2010
Morrisville, North Carolina 27560-2010

Prepared for:
Point Sources Committee
Emission Inventory Improvement Program

METHODOLOGY FOR GATHERING LOCALITY-SPECIFIC EMISSION INVENTORY DATA

June 19, 1996

Final Report

Prepared by:

Jeremy G. Heiken
Barbara S. Austin
Alison K. Pollack

EVIRON Corporation
101 Rowland Way
Novato, CA 94945

Dana L. Coe
Douglas S. Eisinger
Lyle Chinkin

Sonoma Technology, Inc.
5510 Skylane Blvd., Suite 101
Santa Rosa, CA 95403

Prepared for

Greg Janssen
U.S. Environmental Protection Agency
Office of Mobile Sources
2565 Plymouth Road
Ann Arbor, MI 48105

CONTENTS

CONTENTS

Section

FIGURES

TABLES

Page

TABLES

This page intentionally left blank

1 INTRODUCTION AND SUMMARY

As part of the Emission Inventory Improvement Program (EIIP), The U.S. Environmental Protection Agency (EPA) is developing a number of guidance documents to provide information on potential ways of improving emission inventories. Two volumes present guidance on ways to improve the estimation of motor vehicle emissions. This volume presents guidance on the use of three specific data sources for use in developing distributions used in the MOBILE emission factor model. These distributions include registration, mileage accumulation, travel distributions, fuel consumption, and vehicle miles traveled (VMT). Volume II covers the use of transportation models to estimate variables such as VMT, reconciliation of VMT with the highway performance monitoring system (HPMS), speeds, trip durations, and operating mode weighting factors.

In this document, methodologies and example calculations of the estimation of local on-road vehicle fleet characteristics and activity from inspection and maintenance (I/M) program, remote sensing, and fuel sales data are presented. These data resources provide a relatively untapped source for the evaluation of local fleet characteristics and activity data. Specific evaluations include the estimation of mileage accumulation rates, registration distributions, diurnal travel distributions and regional vehicle miles traveled (VMT). This guidance is separated into three topics according to the type of data being evaluated:

- development of registration distributions and mileage accumulation rates from I/M program data;

- development of registration distributions, diurnal travel distributions, VMT mix, fleet registration information, and I/M program status from remote sensing program data; and

- development of fuel consumption and VMT from tax revenue and other data sources.

The provision of guidance on the three areas listed above is not intended to imply there are not other data sources, nor that all areas will be able to utilize the sources discussed. For example, some areas may not have remote sensing programs or have only small pilot programs. Others may have I/M programs that do not collect all the data needed for developing registration or mileage accumulation distributions. As another example, there are data sources such as the Truck Inventory and Use Survey, which is performed by the U.S. Department of Commerce. This survey is a source of data for registration and mileage accumulation for light, medium, and heavy-duty trucks. However, coverage of such sources is not within the scope of this document.

The remainder of this section briefly summarizes the methodologies covered in this report.

DEVELOPMENT OF REGISTRATION DISTRIBUTIONS AND MILEAGE ACCUMULATION RATES FROM I/M PROGRAM DATA

Inspection and Maintenance (I/M) program data can be used to develop registration distributions and mileage accumulation rates for use in the MOBILE model. I/M programs are required in moderate and worse ozone nonattainment areas as well as in carbon monoxide nonattainment areas. The programs generally cover light- and medium-duty vehicles and sometimes include heavy-duty vehicles as well.

Registration distributions describe, for each vehicle class (i.e., light-duty autos), the fraction of vehicles of various ages (i.e., percent of light-duty autos that are 10 years old). The distribution is used in weighting the fleet emission factors. The default registration distribution in the MOBILE model is commonly used. This distribution represents the 1990 national average, derived from registration data maintained by R. L. Polk, and may be inappropriate for some areas of the U.S. Many areas have unique characteristics and may have significantly more older or newer vehicles than the national average. Since the registration distribution affects not only the calculated emission factors but also the calculated effectiveness of mobile source control measures, it is useful to have a locality-specific distribution. For example, a locality with many older vehicles would have a higher benefit from a retired vehicle program than would be estimated for this area if they used a national default registration distribution.

The vehicle mileage accumulation rates also affect calculated emission factors. Mileage accumulation, expressed as annual miles driven is used to calculate deterioration of the emissions as the vehicle ages. The mileage accumulation is also used in the calculation of vehicle travel fractions. There are significant differences between urban and rural areas, as well as among different areas of the U.S. in mileage accumulation by age. The default national mileage accumulation rate used in MOBILE was developed using national travel survey results.

The data collected by I/M programs generally include sufficient information to develop locality-specific distributions. As a part of registration requirements, most light-duty vehicles in nonattainment regions are required to submit to an I/M test. These data provide a logical opportunity for developing locality-specific registration and mileage accumulation distributions. The data collected in I/M programs that may be used to determine mileage and registrations distributions include:

- vehicle class (i.e., truck or auto),
- vehicle model year,
- license plate or Vehicle Identification Number (VIN),
- county of registration,
- date of test (useful since the data recorded on retests are often not as carefully recorded as the initial test), and
- odometer reading.

Some areas have already used I/M data to develop locality-specific mileage accumulation rate. California is an example of one state that has recently developed I/M-based rates. In California, the Bureau of Automotive Repair and the Air Resources Board (ARB) have developed distributions for various purposes, including use with California's emission factor model EMFAC to estimate mobile source emissions. A study was also performed for the EPA in 1985 that developed locality-specific distributions in Arizona, Connecticut, and Washington. Section 2 of this document reviews relevant literature on the use of I/M data for this purpose.

Section 2 also presents a detailed methodology for estimating the registration distributions and mileage accumulation rates, including data collection and initial analysis. This methodology is briefly summarized below. It is relatively straightforward to develop the distributions from the I/M data. For registration distributions, vehicle data from a given time period of an I/M program (i.e. a quarter or a year) are analyzed. Examining either license plates or VIN numbers, vehicles with more than one record in the data (due to retests, for example) are removed. Then, vehicles are grouped together by vehicle class. For each vehicle class, the number of vehicles of that class of a given age is divided by the total number of vehicles of that class to derive the fraction of the vehicles that are of a given age. The set of fractions is the registration distribution for a given vehicle class.

For mileage accumulation distributions similar steps are followed, except that two time periods of I/M data are needed. For example, in a biennial program, at least two years would be needed (for example, 1992 and 1994). After removing retests from each individual year, the number of miles accumulated on the odometer between the two years is divided by the number of days between tests and multiplied by 365 to estimate annual miles driven. Then vehicles of the same age are grouped together. For each group, the annual miles for each vehicle are summed together and this sum divided by the number of vehicles in the group to calculate the average annual miles for the age group.

When calculating annual mileage accumulation rates, there are several ways to handle unusual odometer readings. Many researchers remove odometer readings that indicate very low or very high accumulated mileage in a given time period. There is great variation in what is considered very low or high. It is recommended that the accumulated mileage be graphed before making these decisions to see whether there are obvious break points in the distribution. Odometer rollover may be indicated if the odometer reading from a later test is lower than for an earlier test. If 100,000 miles are added to the later reading and the estimated annual miles accumulated is within the limits decided upon, then rollover can be assumed. More detailed discussion of this issue is provided in Section 2.

DEVELOPMENT OF FLEET CHARACTERISTICS AND ACTIVITY FROM REMOTE SENSING PROGRAM DATA

Data from large-scale remote sensing programs offer a valuable source of information that can be used to generate regional fleet characteristics data for input into the MOBILE emission factor model and mobile source emission inventory development. Vehicle license plates are typically recorded using video cameras as part of remote sensing programs. The license plates can be cross checked with state motor vehicle records providing valuable information on the characteristics of the local fleet. The potential areas of evaluation include vehicle registration distributions, fleet VMT mix, diurnal travel distributions, diesel sales fractions, evaluation of unregistered vehicles, and evaluation of vehicle county/state of origin to determine vehicle I/M status. Combining the license plate data with site location data allows for the evaluation of fleet data by region, subregion, or by facility class (i.e., roadway classification). Although there are physical limitations in the remote sensing apparatus that restrict its use to certain roadway facility and vehicle classes, the data can accurately evaluate the light-duty vehicle classes in several types of on-road locations.

The following data are typically collected in a remote sensing program and can be used to evaluate activity and characteristics of the vehicle fleet:

- site description and location,
- time of day,
- date of measurement,
- vehicle license plate,
- vehicle identification number (VIN),
- vehicle model year,

- vehicle class,
- vehicle fuel type,
- county and/or state of registration, and
- registration expiration date.

Of these, the last five items generally need to be determined from state registration records and/or through decoding the vehicle identification number (VIN). Complications may arise due to variation of state maintained data or due to VIN decoding limitations; these are discussed further in Section 3.

Section 3 of this document presents the methodologies and example calculations of the evaluation of fleet activity and characteristics data from remote sensing programs. In addition, this section also includes background information on related previous work in the form of a literature review. The methodology included in Section 3 provides step-by-step instructions for completing the evaluations and handling potential biases and limitations of the remote sensing database. The example calculations are based on remote sensing data recently collected as part of the California Pilot I/M Program. These data are from a large-scale remote sensing program completed in the city of Sacramento from July to September 1994. Over three hundred sites were evaluated and approximately two million valid remote sensing readings were obtained. Combining the remote sensing data with state records produced 1,329,694 remote sensing records with matched department of motor vehicle records; 47 percent coverage of the eligible Sacramento vehicle population was achieved.

DEVELOPMENT OF FUEL CONSUMPTION AND VMT FROM TAX REVENUE AND OTHER DATA SOURCES

Section 4 describes procedures to estimate VMT based on fuel consumption and fuel economy. VMT estimates are an integral part of any mobile source emission inventory. Unfortunately, VMT estimates are also subject to substantial uncertainty, and can vary depending upon the tools used to develop current estimates and forecast future travel. Therefore, emission modelers may wish to use a simple screening tool in order to independently verify their VMT estimates. The fuels-based screening tool described in Section 4 can help analysts get a “ballpark” estimate for VMT based on how much fuel is consumed in a given area, and the expected fuel economy (in miles-per-gallon) of the vehicle fleet. This estimate can then be compared to VMT estimates produced through transportation demand modeling, the Highway Performance Monitoring System (HPMS), or other methods used locally. Major discrepancies between the ballpark estimate and traditionally generated VMT statistics should signal a need for further review and analysis. It should be noted that this screening method serves as a double-check, not as a replacement, for other estimation techniques.

The methodology to estimate VMT includes five steps:

1. Estimate on and off-road fuels use using Federal Highway Administration (FHWA) data. Fuel use is distinguished between gasoline and other fuels such as diesel.
2. Disaggregate state data for use at the county level using economic and population statistics.

3. Adjust fuel consumption figures to account for refueling losses using EPA or ARB data to make these adjustments. (Since the modifications made by this step are usually small, this step can be eliminated in most cases; the methodology includes this to be complete.)

4. Calculate fleet fuel economy using EPA data unless local data are available.

5. Estimate VMT based on the fuel consumption and fuel economy information developed in steps 1 through 4.

To better describe the methodology, Section 4 includes sample calculations for two metropolitan areas: Sacramento County, California and Maricopa County, Arizona.

2
USE OF LOCAL I/M DATA TO DEVELOP REGISTRATION AND MILEAGE ACCUMULATION DISTRIBUTIONS

Two key inputs to the MOBILE model are the mileage accumulation rates and registration distributions. These distributions can significantly affect the calculated emission factors. Many nonattainment areas use the default registration and mileage accumulation distributions in the MOBILE model when developing emission factors. These default distributions are national and are not necessarily accurate for a given locality. The use of locality-specific registration and mileage accumulation distributions can increase the accuracy of the emission factors calculated by MOBILE. Inspection and Maintenance (I/M) program data can be used to develop area-specific registration distributions and mileage accumulation rates for use as MOBILE inputs. I/M program data bases normally include odometer readings, vehicle age, vehicle class, and data necessary to estimate miles accumulated between annual or biennial I/M tests. This section presents a literature review, describes the methodology for using I/M data to develop locality-specific registration and mileage accumulation distributions, and provides example calculations using a subset of a state I/M data base. While other data sources such as remote sensing data or surveys may also be used, they are not the subject of this chapter.

INTRODUCTION

Mileage accumulation rates describe the average number of miles driven per year for a given vehicle class and age. Typically, vehicles are driven more when they are newer and less frequently as they age. The mileage accumulation rate is the main factor affecting the emission deterioration rate; the more miles on a vehicle the higher the predicted MOBILE5a emission rate.

Registration distributions describe the fraction of vehicles on the road by vehicle class and age. Registration distributions are combined with mileage accumulation rates to determine the weighting factors for each model year required for estimating total emissions by vehicle class. These weighting factors, also called travel fractions, describe the fraction of total vehicle miles traveled (VMT) for each model year and vehicle class. The distribution significantly affects emission rates; the age distribution of a given vehicle class determines the average emission rate from that vehicle class.

I/M data can be used to develop locality-specific registration and mileage accumulation distributions. I/M data from either centralized or decentralized areas can be used. Both maintain I/M data at a central location, and the I/M data collection methods are relatively standard. In this section we first review existing documents that describe such calculations or provide related guidance. We then describe in detail the methods of analysis that can be used to develop the registration and mileage accumulation distributions from local I/M data.

LITERATURE REVIEW

The literature review covers areas that have used I/M data in developing registration data and mileage accumulation rates and other relevant literature. The first three references are summarized very briefly as they do not discuss specifically how to use I/M data in the development of registration distributions and mileage accumulations. However, they do provide a context for this type of analysis.

The following references were reviewed:

1. *Software Requirement Specification for TAS260 Vehicle Miles Traveled (BAR, 1994)*

2. *Analysis of Data from the California Enhanced I/M Program* (EPA, 1995),

3. *On-Road Motor Vehicle Activity Data Volume II: Vehicle Age Distribution and Mileage Accumulation Rate by County* (CARB, 1994),

4. *Methodology for Estimating Emissions from On-Road Motor Vehicles Volume II: Weight (E7FWT) by California Air Resources Board* (CARB, 1993); and

5. *Estimation of Mileage Accumulation Rates and I/M Failure Rates from I/M Program Data (EEA, 1985)*

6. *Volume IV: Mobile Sources, Procedures for Emission Inventory Preparation* (EPA, 1992),

7. *MOBILE5a User's Guide* (EPA, 1994),

8. *Federal Test Procedure Review Project Preliminary Technical Report* (EPA, 1993),

A review of these sources follows.

SOFTWARE REQUIREMENT SPECIFICATION FOR TAS260 VEHICLE MILES TRAVELED (BAR, 1994)

This report discusses how the California Bureau of Automotive Repair (BAR) developed mileage accumulation estimates and registration distributions using I/M data. They use these distributions primarily for developing reports on vehicle miles traveled that summarize the number of vehicles, average miles traveled, standard deviation and error, average current and previous HC and CO readings, and the percent failing. These summaries are provided for five model year groupings for light-duty autos (1955-1971, 1972 - 1974, 1975-1979, 1980-1989, 1990 and newer). Totals are listed for light-duty trucks, medium-duty trucks, heavy-duty trucks.

BAR's methodology was to first match vehicles by license plate number using the earliest and the latest test dates when a license plate had more than one test in a given year. They then calculated the number of days between the tests and eliminated tests less than 300 days apart or more than 1100 days apart. If any of the following conditions occurred the data were also discarded:

- The vehicle was less than two years old and odometer showed more than 60,000 miles traveled in a year.
- The vehicle was two years old and odometer showed more than 50,000 miles traveled in a year.
- The vehicle was more than two years old and odometer showed more than 40,000 miles in a year.

The document does not discuss how this procedure was derived. Presumably it is intended to account for the fact that newer vehicles are driven more than older vehicles and therefore should have higher cut-offs. No tests were conducted for unusually low mileage. However, if the odometer showed fewer miles in a later test, than an earlier test, then the following adjustments were made:

1. 100,000 miles were added to the odometer reading in the current year.
2. The previous year odometer reading was subtracted from the current year reading.
3. If the result in Step 2 was greater than 100,000; 100,000 was subtracted from the result.

In order to calculate the number of miles traveled in a year, the odometer reading (plain or adjusted according to the above three steps) was divided by the number of days between the earliest and latest tests and multiplied by 365 to determine the vehicle miles traveled per year.

ESTIMATION OF MILEAGE ACCUMULATION RATES AND I/M FAILURE RATES FROM I/M DATA (EEA, 1985).

The primary objective of this study was to derive annual mileage accumulation rates of light-duty autos and trucks as a function of age using I/M data. The report notes that the default mileage accumulation rates (in MOBILE2) were derived from a 1979 voluntary survey of families living in single unit households. The study utilized I/M data from Arizona, Washington, and Connecticut, which tested between 1.2 and 4.2 million vehicles per year. Energy and Environmental Analysis (EEA) used data from initial tests (as opposed to retests). Inspections were matched using the Vehicle Identification Number (VIN), vehicle make, license plate, and model year. In most cases the VIN number alone generated matches. In cases where VIN matching failed, the license plate, make, and model were matched. All three of these variables had to match in order for a match to be considered successful. The distinction between light-duty trucks and light-duty autos was made using VIN decoder software developed by the Highway Loss Data Institute called VINDICATOR. In cases where the VIN decoding software could only determine the model year within a range of years, the model year recorded by the I/M program was used to verify the results.

Data editing procedures focused on anomalous odometer readings. It was found that Arizona and Connecticut recorded only three-digit odometer readings in thousands of miles. The third digit was recorded only if the motorist told the inspector that the odometer had rolled over. In addition, it was found that in Seattle a two-digit odometer reading was recorded and listed as 99 if the motorist told the inspector that the odometer had rolled over. Therefore EEA deleted all odometer readings of 99. They also deleted odometer readings greater than 200,000 because they were concerned that higher readings indicated that the reading had potentially been shifted accidentally (as in 700,000 being recorded for 70,000).

In analyzing the data, EEA found high growth in the number of vehicles with low odometer readings as vehicles aged. This was a result of owners not reporting rollover to the inspectors. They also noted that there was a chance that an owner could report an odometer rollover in one year but not the next, potentially leading to an erroneous deduction of a second rollover.

EEA assumed that vehicles driving less than 500 miles per year were suspect, and eliminated vehicles which had over 60,000 miles per year.

They present an example calculation of mileage accumulation, which has the following steps:

1. Subtract latest odometer reading from previous reading.
2. If the difference is negative, add 100.
3. Determine the interval between test dates (in days).
4. Multiply the odometer reading by the ratio of 365 to the number of days calculated in step 3 to calculate the accumulation rate.

EEA found significant differences between states in mileage accumulation rates. For example, new Arizona light-duty autos traveled over 12,000 miles per year; the same vehicle classes in Seattle and Connecticut traveled over 14,000 per year (14,319 and 14,738, respectively). Cars more than twelve years old traveled 6,798 miles per year in Connecticut, 7,981 in Seattle, and 7,428 in Arizona.

ON-ROAD MOTOR VEHICLE ACTIVITY DATA VOLUME II: VEHICLE AGE DISTRIBUTION AND MILEAGE ACCUMULATION RATE BY COUNTY (ARB, 1994).

This study was performed by Valley Research Corporation (VRC) for the California Air Resources Board. The objectives of the study were to develop county level registration distributions and mileage accumulation rates based on California’s I/M data. VRC used California Department of Motor Vehicle (DMV) registration records, BAR’s I/M data, and the DMV’s VIN decoding software in this effort.

Annual mileage accumulations were calculated using a pair of odometer readings from the same vehicle that were at least six months apart. The statewide average mileage accumulation rate was 11,061 miles per year and tended to be higher in urban than in rural areas.

VRC used the VIN number not only to match vehicles but also, whenever possible, to determine vehicle class, weight, and model year. The BAR data do include indicators for model year, whether auto or truck, fuel type, and gross vehicle weight; however, VRC believed the data were not necessarily correct although they do not indicate the reason for this belief. In order to determine the model year, VRC used the VIN number; when the VIN result produced an illogical result, the DMV data on the first year sold was used. For some reason, roughly only half the vehicles in the BAR database were matched and had more than one odometer reading (out of 24 million records, only 13 million were matched).

VOLUME IV: MOBILE SOURCES, PROCEDURES FOR EMISSION INVENTORY PREPARATION (EPA, 1992).

This document discusses procedures for generating emissions for mobile sources, including on-road vehicles, aircraft, and locomotives. It also provides guidance on selecting MOBILE inputs such as registration distributions and mileage accumulation. Locality-specific mileage accumulation rates by age require 200 input values: for each of the 25 model years for each of the eight vehicle classes, the estimated annual mileage accumulation must be supplied. The same number of inputs is required for locality-specific registration distributions: for each vehicle type, a set of 25 values would be used to represent the fraction of all vehicles of that type that are a given age.

This volume cautions that local mileage accumulation data sources can be subject to sampling bias or data entry errors and recommends that local annual mileage accumulation rates should not change from one evaluation year to the next. It does not discuss recommended sample sizes.

MOBILE5A USER'S GUIDE

The MOBILE users guide provides a detailed summary of the record formats for mileage accumulations rates and the registration distribution records. Once locality-specific mileage accumulation rates and registration distributions are developed, an agency may refer to these tables and the text from these sections of the users guide) particularly 2.2.3 in order to incorporate the distributions into the MOBILE model.

FEDERAL TEST PROCEDURE REVIEW PROJECT PRELIMINARY TECHNICAL REPORT (EPA, 1993).

This report discusses the Federal Test Procedure, in-use driving emissions, driving survey methods, and test cycle development methods. It was suggested for review but does not contain information useful for this work effort.

METHODOLOGY FOR CALCULATING REGISTRATION AND MILEAGE ACCUMULATION DISTRIBUTIONS USING INSPECTION AND MAINTENANCE PROGRAM DATA

This section presents methodologies for using Inspection and Maintenance (I/M) data to develop locality-specific registration distributions and mileage accumulation rates by vehicle class. The use of I/M data, which is usually available for virtually all registered vehicles in an area, can provide a significant improvement in accuracy over the use of default values or of distributions developed using sampling methods such as traffic counts. These improvements can result in more accurate emission inventories and provide insight as to the most appropriate control measures. For example, if the locality-specific registration distribution shows a higher fraction of older light-duty vehicles than the defaults, the effect of a vehicle scrappage program will be shown to be higher. These methods are straightforward although there are some variations in the manner in which anomalies in the data can be identified and addressed.

In this section, data collection methods and issues are first described, including issues that should be considered when obtaining the data, data costs, formats, and initial processing. The initial processing is needed both for the development of mileage accumulation rates and registration distributions. Next, the methodology for developing registration distributions, including data processing procedures and the handling of issues such as unregistered or out-of-state vehicles is discussed. Finally, the methodology for developing mileage accumulation distributions, including data processing procedures, and how to identify and address anomalous odometer readings is described. Included in the methodology for calculating mileage accumulation distributions are suggested procedures that may be used to address anomalous odometer readings. A more qualitative evaluation of the odometer data can be substituted for these procedures to obtain reasonable results that are significant improvements over the use of default distributions.

DATA COLLECTION

Obtaining I/M data will usually take at least one month after formally ordering the data from the appropriate agency. In most cases it is best to allow for approximately two to three months to fulfill the data request. It will also take time to identify the correct personnel and procedures for ordering.

There is great variation in the electronic media, format, costs, and staff availability to fill data requests. As these issues can become time consuming to the point that they over-ride technical considerations in actually developing the distributions, they are discussed in some detail below.

Ordering the Data

Start by identifying the agency responsible for the I/M program in the state of interest. Often the best way to proceed will be to contact personnel with responsibility for mobile source issues at the state air agency. In some states the air agency itself will collect the data; in others the state Department of Motor Vehicles; in others, a private contractor responsible for the program. If a private contractor is responsible it will be helpful to alert the state air agency to your efforts to collect the data. This will give the agency the opportunity to monitor the contractor's procedures for transferring data as well as information on who in the state is developing local-specific data.

It is important to identify the appropriate office from which to order the data. For example, in New York state, the DMV Freedom of Information Office is responsible for I/M data.

It is also important to find out exactly what vehicles are tested and what data fields are available. Although I/M programs are legally required only for light-duty vehicles, some states also test medium and heavy duty vehicles. When this is the case, additional information regarding the extent and type of the program, the representation of diesel vehicles, and representation of vehicles registered out-of-state (if any) should be discussed.

Data Costs

There will often be a charge for the data, even if the data is requested by other agencies. In addition, the I/M program may not always collect the information necessary to differentiate between autos and trucks. This issue should also be discussed before ordering the data. Some states charge a flat fee for each month or quarter. For example, Arizona currently charges $185 per month of data. Others charge a nominal flat fee, often less than $10 plus any time used by a programmer to fulfill the request. Other states do not have established procedures and treat each request on a case-by-case basis.

Data Formats

Initially, consideration should be given to the quality of the data. Some areas have long periods for which the tape media upon which the data are recorded has been corrupted and cannot be used at all. Data needed for developing a registration distribution by county include vehicle age, vehicle class (LDA or LDT), and county of residence. Many I/M programs record these items directly. It is critical to be able to differentiate between autos and trucks, and the manner in which this is done should be discussed. If I/M data for vehicle class are missing or suspect, information from the VIN can be used. In 1980, an internationally used VIN number system was adopted by most auto manufacturers and can be used for 1981 and later vehicles. This system readily identifies a vehicle’s model year, vehicle class, and gross vehicle weight. For vehicles manufactured prior to 1981, VIN decoding software such as the California Bureau of Auto Repair’s VINI program or the Highway Data Loss Institute’s VINDICATOR software can be used to determine the model year and vehicle class. However, VIN decoding software can be difficult to acquire and VRC rated its results as unreliable.

Many agencies routinely analyze the I/M data in various ways and therefore have the ability to develop subsets of the data. This can be helpful, as most raw I/M data records have close to one hundred fields and can be very awkward to manipulate. It is not uncommon to test a million cars a month; the volume of data is quite substantial. Given this, it can be very helpful to order only the data needed for the analysis (if the agency has the ability to provide it in this way).

The media upon which the data are supplied needs to be discussed with the providing agency. It is critical to know the precise characteristics of the data and media your agency or company can handle and be sure they are compatible with those of the agency providing the data. It is difficult to provide the data on floppy disks; even when compressed, millions of records can require numerous floppies. Common media are direct file transfer through modem or the Internet, 4 mm DAT tapes, 8 mm exobyte tapes, or 9 track tapes. Electronic mail should not be used to transfer large data sets because most electronic mail software cannot conveniently handle very large file sizes. Unless it is certain that your operating system and software are compatible with that being used by the transmitting agency, data are most easily transferred when written in ASCII format.

It is also important that data formats are understood. For example, some agencies record the odometer reading in thousands rather than the entire reading. A model year may be recorded as a four or a two-digit number. Similarly, missing data codes should be understood, listed in writing, and checked after receipt of the data. It is not uncommon for a missing data code to be useable as actual data (i.e. an odometer reading of "99").

Specific data needed for developing a registration distribution for a given time period (i.e. a year or a quarter) are:

- vehicle class (whether passenger car or truck, and the weight class for trucks),
- model year,
- test date,
- license plate and/or VIN code, and
- county of registration.

It is recommended that only the counties within the nonattainment area of interest be used (as opposed to the entire state). If there is a mix of both rural and urban counties in the nonattainment area it may even be desirable to develop two registration distributions, one for the urban and one for the rural counties.

Conceptually, to develop a registration distribution all that is needed are the vehicle class and model. However, vehicles will be tested more than once in a given time period if they fail a test or are sold. Inclusion of all records would thus bias the distributions. Data from the first test date should be used as records for the first test may often be more carefully completed than for a re-test. The license plate and/or VIN code can be used to identify vehicles that have been tested more than once.

A reasonable registration distribution can often be developed using less than one full year of data. Exceptions are areas where large numbers of vehicles of a certain type are registered at a particular time. For example, trucks in Arizona must be registered by the end of the year so that in November and December there is a disproportionate number of trucks being registered.

To develop a mileage accumulation distribution for two full time periods (i.e., two consecutive years for an annual program or four years for a biennial program), data needs are the same as listed above, with the addition of the odometer reading. For mileage accumulation, the test dates are critical since they establish the number of days between tests and therefore the information needed to calculate the number of miles driven per year. The license plates or VIN numbers are needed to match vehicles at two points in time.

Initial Processing

It is recommended that after obtaining the data set, some basic checks be made to verify the written description of the data format, to verify missing data codes, and to check for biases that may be introduced due to entry practices. For example, it can be very useful to print out data records for retests to see whether vehicle data on retests is entered as completely as for initial tests. It would not be unusual to find the same odometer reading on a retest as on the initial test, even for tests several days or weeks apart. Simply to print out several records of the fields that are most critical for developing the distributions can be very helpful to look at questions such as: How are the odometer readings listed? How are the test dates shown? How do the listed vehicle class and weight classes correspond? Do the model years match the year first sold field?

METHODOLOGY FOR REMOVING VEHICLES TESTED MORE THAN ONCE IN A TEST CYCLE FROM THE DATA SET

Repeated records for the same vehicle in a given test cycle (one year for annual program, two for a biannual program) need to be removed. For the registration distribution, vehicles that have been retested need to be removed because otherwise they will be counted twice in the registration distribution. Since most retests involve a failure of the emission test, not removing retests from the data will over-represent those vehicles which fail the first test and therefore bias the distributions. It is likely that many of these vehicles will be older vehicles. For the mileage accumulation, it is necessary to have a long period between retests in order to evaluate the average miles accumulated over a period of time. Most retests are performed within a month of the first test, which is not long enough to adequately represent annual miles accumulated since there is often several variation in travel. Some retests (i.e., for vehicles that have been sold in between cycles) may occur over six months apart, which may used to calculate a reasonable representation of annual miles, but attempting to retain these data is likely to entail more effort than justified by the benefit of including this small number of vehicles.

In order to determine whether more than one record for the same vehicle is present, the license plate number, the VIN number, or both can be compared. If the license plate alone is compared there is a smaller likelihood of data entry error than for the VIN number, which is much longer. If both the license plate and the VIN number are used to match, the likelihood of one of those being misrecorded is roughly the square of the likelihood that one would be in error.

If the license plate alone is compared, vehicles that have new license plates or license plates that have been entered in error will be counted twice. If there are license plate entries that erroneously duplicate the license plate of another vehicle, one will be mistakenly removed from the data set. If the VIN number alone is compared, vehicles will be counted twice if one of the test records is erroneous. If there are VIN number entries that are erroneously duplicated, the VIN number of another vehicle will be mistakenly removed from the data set. This latter case is less likely to occur (because of the size of the VIN number) than for license plates. If the combination of VIN and license plate numbers is used to indicate unique vehicles, it is very unlikely that any vehicles would be counted twice. However, a large number of vehicles may be removed from the data set due to data entry errors for either the VIN or the license plate. Since the number of data entry errors is highest when matching on both the VIN and license plate number, and higher when matching on the VIN than on the license plate, it is recommended that license plate matching alone be performed. This will minimize the amount of data thrown out when a valid match could have been made.[1]

In the absence of data base software that can generate unique record identifiers, the simplest approach for removing retested vehicles is to write a brief computer program that stores the data for each record. The license plate or VIN number (or both) on each new record is compared to the stored records to check whether that license plate or VIN number has been recorded before. If it has, and the record is for a later test that record would not be used. If the license plate or VIN number has been stored already but the test date is earlier, then the record with the newer test date should be removed from the data set.

[1]Another factor to consider is the availability of “vanity” plates. In states where vanity plates are available, VIN matching may be preferable if vanity plates are popular (because converting from a regular plate to a vanity plate will create a mismatch).

METHODOLOGY FOR DEVELOPING THE REGISTRATION DISTRIBUTION

The registration distribution is a series of percentages of vehicles by age and class. For example, the portion of the registration distribution for light-duty passenger autos will list, for each vehicle age, the percent of light-duty autos that are of that age. Perhaps six percent would be one year old vehicles, eight percent would be two year old vehicles, and so forth; the percentages by age sum up to 100 percent. The methodology for developing the registration distribution is relatively straightforward. The steps to be followed are described here.

Evaluate the Overall Characteristics of the Data

Calculate the total number of light-duty autos and light-duty trucks in the data set. If there are other vehicle classes represented, total these also. From these values, calculate the fraction of light-duty autos, trucks and any vehicle classes that are present. Ensure that these are reasonable before proceeding further. A typical light-duty vehicle class distribution is 75 percent autos and 25 percent trucks although some areas may have higher percentages of trucks, such as Montana or Wyoming. If the distributions seem unreasonable, the vehicle class information in the data may not be accurate. In this case the agency providing the data should be contacted and the data reevaluated in light of the new information obtained.

Calculate the number of vehicles of each age and vehicle class and then divide the total by the number in that vehicle class to obtain the registration distribution. This is shown by the following equation.

$$MYDist_{i,\ unadjusted} = \frac{\sum POP_i}{\sum POP}$$

where:

MYDist is the model year registration distribution for a given vehicle class for model year I, and

POP is the vehicle population for that vehicle class for model year I.

For light-duty vehicles, MOBILE requires a combined registration distribution for diesel and gasoline vehicles. MOBILE has diesel and gasoline sales fractions it will use to adjust these vehicle classes internally to the model. Any vehicles greater than 25 years old should be counted as 25 years old. Also note that MOBILE does not distinguish emission factors for motorcycles more than twelve years old. Therefore, any motorcycles that are more than twelve years old should be included in the count for twelve-year old motorcycles.

Adjust the Distribution to July 1

The distribution needs to be adjusted to a July 1 distribution to reflect vehicle sales in the months prior to the model year and fleet turnover during the year. Specifics of the calculations needed for this adjustment are provided in Section 3. Write out the fractions in the format specified for MOBILE in Table 2-1. Note that there will be 200 values entered: fractions for each of the 25 model years for all vehicle classes except motorcycles. For motorcycles, write ".000" for motorcycles aged 13 to 25. If I/M data are not available for a particular vehicle class, write out the MOBILE defaults for that class unless a registration distribution for that class has been developed through an alternate method. Compare the calculated fractions to the default values in MOBILE or to previously developed distributions to ensure that they appear reasonable.

Unregistered Vehicles

I/M data will of course not be available for non-registered vehicles, out-of-state vehicles, and out-of-country vehicles. Such vehicles could contribute disproportionately to the actual registration distribution and hence to estimated emissions but are difficult to account for.

A recent survey in California showed that roughly nine percent of the vehicles on the road are unregistered (Sierra, 1992). However, this does not mean that all of these vehicles will never be registered. As shown in Figure 2-1, many unregistered vehicles are registered within a few months of expiration and nearly all unregistered vehicles are registered within two years of expiration. This figure does not account for out-of-state vehicles. In areas located along the Mexican border, it may be especially useful to account for out-of-state vehicles.

The only accurate way to account for unregistered and out-of-state vehicles is to perform a survey. The Sierra (1992) survey for California's Air Resources Board looked at the registration status of approximately 30,000 vehicles in parking lots in the Los Angeles area. While this survey focused on registration tags, it would be straightforward to modify a similar survey to include out-of-state vehicle counts. No such studies were identified in the course of preparing this document.

METHODOLOGY FOR DEVELOPING THE MILEAGE ACCUMULATION DISTRIBUTION

Development of the mileage accumulation distribution is more complex than development of the registration distribution. Data required include vehicle age, vehicle class, and two or more odometer readings from each vehicle. This methodology focuses on mileage accumulation by vehicle age (i.e., a ten-year-old vehicle is assumed to drive the same number of miles in the year 2000 as a ten-year-old vehicle in 1995), as is currently required in MOBILE5a. In California, county-specific mileage accumulation rates that are also model-year specific were developed for use in the EMFAC7F model (which calculates California-specific motor vehicle emission factors). Development of model-year specific mileage accumulation distributions can capture important variability in specific years as well as any trends in driving patterns (for example average driving distances to work becoming longer). However, this approach would necessitate updating mileage accumulation distributions annually. Further, model-year specific mileage accumulation rates cannot be readily incorporated into the current version of the MOBILE model.

Briefly, the methodology entails averaging annual miles accumulated by vehicles of a given age and listing out the average for each vehicle age in the format specified in Table 2-2. The basic procedure is described below. Procedures for addressing anomalous odometer readings are then described. Note that the steps described should not be applied until anomalous odometer readings have been either adjusted or removed from the data set.

Developing Data Sets of Unique Vehicles and Initial Processing

Develop, for each year of data, a data set containing only unique vehicles using one of the approaches described above. It is recommended that a count of vehicles by model year and vehicle class also be made at this time to evaluate whether the number of vehicles of a given age is large enough to yield a reliable estimate of average annual miles driven by a vehicle of that age: the more vehicles in a given age group, the more reliable the average annual miles will be. Typically the oldest age groups will have the smallest fraction of vehicles; it is recommended that the data base be large enough to have at least hundred vehicles in each of the age groups.[2]

Additional data summaries should also be examined, including:

- the range of raw odometer readings,
- a frequency distribution of raw odometer readings, and
- average odometer readings by vehicle age

This summary information can be valuable in determining whether the data set will be adequate to develop the distribution, and whether there are particular vehicle ages in the data set that have a high frequency of unusual odometer readings or significantly higher variation

[2]Based on our analysis of the example Arizona I/M data base, approximately 200 vehicles would be required in each age group in order to estimate the average annual mileage for each group within an uncertainty range of plus or minus 1000 miles.

in the odometer readings than other vehicle ages. Note that the variation in odometer readings is likely to be larger for older vehicles.

Calculate the Annual Miles Driven by Each Vehicle

Subtract the earliest odometer reading from the later reading and divide by the number of days between the two readings. If the difference between the odometer readings is negative, add 100,000 to account for odometer rollover. The value of the average daily miles is equal to the number of miles driven between retests divided by the number of days between tests.

Subtracting the earliest odometer reading from the latest odometer reading and subtracting the corresponding test dates (counting January one of the first test year as day one and January 1 of the second test year as day 366 as long as the first test year was not a leap year), then multiplying by 365 will yield the annual miles for a vehicle. If annual miles are less than a specified minimum or maximum (see section on anomalous odometer readings) the record can be removed from the data.

To calculate average annual miles across vehicles within an age group, first group the vehicles by age. It is recommended that vehicles older than 25 years be included in the set of 25-year old vehicles.

Next, sum the annual miles for all vehicles in a group (i.e., sum annual miles for all 10-year-old vehicles). Lastly, divide the sum of the annual miles by the number of vehicles in the age group.
Write out mileage accumulation to input into MOBILE according to the format specified in Table 2-2. Compare them to the default mileage accumulation rates currently used in MOBILE5 are listed in Table 2-3. For comparison, mileage accumulation rates calculated using California I/M data are presented in Table 2-4. Note: mileage accumulation rates for one-year old vehicles cannot be calculated. However, they may be estimated by fitting an exponential curve to the data for the 2 year and older vehicles.

ANOMALOUS ODOMETER READINGS

When developing the mileage accumulation rates it is likely that some of the odometer readings will not make sense. Some readings will imply unusually high or low miles per year; some are recorded incorrectly; some individuals tamper with the odometers; and some odometers only have five digits and do not register rollover after 100,000 miles. Table 2-5 presents some examples of how researchers have addressed these issues. It is recommended that three general factors be considered when addressing anomalous odometer readings:

1. The annual mileage accumulation will be an average for a large number of vehicles and should not be unduly influenced by a few outliers that are

incorrectly handled due to the procedures used to address outliers. Any reasonable procedure for identifying unusual odometer readings will be able to correct most of them. No procedure will be able to identify and correct for all the anomalous readings.

2. The specific procedure chosen should be guided by the data itself. It is useful to graph (at a minimum for all vehicle ages combined) annual mileage accumulation at different ranges (for example in 2,000 mile increments) and the percentage (and/or absolute number) of vehicles driving at each range. In most cases it will be obvious what the outliers are; the graph will tail off smoothly at the higher accumulations and suddenly there will be a jump. When there is an obvious point at which the distribution seems to "jump", which will generally be in the neighborhood of 50,000 miles per year, that can be used as the maximum annual mileage allowed in the data set.

3. In a few cases, more detailed analysis may be helpful. If the distribution does not break at any particular point and there are many vehicles which drive unusually high numbers of miles, one can fit a distribution to the data and use the fitted distribution to approximate at least the upper tail. An example of data where there is no obvious place to "break" the distribution is presented in Figure 2-2.

Later in this chapter an example mileage accumulation calculation is presented using each of the four sets of decision criteria listed in Table 2-5. It is noted there that using a mileage cut-off of 100,000 miles resulted in a distribution significantly higher from all the others. In addition, the distribution implied that as vehicles aged, they drove more miles per year than newer vehicles.
The use of such a high cut-off can introduce bias since it is so much higher than the average miles driven (approximately 12,000 miles per year). The other three sets of decision criteria resulted in distributions that were similar in shape to each other and to the MOBILE default. Therefore it is recommended that one of these be used.

Odometer Rollover

Odometer rollover is likely when a later odometer reading is lower than the earlier reading. It is also possible that a lower odometer reading indicates odometer tampering, or there could be a data entry error. If after adding 100,000 to the difference in odometer readings the annual miles calculated falls within the limits decided upon in the steps above, it is recommended that the reading be assumed reliable.

EXAMPLE APPLICATION

This subsection presents examples of calculating registration and mileage accumulation distributions from I/M data using data from Arizona's I/M program. Several states and both centralized and decentralized programs were considered for use in the example. The decision to use Arizona data for the example was based on the fact that it was the first state for which two usable years of I/M data were obtained and not because of any lack of data quality in the other states considered.

DESCRIPTION OF THE ARIZONA I/M DATA

Arizona has maintained an annual basic I/M program for many years. Arizona tests medium and heavy duty vehicles in addition to light duty autos and trucks. In 1995 they began a biennial enhanced I/M program for 1981 and newer vehicles, run by a contractor. Electronic files are maintained by the contractor for 12 months after collection and are then transferred to the Arizona Department of Environmental Quality (ADEQ) for storage in archives.

Data were obtained for March, 1994 and 1995 for the example. If the data were intended for use in actual inventory development it would be advisable to use at least three consecutive months in each of two years to increase the number of vehicles in the database. This will increase the sample size by including I/M tests made in the month before or after the month for which they are scheduled. In this example this is particularly important because of the 1995 implementation of the biennial enhanced program, which tested 50 percent of 1981 and newer vehicles in 1995. The effect was to lessen the sample size of 1981 and newer vehicles. In states where a biennial program has been in place for a number of years, it would be preferable to obtain three years worth of data; in Arizona two consecutive years were useable since the program had just begun and since 1981 and older vehicles were still subject to annual tests.

Data used in this example included the test date, vehicle identification number, license plate, vehicle make, vehicle model year, vehicle style, vehicle fuel code, indicator for initial or retest, weight class, number of cylinders, and odometer reading. It was initially expected that these data would be sufficient to differentiate between light duty autos and light duty trucks but this was not actually the case. In a real-world application, VIN decoding software or access to DMV data would need to be utilized in order to differentiate between autos and trucks as there may be significant differences in mileage accumulation and often in the registration distributions as well. For the purpose of this example calculation this step was not taken.

There were 196,132 records in the March, 1994 data and 159,028 in the March, 1995 data. After deleting retested vehicles there were roughly 164,000 records for 1994 and 131,000 for 1995. 62,007 vehicles, or approximately 47 percent, were matched based on the VIN number. This match rate would be increased if more than a single month were used[3]. The match rate was also lowered due to the implementation of the biennial program so that only half the 1981 and newer vehicles were tested in 1995, whereas all had been required to be tested in 1994. Of these 62,000 matches, 4,311, or seven percent, had invalid odometer readings (defined as zero or blank entries in the odometer field).

EXAMPLE APPLICATION FOR REGISTRATION DISTRIBUTION

A registration distribution for all light duty vehicles was estimated for 1993 and older vehicles using the 1994 data. The 1995 data was not used for developing a registration distribution as only 50 percent of the 1981 and newer vehicles were present in the data. In a real-world application involving a partially annual and partially biennial program it would be necessary to calculate distributions separately for vehicles subject to the annual program and to the biennial program and then normalize them in order to combine them into one distribution.

Before developing the registration distribution, obvious anomalies in the data were identified and deleted. For example, there were two vehicles with obviously invalid model years of 1995 and 1999 in the March, 1994 data. Repeated records for the same vehicle, such as retests were also removed.

As discussed previously, the Arizona I/M program data does not explicitly differentiate between autos and trucks. This is quite common; an informal survey revealed that most states rely on VIN decoding software to distinguish between autos and trucks although some are planning to change the I/M program to require that a distinction be made on the I/M record. It is reportedly not always possible to make this distinction through visual inspection alone.

MOBILE calculates emission factors for vehicles up to 25 years old. Vehicles older than 25 years are placed in the 25 year old category. In keeping with this, all vehicles older than 25 years were counted as 25 year old vehicles.

[3] For example, an analysis of two full years of British Columbia Inspection and Maintenance Program data conducted for this document showed that roughly 50 percent of vehicles are matched when one month is used and 75 percent when neighboring months are used. For example, a 75 percent match rate was observed when vehicles from February-April were matched against vehicles tested in March of a different year. This is the result of individuals having their vehicles tested earlier or later than the deadline.

Figure 2-3 presents a comparison of the MOBILE5 registration distribution for all light duty vehicles with the registration distribution calculated using the Arizona data. The MOBILE light duty autos and trucks were combined together by weighting them according to the default light duty auto and light duty truck travel fractions assumed in MOBILE. In this example, the differences are significant in the older vehicles which also have the highest prevalence of high emitting vehicles. There are striking differences in the later model years, with the Arizona registration distribution being either much higher or much lower than the default MOBILE distribution for corresponding years.

Note that the distribution does not include 1993. In Arizona, as well as many other states, vehicles are not required to undergo an I/M test in their first year unless they are resold.

EXAMPLE APPLICATION FOR MILEAGE ACCUMULATION

The data were first examined for odometer changes that were clearly anomalous. A simple anomaly was when the difference in odometer readings between the two years was negative or zero. Fully seventeen percent of the records had such differences. A visual inspection showed that some were clearly rollovers while others were apparently errors or the result of tampering.

The odometer readings were then processed so that any mileage difference that was negative was assumed to be due to odometer rollover and 100,000 miles were added to the reading. After doing this, 1.1 percent were still negative and 3.7 percent were still greater than or equal to 100,000 miles.

The annual miles calculated after adding 100,000 miles to negative mileages are presented in Figure 2-4. Most of the vehicles are seen to drive near the national average annual miles driven although the distribution has a long tail showing very high annual miles.

Mileage accumulation distributions were then calculated using each of the sets of decision criteria listed in Table 2-5. These are graphed in Figures 2-5 and 2-6. Figure 2-5 shows all four distributions along with the MOBILE default distribution. Each distribution shows higher miles per year than the MOBILE default. The VRC truncation approach, which accepts as valid vehicles listed as driving up to 100,000 miles per year, is significantly higher that all the other distributions. The average miles driven per year with the VRC approach is 20,000 miles. The average of the next highest approach, that of BAR, is 11,300. The high number of older vehicles with calculated high annual miles is thought to be due to a higher incidence of tampering and errors in entering the odometer reading. Since odometers are more likely to be tampered in older vehicles, the fraction of high calculated mileages (i.e., between 40,000 and 100,000 miles) for older vehicles would also increase; this is what was observed in the Arizona data.

Figure 2-6 presents the EEA, BAR, and CARB distributions graphed on a scale that allows the differences between them to be more clearly seen. The MOBILE default distribution forms a smooth curve downward, starting at approximately 14,000 miles per year. The CARB distribution is lower than the MOBILE default for newer vehicles up to five years old and is higher for all later model vehicles. The EEA and BAR distributions are higher than the MOBILE distribution for all years. The BAR distribution is slightly higher in the first year than the EEA distribution because the BAR distribution allows for higher mileage the first year than the EEA distribution does.

Table 2-1
Summary of registration distribution by age Records
(required if MYMRFG = 3 or 4)

Record	Field	Content and Description	Format	Allowable Value
1	1-10	Registration distribution fractions[2] for LDGVs[3] of ages 1,2,...,10	10F5.3	0.0 - 1.0
2	1-10	Registration distribution fractions[2] for LDGVs[3] of ages 11,12,...,20	10F5.3	0.0 - 1.0
3	1-5	Registration distribution fractions[2] for LDGVs[3] of ages 21,22,...,25+	5F5.3/	0.0 - 1.0
4	1-10	Registration distribution fractions[2] for LDGTs[3] of ages 1,2, ..., 10	10F5.3	0.0 - 1.0
5	1-10	Registration distribution fractions[2] for LDGTs[3] of ages 11,12, ..., 20	10F5.3	0.0 - 1.0
6	1-5	Registration distribution fractions[2] for LDGTs[3] of ages 21,22, ...,	5F5.3/	0.0 - 1.0
7-9[4]		Registration distribution fractions[2] for LDGT2s		
10-12[4]		Registration distribution fractions[2] for HDGVs		
13-15[4]		Registration distribution fractions[2] for LDDVs[3]		
16-18[4]		Registration distribution fractions[2] for LDDTs[3]		
19-21[4]		Registration distribution fractions[2] for HDDVs3		
22-24[4.5]		Registration distribution fractions[2] for MCs		

Source: “MOBILE5 User’s Guide”

1 If both annual mileage accumulation rates and registration distributions by age are being input (MYMRFG=4), the two sets of values must be internally consistent (see section 2.2.3.4).
2 Values must sum to 1.0 for each vehicle type. The registration distribution entered as data for MOBILE5 should be based on July 1; MOBILE5 will convert them to a January 1 distribution if emission factors for January 1 are requested.
3 The same set of registration distribution fractions must be entered for LDGVs and LDDVs, and the same set of LDGT1s and LDDTs (see section 2.2.3.4).
4 Record continue in sets of three per vehicle type, following the structure shown above for LDGVs and LDGT1s.
5 For motorcycles only, values of .000 should be used for ages 13 through 25+ (see section 2.2.3.4).

Table 2-2
Summary of annual mileage accumulation rates Records
(required if MYMRFG = 2 or 4[1])

Record	Field	Content and Description	Format	Allowable Value
1	1-10	Average Annual mileage accumulation [2] for LDGVs of ages 1,2, ..., 10	10F7.5	≥0.0
2	1-	Average Annual mileage accumulation[2] for LDGVs of ages 11,12,..., 20	10F7.5/	≥0.0
3	1-5	Average Annual mileage accumulation[2] for LDGVs of ages 21,22,..., 25+	5F5.3/	≥0.0
4	1-10	Average Annual mileage accumulation[2] for LDGT1s of ages 1,2,..., 10	10F7.5/	≥0.0
5	1-10	Average Annual mileage accumulation[2] for LDGT1s of ages 11,12,..., 20	10F7.5/	≥0.0
6	1-5	Average Annual mileage accumulation[2] for LDGT1s of ages 21,22,..., 25+	5F7.5/	≥0.0
7-9[3]		Average annual mileage accumulation[2] for LDGT2s		
10-12[3]		Average annual mileage accumulation[2] for HDGVs		
13-15[3]		Average annual mileage accumulation[2] for LDDVs		
16-18[3]		Average annual mileage accumulation[2] for LDDTs		
19-21[3]		Average annual mileage accumulation[2] for HDDVs		
22-24[3,4]		Average annual mileage accumulation[2] for MCs		

Source: “MOBILE5 User’s Guide”

1 If both annual mileage accumulation rates and registration distributions by age are being input (MYMRFG=4), the two sets of values must be internally consistent (see section 2.2.3.4).
2 Values as input as miles/100,000 (e.g.,24,358 miles in input as .23458
3 Record continue in sets of three per vehicle type, following the structure shown above for LDGVs and LDGT1s.
4 For motorcycles only, values of .00000 should be used for ages 13 through 25+ (see section 2.2.3.3).

Table 2-3
MOBILE5 default mileage accumulation rates per vehicle by vehicle class

Age	LDGV	LDGT1	LDGT2	HDGV	LDDV	LDDT	MC
1	.14390	.15442	.14779	.17251	.17825	.21004	.04786
2	.13612	.14508	.14259	.16185	.16478	.19125	.04475
3	.12875	.13631	.13758	.15185	.15233	.17415	.04164
4	.12180	.12807	.13275	.14246	.14081	.15858	.03853
5	.11522	.12032	.12809	.13365	.13017	.14440	.03543
6	.10899	.11305	.12359	.12539	.12033	.13149	.03232
7	.10310	.10621	.11924	.11764	.11124	.11973	.02921
8	.09751	.09979	.11505	.11037	.10283	.10902	.02611
9	.09225	.09376	.11101	.10355	.09506	.09927	.02300
10	.08726	.08809	.10711	.09715	.08788	.09040	.01989
11	.08254	0.8276	.10335	.09114	.08123	.08231	.01678
12	.07807	.07776	.09972	.08551	.07509	.07495	.01368
13	.07386	.07306	.09621	.08022	.06942	.06825	.00000
14	.06987	.06864	.09283	.07526	.06417	.06215	.00000
15	.06608	.06449	.08957	.07061	.05932	.05659	.00000
16	.06251	.06059	.08642	.06625	.05484	.05153	.00000
17	.05913	.05693	.08339	.06215	.05069	.04692	.00000
18	.05594	.05348	.08046	.05831	.04686	.04272	.00000
19	.05291	.05025	.07763	.05471	.04332	.03890	.00000
20	.05005	.04721	.07490	.05132	.04005	.03543	.00000
21	.04735	.04436	.07227	.04815	.03702	.03226	.00000
22	.04478	.04168	.06973	.04517	.03422	.02937	.00000
23	.04237	.03916	.06728	.04238	.03163	.02675	.00000
24	.04007	.03679	.06492	.03976	.02924	.02435	.00000
25+	.03790	.03456	.06264	.03730	.02703	.02218	.00000

Table 2-4
Mileage accumulation rates calculated in California using I/M and DMVdata

Age	Year	Mileage Accumulation		
		Autos	LDTs	MDVs
0	92	13002	14348	9602
1	92	15772	14954	13582
2	91	13826	15448	20083
3	90	13872	15999	16859
4	89	13447	15387	15534
5	88	12942	14550	15259
6	87	12300	13635	13911
7	86	11634	12921	12314
8	85	11046	12072	12264
9	84	10551	11362	11722
10	83	10189	10724	11497
11	82	9906	10592	10318
12	81	9771	10330	9438
13	80	9804	10097	10407
14	79	9187	9729	9771
15	78	8912	9516	10552
16	77	8586	9323	10147
17	76	8322	9276	8644
18	75	8057	9400	8819
19	74	8221	9160	9535
20	73	7770	8859	8593
21	72	7654	8646	9060
22	71	7875	8322	10373
23	70	7803	8159	8680
24	69	7720	8207	7542
25	68	7516	7994	5462

Table 2-5
Summary of decision criteria used in development of mileage accumulation rates from I/M data

Study	Vehicle Matching	**Low mileage cut-off (miles per year)**	**High Mileage Cut-off (miles per year)**	**Minimum Time Between Tests**
BAR, 1994	License plate	None	60,000 for vehicles < 2 years old; 50,000 for vehicles 2 years old and 40,000 for vehicles > 2 years old	300 days
EEA, 1985	VIN number or combination of license, make and model year	500	40,000	none
VRC, 1994	VIN number	0	100,000	6 months
ARB, 1993	License Plate	0	30,000	Not discussed

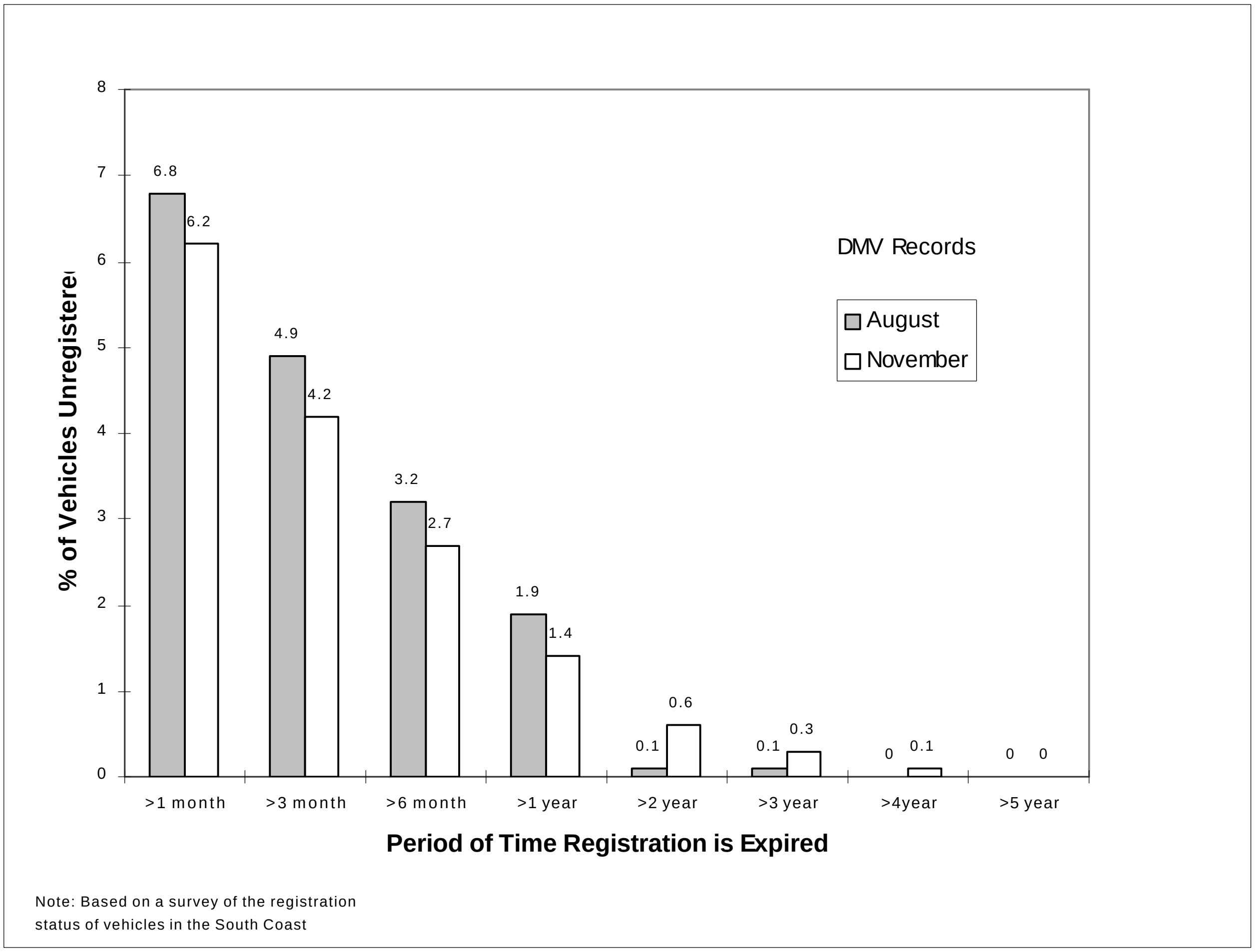

Figure 2-1
Relationship between registration status of California vehicles and period of time with expired registrations. Source: "Status Report - Unregistered Vehicle Survey," ARB, July 25, 1990

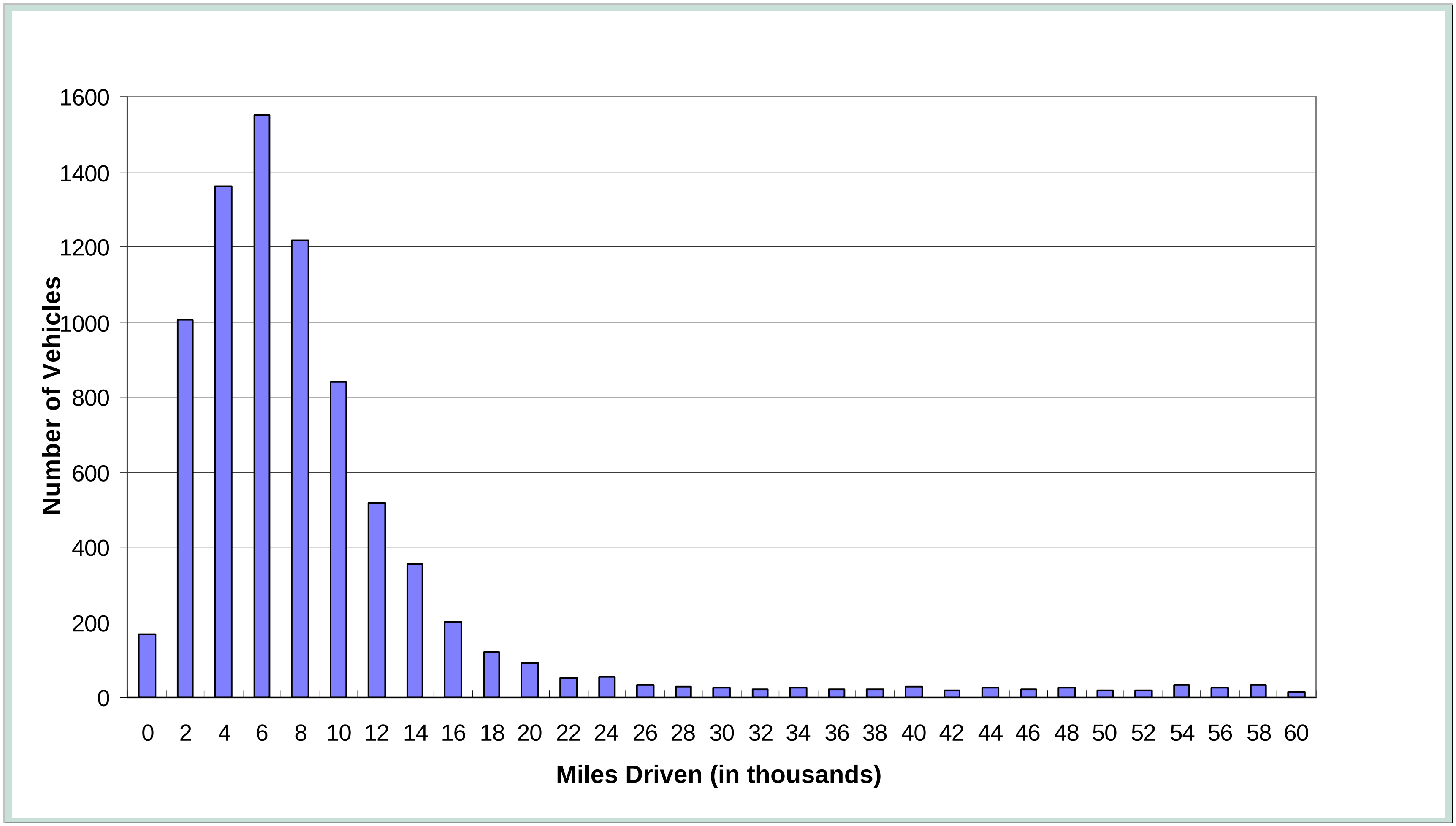

Figure 2-2
Example mileage accumulation distribution

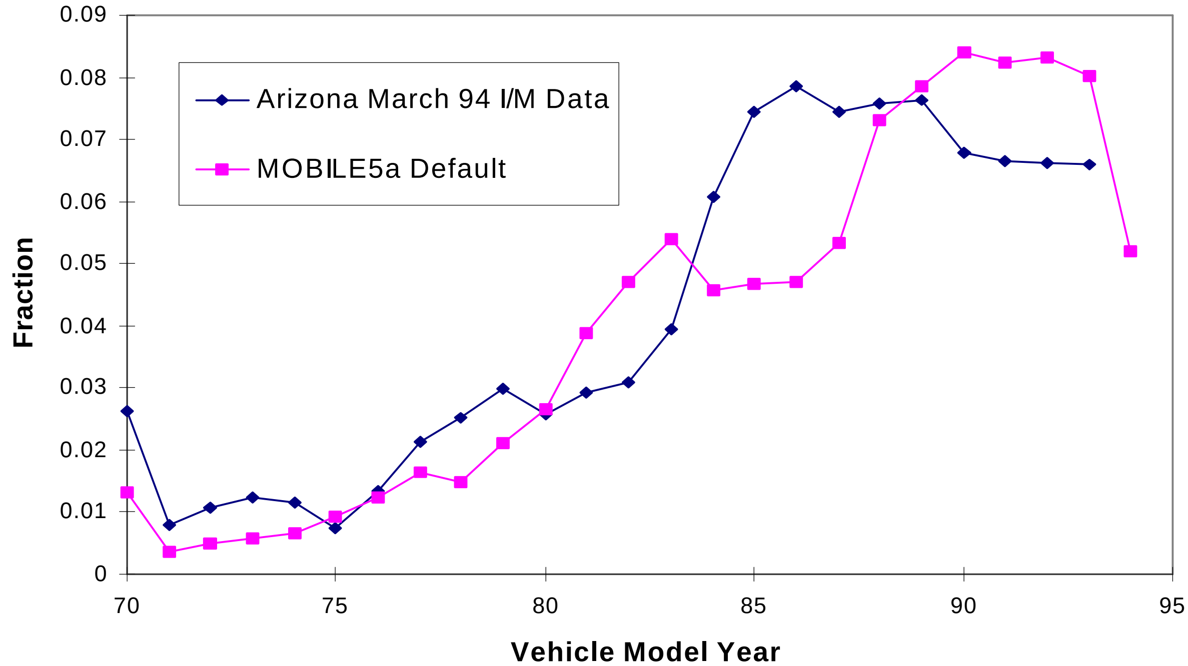

Figure 2-3
Comparison of MOBILE5 and ArizonaI/M data base registration distributions for all light duty vehicles

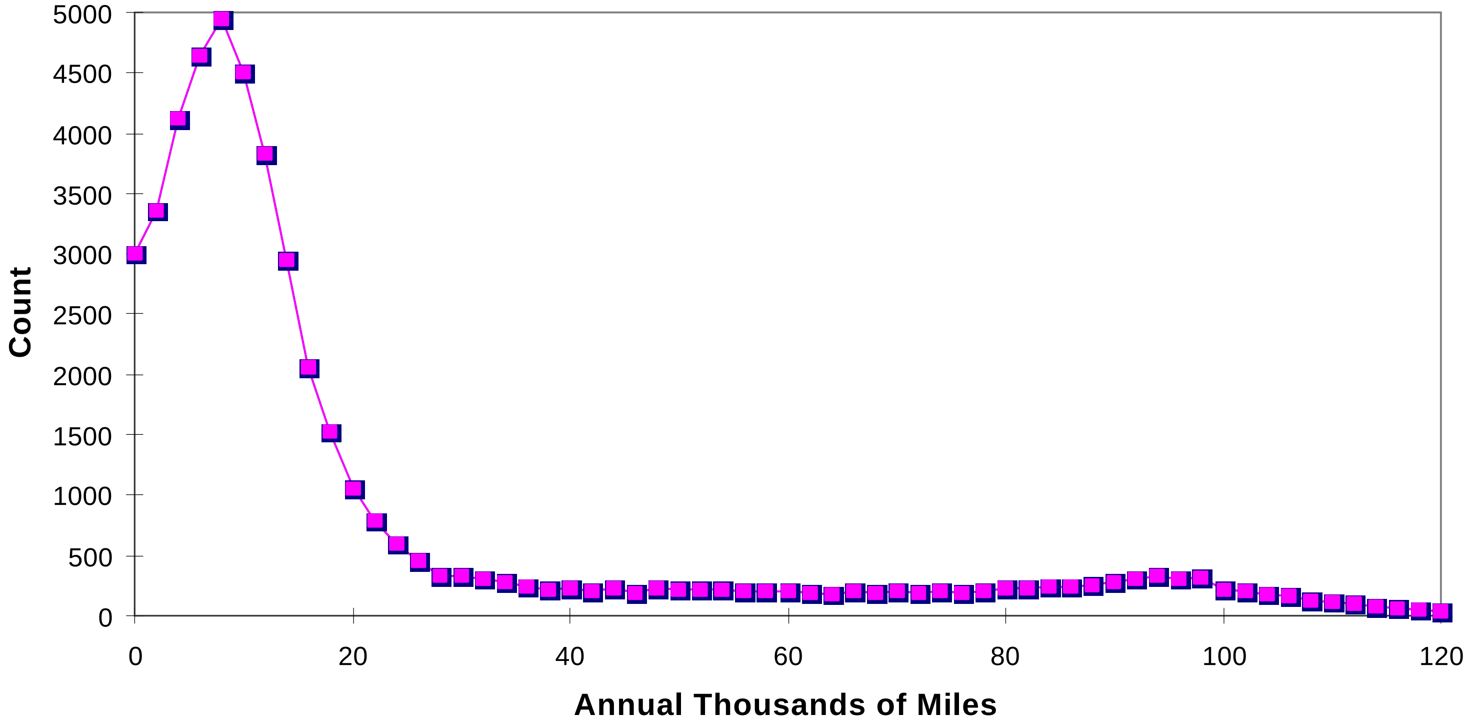

Figure 2-4
Arizona I/M mileage distribution across all vehicle ages

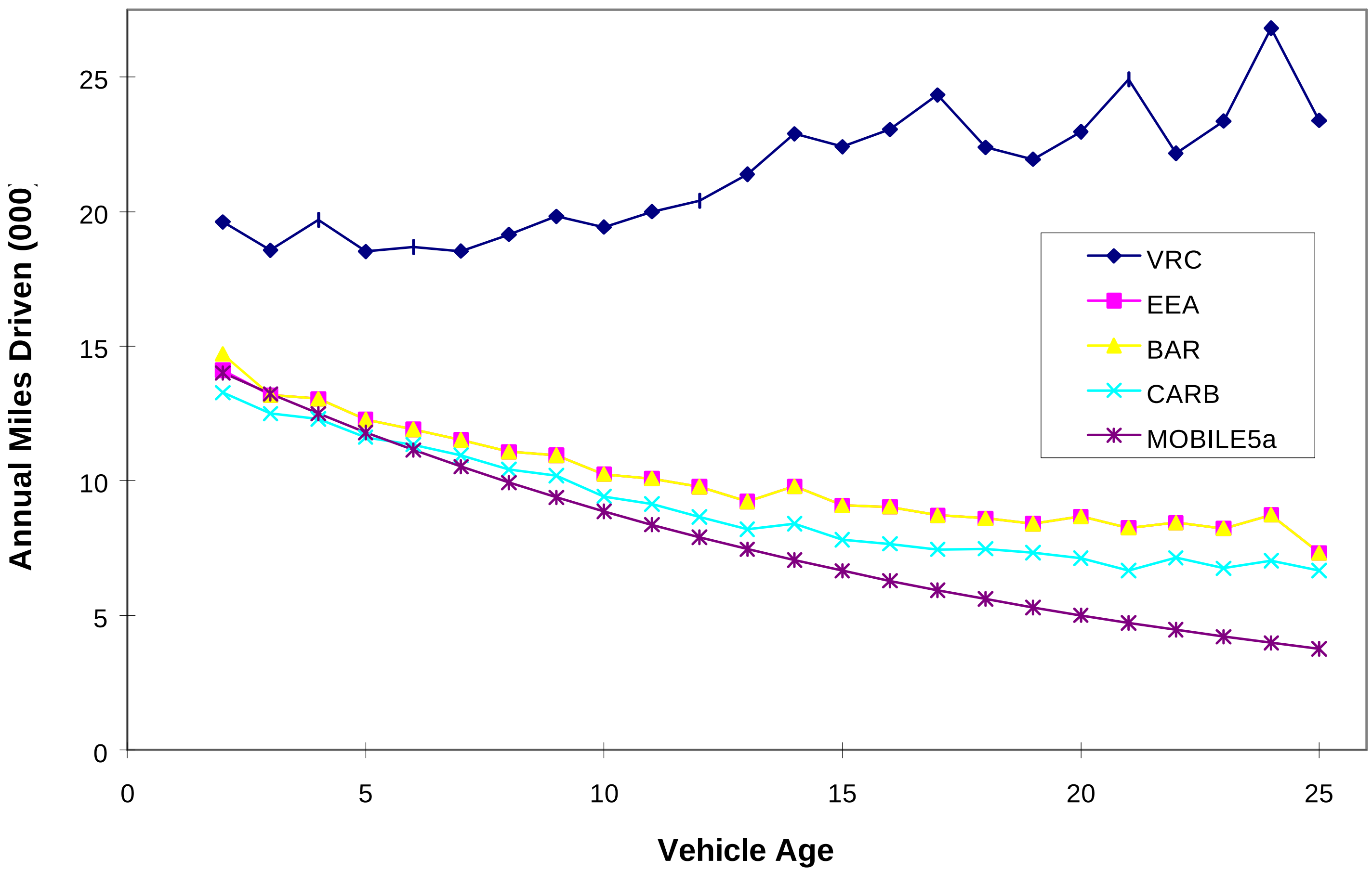

Figure 2-5
Comparison of Arizona I/M mileage accumulation distributions under different truncation schemes

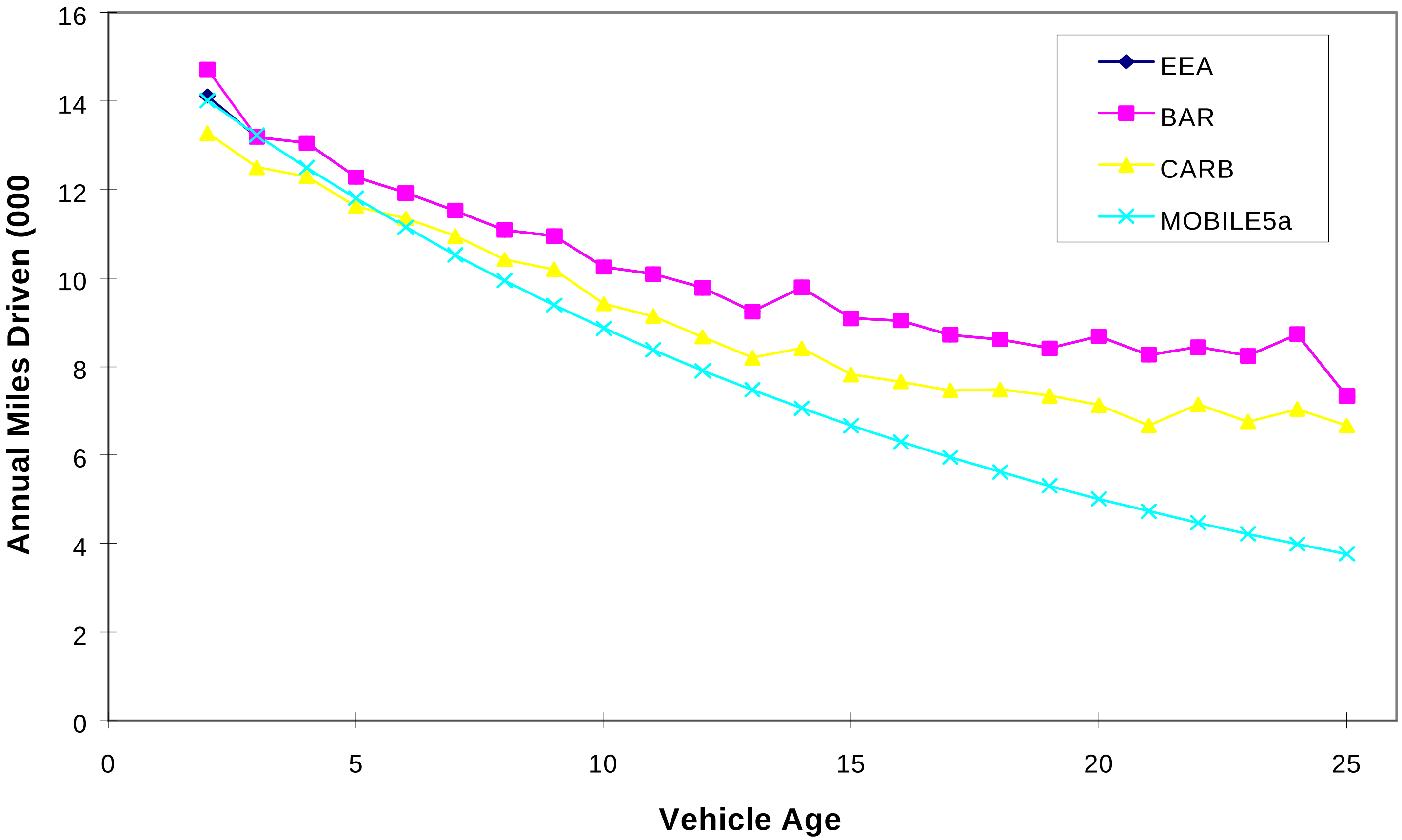

Figure 2-6
Comparison of mileage accumulation distributions under three different truncation schemes

3
USE OF REMOTE SENSING DATA TO GENERATE VEHICLE ACTIVITY CHARACTERISTICS

Data from large-scale remote sensing programs can be used to generate regional fleet characteristic data for input into to mobile source emission inventory development and the MOBILE emission factor model. This section describes how the following vehicle activity characteristics can be derived or approximated from remote sensing data for use in emission inventory development:

- registration distributions,
- VMT mix,
- diurnal travel distributions,
- diesel sales fractions,
- evaluation of unregistered vehicles, and
- evaluation of vehicle county/state of origin to determine vehicle I/M status.

The remote sensing data can provide information at a regional level which can be significantly different from default data currently used to generate many mobile source emission inventories. In addition, if sufficient data are available, it may be possible to evaluate data at the facility class level (i.e., roadway classification) or the transportation model zonal level allowing for additional refinement of emissions estimates. Note that for such uses, the scale of the remote sensing program needs to be large enough to capture a significant portion of the vehicle fleet, such as the one completed as part of the California Pilot I/M program.

Although the primary focus of most remote sensing programs is to collect emissions information, fleet characteristics can be derived from license plate data typically recorded during a remote sensing program. The license plate data, cross checked with state motor vehicle records, can yield valuable information on the characteristics of the local fleet. These local data can offer an improvement over generally used national default or state-level fleet characteristics data.

The remainder of this section covers (1) an overview of fleet data from remote sensing programs useful in understanding how these data are used, (2) a literature review of previous studies using remote sensing to develop fleet data, (3) the methodology for estimating fleet characteristics from remote sensing data, and (4) example applications of this methodology.

OVERVIEW OF FLEET DATA FROM REMOTE SENSING PROGRAMS

The methodologies described in this document summarize how fleet characteristics and activity data can be distilled from the vehicle data of remote sensing programs (note that the emissions data of these programs are not used). For example, vehicle counts, roadway type, and time of day data can yield information on hourly variation of travel by roadway classification. Moreover, when remote sensing data are combined with license plate data that are cross-checked with state motor vehicle records, they can provide valuable information on the age and vehicle class make up of the fleet.

In brief, the remote sensor apparatus is triggered when a vehicle blocks the path of an infra-red beam set across a lane of traffic. Once the vehicle clears the path, the sensor takes a measurement of the exhaust plume. In general, the light-duty cars and trucks are the vehicle group of interest, thus the sensor is located about ten to twelve inches off the ground. A video camera is used to record the license plate of the passing vehicle. License plates are determined either through manual review of video recordings or through electronic license plate readers. License plate data are matched with state records to produce the data needed for the evaluation of fleet characteristics.

The data generally collected that are useful for the evaluation of activity and characteristics of the fleet include the following:

- site description and location,
- time of day,
- date of measurement, and
- vehicle licence plate.

From the license plate data, the following can be determined from state registration records:

- vehicle identification number (VIN)
- vehicle model year,
- vehicle class,
- vehicle fuel type,
- county and/or state of registration, and
- registration expiration date.

Except for the vehicle identification number (VIN), not all states keep each of these on record. If this is the case, these items can be determined from the VIN that is unique to each vehicle.

VIN decoding can be through VIN decoding software supplemented with manual decoding using VIN decoding books where needed.

Noteworthy considerations in the evaluation of fleet and activity data from remote sensor data include (1) regional variation and fleet coverage, (2) facility (roadway) class limitations, (3) vehicle class biases, (4) VIN decoding considerations, and (5) using vehicle counts to estimate travel, population and sales. These are discussed individually below.

REGIONAL VARIATION AND FLEET COVERAGE

Variation of socioeconomic conditions and land-use categories (e.g., commercial, industrial, and residential) result in variation in fleet characteristics. In order to accurately evaluate the vehicle fleet as a whole, it is important that remote sensing be completed at several locations to ensure fleet coverage. As a benchmark, the Enhanced I/M Final Rule includes requirements that on-road testing of 0.5 percent of the fleet through remote sensing be completed. It is programs of this magnitude that can utilize remote sensing data for the evaluation of fleet characteristics. If a program is sufficiently large, such as the California Pilot I/M Program where 47 percent of the eligible fleet was measured at over three hundred locations (Radian, 1995), subregional analyses can be completed. Subregional information can be used to evaluate data by zone for use with travel demand models or by roadway facility class.

FACILITY CLASS LIMITATIONS

Classification by facility class can be made provided that individual sites are identifiable within the vehicle database and that each sites is assigned a roadway classification. This is generally the case for most programs. Remote sensor technology is currently limited to use on single-lane roadways. Some initial work has been done using remote sensing on two traffic lanes and has been generally successful as long as travel densities are low (two seconds between passing vehicles) (Bishop et. al., 1994). Commonly used remote sensor sites include single-lane roads, multi-lane roads where travel is restricted to one lane, expressway on and off ramps, and single-lane expressway interchanges. Because of the restriction on roadway types, the remote sensing sample may not accurately reflect the distribution of travel on each facility class. It is reasonable to expect that the characteristics of limited-access roadways (e.g., expressways and freeways) may be different from surface streets. For example, limited-access roadways may have more out-of-state vehicles or heavy-duty vehicles. For this reason, it may be desired to evaluate the characteristics of limited-access roadways separately from surface streets. The data for surface streets and limited-access roadways can be combined into an overall value according the relative VMT proportion of each roadway from the Highway Performance Monitoring System (HPMS) VMT data.

VEHICLE CLASS BIASES

Remote sensing is generally used to evaluate the light-duty fleet. The larger the vehicle, the less likely the apparatus will obtain a valid measurement. For example, the wheels or the cab of large trucks can trigger the remote sensor device resulting in an invalid emissions and license plate reading. For the purposes of evaluating fleet characteristics, retaining the records of license plates even if valid emissions data are not recorded will enhance the representativeness of larger vehicle classes in the database. It is therefore recommended that the evaluation of fleet characteristics from remote sensing data be restricted to the light-duty vehicle classes[1] unless review of license plate data show that heavier vehicles are captured regularly (e.g., manual review of license plate data may capture the heavier vehicle classes).

VIN DECODER LIMITATIONS

Vehicle identification number (VIN) decoding may be required if state records do not include vehicle class, vehicle model year, and/or vehicle fuel type data. The VIN is a 12 to 17 digit alpha-numeric field which, by law, is unique to each vehicle. Since the 1981 model year vehicle, the format of the VIN has been standardized. Most states maintain VIN decoders in some form. VIN decoding to obtain vehicle data can be done manually although this is not recommended especially for pre-1981 vehicles. VIN decoding software can be obtained from various sources (e.g., the Highway Loss Data Institute, R. L. Polk, and Radian Corporation) that includes the complex algorithms required to evaluate older model year vehicles. There are some limitations to VIN decoding even when utilizing decoding software. For example, for older model years, a VIN may only indicate a range of model years instead of a specific value. For increased accuracy, vehicle data can be obtained through VIN decoding in combination with state records. A previous example of how this was completed is included in an EEA study of I/M data for the EPA (EEA, 1985).

USING VEHICLE COUNTS TO ESTIMATE THE FRACTION OF TRAVEL, POPULATION AND SALES

Each record of the remote sensing database represents a single vehicle. Classifying each record, for example, by time of day or age allows for estimating counts by each classification. The fraction of a particular activity, such as vehicle travel, is determined by the total count by each classification divided by the total count of the database. In completing this calculation, however, it is important to understand the distinction between different types of activity

[1] There are five light duty classes defined by MOBILE5a, light-duty gasoline vehicles (LDGV), light-duty gasoline trucks less than 6,000 pounds gross vehicle weight (LDGT1), light-duty gasoline trucks more than 6,000 pounds gross vehicle weight (LDGT2), light-duty diesel vehicles (LDDV), and light-duty diesel trucks (LDDT).

(vehicle miles traveled, vehicle population, and vehicle sales) that are discussed in this section.

Vehicle miles traveled (VMT) represents the distance traveled by each vehicle. In this document, it assumed that VMT is proportional to the vehicle counts of the remote sensing database. In other words, the more a vehicle travels, the greater the chance it will pass by and be measured by a remote sensor. The fraction of VMT used to determine VMT mixes and diurnal travel distributions (discussed in the methodology portion of this section) can be determined from the fraction of vehicle counts from the remote sensor database assuming that travel is proportional to the vehicle counts.

Vehicle population or sales represent the number of vehicles in the fleet independent of how much each vehicle is driven. For the purpose of evaluating vehicle population or sales from the remote sensing database, one should remove duplicate vehicle records (i.e., vehicles that appear more than once). This is because for population and sales, it is the count of unique vehicles which best represents these activities and because the probability of multiple measurements is proportional to how much a vehicle is driven.[2] The fraction of population or sales (used to define registration distributions and diesel sales fractions) is then determined from the fraction of vehicle counts after duplicate vehicle records have been removed.

Remote sensing should be used with great caution to estimate travel fractions, population, and sales distributions. As mentioned above, fleet characteristics can vary greatly with neighborhood socioeconomic status and landuse (e.g., residential vs. Industrial). Accurate portrayal of fleet characteristics would require a well-designed mix of remote sensing locations.

LITERATURE REVIEW

The literature review summarized below is separated into two sections covering: (1) a document review of key references; and (2) verbal communications with regulators and scientists experienced in the evaluation of remote sensing data. The review focuses on methods of identifying fleet characteristics data such as vehicle class and model year in the remote sensing databases, as well as example calculations of fleet characteristics data, and description of potential biases of remote sensing databases and methods of adjusting for these biases.

[2] Note that it is common to have multiple measurements of vehicles in the remote sensing database, because measurements are often taken at the same or nearby sites on different days. In the example applications included in this section, 56 percent of the vehicles in the database were measured more than once.

DOCUMENT REVIEW

The documents covered in this review are:

1. *Evaluation of the California Pilot I/M Program* (Radian, 1995),

2. *Analysis of Data from the California Enhanced I/M Program* (Sierra, 1995),

3. *Comparison of Remote Sensing Data and Emission Factor Models: The Proportion of Emissions From High Emitting Vehicles* (Pollack et. al., 1992),

4. *Estimation of Mileage Accumulation Rates and I/M Failure Rates from I/M Program Data* (EEA, 1985), and

5. *On-Road Motor Vehicle Activity Data, Volume II: Vehicle Age Distribution and Mileage Accumulation Rate by County* (Valley Research, 1994).

The following documents were also reviewed but did not offer sufficient information relative to use of remote sensing data to generate vehicle activity inputs:

1. *Evaluation of a Remote Sensing Device at a Centralized I/M Lane* (EPA, 1992b),

2. *Identifying Excess Emitters with a Remote Sensing Device: a Preliminary Analysis* (EPA, 1991), and

3. *Methodology for Estimating Emissions From On-Road Motor Vehicles, Volume II: Weight (E7FWT)* (ARB, 1993).

Evaluation of the California Pilot I/M Program (Radian, 1995)

This report, completed for the State of California, describes the California Pilot I/M Program and presents the results of an initial evaluation of the data collected. A portion of the pilot program was a large-scale remote sensing program conducted in the Sacramento region in the summer of 1994. In this program, remote sensors were placed at more than three hundred locations resulting in more than two million measurements covering 47 percent of the registered Sacramento fleet.

Detailed information is provided describing the fleet coverage of the remote sensing program and how to optimize coverage. Also, a detailed model year analysis is presented examining the differences between the model year distribution of the state registration data for Sacramento and the model year distribution of the fleet detected by the remote sensor. Notably, the Radian report adjusted state registration data to reflect the model year variation in annual travel (i.e., new vehicles are driven more than older vehicles) for a more equitable comparison to the remote sensing data. This adjustment consisted of:

$$MYDist_{i,\ adjusted} = MYDist_{i,\ unadjusted} * \frac{MAR_i}{MAR_{fleet\ average}}$$

where:

MYDist is the model year registration distribution,
MAR is the mileage accumulation rate,
the subscript *i* indicates each model year, and
the subscript *fleet average* is the average over all model years.

In this adjustment, the mileage accumulation rate (MAR) data were taken from the California Air Resources Board (ARB) EMFAC7F emission factor model. For light-duty automobiles, these MAR data are based on analysis of California I/M Program odometer data.

It should be noted that the Radian equation above can be used more generally to adjust for any model year bias as long as the variation of each model year versus the fleet average is known. This section discusses adjusting the remote sensing data to more accurately reflect vehicle registration which is the opposite of the Radian adjustment. Therefore the inverse of the adjustment ratio shown above will be used.

It should also be noted than even after the Radian adjustment of the model year variation in mileage accumulation rates was completed, the remote sensing data still appear to underrepresent older model years (1979 and earlier). This may be due to a combination of two effects: (1) the mileage accumulation rates used in the adjustment noted above may not reflect local conditions, and/or (2) additional biases may be present, e.g., older license plates may be more difficult to read. Radian, however, reports the differences in the model year distributions of the registration data and the remote sensing data to be statistically insignificant.

Analysis of Data from the California Enhanced I/M Program (Sierra, 1995)

This report, completed for the U.S. EPA, reviews and evaluates the data of California Pilot I/M Program. Of importance to this work assignment is the discussion of the coverage of the remote sensing program. Sierra reports that the remote sensing program coverage of the registered fleet varies significantly by model year as summarized by the following capture percents of four model year groups.

1983-and-later	48%
1981-1982	35%
1975-1980	27%
pre-1975	19%

Sierra reports that this model year variation is likely due to lower mileage accumulation rates of the older vehicles, different driving patters of the older fleet, and/or representativeness of the remote sensing sites. Sierra did not attempt to correct for these biases as was done in the Radian report.

Note that there is an apparent discrepancy between the 47 percent coverage reported by Radian and the model year coverage reported by Sierra (i.e., if all of the Sierra-reported model year groups were combined, they would not yield 47 percent). This may be due to an examination of a subset of data, for example, only light-duty automobiles.

Comparison of Remote Sensing Data and Emission Factor Models: The Proportion of Emissions From High Emitting Vehicles (Pollack et. al., 1992)

This paper, presented and the 85th Annual Air and Waste Management Association Meeting, examines the emissions distribution of remote sensing data versus that of the EMFAC and MOBILE emission factor models. Registration distributions (weighted for model year differences in annual miles traveled) from remote sensing data were used with the emissions factor models for comparative purposes. The paper also describes why distributions are adjusted for model year differences in annual miles traveled.

Estimation of Mileage Accumulation Rates and I/M Failure Rates from I/M Program Data (EEA, 1985)

This report summarized work completed by EEA for the U.S. EPA in 1985. Of particular interest is a detailed discussion of the method used to match motor vehicle data records. The report describes how the VIN was used to verify state I/M data for vehicle class and model year. EEA utilized VIN decoding software called VINDICATOR produced by the Highway Loss Data Institute and supplemented VINDICATOR with additional decoding data on imported light-duty trucks. There are some noted deficiencies in the VIN decoding for older vehicles (pre-1980 model years) because of lack of industry agreement on codes until the 1980 model year. For example, VIN decoding can only identify a range of model years (instead of a specific model year) in some cases. For 1981 and later model years, VINs were standardized making VIN decoding much more reliable.

The information in this report would be of use to states that do not maintain model year or vehicle class information in the state records. In this case, VIN decoding would be required to get the information required to estimate fleet characteristics data.

On-Road Motor Vehicle Activity Data, Volume II: Vehicle Age Distribution and Mileage Accumulation Rate by County (Valley Research, 1994)

This report, completed by Valley Research Corporation for the ARB, evaluated county-specific fleet characteristics data from the state department of motor vehicles (DMV) data and state I/M data. The relevance of this report is that the state DMV data provide an example of how DMV data could be used to identify vehicle coverage of a remote sensing program. Notable aspects of this study for the the issues at hand are:

- Valley Research utilized an electronic VIN decoder, called VINA (developed by R. L. Polk and used by the State of California), to determine a vehicle's model year instead of relying of the model year data included in the records of the California Department of Motor Vehicles (DMV). The VIN-decoded model year was considered more reliable than that input into the DMV database.

- Wide variation was found in model year distributions by county. For example, the mean age of automobiles ranges from 5.1 years to 13.1 years. This indicates the importance of using region-specific data in place of statewide averages where possible.

VERBAL COMMUNICATIONS

Personal contacts were made with individuals experienced in processing remote sensing data. The agencies contacted included the California Bureau of Automotive Repair (BAR), the Arizona Department of Environmental Quality (ADEQ), the Department of Chemistry of the University of Denver, and the U.S. EPA Motor Vehicle Emission Laboratory.

California BAR

The California Pilot I/M program was discussed with the Bureau of Automotive Repair. They noted that there are obvious limitations in measuring the vehicle fleet accurately. For example, the remote sensing apparatus has the potential to measure all vehicles; however, in general, the setup is designed to measure the light-duty fleet thereby missing the license plate information of the majority of buses, articulated trucks and motorcycles. With respect to model year or age biases of remote sensing programs, it was noted that the plate recorder had a higher success rate on the newer California licence plates, which have a reflective coating added to the plate. There are no data on how this affected the results.

Arizona DEQ

The specifics of the current Arizona program were discussed, as well as potential biases of the license plate recording apparatus with DEQ staff. The automatic recorder used by Arizona decodes the information based on the font shape and spacing to distinguish an Arizona plate from other states. In a couple of instances a Michigan plate has been misread as an Arizona plate.

The sensor apparatus and plate reader are located about 12 inches off the ground. There is a problem measuring jacked-up pickups, but in general most light-duty vehicles are measured. The plate recorder has about a 75 percent success rate; this can go up to about 90 percent if the remote sensing operator reviews the recorded images and fills in missing plate data. In the Arizona program, this back filling of missing data is completed.

University of Denver, Department of Chemistry

Issues related to obtaining vehicle information from state DMVs were discussed. It was noted that DMV VIN transcription errors are typically one to two percent. However, these errors are generally slight and do not statistically alter the results. It was also noted that VINs are very unreliable prior to the 1981 model year. They use several older VIN decoder books to assist in looking at older model years. The University has had some difficulty in the past with some states in cross referencing remote sensing data with state records; however, most states have the capabilities and are responsive to their requests.

METHODOLOGY

This subsection summarizes the methodology for using remote sensing data to develop the following mobile source emission inventory modeling inputs:

- registration distributions,
- VMT mix,
- diurnal travel distributions,
- diesel sales fractions,
- evaluation of unregistered vehicles, and
- evaluation of vehicle county/state of origin to determine vehicle I/M status.

Each of these items is discussed in detail in the following.

EVALUATION OF VEHICLE REGISTRATION DISTRIBUTIONS

Vehicle registration distributions are the fractional distribution of vehicle class population by model year. The registration distribution accounts for the age of the fleet and has a significant impact on emissions calculations. For each model year of a vehicle class, the registration distribution is determined by the population of the model year divided by the population of the entire vehicle class. For MOBILE5a, a twenty-five year distribution is used where the 25th year represents all model years of 25 years and older.

The steps required for the evaluation vehicle registration distributions in the format required by MOBILE5a are:

1. Separate the data by vehicle class and remove duplicate vehicle records.
2. Calculate the unadjusted distribution.
3. Adjust the data for age-dependent mileage.
4. Adjust the data for other biases.
5. Adjust the data to represent July 1st of the calendar year.

Each of these steps is discussed here.

Separate the Data by Vehicle Class and Remove Duplicate Records

Registration distributions are input into MOBILE5a for each vehicle class requiring the separation of data by vehicle classes. Note that for light-duty vehicles and trucks, MOBILE5a requires a single registration distribution that includes both diesel- and gasoline-powered vehicles (i.e., one distribution for LDGV and LDDV combined and one distribution for LDGT1 and LDDT combined); the model will not run unless the distributions of these classes are equivalent. In general, remote sensing data will only capture the light-duty fleet accurately (see Section 2.2.3 of the MOBILE5a User's Guide). Thus for the heavier vehicles, other data sources or the model default values should be used.

Calculate the Unadjusted Distribution

The unadjusted distribution is calculated as the sum of population of each model year divided by the sum of population of all model years of the vehicle class. This is shown by the following equation.

$$MYDist_{i,\ unadjusted} = \frac{\sum POP_i}{\sum POP} \tag{3-1}$$

where:

MYDist is the model year registration distribution,
POP is the vehicle population, and
the subscript *i* indicates each model year.

For consistency with MOBILE5a format requirements, the 25th model year includes all vehicles 25 years old and older.

Adjust the Data for Age-dependent Mileage

The unadjusted registration distribution should be adjusted to account for age-dependent mileage accumulation. Because older vehicles travel less than newer vehicles, older vehicles are less likely to have been included in the remote sensing database and the population of older vehicles in the unadjusted data are likely underrepresented. This adjustment for age-dependent mileage is shown by the following equation.

$$MYDist_{i,\ adjusted} = MYDist_{i,\ unadjusted} * \frac{MAR_{fleet\ average}}{MAR_i} \tag{3-2}$$

where:

MYDist is the model year registration distribution,
MAR is the mileage accumulation rate,
the subscript *i* indicates each model year, and
the subscript *fleet average* is the average over all model years.

Ideally, MAR data used in this adjustment should be representative of the local fleet. In the absence of data, the default mileage accumulation rates from MOBILE can be used. After applying this equation, the adjusted model year distribution should be renormalized to sum to one. From the equation above, the adjustment is inversely proportional to the model year MAR thus adjusting the data downwards for newer model years with a MAR greater than the fleet average and adjusting the data upwards for older model years with a MAR less than the fleet average. Note that this equation is the inverse of the adjustment used by Radian (Radian, 1995). It is the inverse because in the Radian report, the registration data were adjusted to more accurately reflect the remote sensing data whereas in this analysis, the remote sensing data are adjusted to more accurately reflect the registration data.

Adjust the Data for Other Biases

The equation shown above to adjust for age-dependent mileage can also be applied more generally to account for any model year bias in the database as long as the variation of each model year versus the fleet average is known. For example, the California BAR noted that the accuracy of the automatic license plate reader to interpret different types of license plates varied. Newer license plates have a reflector coating aiding in the ability of the reader to evaluate the plate. However, the effect of varying license plate types was not quantified in the California Pilot I/M Program.

Adjust the Data to Represent July 1 of the Calendar Year

The MOBILE5a registration distribution data represent July 1 of the calendar year of evaluation. For input into the MOBILE5a model, the data need to be adjusted to represent July 1. MOBILE5a assumes that light-duty vehicle sales begin on October 1 of the year prior to the model year (i.e., 1995 model year vehicles begin selling on October 1, 1994). For other vehicle classes, the model assumes that sales begin on January 1 of the model year. MOBILE5a also assumes that sales are uniform throughout a year.

The process to adjust light-duty vehicles to represent July 1 can be separated into two cases:

1. For remote sensing data collected after July 1 and before October 1, too many current model year vehicles are included in the database. The excess percent of current model year vehicles is the number of days past July 1 divided by 365. For example, if the remote sensing data were collected on September 1, the current model year is 16.7 percent (61 days divided by 365 days) too large. Excess percent of the current model year are removed from the distribution. The registration distribution is then renormalized to sum to one.

2. For remote sensing data collected before July 1 and after October 1, too few current model year vehicles are included in the database. The percent missing current model year vehicles is the number of days prior July 1 divided by 365. The current model year distribution is adjusted upward by the percent of vehicles missing. The registration distribution is then renormalized to sum to one.

EVALUATION OF DIURNAL TRAVEL DISTRIBUTIONS

Diurnal travel distributions are the hourly variation of vehicle traffic. The diurnal distribution of travel is important in air quality modeling because it affects the time of day emissions are predicted to be released. Remote sensing data can provide data on diurnal travel patterns during the period the sensor is operating. Emissions modeling generally employs the use of separate diurnal distributions for weekday and weekend travel. Generally, diurnal distributions are not developed for individual vehicle classes. Since identification of vehicle classes is not required, the remote sensing data do not need to be matched with state records. The following are the steps required to evaluate diurnal travel distributions from remote sensing data:

1. Separate the data by hour of day and day of week.
2. Evaluate the representation of directions of travel.
3. Evaluate the diurnal distribution of the portion of the day not covered by remote sensing.

Each of these steps is discussed below. Note that diurnal distribution data for the hours not covered by the remote sensing program will be required to complete the evaluation of a 24-hour diurnal distribution.

Separate the Data by Hour of Day and Day of Week

Diurnal travel patterns vary by each day of the week. In general, weekday travel is similar enough so that one distribution can be used to represent all five weekdays. Sunday and Saturday diurnal distributions can be determined separately or together depending on how much variation between days is observed. The general equation for determining the diurnal distribution is the sum of travel for each hour divided by the sum over all hours:

$$Diurnal = \frac{\sum Travel_i}{\sum Travel} \qquad (3\text{-}3)$$

where:

Diurnal is the diurnal distribution,
Travel is the vehicle count, and
the subscript *i* indicates each hour.

Evaluate the Representation of Directions of Travel

It is important when using remote sensing data to make sure travel in both directions of roadways are included to capture commute travel patterns. In the California Pilot I/M

Program, remote sensing was generally performed for both directions of traffic for each roadway site in the database, so the representation of travel should be accurate. To illustrate this point, two sites (Site 93 and Site 34) representing Highway 50 in opposite directions of travel are shown in Figure 3-1. Site 93 is dominated by the afternoon commute traffic; Site 34 is dominated by the morning commute traffic. Thus, it is essential that equal representation of directions of travel be present when determining the diurnal distribution from remote sensing data.

Evaluate Diurnal Distribution of the Portion of the Day Not Covered by Remote Sensing

Remote sensing is generally not performed across all 24 hours of a day. A second source of diurnal distribution data, such as that used in EPS2 (EPA, 1992a) or locally available data, is required to complete the 24-hour diurnal distribution. The steps to do this are (1) calculate the fraction of travel covered by the remote sensing program from the secondary data source, (2) multiply each hour of the diurnal distribution calculated from the remote sensing program data by this fraction, and (3) use the diurnal data from (2) for the hours covered by the remote sensing program and use the diurnal data from the secondary source for the remaining hours.

EVALUATION OF VMT MIX

VMT mix is the fraction of travel among the individual vehicle classes. The MOBILE model uses VMT mix to estimate a fleet average emission factor. VMT mix can vary by roadway classification, thus vehicle class data derived from a remote sensing program can offer a method for estimated VMT mix on various roadways. The following are the steps required to evaluate the VMT mix from remote sensing data.

1. Separate the data by vehicle class.
2. Combine with data for vehicle classes not captured by the remote sensing apparatus.

These two steps are discussed below.

Separate the Data by Vehicle Class

VMT mix is calculated by separating the DMV-matched database into individual vehicle classes. The mix is calculated from the sum of each vehicle over the total records in the database:

$$VMT\ \ Mix\ = \frac{\sum Travel_i}{\sum Travel} \tag{3-4}$$

where:

VMT Mix is the calculated VMT mix,

Travel is the vehicle count, and
the subscript *i* indicates each vehicle class.

If sufficient data are available, VMT mix by facility class (i.e., roadway classification) can be estimated. In this case, the data would be separated into facility classes as well as vehicle classes prior to calculating the VMT mix.

Combine with Data for Vehicle Classes Not Captured by the Remote Sensing Apparatus

As was noted above, the remote sensing apparatus is less likely to accurately capture the heavier vehicles. MOBILE5a requires that all vehicle classes be included in a user-supplied VMT mix. An alternate source of VMT mix data from the MOBILE model or from local sources should be used to fill in the remaining vehicle classes not captured by the remote sensing program. The steps to complete this are (1) calculate the fraction of VMT covered by the vehicle classes of the remote sensing program, (2) multiply the VMT mix of each vehicle class by this fraction, and (3) combine the revised VMT mix from (2) with the remaining vehicle classes of the alternative source to complete the VMT mix. An example of these steps follows.

Let's assume that the remote sensing program accurately identifies the LDGV, LDGT1, and LDGT2 vehicle classes. It is determined from the existing inventory data that 75 percent of the VMT is covered by these vehicle classes (step 1). From the remote sensing data, the VMT mix for each of these vehicle classes is determined to be:

LDGV	LDGT1	LDGT2
0.650	0.250	0.100

In order to combine these values with those not covered by the remote sensing program, multiply the values above by the total contribution of these vehicle classes (0.75 or 75 percent) to yield:

LDGV	LDGT1	LDGT2
0.488	0.187	0.075

These values can then be combined with those of the remaining vehicle classes so that the total over all vehicle classes sums to one.

EVALUATION OF DIESEL SALES FRACTIONS

Diesel sales fractions are the fraction of each model year vehicles which are diesel powered. Diesel sales fractions are used in MOBILE for the light-duty vehicles to calculate the diesel travel fraction of light-duty diesel vehicles and trucks (LDDVs and LDDTs). These vehicle classes differ from their gasoline counterparts in the chemical speciation of exhaust VOC emissions (diesel exhaust VOC is generally more reactive than gasoline exhaust VOC), the quantity of evaporative emissions (diesel vehicles have little or no evaporative emissions), the quantity of particulate emissions (diesels generally emit more particulate emissions). The steps required to estimate diesel sales fractions are as follows.

1. Separate the data by vehicle class/model year and remove duplicate vehicle records.
2. Calculate the diesel sales fraction.

Duplicate vehicles are removed from the calculation since vehicle population is to be estimated. The procedure to calculate diesel sales fractions is to sum the population of diesel vehicles over the population of all vehicles for a given model year and light-duty vehicle class as shown below.

$$Diesel\ \ Sales = \frac{\sum Pop_{Diesel}}{\sum Pop_{Total}} \tag{3-5}$$

where:

Diesel Sales is the calculated diesel sales fraction for each model year, and
Pop is the model year population for diesel and total vehicles.

This calculation is repeated for each model year and for the two vehicle classes LDDV and LDDT.

EVALUATION OF UNREGISTERED VEHICLES

Data from remote sensing programs can be used to evaluate the characteristics of the unregistered fleet. In this discussion, we use "unregistered" to refer to vehicles with in-state license plates that have expired registrations[3]. The license plate data, cross checked with state motor vehicle records, can provide valuable information on the fraction of fleet unregistered and the length of period the vehicle is unregistered. Unregistered vehicles are not generally accounted for in mobile source modeling and it may be worthwhile to examine the remote sensing database to determine the pervasiveness of unregistered vehicles. One benefit of using remote sensing data to estimate fleet characteristics is that unregistered vehicles are

[3]Evaluating vehicles registered out-of-state or out-of-country is discussed later in this section.

included in the remote sensing data base. The steps required to estimate the unregistered vehicle fraction are as follows.

1. Remove duplicate vehicle records from the DMV-matched database.
2. Calculate the unregistered vehicle fraction.

Duplicate vehicle records are removed from the data base to obtain the unique vehicle population. To evaluate the fraction of unregistered vehicles in the remote sensing database, the sum of the population of unregistered vehicles are divided by the total population of vehicles as shown below.

$$Unregistered = \frac{\sum Pop_{Unregistered}}{\sum Pop_{Total}} \tag{3-6}$$

where:

Unregistered is the fraction of unregistered vehicles, and
Pop is the population of unique vehicles.

This calculation can be completed by vehicle class or for the fleet total. Note that the date by which active vehicle registration should be determined is the date that the DMV data were obtained. For example, if the DMV data were obtained on March 1, active registration should be determined relative to March 1.

It is also worthwhile to examine the length of period unregistered. Most unregistered vehicles may be late in registering and will renew registration within a few months, as described in Section 2. Thus it would be worthwhile to note the fraction of unregistered vehicles by the length of period unregistered to determine the fraction of vehicles that have been unregistered for a considerable period. If it is determined that a significant portion of the fleet are unregistered for a long period, it would be informative to examine the differences between the registered and unregistered fleet (e.g., model year distribution or average vehicle age).

EVALUATION OF VEHICLE COUNTY/STATE OF ORIGIN TO DETERMINE VEHICLE I/M STATUS

Data from a remote sensing program can be used to evaluate the origin of travel within the region covered by the program. For example, the fraction of the local travel which is out-of-state can be determined from review of license plate data[4]. Moreover, from the DMV data of in-state vehicles, the county of registration can be determined to evaluate whether a vehicle is subject to an I/M program or not. This information would be useful in determining the fraction of local travel by vehicles not subject to an I/M program. It is common in mobile source modeling practices to assume that all local travel is subject to the local I/M program, which can lead to an overestimate of the benefits of an I/M program.

Calculating the fraction of vehicles not subject to an I/M program can be done in three steps:

1. Separate out vehicles classes covered by the I/M program.
2. Separate vehicles by whether or not the vehicle is subject to an I/M program.
3. Calculate the fraction not subject to the I/M program.

The equation to calculate the fraction of vehicles not subject to an I/M program is the sum of vehicles not subject divided by the total number of vehicles.

$$Fraction_{no-I/M} = \frac{\sum Travel_{no-I/M}}{\sum Travel_{Total}} \qquad (3\text{-}7)$$

where:

Fraction $_{no\text{-}I/M}$ is the calculated fraction of vehicles not subject to an I/M program, and
Travel is the vehicle count (including duplicate vehicle records).

EXAMPLE APPLICATIONS

This subsection provides example applications of the methods described above using the California Pilot I/M Program remote sensing database. The California Pilot I/M Program data base is first described, and is followed by example calculations of:

- registration distributions,
- diurnal travel distributions,
- VMT mix,
- unregistered vehicles, and
- vehicle county of origin (to determine I/M status).

[4] This methodology can be applied to determine the fraction of travel by vehicles registered out-of-state or out-of-country in case special consideration of these vehicles is desired in the emission inventory development.

Because diesel vehicles were removed from the California Pilot I/M Program remote sensing database, an estimate of diesel sales fractions was not completed.

DESCRIPTION OF THE CALIFORNIA PILOT I/M PROGRAM DATA USED IN THE EXAMPLES

The California Bureau of Automotive Repair (BAR) sponsored the California Pilot I/M Program, which included a large-scale remote sensing program in the city of Sacramento in the summer of 1994. During the months of July, August and September, remote sensing was performed at 337
sites and approximately two million valid remote sensing readings were obtained. The evaluation of the remote sensing data was completed by Radian Corporation for the California BAR. According to Radian, the program produced 1,329,694 remote sensing records with matched DMV records, covering 47 percent of the eligible Sacramento vehicle population (Radian, 1995).

The remote sensing data are available to the public through the California BAR. The data base includes two subsets consisting of 1,969,044 valid remote sensing readings and 1,113,939 remote sensing readings with matched DMV data. The differences between the Radian DMV-matched databases (1.3 million records) and the publicly available DMV-matched data (1.1 million records) are not documented. The DMV-matched data include model year data, fuel type data (although diesel-powered vehicles were removed from the databases), vehicle weight data (not complete), and vehicle model description (e.g., sedan, pick-up). From the model description, vehicle weight data and information received from the BAR, vehicle classes were identified.

Of the 1,113,939 records in the DMV-matched database, more than 95 percent – 1,062,073 records representing 503,838 unique vehicles – were used in the examples summarized in this section. Records were removed from the database for two reasons. First, because model descriptions were not standardized (more than a thousand variations), not all vehicles could be assigned to a vehicle class and were dropped from this evaluation resulting in a loss of about 3 percent of the total records. In addition, the number of trucks over 6,000 pounds gross vehicle weight (classified as LDGT2 in MOBILE5a) did not appear to be completely represented in the database, and the records for these vehicles were dropped (of the trucks with valid vehicle weight data, less than one percent were classified as over 6,000 lbs gross vehicle weight). Records for buses and motor homes were also dropped as their representativeness in a remote sensing database are questionable. Total records dropped due to vehicle class considerations were 50,509. Second, an additional 1,357 records were eliminated due to missing or undecipherable data.

ESTIMATION OF REGISTRATION DISTRIBUTIONS

The following presents the example estimation of registration distributions from the remote sensing data of the California Pilot I/M Program; a comparison of estimated distributions with those of the emissions factor models (MOBILE5a and EMFAC7F) and an example of site-by-site variation of the California Pilot I/M Program data are then presented.

Duplicate records are removed from the database (as the registration distribution represents the population of unique vehicles by model year) and the registration distributions were calculated for the LDGV and LDGT1 vehicle classes. In the examples that follow, 399,309 and 104,529 unique LDGVs and LDGT1s were used in the calculations. The unadjusted registration distribution calculated from the California Pilot I/M database are shown in Tables 3-1a and 3-1b for LDGVs and LDGT1s, respectively.

The California Pilot I/M Program data were adjusted for age-dependent mileage accumulation using mileage accumulation rate (MAR) data from the California ARB EMFAC7F emission factor model. For LDGVs, these MAR data are based on analysis of California I/M Program odometer data; for LDGT1s, these MAR data are based on the MOBILE4 emission factor model (ARB, 1993). The mileage-adjusted data for LDGVs and LDGT1s are also shown in Tables 3-1a and 3-1b, respectively.

For the purposes of this evaluation, no further adjustments to the registration distributions determined from the California Pilot I/M Program were made other than to adjust the data to July 1. In the California Pilot I/M Program the mean date of the remote sensing test data was July 26, 1994. Thus to adjust the data to represent July 1, the current model year vehicles were reduced by 7.1 percent (25 divided by 365) and the distributions were renormalized. The resulting LDGV and LDGT1 registration distributions adjusted to July 1 are also shown in Tables 3-1a and 3-1b, respectively. These distributions are now in the format required by the MOBILE model.

Comparison of the California Pilot I/M Program Registration

The registration distribution of the California Pilot I/M Program along with the 1994 default distributions of MOBILE5a and EMFAC7f are shown in Figures 3-2a and 3-2b for LDGVs and LDGT1s, respectively. For MOBILE5a, the registration data are national average 1990 values determined from R. L. Polk data. For EMFAC7F, the LDGV distribution is 1994 data determined from California DMV data; the LDGT1 data are the averages of 1976 through 1983 calendar years of R. L. Polk data for the State of California (ARB, 1993). Clearly, there are significant differences between these and the distribution calculated from the remote sensing data. The average fleet age can be used to quantify the differences between these distributions and are presented in Table 3-2 for each of the distributions. Table 3-2 shows that the age of the fleet in Sacramento is, on average, more than one year older than the fleet represented in the emission factor models for 1994. This is significant considering that a year of fleet turnover can lower emission rates by more than 10 percent (the degree to which fleet turnover affects emissions depends on the fleet makeup and the calendar year of evaluation).

Site-by-site Variation of the California Pilot I/M Program Data

Significant site-by-site variation of registration distributions can be expected as the fleet characteristics of various neighborhoods are different. This illustrates the problem with using only one or two remote sensing sites to characterize the whole fleet. However, when several hundred sites are measured as was done in the California Pilot I/M Program, sites can be grouped into subregions for the purposes of calculating fleet characteristics at the subregion level. For illustrative purposes, we evaluated twelve random sites (chosen by every thirtieth location in order of appearance in the DMV-matched database). The variation of fleet age for each site was determined by the population-weighted average age for each vehicle class. The results of this evaluation are shown in Table 3-3. Average age can vary by more than two years showing considerable variation within Sacramento region.

ESTIMATION OF DIURNAL TRAVEL DISTRIBUTIONS

Diurnal travel distributions were determined from the remote sensing database prior to matching with DMV records. DMV matched data were not required because vehicle information such as vehicle class was not needed to estimate diurnal travel distributions. In order to maximize the number of hours included in this evaluation, the diurnal distributions from the California I/M Pilot Program remote sensing data were determined using sites which began operation before 7 am and finished after 6 pm. This resulted in 101 sites (271,886 records) used to determined the weekday distribution and 10 sites (41,359 records) used to determine the Saturday diurnal distribution; no measurements were taken on Sundays. Note that for the purposes of comparison and for filling in hours not covered by the remote sensing program, the default diurnal distribution of the EPA's Emission Preprocessor System Version 2.0 (EPS2) was used (EPA, 1992a).

Figures 3-3a and 3b show the diurnal distribution determined from the California Pilot I/M Program data for weekdays and Saturdays, respectively. Included in the weekday profile is the default distribution for the same period from EPS2. Figure 3-3a shows that the diurnal distribution from the California Pilot I/M Program has significantly less travel during the weekday morning and afternoon peak periods relative to the mid-day travel than the EPS2 distribution. Comparing Figures 3-3a and 3-3b reveals considerable differences between the distribution of travel on weekdays and on Saturdays.

From the EPS2 distribution, 73.9 percent of the travel is included in the hours evaluated from the California Pilot remote sensing program (hours ending 8 through 18). In order to get the 24-hour distribution, each hour of the fractional travel from the remote sensing data are multiplied by 73.9 percent, and the remaining hours are taken from the EPS2 distribution. The results of this evaluation are shown in Table 3-4 and in Figure 3-4. Again, the diurnal distribution from the California Pilot I/M Program has significantly less travel during the weekday morning and afternoon peak periods relative to the mid-day travel than the EPS2 distribution.

ESTIMATION OF VMT MIX

VMT mix from the remote sensing data of the California Pilot I/M program is only calculated for two vehicle classes (LDGVs and LDGT1s) because heavier vehicles were not accurately represented and all diesel-powered vehicles were removed from the data base. VMT mix for the entire database (1,062,073 total records) and for twelve randomly selected sites are shown in Table 3-5 for these two vehicle classes. The data shown in Table 3-5 illustrate that VMT mix can vary widely resulting in regional differences in estimated emissions.

The MOBILE5a 1994 VMT mix was used to fill in the VMT mix of the remaining vehicle classes. From the MOBILE5a data, the LDGVs and LDGT1s account for 81.3 percent of the fleet total VMT. Using 81.3 percent to weight the values for LDGVs and LDGT1s calculated from the California Pilot I/M Program, the overall VMT mix is shown in Table 3-6 for the entire

database along with the MOBILE5a default VMT mix. Table 3-6 shows that LDGVs are present in a larger proportion to LDGT1s in the California Pilot I/M Program database than is assumed in MOBILE5a.

ESTIMATION OF UNREGISTERED VEHICLES

The fraction of the fleet with lapsed registration was determined from the California Pilot I/M program database since the database contained the expiration date of each matched vehicle. Unfortunately, the California BAR was uncertain as to the exact date of the acquisition of the DMV registration data, but the BAR indicated that the data were obtained at the initiation of the project. For the purposes of this example, an acquisition date of July 1, 1994 was assumed. After the removal of duplicate vehicle records, 399,309 and 104,529 unique LDGVs and LDGT1s were used resulting in a lapsed registered vehicle population of 12.5 percent and 10.9 percent for LDGVs and LDGT1s, respectively. However, if one examined the population of vehicles more than one year unregistered, the populations fell to 1.5 percent and 1.4 percent for LDGVs and LDGT1s, respectively. This indicates that the vast majority of unregistered vehicles are eventually registered within a year.

ESTIMATION OF VEHICLE COUNTY OF ORIGIN (TO DETERMINE I/M STATUS)

The county of registration was examined to determine the type of I/M program the fleet of the California Pilot I/M Program was subject to. No data were provided on the fraction of travel by out-of-state vehicles in the database. In 1996, the enhanced I/M program will be initiated in California, and this evaluation determined the fraction of the fleet which would be subject to the enhanced I/M program based on county of registration. Of the 1,062,073 records of the DMV-

matched database, 78.2 percent of the travel was by vehicles registered in Sacramento County. This shows that the majority of travel was made by the local fleet. When including all of the counties to be incorporated into the California enhanced I/M program, 81.3 percent of the travel
was made by vehicles subject to the enhanced I/M program. This illustrates that about one-fifth of the travel (for these two vehicle classes) will be made by vehicles subject to the existing California I/M program.

Table 3-1a
Unadjusted and adjusted LDGV registration distribution data from the California Pilot I/M Program

Model Year by Age (1994 = 0)	Unadjusted Model Year Distribution	Distribution Adjusted for Age-dependent Mileage	Distribution Adjusted to July 1st
0	0.059	0.040	0.037
1	0.075	0.052	0.053
2	0.069	0.051	0.051
3	0.083	0.064	0.064
4	0.081	0.065	0.066
5	0.086	0.073	0.073
6	0.080	0.072	0.072
7	0.080	0.077	0.077
8	0.071	0.072	0.072
9	0.063	0.069	0.069
10	0.053	0.062	0.062
11	0.035	0.044	0.044
12	0.027	0.036	0.036
13	0.024	0.034	0.034
14	0.021	0.031	0.031
15	0.022	0.033	0.033
16	0.018	0.029	0.029
17	0.012	0.019	0.020
18	0.007	0.012	0.012
19	0.004	0.006	0.006
20	0.004	0.007	0.007
21	0.004	0.008	0.008
22	0.004	0.007	0.007
23	0.003	0.005	0.005
24	0.017	0.032	0.032

Table 3-1b
Unadjusted and adjusted LDGT1 registration distribution data from the California Pilot I/M Program

Model Year by Age (1994 = 0)	Unadjusted Model Year Distribution	Distribution Adjusted for Age-dependent Mileage	Distribution Adjusted to July 1st
0	0.068	0.039	0.037
1	0.092	0.057	0.057
2	0.081	0.054	0.054
3	0.093	0.067	0.067
4	0.072	0.055	0.055
5	0.083	0.068	0.068
6	0.070	0.062	0.062
7	0.073	0.069	0.069
8	0.078	0.079	0.079
9	0.052	0.056	0.057
10	0.042	0.049	0.049
11	0.023	0.028	0.029
12	0.019	0.026	0.026
13	0.015	0.022	0.022
14	0.015	0.023	0.023
15	0.021	0.035	0.035
16	0.018	0.032	0.032
17	0.016	0.030	0.030
18	0.010	0.021	0.021
19	0.007	0.016	0.016
20	0.007	0.016	0.016
21	0.008	0.018	0.018
22	0.007	0.015	0.016
23	0.005	0.011	0.011
24	0.024	0.051	0.052

Table 3-2
Population-weighted average fleet ages based on registration distribution data. The age of current model years is assumed to be 0; the age of vehicles 24 years old and older is assumed to be 24

Vehicle Class	Population-weighted Average Vehicle Age (years) by Source of Registration Distribution Data		
	California Pilot I/M Program	EMFAC7f	MOBILE5a
LDGV	8.5	7.3	7.2
LDGT1	9.4	7.6	7.8

Table 3-3
Population-weighted fleet average ages for twelve randomly selected sites of the California Pilot I/M Program database. The age of current model years is assumed to be 0; the age of vehicles 24 years old and older is assumed to be 24

Site ID	Site Type	LDGV Average Age (years)	LDGT1 Average Age (years)
514	Limited Access	8.1	9.2
12	Limited Access	7.8	8.2
66	Limited Access	8.0	8.7
223	Limited Access	8.2	8.2
151	Surface Street	8.5	9.3
586	Surface Street	8.0	8.6
606	n/a	7.9	8.6
B521	Surface Street	10.1	10.6
B129	Surface Street	7.5	8.3
B57	n/a	8.1	8.8
557	Limited Access	7.6	7.8
190	n/a	8.5	9.9

n/a = site description not included in BAR-released data.

Table 3-4
24-hour weekday diurnal distributions from EPS2 (EPA, 1992) and from the remote sensing data of the California Pilot I/M Program

Hour Ending	EPS2 Diurnal Distribution (EPA, 1992b)	CA Pilot I/M Program Diurnal Distribution
0	0.010	0.010
1	0.010	0.010
2	0.010	0.010
4	0.010	0.010
5	0.010	0.010
6	0.010	0.010
7	0.063	0.063
8	0.104	0.062
9	0.063	0.058
10	0.052	0.052
11	0.052	0.055
12	0.052	0.061
13	0.052	0.066
14	0.052	0.066
15	0.052	0.068
16	0.063	0.075
17	0.104	0.083
18	0.083	0.086
19	0.063	0.063
20	0.042	0.042
21	0.010	0.010
22	0.010	0.010
23	0.010	0.010
24	0.010	0.010

Table 3-5
VMT Mix data for the sum of all sites and for twelve randomly selected sites of the California Pilot I/M Program Database

Site ID	Site Type	LDGV VMT Mix	LDGT1 VMT Mix
All Sites		0.804	0.196
514	Limited Access	0.747	0.253
12	Limited Access	0.855	0.145
66	Limited Access	0.677	0.323
223	Limited Access	0.852	0.148
151	Surface Street	0.794	0.206
586	Surface Street	0.823	0.177
606	n/a	0.810	0.190
B521	Surface Street	0.834	0.166
B129	Surface Street	0.721	0.279
B57	n/a	0.780	0.220
557	Limited Access	0.725	0.275
190	n/a	0.824	0.176

n/a = site description not included in BAR-released data.

Table 3-6
MOBILE5a 1994 default VMT mix data and VMT mix data calculated from the California Pilot I/M Program database. Only LDGVs and LDGT1s are included

Vehicle Class	MOBILE5a 1994 Default	California Pilot I/M Program
LDGV	0.636	0.654
LDGT1	0.177	0.159
LDGT2	0.083	0.083
HDGV	0.031	0.031
LDDV	0.004	0.004
LDDT	0.002	0.002
HDDV	0.059	0.059
MC	0.007	0.007

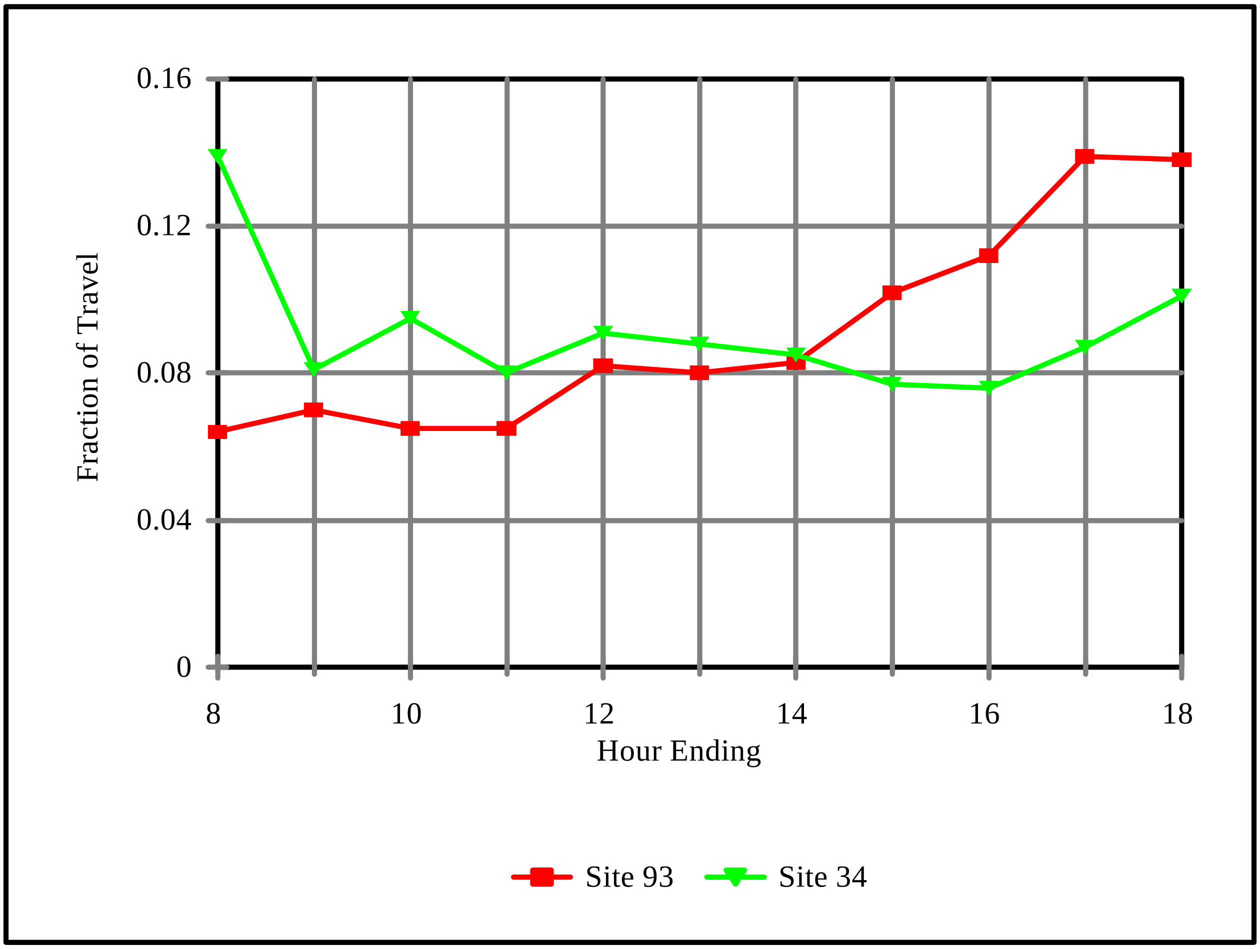

Figure 3-1
Two sites on opposing directions of Highway 50 in Sacramento showing weekday morning and afternoon commute patterns

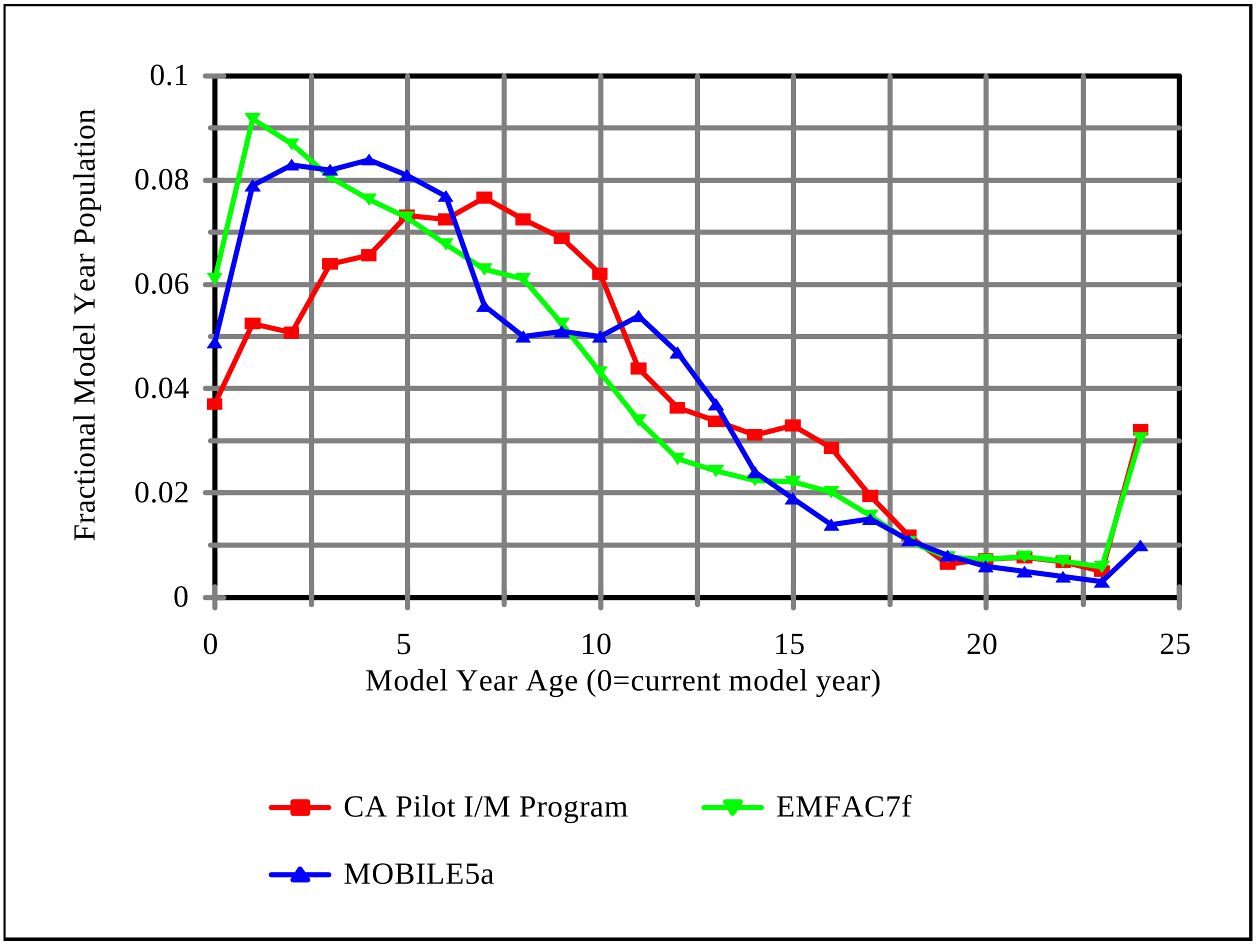

Figure 3-2a
LDGT1 registration distribution data determined from the California Pilot I/M Program remote sensing data and the emission factor models, MOBILE5a and EMFAC7F

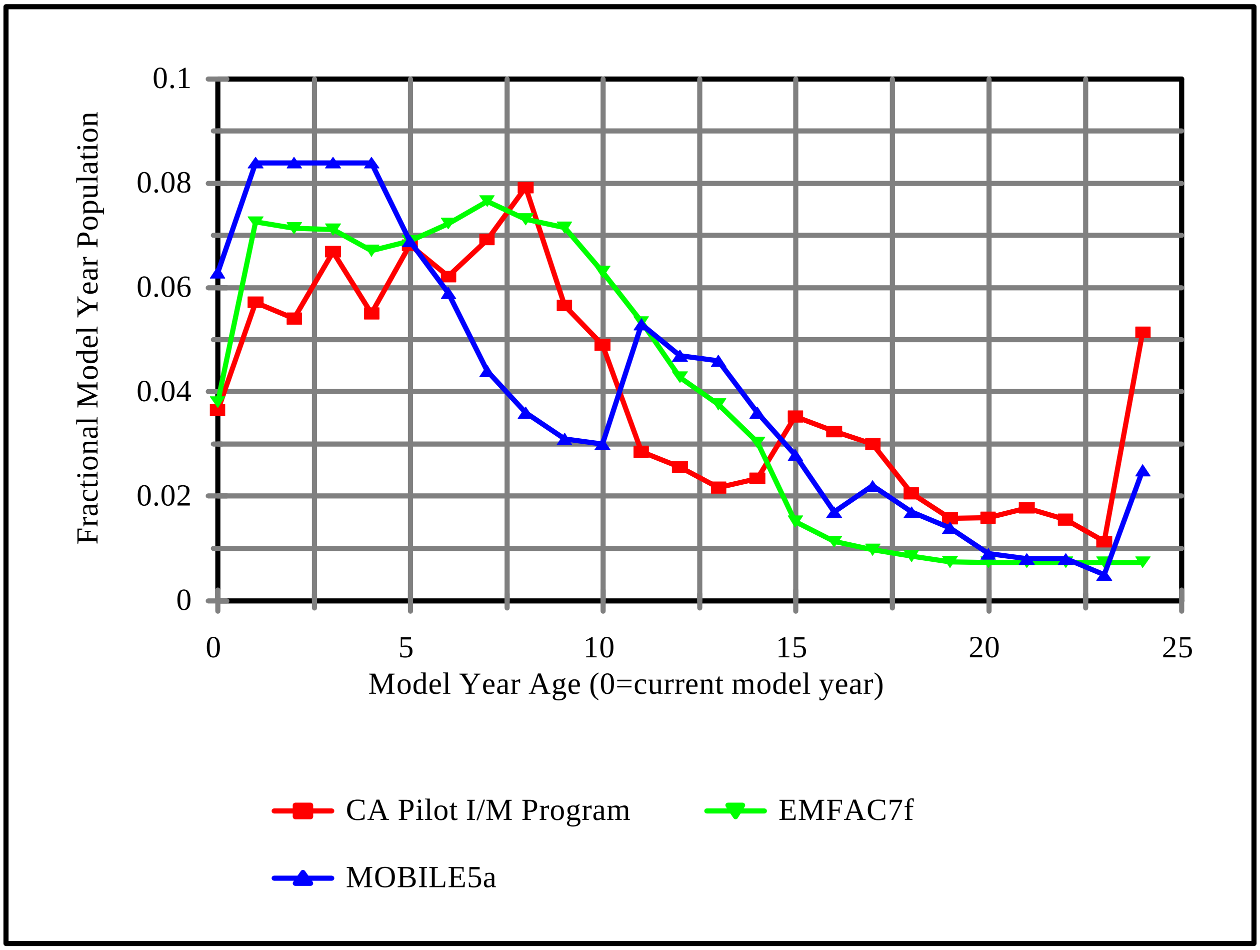

Figure 3-2b
LDGV registration distribution data determined from the California Pilot I/M Program remote sensing data and the emission factor models, MOBILE5a and EMFAC7F

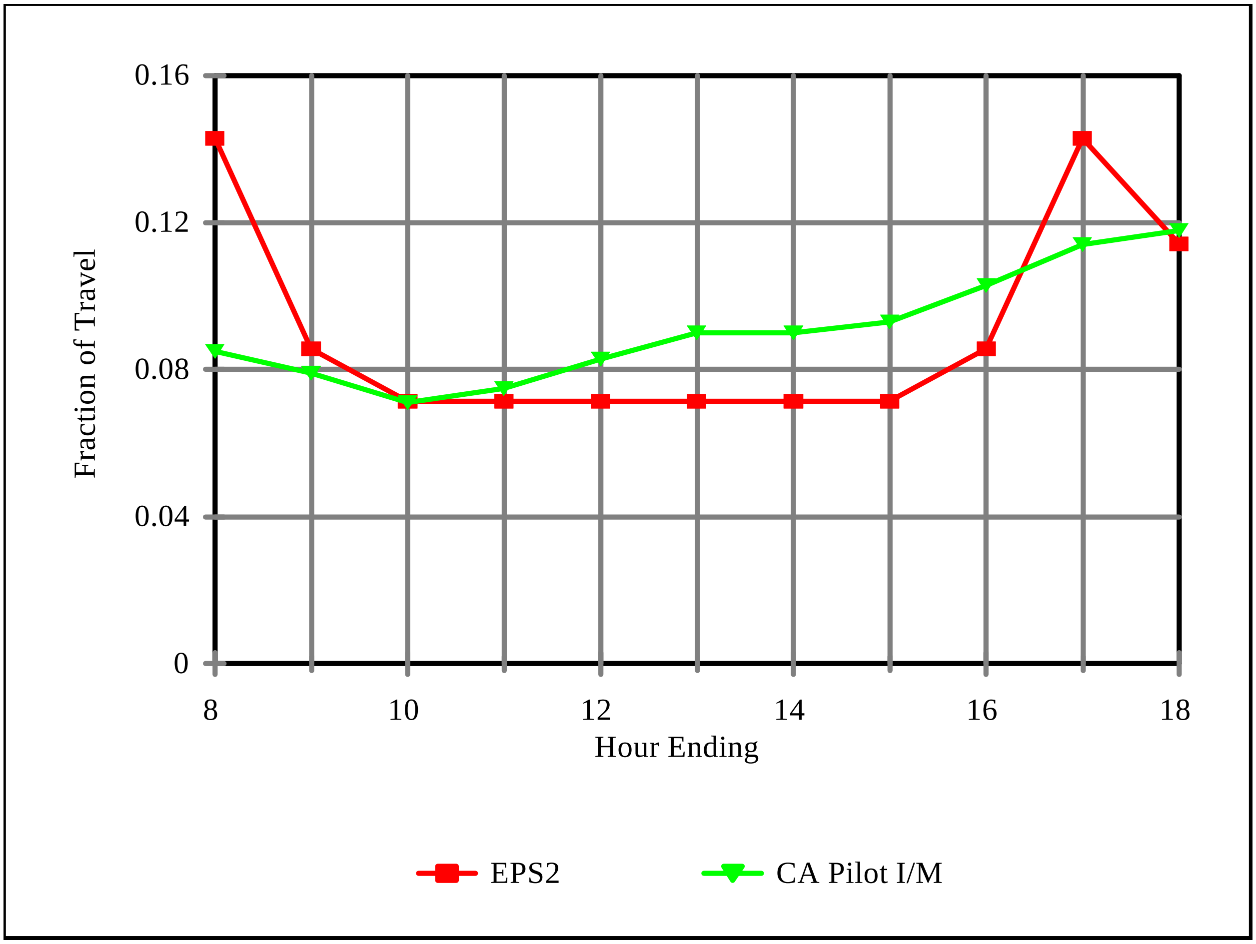

Figure 3-3a
Weekday diurnal travel distributions from EPS2 (EPA, 1992) and from the remote sensing data of the California Pilot I/M Program

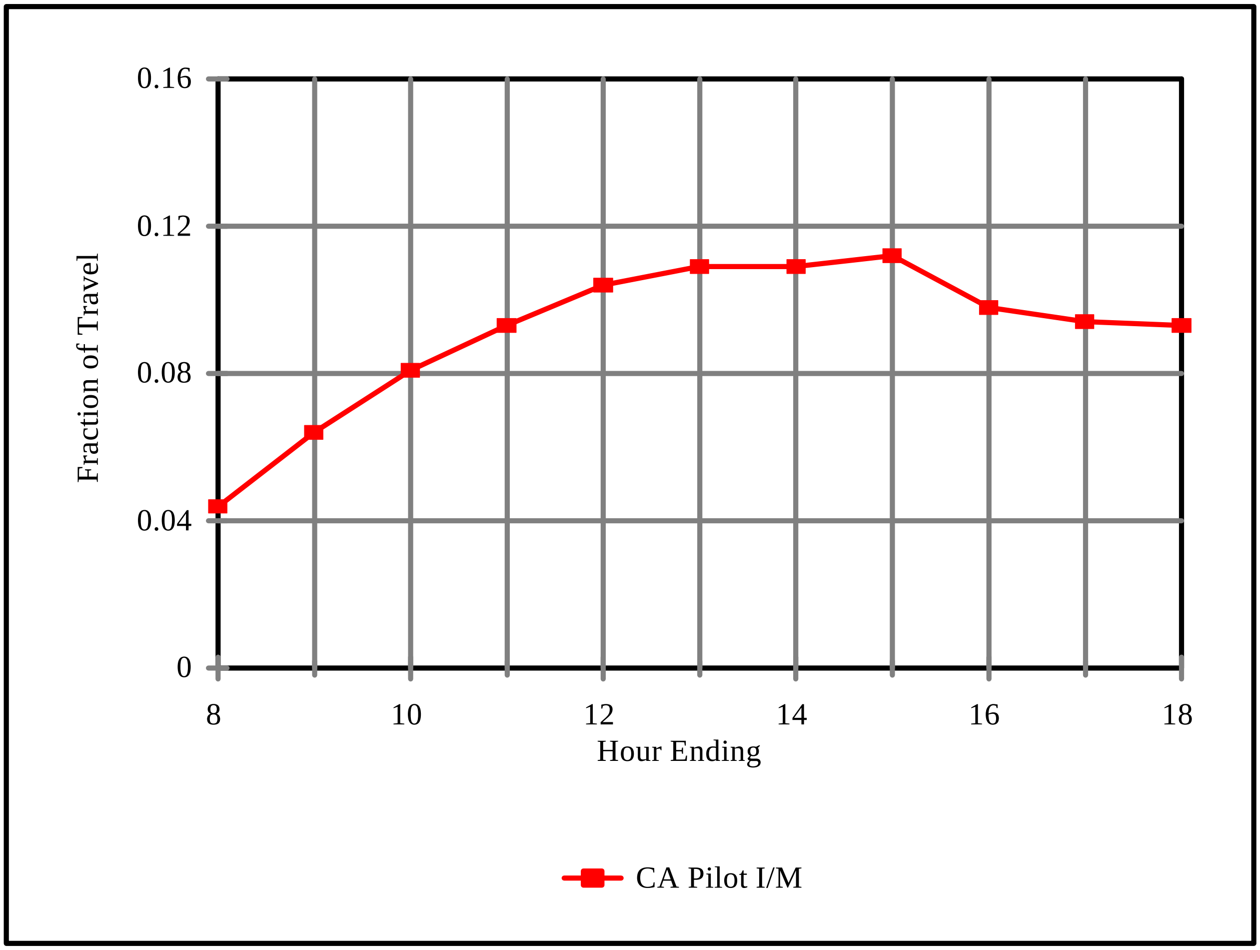

Figure 3-3b
Saturday diurnal travel distribution calculated from the remote sensing data of the California Pilot I/M Program

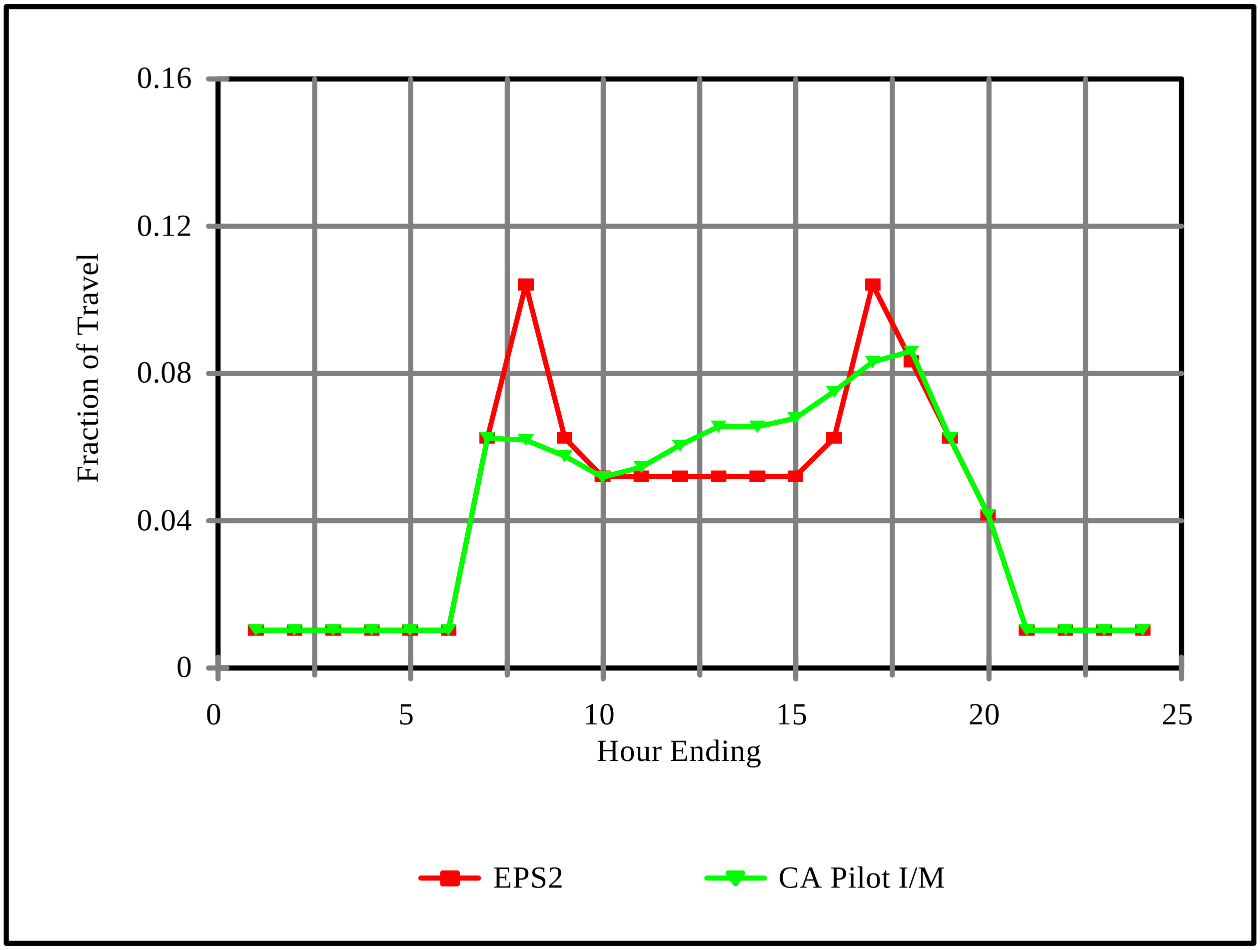

Figure 3-4
24-hour weekday diurnal travel distributions from EPS2 (EPA, 1992) and from the remote sensing data of the California Pilot I/M Program

4
ESTIMATION OF MOBILE SOURCE FUEL CONSUMPTION AND AREA VMT

INTRODUCTION

This section presents the methodology for the estimation of mobile source fuel consumption and vehicle miles traveled (VMT) from tax revenue and other data sources. VMT estimates are usually based upon data compiled for transportation network planning purposes. This section presents an alternate methodology to independently estimate VMT and compare this estimate to others calculated by traditional methods. The methodology calculates VMT from estimates of area fuel consumption, fleet fuel economy and refueling loss rates. It is suggested that if the two VMT estimates differ substantially, further investigation may be warranted. This section begins with a literature review of relevant studies, then provides details of the methodology to estimate VMT. Example calculations are then provided for Sacramento County, CA and for Maricopa County, AZ.

LITERATURE REVIEW

This literature review summarizes previous work or related studies on how to estimate mobile source fuel consumption in a nonattainment area. Each of these documents discusses issues that should be considered in calculations of VMT from fuel consumption estimates. The following documents are reviewed in this section.

- *A Determination of Motor Vehicle Activity Factors for Atlanta, Georgia through Fuel Consumption Analysis* (Hayes, 1993).
- *The Mobile4 Fuel Consumption Model* (EPA, 1991).
- *Procedures for Emission Inventory Preparation - Volume IV: Mobile Sources* (EPA, 1992).
- *User's Guide to MOBILE5 (Mobile Source Emission Factor Model)* (EPA, 1994).
- *Emission Inventory Procedural Manual - Volume III: Methods for Assessing Area Source Emissions* (ARB, 1995).

In addition to the document review, several contacts with state agencies have provided useful information regarding estimates of fuel sales or consumption. Two that were of particular importance are also summarized below.

A Determination of Motor Vehicle Activity Factors for Atlanta, Georgia through Fuel Consumption Analysis (Hayes, 1993).

Adam Hayes completed this master's thesis under the direction of Professor Michael O. Rodgers at the Georgia Institute of Technology. Hayes' objective was to develop alternate means to calculate VMT in Atlanta, Georgia. He discussed traditional VMT estimation methods and two alternative methods based on estimates of fuel consumption and fuel economy. Hayes also discussed the limitations of the various approaches.

Traditional VMT estimation methods utilize traffic count data, random sampling, historical data, and statistical extrapolation. Transportation modelers developed these methods in order to evaluate the performance of the transportation network rather than to provide an accurate estimate of vehicle emissions. Hayes identified a number of biases (listed below) that may associated with these traditional techniques.

- The Department of Transportation and travel demand modelers use road classifications in order to extrapolate traffic flows. Road sections may need to be re-classified more frequently than current practices allow (e.g. urban vs. rural), especially during periods of rapid growth.

- Traffic count data are incomplete because (1) many roadways are sampled only once in several years, and (2) sample locations are often sited to target traffic flow problems of local concern.

- Traffic counts usually last only 24-48 hours. At this time scale, many short-term factors may confound results (e.g., weather, traffic accidents, traffic congestion).

- Seasonal variations in traffic flow at the sample site must be accounted for properly.

- Many data collection efforts are costly, therefore, outdated information is frequently used. For example, origin-destination roadside questionnaires provide extremely valuable data, but are expensive to administer. The Census Bureau's Journey to Work data are often used, however, these are only updated every 10 years.

- Analysts adjust transportation demand models until the model results match traffic count data. Some analysts use arbitrary adjustment techniques that do not provide an accurate representation of VMT, particularly if VMT counts are not the analysts' primary focus.

- Transportation demand models require large quantities of descriptive data regarding the transportation network. Human coding errors sometimes occur as a result.

In order to circumvent these biases, Mr. Hayes proposed two alternative methods to estimate fuel consumption and calculate VMT. The first method utilized fuel shipment records, and the second method applied fuel tax revenue data.

Data representing the total fuel shipped to Atlanta, Georgia in 1986 were gathered from two pipeline companies. These data were treated as an upper limit estimate of fuel consumed on-road in Atlanta. Error may be introduced in this estimate if significant amounts of fuel were (a) transported overland for use in other communities following dispersal from the pipelines, (b) transported from other communities for use in Atlanta, (c) used for off-road purposes, or (d) lost due to volatilization. Fuel shipments to Atlanta are limited to two pipeline carriers, thus data acquisition is relatively simple. Mr. Hayes points out that this method is difficult to apply to other U.S. cities because numerous overland or waterway shipment routes are difficult to track and quantify. Furthermore, recent data are very difficult to obtain because shippers lack incentive to share information.

A second estimate of fuel consumption was based upon fuel tax revenue information. Tax revenue figures are available only at the state level. Mr. Hayes disaggregated the state tax revenue figure according to the proportion of VMT that occurred in Atlanta during 1986 (estimated by the Georgia Department of Transportation).

Mr. Hayes then calculated VMT from these estimates of fuel consumption. He demonstrated sensitivity to fleet fuel economy, off-road fuel use, and out-of-area fuel use. Assuming 1986 fleet fuel economy between 14 and 16.7 mpg and more than 75 percent on-road fuel use in the area, Mr. Hayes' estimates of VMT differed from those made by the Georgia DOT and HPMS by less than 30 percent.

***Procedures for Emission Inventory Preparation - Volume IV: Mobile Sources* (EPA, 1992), and the *User's Guide to MOBILE5 (Mobile Source Emission Factor Model)* (EPA, 1994).**

These two documents are complementary and are discussed here jointly. They describe current EPA procedures for completing the mobile source emissions inventory and provide a context for the calculation of VMT. A detailed method to estimate VMT from fleet characteristics is fully described. Additionally, methods to quantify resting losses and emissions from non-road sources are described. These documents provide helpful guidance for segregating on-highway fuel consumption from total fuel consumption.

The Mobile4 Fuel Consumption Model **(EPA, 1991).**

The purpose of the MOBILE4 Fuel Consumption Model (M4FC) is to calculate fuel consumption using fleet VMT and fleet fuel economy. Note that this is the reverse of the methodology presented in this document (i.e., calculating VMT from fuel economy and fuel consumption). The M4FC guide demonstrates that these calculations are sensitive to fuel economy, fleet distribution, and mileage accumulation rates.

Emission Inventory Procedural Manual - Volume III: Methods for Assessing Area Source Emissions **(ARB, 1995).**

This document presents the California ARB's methods for calculating vaporization losses associated with refueling stations. Emission factors as a function of fuel sales are given for underground storage tank working and breathing losses, and for refueling losses due to vapor displacement and spillage. These types of losses should be considered during calculations of on-highway fuel consumption.

CALIFORNIA DEPARTMENT OF TRANSPORTATION, ECONOMICS ANALYSIS UNIT

The Economics Analysis Unit of the California Department of Transportation (CalTrans) has performed a regression analysis to verify correlations of several factors with historical state fuel consumption. These factors include the total number of driver's licenses, total number of registered vehicles, total sales, service station sales, and sales of eating and drinking establishments. The relationship of retail sales to fuel consumption is not intuitive. However, the rationale was that sales are inflated when a heavy tourist population exists, and so too, fuel consumption must increase.

In order to calculate fuel consumption in California localities, CalTrans extrapolates the above regression to the county level. The statewide fuel consumption figure, which is known, is disaggregated according to each county's proportion of the above factors.

CALIFORNIA BOARD OF EQUALIZATION, FUEL TAXES DIVISION

The Fuel Taxes Division of the California Board of Equalization also estimates fuel consumption at the county level. The statewide fuel consumption figure is disaggregated according to the proportion of service station taxable sales in each county. Agency employees estimate that 80 to 95 percent of service station sales are attributable to fuel.

METHODOLOGY FOR ESTIMATING VMT

Emissions associated with on-road motor vehicle use are an important element of any emission inventory. The USEPA provides a mobile source emission factor model (MOBILE) that estimates vehicle emission factors. These emission factors are combined with locale-specific estimates of vehicle miles traveled (VMT) to estimate mobile source emissions. VMT estimates are available from state departments of transportation (DOT) or county planning agencies, such as metropolitan transportation planning organizations (MPO). While readily available at the resolution needed for air quality planning purposes (e.g., county or even sub-county levels), these VMT estimates are usually compiled for transportation network planning purposes, not for air quality planning. The adequacy of these estimates for air quality planning purposes can vary depending upon the area and the VMT estimation methods used.

This subsection presents a method to independently derive VMT for comparison with transportation planners' estimates of VMT. Specifically, this section presents a means to estimate VMT from fuel consumption and fleet fuel economy. Locale-specific or national average data may be used, depending upon availability. If the results of this fuels-based method differ substantially from transportation planners' estimates, further investigation may be warranted.

In order to allow a valid comparison of results, the fuels-based method must be based upon data that are independent of transportation planners' estimates of VMT. Tax revenue data and demographic statistics are used to disaggregate statewide figures of on-road fuel consumption provided by the Federal Highway Administration (FHWA). On-road fuel consumption is also corrected for refueling losses. Locally determined fleet fuel economy is suggested for use, although national average fuel economy may be used if no other data are available.

OVERVIEW OF CALCULATION PROCEDURES

The basis of this methodology is a five-step procedure to estimate VMT. Before proceeding, the following data must be obtained:

- Statewide gasoline and special fuels consumption data which are published annually in *Highway Statistics* by the Federal Highway Administration.

- Information regarding the extent of Stage I and Stage II refueling control measures in the area of interest; these are available from state agencies or the EPA.

- Registration distributions and mileage accumulation rates should be used to determine local fuel economy if possible; otherwise, national averages contained in the MOBILE model can be used.

- Economic and population statistics which are maintained by state and federal agencies for the U.S., state, and counties of interest (see List of Contacts, Appendix A).

The five steps in the methodology, described in detail below, are as follows:

Step 1: Adjust statewide gasoline and special fuels distributions to account for off-road use.

Step 2: Geographically disaggregate fuel distribution from Step 1 and assign fuel volumes to the counties of interest according to each county's share of economic and population indicators.

Step 3: Adjust the fuel distributions obtained in Step 2 for refueling losses. The result represents the counties' on-road fuel consumption.

Step 4: Calculate diesel- and gasoline-powered fleet fuel economies. This is achieved by weighting on-road fuel economies according to vehicle stock (numbers of vehicles) and mileage accumulation rates. If possible, local data should be used. However, a method to use national averages extracted from MOBILE 5 is described for cases where no other data are available.

Step 5: Multiply on-road fuel consumption by fleet fuel economies to calculate VMT.

DESCRIPTION OF CALCULATION PROCEDURES

Step 1: Estimate On-Road Fuel Consumption

On-road gasoline consumed in a state (G_{state}) includes all gasoline and gasohol used on state roads, taxed and untaxed. G_{state} can be calculated for the 50 states and the District of Columbia from data that are published annually by the Federal Highway Administration (FHWA) in *Highway Statistics*. In order to calculate G_{state}, total gasoline consumption must be corrected for refueling facility losses, and for the following non-road uses: agriculture, aviation, industry, construction, and marine. Table 4-1 shows 1994 data corrected for non-road use for selected states and the nation as a whole.

The FHWA also tabulates on-road private and commercial use of special fuels (D_{state}) (Table 4-2). Special fuels consist principally of diesel fuel with minor amounts of liquefied petroleum gases and other fuels. In this analysis, special fuels are treated as a surrogate for diesel.

Table 4-1
Gasoline distributed in selected states for 1994 (FHWA, 1994)

State	Total Gasoline (10^6 gallons)	Non-Highway Gasoline[1] (10^6 gallons)	Highway Gasoline[1] (10^6 gallons)
Arizona	1935	35	1900
California	13162	229	12933
Connecticut	1403	67	1336
New York	5543	111	5433
United States	118531	2955	115576

Not yet corrected for volatilization and handling losses. In *Highway Statistics*, FHWA subtracts losses reported by states due to handling and volatilization. The methods used by states are not uniform, therefore, a different method to calculate these losses is used later in this analysis. Values in the table do not perfectly match those in the FHWA publication because FHWA-reported losses were re-distributed on a percent basis to highway and non-highway data.

Table 4-2
On-road special fuels consumption in selected states for 1994 (FHWA, 1994)

State	Special Fuels On-Road Use (10^6 gallons)
Arizona	463.5
California	2,035.6
Connecticut	186.5
New York	917.0
UnitedStates	25,123.6

Special fuels consist principally of diesel fuel with minor amounts of liquefied petroleum gases and other fuels. In this analysis, special fuels are treated as a surrogate for diesel.

Step 2: Disaggregate to Counties

Economic and population statistics can be used to proportionally disaggregate statewide on-road fuel consumption (F_{state}) to counties. (Note that F_{state} is used here interchangeably with D_{state} or G_{state}.) These statistics include:

- service station taxable sales (ss),
- number of registered vehicles (rv),
- number of valid driver's licenses (dl),
- total taxable sales (ts), and
- population (pop).

Records of these statistics are maintained by state agencies (see Appendix A for an example list of agencies including those contacted in this study).

County proportions (p) are calculated for each of the five statistics. For example, the proportion of statewide service station sales for a county, p_{ss}, is calculated as

$$p_{ss} = ss_{county} / ss_{state} \tag{4-1}$$

The average proportion is calculated as

$$p_{avg} = (p_{ss} + p_{rv} + p_{dl} + p_{ts} + p_{pop}) / 5 \tag{4-2}$$

It is expected that the five proportions should be very similar. If the five proportions are approximately equal, then the average proportion may be used to calculate F_{county}. If any one of the proportions above differs substantially from p_{avg}, say by more than 10 percent, a simple regression on historical state/country proportions can be used to determine weighting factors for the five proportions. In such a case, state and U.S. data for at least 20 time periods should be used (e.g. 20 years, 20 months, etc.). This approach assumes that state/U.S. correlations reflect county/state correlations reasonably well. Details of estimating county fuel consumption (diesel or gasoline) using weighted averages of the five proportions are provided in Appendix D.

$$F_{county} = F_{state} \times p_{avg} \tag{4-3}$$

Step 3: Adjust for Refueling Emissions

The California ARB and the U.S. EPA have developed spillage and volatilization emission factors for gasoline dispensing facilities (see Table 4-3). These factors depend on the use of evaporative control systems, and are expressed in lb/1000 gallons throughput. Loss rates expressed in gallons lost/gallons throughput are needed to adjust fuel consumption figures for refueling losses. The liquid density of gasoline is approximately 7.5 lb/gallon. Division of the emission factors listed in Table 4-3 by this density yields the required loss rates, as shown in Table 4-4. Note that although it is theoretically correct to subtract the amount of fuel loss due to refueling, the quantity of fuel lost is generally a very small fraction of the total fuel consumption, and this step can be skipped without affecting the resulting calculations. This point is demonstrated in the example calculations that follow.

Diesel fuel is relatively non-volatile, therefore, its volatilization losses are neglected here. The refueling spillage rate for diesel is assumed to be equal to that for gasoline.

Table 4-3
Emission factors for gasoline dispensing facilities (lb/1000 gallons throughput).
Source: ARB (1995)

Source of Loss	Without Stage I/Stage II Control	With Stage I/Stage II Control
Underground Storage		
Tanks Working Losses	9.5	0.475
Breathing Losses	1.0*	0.1*
Refueling:		
Vapor Displacement	10.0	0.5
Spillage	0.7	0.7

* 1.0 if no control or only Stage I control; 0.1 if both Stage I and Stage II controls.

Table 4-4.
Loss rates for gasoline dispensing Facilities
(gallons lost/1000 gallons throughput).

Source of Loss	Without Stage I/Stage II Control	With Stage I/Stage II Control
Underground Storage Tanks:		
Working Losses	1.27	0.06
Breathing Losses	0.13[*]	0.01[*]
Refueling:		
Vapor Displacement	1.33	0.07
Spillage	0.09	0.09

[*] 0.13 if no control or only Stage I control; 0.01 if both Stage I and Stage II controls (derived from Table 3-3).

Step 4: Calculate Fleet Fuel Economy

Fleet fuel economy (mpg_f) is averaged across vehicle classes and ages, weighted by mileage accumulation rate (ma) and vehicle stock (vs). There are eight vehicle classes used in MOBILE5a, defined by EPA (1992) as:

Class	Description
LDGV	Light duty gasoline-powered vehicles
LDGT1	Light duty gasoline-powered trucks (up to 6000 lb. GVW)
LDGT2	Light duty gasoline-powered trucks (6001 - 8500 lb. GVW)
HDGV	Heavy duty gasoline-powered vehicles
LDDV	Light duty diesel-powered vehicles
LDDT	Light duty diesel-powered trucks
HDDV	Heavy duty diesel-powered trucks
MC	Motorcycles

In the MOBILE5 model and the ARB mobile source emissions factor model (EMFAC7F), estimates of on-road fuel economies are provided for these eight categories (or their equivalents) according to vehicle model year (EPA, 1994; ARB, 1993). These values are tabulated in Appendix B.

If available, local mileage accumulation data and registration distributions can be used to calculate fleet fuel economy (mpg_f) according to the following equation.

$$Fleet\ Fuel\ Economy = mpg_f = \frac{\sum mpg_{ij} \times ma_{ij} \times vs_{ij}}{\sum ma_{ij} \times vs_{ij}} \tag{4-4}$$

Where

- Vehicle stock (vs) is the total number of vehicles for class *i* and age *j*,
- mileage accumulation rate (ma) is the average number of miles per year traveled by a vehicle for class *i* and age *j*

Separate fuel economies should be calculated for the diesel- and gasoline-powered fleets (mpg_{fd} and mpg_{fg}).

If local mileage and registration data are unavailable, national average fleet fuel economy may be used. Equation 4-4 is equivalent to:

$$mpg_f = \sum \frac{vmt_{ij}}{VMT} \times mpg_{ij} \tag{4-5}$$

Where

vmt_{ij} = annual average vehicle miles traveled for cars of class *i* and age *j*.
VMT = annual total vehicle miles traveled for all classes
mpg_{ij} represents the fuel economy of a vehicle of class *i* and age *j*

The ratio, vmt/VMT, is defined as the VMT mix in MOBILE. MOBILE5a includes national average VMT mixes as default values based on historical mileage accumulation rates and national registration distribution data (EPA, 1994). The 1994 predicted VMT mix by class (vmt_i/VMT) using MOBILE5a default settings is shown in Table 4-5.

Table 4-5.
1994 VMT Mix by class predicted by MOBILE5a.

Fleet	Vehicle Class	vmt_i/VMT	vmt_i/VMT Normalized to gas and diesel fleets
Gasoline	LDGV	0.636	0.681
	LDGT1	0.177	0.190
	LDGT2	0.083	0.090
	HDGV	0.031	0.033
	MC	0.007	0.008 SUM=1.0
Diesel	LDDV	0.004	0.062
	LDDT	0.002	0.031
	HDDV	0.059	0.908 SUM=1.0

However, for input to Equation 4-5, the travel mix must be further broken down by vehicle age. The travel fraction (tf) is defined as the average proportion of VMT within a vehicle class traveled by vehicles of a given age. MOBILE5a contains travel fractions estimated from national average historical data. The necessary travel fractions may be extracted from the MOBILE5 source code as described in Appendix E. Using these travel fractions, the VMT mix shown in Table 4-5 may be further broken down by vehicle age as follows:

$$vmt_{ij}/VMT = tf_{ij} \times vmt_i/VMT \tag{4-6}$$

The national average VMT mix by age and class is calculated for 1994 in Appendix C. Note that in Section 3 of this report, example California data showed a VMT mix that differs from national defaults. It is recommended that local VMT mix data be used whenever possible.

Step 5: Calculate Area VMT

After the previous calculations are carried out, county-wide annual VMT may be estimated.

$$\textit{Diesel fleet VMT} = D_{county} \times mpg_{fd} \tag{4-7}$$

$$\textit{Gasoline fleet VMT} = G_{county} \times mpg_{fg} \tag{4-8}$$

D_{county} and G_{county} represent the estimated on-road diesel and gasoline consumption in a county (from Equation 4-3). The variables mpg_{fd} and mpg_{fg} represent the diesel- and gasoline- powered fleet fuel economies (from Equations 4-4 or 4-5).

If monthly VMT is desired, yearly VMT may be scaled and distributed throughout the year according to the monthly distributions of fuel usage reported in the *Highway Statistics*, and/or periodic sales patterns reported by state agencies. The Federal Highway Administration utilizes monthly adjustment factors to seasonally allocate VMT (US Department of Transportation, 1981).

EXAMPLE APPLICATIONS

In this subsection the methodology described above is applied to Sacramento County, California and Maricopa County, Arizona using 1994 data to estimate 1994 VMT.

SACRAMENTO COUNTY, CALIFORNIA

Table 4-6 lists data that were obtained from the FHWA report, *Highway Statistics-1994*, and from various California state agencies (listed in Appendix A).

Table 4-6
Sacramento County and California 1994 relevant statistics

	Sacrament	California	units
Total Gasoline Consumption		13.16	billion
Total Special Fuels (i.e. Diesel)		2.04	billion
Active Drivers Licenses	727	20156	thousand
Registered Vehicles	779	22339	thousand
Population	1.130	31.95	million
Total Taxable Sales	10.97	285.98	billion $
Service Station Taxable Sales	0.5470	16.61	billion $

Step 1. Adjust fuel consumption for off-road use.

FHWA reports that 0.229 billion gallons of gasoline were used off-road in California in 1994, and that the state did not account for handling losses. For states that track handling losses, these losses should be redistributed to on- and off-road FHWA consumption figures on a percent basis before continuing with this procedure. These losses are inconsistently calculated for the FHWA report, and an alternate means to account for these losses is included in Step 3. Special fuels consumption was reported as on-highway use, therefore, no adjustment is necessary. Adjusted gasoline consumption is simply calculated as total gasoline consumption (13.16 billion gallons) minus off-road gasoline consumption (0.229 billion gallons), or 12.93 billion gallons.

Step 2. Geographically disaggregate fuel consumption from Step 1.

Calculate county/state proportions (p) for numbers of drivers licenses (dl), numbers of registered vehicles (rv), population (pop), total taxable sales (ts), and service station taxable sales (ss). Also calculate the average proportion (p_{avg}) and percent differences from the average (see Table 4-7).

For instance, the county/state proportion for the number of drivers' licenses is calculated as follows: $p_{dl} = dl_{county}/dl_{state} = 727 / 20156 = 0.0361$.

Table 4-7
Sacramento County/state proportions for various factors

Proportion	Value	% Difference from p_{avg}
p_{dl}	0.0361	+2%
p_{rv}	0.0341	-4%
p_{pop}	0.0354	+0%
p_{ts}	0.0384	+8%
p_{ss}	0.0329	-7%
p_{avg}	0.0354	na

Since the proportions are roughly equal (none differs from the average by more than 10 percent), p_{avg} may be used to geographically disaggregate fuel consumption.

Disaggregate fuel consumption, and assign volumes to the counties of interest:

Gasoline volume disaggregated for Sacramento = State gasoline volume × p_{avg}
= 12.93 × 0.0354 = 0.458 billion gallons = 458 million gallons

Diesel volume disaggregated for Sacramento = State special fuels volume × p_{avg}
= 2.035 × 0.0354 = 0.072 billion gallons = 72 million gallons

Step 3. Adjust for refueling losses.

Sacramento county has implemented both Stage I and Stage II refueling losses controls. From Table 4-4, the loss rate for gasoline = 0.06 + 0.01 + 0.07 + 0.09 gallons lost per 1000 gallons throughput = 0.23 gallons/1000 gallons. The loss rate for diesel is simply due to spillage, 0.09 gallons/1000 gallons. Therefore,

Adjustment for gasoline = -458 million × 0.23/1000 = -105 thousand gallons

Adjustment for diesel = -72 million × 0.09/1000 = -6.5 thousand gallons

On-road gasoline use in Sacramento County, G_{county} = 458 million - 105 thousand
$\cong$ 458 million gallons

On-road diesel use in Sacramento County, D_{county} = 72 million - 6.5 thousand
$\cong$ 72 million gallons

These loss rates generally do not significantly affect the results of these calculations. However, if loss rates are improved and revised in the future, or if there is a substantial future increase in automobile fuel economies, it is possible that this fraction of total fuel consumption may become more important.

Step 4. Calculate diesel- and gasoline-powered fleet fuel economies (mpg_{fd} and mpg_{fg}).

Local data could not be obtained, therefore, estimates based on national averages were used. Values of $vmt_{ij}/VMT \times mpg_{ij}$ (see Equation 4-5) are tabulated in Table 4-10 (at the end of this section). Summing over vehicle class *i* and vehicle age j, the following estimates for fleet fuel economies were calculated.

mpg_{fd} = 7.5 miles/gallon (for the diesel fleet)
mpg_{fg} = 20.7 miles/gallon (for the gasoline fleet)

The low fuel economy for the diesel fleet reflects the large proportion of vehicle miles traveled by heavy duty diesel vehicles (approximately 91% from Table 4-5).

Step 5. Calculate area annual VMT for the diesel- and gasoline- powered fleets.

VMT-gas = 458 million gallons × 20.7 miles/gallon = 9.48 billion miles

VMT-diesel = 72 million gallons × 7.5 miles/gallon = 0.54 billion miles

These calculated values compare well to data obtained from the FHWA's Highway Performance Monitoring System (HPMS) database, and from BURDEN7F model output:

	HPMS	BURDEN7F	Units
VMT-gas	8.25	8.89	billion miles
VMT-diesel	0.65	0.49	billion miles

However, they compare poorly to the California Department of Transportation total VMT estimate for Sacramento County, 4.11 billion miles (California Department of Transportation, 1994).

MARICOPA COUNTY, ARIZONA

Table 4-8 lists data that were obtained from the FHWA report, *Highway Statistics-1994*, and from various Arizona state agencies (listed in Appendix A). Service station taxable sales for the state could not be calculated. Confidentiality restrictions have been placed on this data for several Arizona counties. However, this figure is at least 0.198 billion dollars, the sum of taxable service station sales in the unrestricted counties. Judging by the impact the restricted counties have on the state economy, this figure probably lies between 0.20 and 0.24 billion dollars.

Table 4-8
Maricopa County data obtained for 1994

	Maricopa County	Arizona	Units
Total Gasoline Consumption		1.935	billion gallons
Total Special Fuels Consumption (On-Highway)		0.463	billion gallons
Active Drivers Licenses	1514	2631	thousand
Registered Vehicles	1532	2762	thousand
Population	2.36	4.08	million
Total Taxable Sales	31.9	49.3	billion $
Service Station Taxable Sales	0.149	0.20-0.24	billion $

Step 1. Adjust fuel consumption for off-road use.

FHWA reports that 0.035 billion gallons of gasoline were used off-road in Arizona in 1994, and that the state did not account for handling losses. For states that track handling losses, these losses should be redistributed to on- and off-road FHWA consumption figures on a percent basis before continuing with this procedure. These losses are inconsistently calculated for the FHWA report, and an alternate means to account for these losses is included in step 3. Special fuels consumption was reported for on-highway use, therefore, no adjustment is necessary.

Adjusted gasoline consumption = 1.935 - 0.035 = 1.90 billion gallons.

Step 2. Geographically disaggregate fuel consumption.

Calculate county-state proportions (p) for numbers of drivers licenses (dl), numbers of registered vehicles (rv), population (pop), total taxable sales (ts), and service station taxable sales (ss). Also calculate the average proportion (p_{avg}) and percent differences from the average. For instance, the county-state proportion for the number of drivers licenses is calculated as follows: $p_{dl} = dl_{county}/dl_{state} = 727 / 20156 = 0.0361$. The five proportions and the average proportion are shown in Table 4-9. Due to uncertainty in p_{ss}, it was excluded from the calculation of the average in this case.

Table 4-9
Maricopa county/state proportions for various factors

Proportion	Value	% Difference from p_{avg}
p_{dl}	0.575	-7%
p_{rv}	0.555	-10%
p_{pop}	0.578	-6%
p_{ts}	0.647	+5%
p_{ss}	0.62 to 0.75	+1% to +20%
p_{avg}	0.617	na

Since the proportions are roughly equal (none differs from the average by more than 10 percent, with the possible exception of p_{ss}) p_{avg} may be used to geographically disaggregate fuel consumption:

Gasoline volume desaggregated for Maricopa = State gasoline volume × p_{avg}
$= 1.90 \times 0.617 = 1.17$ billion gallons

Diesel volume disaggregated for Maricopa = State special fuels volume × p_{avg}
$= 0.463 \times 0.617 = 0.286$ billion gallons

Step 3. Adjust for refueling losses.

Maricopa county has implemented both Stage I and Stage II refueling losses controls. From Table 4-4, the loss rate for gasoline = 0.06 + 0.01 + 0.07 + 0.09 gallons lost per 1000 gallons throughput = 0.23 gallons/1000 gallons. The loss rate for diesel is simply due to spillage, 0.09 gallons/1000 gallons.

Adjustment for gasoline = -1.17 billion × 0.23/1000 = -270 thousand gallons

Adjustment for diesel = -0.286 billion × 0.09/1000 = -26 thousand gallons

On-road gasoline use in Maricopa County, G_{county} = 1.17 billion - 270 thousand
≅ 1.17 billion gallons

On-road diesel use in Maricopa County, D_{county} = 0.286 billion - 26 thousand
≅ 0.286 billion gallons

Step 4. Calculate diesel- and gasoline-powered fleet fuel economies (mpg_{fd} and mpg_{fg}).

Local mileage accumulation and registration data could not be obtained, therefore, fleet fuel economy estimates based on national averages were used. The results are the same as calculated above for Sacramento county:

mpg_{fd} = 7.5 miles/gallon (i.e., 7.5 mpg for all diesel vehicles)
mpg_{fg} = 20.7 miles/gallon (i.e., 20.7 mpg for all light and heavy duty trucks and cars).

Step 5. Calculate area annual VMT for the diesel- and gasoline- powered fleets.

VMT-gas = 1.17 billion gallons × 20.7 miles/gallon = 24.2 billion miles
VMT-diesel = 0.286 billion gallons × 7.5 miles/gallon = 2.15 billion miles

These figures were compared to the 1994 HPMS database, and were more than an order of magnitude larger. However, this appears to be due to a typographical error in the HPMS database. This example highlights the value of performing a double-check on VMT estimates. Other sources of information estimate that total VMT in Maricopa County was approximately equal to 20 billion miles in 1993. This figure represents VMT by both gas- and diesel-powered vehicles. The 1994 estimate calculated above is 30 percent greater, while an annual growth rate of only five to seven percent is expected. This suggests that it may be useful to reevaluate Maricopa County VMT estimates calculated by more traditional means.

Table 4-10
Calculation of national average fleet fuel economies from MOBILE5a Defaults

	vmtij/VMT*mpgij by vehicle class and model year										
Class	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981
Gas-powered fleet											
LDGV	0.033	0.011	0.015	0.021	0.030	0.044	0.069	0.100	0.112	0.168	0.230
LDGT1	0.019	0.004	0.007	0.007	0.009	0.016	0.023	0.030	0.024	0.050	0.073
LDGT2	0.015	0.004	0.005	0.005	0.006	0.011	0.015	0.020	0.015	0.032	0.046
HDGV	0.014	0.002	0.003	0.003	0.003	0.005	0.007	0.008	0.010	0.010	0.014
MC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Sum =	**20.66**	**=mpg-fg**								
Diesel-powered fleet											
LDDV	0.001	0.000	0.001	0.001	0.001	0.003	0.007	0.018	0.046	0.142	0.263
LDDT	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.008	0.033	0.077
HDDV	0.033	0.011	0.016	0.022	0.028	0.044	0.069	0.085	0.074	0.065	0.162
	Sum =	**7.54**	**=mpg-fd**								

vmt_{ij} = annual average vehicle miles traveled for cars of class *i* and age *j*.
VMT = annual total vehicle miles traveled for all classes
mpg_{ij} represents the fuel economy of a vehicle of class *i* and age *j*

Table 4-10. Continued.

	vmtij/VMT*mpgij by vehicle class and model year									
Class	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991
Gas-powered fleet										
LDGV	0.380	0.531	0.667	0.680	0.745	0.782	0.911	1.309	1.436	1.573
LDGT1	0.097	0.114	0.138	0.085	0.098	0.122	0.155	0.217	0.271	0.347
LDGT2	0.058	0.112	0.112	0.042	0.050	0.091	0.058	0.087	0.093	0.132
HDGV	0.017	0.017	0.022	0.012	0.011	0.011	0.009	0.016	0.019	0.026
MC	0.000	0.000	0.019	0.005	0.008	0.011	0.016	0.022	0.030	0.041
Diesel-powered fleet										
LDDV	0.347	0.214	0.220	0.120	0.045	0.048	0.000	0.000	0.000	0.035
LDDT	0.182	0.106	0.063	0.019	0.014	0.008	0.007	0.010	0.013	0.017
HDDV	0.218	0.223	0.283	0.241	0.248	0.214	0.224	0.366	0.471	0.465

vmt_{ij} = annual average vehicle miles traveled for cars of class *i* and age *j*.
VMT = annual total vehicle miles traveled for all classes
mpg_{ij} represents the fuel economy of a vehicle of class *i* and age *j*

Table 4-10. Concluded

	vmtij/VMT*mpgij by vehicle class and model year		
Class	1993	1994	1995
Gas-powered fleet			
LDGV	1.733	1.742	0.365
LDGT1	0.391	0.415	0.105
LDGT2	0.140	0.145	0.036
HDGV	0.030	0.032	0.000
MC	0.072	0.096	0.000
Diesel-powered fleet			
LDDV	0.041	0.042	0.089
LDDT	0.020	0.022	0.006
HDDV	0.542	0.588	0.000

vmt_{ij} = annual average vehicle miles traveled for cars of class *i* and age *j*.
VMT = annual total vehicle miles traveled for all classes
mpg_{ij} represents the fuel economy of a vehicle of class *i* and age *j*

5
REFERENCES

ARB. 1993. "Methodology for Estimating Emissions from On-Road Motor Vehicles Volume II: Weight (E7FWT)." CARB, Technical Support Division, Mobile Source Emission Inventory Branch. June 1993.

ARB. 1993. "Methodology for Estimating Emissions from On-Road Motor Vehicles: Volume III: Burden 7F." CARB, Sacramento California, June 1993.

ARB. 1995. "Emission Inventory Procedural Manual - Volume III: Methods for Assessing Area Source Emissions." CARB, Sacramento, California.

Attaran, Kazan. 1995. California Department of Transportation, Transportation Economics Division. Personal communication.

Bishop et. al. 1994. "Enhancements of Remote Sensing for Vehicle Emissions in Tunnels." *J. Air & Wast Manage. Assoc.* **44**: 169-175.

California Department of Transportation. 1994. *Travel and Related Factors in California*. California Department of Transportation, Sacramento, California.

EEA. 1985. "Estimation of Mileage Accumulation Rates and I/M Failure Rates from I/M Program Data." Prepared for U.S. EPA, MVEL. Energy and Environmental Analysis, Inc., Arlington, VA. (Contract No. 68-03-1865, Work Assignment No. 12).

EPA. 1991. "Identifying Excess Emitters with a Remote Sensing Device: A Preliminary Analysis." Presented to the Society of Automotive Engineers, Paper #911672.

EPA. 1992a. "User's Guide for the Urban Airshed Model. Volume IV: User's Manual for the Emissions Preprocessor System 2.0." EPA-450/4-90007D(R), U.S. EPA Office of Air Quality Planning and Standards.

EPA. 1992b. "Evaluation of a Remote Sensing Device at a Centralized I/M Lane." Presented at the Society of Automotive Engineers, International Fuels and Lubricants Meeting and Exposition, Paper #922315, San Francisco, CA., October 19-22, 1992.

Federal Highway Administration. 1994. *Highway Statistics* (published annually), Office of Highway Information Management, Washington D.C.

Hayes, A. 1993. *A Determination of Motor Vehicle Activity Factors for Atlanta, Georgia* *through Fuel Consumption Analysis.* Master's thesis prepared under the direction of Michael O. Rodgers, Georgia Institute of Technology, Atlanta, GA.

Mendenhall, William, Richard L. Scheaffer, and Dennis D. Wackerly. 1986. *Mathematical Statistics with Applications*, Duxbury Press, Boston.

Pollack, A. K. et al. 1992. "Comparison of Remote Sensing Data and Vehicle Emission Models: The Proportion of Emissions from High Emitting Vehicles." Presented at Air and Waste Management Association 85th Annual Meeting, Kansas City, June 21-26, 1992.

Radian. 1995. "Evaluation of the California Pilot Inspection/Maintenance (I/M) Program)." Prepared for California Bureau of Automotive Repair by Radian Corporation and de la Torre Klausmeier Consulting, Inc., Austin, TX.

Sierra. 1995. "Analysis of Data from the California Enhanced I/M Pilot Program." Prepared for the U.S. EPA by Sierra Research, Inc., Sacramento, CA. (Contract No. 68-C4-0056, Work Assignment No. 0-03).

US Department of Transportation. 1984. *Guide to Urban Traffic Volume Counting*. FHWA/L/81/019. US Department of Transportation, Washington, DC.

USEPA. 1991. *The Mobile4 Fuel Consumption Model.* EPA-AA-TEB-EF-91-x (draft for internal circulation), U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

USEPA. 1992c. *Procedures for Emission Inventory Preparation - Volume IV: Mobile Sources.* EPA-450/4-81-026d, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

USEPA. 1994. *User's Guide to Mobile5 (Mobile Source Emission Factor Model).* EPA-AA-AQAB-94-01, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Valley Research. 1994. "On-Road Motor Vehicle Activity Data Volume II - Vehicle Age Distribution and Mileage Accumulation Rate by County." Prepared for the California Air Resources Board, Research Division by Valley Research Corporation, Northridge, CA. (Contract No. A132-182).

APPENDIX A

LIST OF CONTACTS IN FEDERAL AND STATE AGENCIES

All States

Office of Highway Information Management, Federal Highway Administration, Washington, D.C.

California

For taxable sales:
(Robert Rossi) Statistics Section, Agency Planning and Research Division, State Board of Equalization, 450 N. Street, Sacramento, California 95814.

For statewide fuel sales, drivers' license, vehicle registration, and population:
Travel and Related Factors in California. published annually by the Transportation System Information Program, California Department of Transportation, Sacramento, California.

Arizona

For taxable sales:
(Karen Walker) Econometrics, Arizona Department of Revenue, 1600 W. Monroe, Phoenix, Arizona 85007.

For drivers' license, vehicle registrations:
(Marv Dobson) Motor Vehicle Division, Arizona Department of Transportation, 1801 West Jefferson Street, Phoenix, Arizona.

For population:
Arizona Department of Economic Security, Phoenix, Arizona.

New York

For taxable sales:
(Steven Zych) Revenue Analysis and Data Bureau, Office of Tax Policy Analysis, Department of Taxation and Finance, W.A. Harriman Campus, Albany, New York 12227.

For drivers' license, vehicle registrations:
(Brian Ginett) Department of Motor Vehicles, Albany, New York 12228.

This page intentionally left blank.

APPENDIX B

ESTIMATED VEHICLE FUEL ECONOMIES BY VEHICLE CLASS AND MODEL YEAR

Fuel Economies from Emissions Models Defaults

Class	Fuel Economies by Vehicle Class and Model Year (mpg)										
	1941-1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
Gas-powered fleet											
LDGV*	12.46	12.46	12.06	12.15	11.95	13.20	14.58	15.32	16.55	16.83	19.42
LDGT1*	10.64	10.64	10.64	10.64	10.45	10.45	11.45	12.05	13.09	12.78	12.31
LDGT2*	10.65	10.65	10.65	10.65	10.46	10.46	11.45	12.06	13.10	12.78	12.34
HDGV*	7.76	7.88	7.88	7.88	7.90	8.06	8.22	8.46	8.68	8.87	9.01
MC***	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
Diesel-powered fleet											
LDDV**	21.56	21.56	20.86	21.02	20.64	22.84	25.22	24.82	24.99	23.90	26.22
LDDT**	18.42	18.42	18.42	18.42	18.09	18.09	19.81	19.53	19.77	18.15	16.64
HDDV***	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50

*From Mobile5a
*From diesel advantage factors given in the Mobile4 Fuel Consumption model.
Fuel econ (diesel) = adv factor x Fuel econ (gas)
***From EMFAC7F

Fuel Economies from Emissions Models Defaults

Class	Fuel Economies By Vehicle Class and Model Year (mpg) 1991	1992	1993	1994	1995
Gas-powered fleet					
LDGV*	22.71	22.68	22.66	22.64	22.64
LDGT1*	17.10	17.05	16.99	16.93	16.87
LDGT2*	16.95	16.86	16.77	16.68	16.58
HDGV*	10.48	10.49	10.50	10.51	10.52
MC***	50.00	50.00	50.00	50.00	50.00
Diesel-powered fleet					
LDDV**	27.25	27.22	27.19	27.17	27.17
LDDT**	20.43	20.35	20.26	20.17	20.07
HDDV***	5.81	5.86	5.96	6.07	6.17

*From Mobile5a

**From diesel advantage factors given in the Mobile4 Fuel Consumption model.

Fuel econ (diesel) = adv factor x Fuel econ (gas)

***From EMFAC7F

APPENDIX C

ESTIMATED NATIONAL AVERAGE VMT MIX FOR 1994

Vehicle Mix for 1994 - Estimated Using Mobile5 Defaults

	Vehicle Mix by Class	Vehicle Mix by Class and Model Year									
Class	vmti/VMT	1941-1971*	1972	1973	1974	1975	1976	1977	1978	1979	1980
Gas-powered											
LDGV	0.681	0.00272	0.00089	0.00123	0.00157	0.00204	0.00286	0.00415	0.00592	0.00579	0.00810
LDGT1	0.190	0.00174	0.00038	0.00063	0.00066	0.00080	0.00133	0.00172	0.00235	0.00191	0.00330
LDGT2	0.089	0.00138	0.00033	0.00052	0.00047	0.00055	0.00095	0.00118	0.00156	0.00124	0.00211
HDGV	0.033	0.00173	0.00022	0.00037	0.00039	0.00040	0.00059	0.00076	0.00095	0.00106	0.00100
MC	0.007	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Diesel-powered											
LDDV	0.062	0.00003	0.00002	0.00003	0.00006	0.00006	0.00011	0.00028	0.00075	0.00176	0.00542
LDDT	0.031	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00025	0.00048	0.00170
HDDV	0.908	0.00608	0.00200	0.00300	0.00399	0.00517	0.00799	0.01262	0.01552	0.01343	0.01180

*approximated

Vehicle Mix for 1994 - Estimated Using Mobile5 Defaults

	Vehicle Mix by Class	Vehicle Mix by Class and Model Year									
Class	vmti/VMT	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
Gas-powered	0.681	0.01069	0.01764	0.02431	0.02969	0.02928	0.03180	0.03296	0.03909	0.05686	0.06319
LDGV	0.190	0.00442	0.00576	0.00658	0.00811	0.00495	0.00546	0.00677	0.00881	0.01258	0.01567
LDGT1	0.089	0.00279	0.00349	0.00652	0.00660	0.00243	0.00279	0.00506	0.00331	0.00504	0.00543
LDGT2	0.089	0.00279	0.00349	0.00652	0.00660	0.00243	0.00279	0.00506	0.00331	0.00504	0.00543
HDGV	0.033	0.00142	0.00173	0.00171	0.00216	0.00122	0.00112	0.00107	0.00092	0.00154	0.00189
MC	0.007	0.00000	0.00000	0.00000	0.00039	0.00011	0.00016	0.00022	0.00031	0.00043	0.00059
Diesel-powered											
LDDV	0.062	0.00994	0.01320	0.00810	0.00815	0.00433	0.00159	0.00169	0.00000	0.00000	0.00000
LDDT	0.031	0.00384	0.00895	0.00507	0.00307	0.00091	0.00066	0.00036	0.00032	0.00048	0.00061
HDDV	0.908	0.02950	0.03958	0.04048	0.05147	0.04384	0.04475	0.03821	0.03958	0.06417	0.08178

Vehicle Mix for 1994 - Estimated Mobile5 Defaults

	Vehicle Mix by Class	Vehicle Mix by Class and Model Year				
Class	vmti/VMT	1991	1992	1993	1994	1995
Gas-powered						
LDGV	0.681	0.06925	0.07143	0.07647	0.07695	0.01614
LDGT1	0.190	0.02030	0.02160	0.02299	0.02448	0.00622
LDGT2	0.089	0.00779	0.00808	0.00837	0.00867	0.00219
HDGV	0.033	0.00249	0.00265	0.00282	0.00304	0.00000
MC	0.007	0.00081	0.00109	0.00145	0.00192	0.00000
Diesel-powered						
LDDV	0.062	0.00129	0.00137	0.00150	0.00154	0.00326
LDDT	0.031	0.00082	0.00090	0.00098	0.00108	0.00028
HDDV	0.908	0.07997	0.08514	0.09086	0.09694	0.00000

This page intentionally left blank

APPENDIX D

MULTIVARIATE LINEAR REGRESSION PROCEDURE

INTRODUCTION

This appendix describes how to use standard statistical regression analysis to determine weighting factors for calculating county fuel consumption from related variables as discussed in Section 4. We also describe here how to test whether particular variables are useful for predicting a dependent variable. The following state and U.S. variables that are correlated with fuel consumption are used:

On-road fuel consumption
Population
Number of registered vehicles
Number of active drivers licenses
Total taxable sales (dollar amount)
Total service station sales (dollar amount)

These state and national figures should be obtained for 20 time periods (e.g. 20 years, 20 months, 20 quarters, etc.).

Procedure Steps

1. Calculate state/U.S. proportions of the above quantities.
2. Perform a linear least-squares regression analysis using a spreadsheet or statistical software.
3. Perform a standard error analysis to determine which of the proportions are useful for predicting state fuel consumption.
4. Eliminate those proportions that are not useful.
5. Repeat Step 2, with useful proportions only.

The linear model to predict county fuel consumption (F_{county}) is

$$F_{county}/F_{state} = a \times p_{ss} + b \times p_{ts} + c \times p_{pop} + d \times p_{rv} + e \times p_{dl} \quad \text{D-1}$$

In Equation D1, F_{county}/F_{state} is termed the dependent variable, the five proportions are the independent variables, and a though e are regression coefficients. For equation D1, it is assumed that the y-intercept is equal to 0, although this is not absolutely necessary. If prediction of F_{county}/F_{state} is improved by allowing a non-zero y-intercept, then equation D1 should be written as follows, where f represents the y-intercept.

$$F_{county}/F_{state} = a \times p_{ss} + b \times p_{ts} + c \times p_{pop} + d \times p_{rv} + e \times p_{dl} + f \quad \text{D-2}$$

Ideally, a regression analysis would be performed using Equation D1 or D2 as the linear model. However, fuel consumption statistics are only known at the statewide and nationwide levels.

Therefore, historical state and U.S. data must be used to estimate the regression coefficients. Equation D3 is the linear model that will be used for this approach, where the proportions are calculated on a state/U.S. basis, rather than a county/state basis.

$$F_{state}/F_{U.S.} = a \times p_{ss} + b \times p_{ts} + c \times p_{pop} + d \times p_{rv} + e \times p_{dl} + f \qquad \text{D-3}$$

This approach assumes that state/U.S. fuel consumption patterns accurately reflect county/state patterns. Historical data for at least 20 time periods should be used to calculate the proportions in order to obtain a statistically robust sample size.

PROCEDURE

Step 1. Calculate state/U.S. proportions for population (pop), number of registered vehicles (rv), number of active drivers licenses (dl), total taxable sales (ts), total service station sales (ss), and on-road fuel consumption (F) for 20 time periods. For example, the state/U.S. proportion of service station sales, p_{ss} , would be calculated as

$$p_{ss} = ss_{state}/ss_{U.S.} \qquad \text{D-4}$$

Similarly, calculate p_{pop} , p_{rv} , p_{dl}, p_{ts} , and $F_{state}/F_{U.S.}$

Step 2. Perform a linear least-squares regression analysis on the values obtained in Step 1. A spreadsheet computer program (such as Quattro Pro, Excel, or another) or a simple statistical program (such as Minitab or StatMost) is needed. $F_{state}/F_{U.S.}$ should be treated as the dependent variable, and the other proportions should be treated as five independent variables. Output will include the regression coefficients (a though e) and their standard errors (SE_a though SE_e), the y-intercept (f) and its standard error (SE_f), and a correlation coefficient (R-squared).

Step 3. Perform a standard error analysis to determine which of the five proportions are useful for predicting state fuel consumption. This analysis is based on the Student's t-Distribution. See Woolon (1987), Mendenhall et. al. (1986), or another statistics text for a complete discussion of the t-Distribution. A statistical level of significance (α) must be selected as a decision criteria, which defines the probability that this analysis will be in error. Frequently used levels of significance are 0.01, 0.05, or 0.10.

t-Statistics are then calculated for the intercept, f, and the correlation coefficients a though e. For example, the t-statistic for coefficient b is calculated as

$$t\text{-}statistic_b = b \;/\; Se_b \qquad \text{D-5}$$

Each value of the t-statistic corresponds to a probability (p-value), tabulated in tables of the student's t-distribution (available in most statistics texts). If the p-value for a coefficient is greater than α, its corresponding proportion is considered to be not useful for predicting fuel consumption, and it may be eliminated from the linear model. If the p-value for the intercept is greater than α, the regression may be forced through zero ($f = 0$).

Step 4. Eliminate the proportions that are not useful, and repeat Step 2 with useful proportions only.

This page intentionally left blank

APPENDIX E

TECHNIQUE TO EXTRACT NATIONAL AVERAGE TRAVEL FRACTIONS AND FUEL ECONOMICS FROM THE MOBILE5a SOURCE CODE

In order to estimate VMT, travel fractions must be defined by vehicle type and age. National average values may be extracted from the source code of MOBILE5, the EPA-approved model for calculating emission factors. The following information is provided because the methods to extract these data are not readily apparent in the MOBILE documentation.

Travel fractions may be obtained by generating a certain output file. This file is generated by setting the OUTFMT flag (*) to 5 and including the shown set of flags int he "One-Time Data" section of the MOBILE5 input file (**). The One-Time Data" record flags shown below reference the various motor vehicle categories included in the MOBILE model. For more detailed information regarding user inputs, the user's guide to MOBILE should be consulted (USEPA, 1994). The following example input file:

	1	PROMPT -	No prompting.
	Project ID: *User-Defined Text*		
	1	TAMFLG -	Use MOBILE5 default tampering rates.
	1	SPDFLG -	User supplies one value of average speed for all vehicle types.
	1	VMFLAG -	Use MOBILE5 VMT mix.
	1	MYMRFG -	Use default mileage accumulation rates and registration data.
	1	NEWFLG -	MOBILE5 basic exhaust emissions rates are used.
	1	IMFLAG -	No I/M program.
	1	ALHFLG -	Do not apply additional correction factors.
	1	ATPFLG -	No ATP program.
	1	RLFFLAG -	Use uncontrolled refueling emission rates.
	2	LOCFLG -	LAP record will appear once, in one-time data section.
*	1	TEMFLG -	Min/Max temperatures will be used.
	5	OUTFMT -	By model year, 112-column descriptive format.
	4	PRTFLG -	Print exhaust HC, CO, and NO_x results.
	1	IDLFLG -	No idle emission factors calculated.
	3	NMHFLG -	Calculate emissions for volatile organic hydrocarbons.
	3	HCFLAG -	Print sum and component emissions.
	User-Defined. C 72. 92. 11.5 08.7 92 1 1 2		Local Area Parameter record
	22222222 1		By-Model Year Vector
**	1 95 19.6 75.0 20.6 27.3 20.6 01		Scenario description record

The output file produced by running MOBILE5 with the above inputs would provide the travel fractions by model type and age.

This page intentionally left blank.