1. Report No. TX-02/4142-9	2. Government Accession No.	Recipient's Catalog No.
4. Title and Subtitle Forecasting Traffic Characteristics for Air Quality Analyses		5. Report Date Aug-02
		6. Performing Organization Code
7. Author(s) Lei Yu, Fengxiang Qiao, Guangchun Li and Hong Oey		8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Transportation Studies Texas Southern University 3100 Cleburne Avenue Houston, TX 77004		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. 0-4142
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Transfer Office P.O. Box 5080 Austin, Texas 78763-5080		13. Type of Report and Period Covered Research Report 9/1/00-8/31/02
		14. Sponsoring Agency Code

15. Supplementary Notes

Project conducted in cooperation with Federal Highway Administration
Research project title: Forecasting Traffic Characteristics for Air Quality Analyses

16. Abstract

This research intends to develop techniques for estimating and forecasting three critical mobile source emission related travel indicators: vehicle age distribution, mileage accumulation rates by vehicle type, and VMT & mix. As for modeling the vehicle age distribution, two model types were used; each of which contains the linear model, nonlinear model and time series model. Age distributions for the 8 counties in HGAC area and in El Paso in Texas were used to validate the model structures and parameters. As for modeling the mileage accumulation rate, a practical algorithm for adjusting the default mileage accumulation rate for site-specific application based on a small sample field survey was developed. Applications in two Texas cities (Houston and El Paso) were conducted and the whole operation process and its impacts on the estimates of emission factors were presented. As for VMT & mix, extensive efforts were made on collecting information on VMT & mix estimation. A national-wide survey through e-mail was conducted to ascertain what kinds of methodologies were used by other states. Accordingly, a practical improvement to the VMT & mix estimation methodologies was developed. Link volume estimates were modeled as the function of both the traffic count data and the link attributes. A case study in southwest Houston was illustrated to show the estimation process and the effectiveness of the proposed approach. Impact analysis shows that the emission factors generated by local improved VMT estimation are closer to the ideal one, and better than both the nationwide default and the Traffic Count Method.

17. Key Words	18. Distribution Statement
Vehicle Emission, Transportation Foresting, MOBILE, Vehicle Age Distribution, Mileage Accumulation, VMT, Volume Estimation	No restriction. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classify (of this report)	20. Security Classify (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	122	

Forecasting Traffic Characteristics for Air Quality Analyses

By

Lei Yu, Ph.D., P.E., Fengxiang Qiao, Ph.D., Guangchun Li and Hong Oey

Report Number 4142-9
Research Project Number 0-4142
Project Title: Forecasting Traffic Characteristics for Air Quality Analyses

Sponsored by the Texas Department of Transportation
In Cooperation with
U.S. Department of Transportation
Federal Highway Administration

August 2002

Texas Southern University
3100 Cleburne Avenue
Houston, Texas 77004

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the Texas Department of Transportation. This report does not constitute a standard, specification, or regulation.

Notice

The United States Governments and the state of Texas do not endorse products or manufacturers. Trade or manufactures' names appear herein solely because they are considered essential to the subject of this report.

Acknowledgments

The authors would like to express their sincere gratitude for the support and valuable comments that they received from Project Director Mr. Charlie Hall, Texas Department of Transportation through the course in conducting this project. The authors would like to express their thanks to the Project Monitoring Committee, other TxDOT, TTI and HGAC personnel for any direct or indirect assistance that they received.

The authors also express their sincere thanks to those who responded the e-mail survey on VMT related variables. The authors would like to thank the research assistants of both Department of Transportation Studies at Texas Southern University (TSU) and University of El Paso for collecting the data. Also the authors would like to express thanks to all personnel at the two universities who have directly or indirectly contributed to this project or have provided various assistances.

Credits for Sponsor

Research performed in cooperation with the Texas Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration.

Table of Contents

List of Figures

List of Tables

Summary

Accurate and reliable quantification of mobile source emissions is very important in the conformity determination process. In order for each state to determine conformity in a consistent manner, the U.S. Environmental Protection Agency (EPA) requires that all states employ MOBILE (the previous version is MOBILE5 and the newest version is MOBILE6) emission factor model (EMFAC in California). MOBILE is a computer program that estimates hydrocarbon (HC), carbon monoxide (CO), and oxides of nitrogen (NO_x) emission factors for gasoline-fueled and diesel highway motor vehicles, and for certain specialized vehicles such as natural gas fueled or electric vehicles that may replace them (Environmental Protection Agency, 2001a).

MOBILE6 calculates emission factors for 28 individual vehicle types in low- and high-altitude regions of the United States. MOBILE estimates emission factors for any calendar year between 1952 and 2050, inclusive. Vehicles from the 25 most recent model years are considered to be in operation in each calendar year.

MOBILE6 emission factor estimates require inputs of various conditions such as ambient temperatures, travel speeds, operating modes, fuel volatility, and mileage accrual rates.

A crucial part in using MOBILE is the input of reliable travel indicators such as the vehicle age distribution, mileage accumulation rates by vehicle type, vehicle miles traveled (VMT) related variables, compositions of traffic, average speeds, etc.

This research intends to develop techniques for estimating and forecasting the three critical travel indicators related to mobile source emission: vehicle age distribution, mileage accumulation rates by vehicle type, and VMT & mix.

As for estimating vehicle age distribution, two types of models were developed. Model Type I (**MT I**) models the number of vehicles for the particular vehicle type in particular age, and then transfers the results to project the future age distribution. Model Type II (**MT II**) models the future age distribution directly. Both model types contain a family of linear models, nonlinear models and time series models. Based on a certain kind of criteria, the "best" model can be chosen from the two model families. Examples for the eight counties in Houston-Galveston Area Council (HGAC) area and in El Paso area are presented. In addition, the differences between the emission factors generated by MOBILE based on the default age

distribution values and the forecasted values by the proposed model are compared. Results show that the differences are big, which implies that the proposed model should be used to generate locality-specific MOBILE emission factors.

As for mileage accumulation rates, extensive efforts were made to collect vehicle mileage accumulation data in the Houston area and El Paso area. The survey results were used for building the site-specific model for estimating vehicle mileage accumulation rates in the corresponding local area. The modeling of the adjusting process for mileage accumulation was developed mathematically in this report. The adjusted local mileage accumulation can be obtained by the combined usage of the real survey data and the default nationwide data. To illustrate this process, the adjusting factors as well as the final adjusted mileage accumulation rates for Houston area and El Paso were calculated. As shown from the results, the real mileage accumulation rate in Houston area is 1.34 times higher than the nationwide default value, while is 0.58 times lower than the national-wide default value in El Paso area.

As for VMT related variables, currently there are several estimation methodologies. However, none of the existing approaches can be directly used for MOBILE6 in Texas. In this research, the improvements to VMT estimation were proposed which considered both the link attributes and the traffic count information. The proposed model for volume estimation is easy to be calibrated. Case study and model calibration in southwest Houston show that the improved approach is better than both the EPA Traffic Count Method and the nationwide MOBILE6 defaults in terms of the estimation of both VMT related variables and emission factors.

CHAPTER 1

INTRODUCTION

1.1 Background of Research

A number of Texas cities have been designated as non-attainment areas in the past years due to the stringent air quality set by the Environmental Protection Agency (EPA) and federal regulations. These designations are accompanied by a set of planning requirements, a State Implementation Plan (SIP) mandate, and potential retributions for failure to comply with the conditions. TxDOT and State MPOs must work with TNRCC to assess trade-offs between mobile- and other-source-emission reduction programs and adopt a specific set of SIP strategies that are feasible and achievable to reach air quality attainment status. If large emission reduction targets are assigned to mobile sources and included in the SIP unrealistically, conformity demonstrations will be difficult to make. Therefore, accurate and reliable quantification of mobile source emissions is very important in the conformity determination process. In order for each state to determine conformity in a consistent manner, EPA requires that all the states employ MOBILE emission factor model (EMFAC in California) to generate mobile source emission factors for different vehicle types.

A crucial part in properly running MOBILE, or some other emission models, is the availability of reliable travel indicators related to mobile source emissions , such as the vehicle age distribution, mileage accumulation rates by vehicle type, vehicle miles traveled (VMT) related variables, compositions of traffic, average speeds, ambient temperature, and etc. MOBILE is used to generate emission factors for each emission species, which can be combined with travel demand models to calculate the mobile source emissions factors. Specifically, MOBILE calculates the emissions such as HC, CO, and NO_x in grams per mile, a travel demand model supplies an estimate of Vehicle Miles Traveled (VMT), and the total grams of pollutants emitted by vehicles can be produced by multiplying the emission factors by the VMT.

In practice, the level of detail at which the emissions analysis is conducted varies substantially among different metropolitan regions. But the EPA requires that metropolitan planning areas categorized as serious or higher in non-attainment designation for ozone and CO estimate their mobile source emissions using network-based transportation models. The planning organizations in these areas, in general, conduct their emissions analysis at an individual link level. This involves the estimation of volumes and speeds on each network link in the metropolitan area from travel demand models such as EMME/2 and TRANSPLAN, followed by the computation of link-specific emissions factors based on a) link VMT, b) vehicle speed on the link, c) the vehicle class-specific emissions factors, and d) VMT mix fractions in vehicle classes. Of all of these, the link VMT and link speeds are obtained directly from the network-based travel demand models. The vehicle class-specific emissions factors are obtained from the emissions factor models based on the various inputs listed earlier.

1.2 Objectives of Research

This research intends to develop techniques for estimating and forecasting three critical mobile source emission related travel indicators: vehicle age distribution, mileage accumulation rates by vehicle type, and VMT related variables.

As a final product, the study will develop a guidebook containing techniques and models for estimating and forecasting mobile source emissions related travel indicators.

1.3 Outline of This Report

The next chapter of this report presents the extensive review of the state-of-the-art/practice of the modeling and forecasting of the three mobile source emissions related travel indicators. Chapter 3 describes the modeling process and computer programming for estimating vehicle age distribution, as well as the real case study in the Houston-Galveston Council Area (HGAC). Chapter 4 subsequently introduces the survey process of mileage accumulation rate in Houston and El Paso areas, and also describes the mathematical modeling of the correcting process for mileage accumulation. Chapter 5 presents the information collected on VMT related variables' estimation within and outside Texas. Chapter 6 proposes the improvements to VMT estimation. Finally, Chapter 7 gives conclusions for this report.

CHAPTER 2

REVIEW STATE-OF-THE-ART AND STATE-OF-THE-PRACTICE

This chapter intends to explore state-of-the art/practice on the estimation of vehicle age distribution, mileage accumulation rates and VMT related variables. A review has demonstrated, however, that reliable and consistent techniques for estimating the necessary travel indicators either do not exist or need to be substantially improved.

It is noted that MOBILE6 was released while the project was underway. In MOBILE5 only VMT mix is required. In MOBILE6 there are several VMT related variables that need to be provided including VMT by Facility, VMT by Hour, Speed VMT, and VMT mix. Therefore, the review of state-of-the-art/practice will include not only VMT mix but also the other VMT related variables.

2.1 Vehicle Age Distribution

MOBILE's emission factor calculations rely in part on travel fractions for vehicles of each given age and type, which in turn are based on estimates of the average annual mileage accumulation by age (first year to 25th - and - greater years of operation) for each of the eight vehicle types, and the registration distribution by age (age 0 - 1 to age 24 - 25+) for each vehicle types, except motorcycles, for which annual mileage accumulation rates and registration distribution are only provided for the 12th - and - later years of operation (age 0 - 1 to 11 - 12+). MOBILE uses national average annual mileage accumulation rates and registration distributions by age, and has provisions allowed the input of alternate data for either or both of these. The national annual mileage accumulation rates are based on analyses of information developed over a long period of time, and the registration distributions are based on analysis of calendar year 1990 registration.

Besides using the national average values for vehicle age distribution and mileage accumulation rates, there exist methodologies for forecasting vehicle age distribution. For example, Pearson and Frankel (1993) in Texas Transportation Institute (TTI) proposed the methodology for the development of projections of vehicle distributions by age that was actually two independent techniques, one being a cohort survival method and the other being a theoretical curve fit method. By these two methods, the vehicle age distribution for the year 2000, 2010, and 2020 was estimated and used to evaluate the potential impact on vehicle emissions and energy consumption.

For cohort survival method, the information needed was scrappage rates, migration factors, and a method for estimating new vehicle registrations. The new vehicle registrations within any given county were hypothesized to be dependent on a number of factors including population growth, local economic conditions, and others. The data used for their study included population projection in Texas counties. A linear regression was performed with new vehicles per capita as the dependent variable and total vehicles per capita as the independent variable.

For theoretical curve fit methodology, the hypothesis was that using the average vehicle age to non-dimensionalize the vehicle age groups creates a normalized distribution, which is approximately the same for every year. The cumulative percentage of vehicles of age i is an exponential function of a variable x, where x is the vehicle age i divided by average vehicle age.

The two methods are practical in real applications. Cohort survival technique realized the impact of socio-economic factors to new vehicle registration, and therefore to vehicle age distribution. However, only population was used in the final application. Moreover, the application of cohort survival technique in a single year steps could produce continually declining or increasing values which, over a 20-year period, could yield unreasonable results (Pearson and Frankel 1993, P. 29.) The theoretical curve fit method is straightforward. But it simply extends the current trends to more than 20 years later, which ignores the possible impacts of socio-economic factors.

It is a common knowledge that transportation characteristics are greatly influenced by its socio-economic environments. This is also true to vehicle age distribution. Therefore, it might be interesting to develop a methodology for estimating vehicle age distribution that emphasizes the impacts of socio-economic factors. This could improve the practice mentioned above.

2.2 Mileage Accumulation

Vehicles accumulate mileage at different rates depending on the type and age of the vehicle. Trucks tend to be driven more miles per year than cars. Older vehicles tend to be driven fewer miles per year than newer ones. Annual mileage accumulation affects the rate at which vehicle emission controls deteriorate and affect the relative emissions contributions of newer and older vehicles (United States Environmental Protection Agency, 2002, p.12.)

MOBILE6 assigns separate default mileage accumulation rates and allows user input for all vehicle classes. For annual mileage accumulation rates, MOBILE6 treats gasoline and diesel vehicles separately for a total of 28 separate vehicle types. MOBILE6 allows users to specify the vehicle classes for which new annual mileage accumulation rates are entered. All other classes are assumed to have the default rates (United States Environmental Protection Agency, 2002, p.12.)

Vehicle mileage rates may vary across years and with different local areas. For example, the odometer readings taken in Ohio, Wisconsin IM240 data and inspection and maintenance (I/M) data in Nashville, Tennessee indicate that average mileage accumulations may be much less than those used in MOBILE6 model in those area (Miller, et al. 2001.)

Methods for estimating annual mileage accumulations from I/M program data are given in the report "Methodology for Gathering Locality-Specific Emission Inventory Data" which is incorporated in "Volume IV: Chapter 1, Preferred and Alternative Methods for Gathering and Locating Specific Emission Inventory Data" (Heiken J. G. and et al., 1996.)

Heiken et al (1996, p.2-1:31) summarized methodology for gathering locality-specific mileage accumulation. 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 required in MOBILE5a. 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 new versions of the MOBILE model.

Some pioneer work in increasing the accuracy of local mileage accumulation rate has been conducted with the releasing of MOBILE6. Miller et al. (2001) gave a simple model accounting for scrappage of old vehicles as a function cumulative mileage (Miller, et al. 2001.) This model was calibrated based on the I/M data in Nashville, TN.

However, in some juristic area, the I/M data is not always available, and large scaled survey is also not feasible. For these areas, it is necessary to have a better way to incorporate the local information on mileage accumulation into MOBILE6. In this context, the research in this paper is intended to develop a practical algorithm for adjusting the default mileage accumulation rate for local use based on the small sample field survey. The small sample survey may contain some incomplete information such as small sample size, incomplete surveyed vehicle types.

2.3 Estimation of *VMT Related Variables*

Information on estimation of VMT related variables will be discussed in Chapter 5 in detail. The vehicle miles traveled (VMT) mix specifies the fraction of total highway VMT that is accumulated by each of the vehicle types. In MOBILE5, the VMT mix is used only to calculate the composite (all vehicle, or fleet wide) emission factors. MOBILE calculates a typical urban area VMT mix based on national data characterizing registration distributions and annual mileage accumulation rates by age for each vehicle type, the fraction of travel by each vehicle type that is typical of urban areas, and total vehicle counts (fleet size) by vehicle type.

The emissions factors for each of the three pollutants CO, VOC, and NOx vary quite widely among the different vehicle classes. Consequently, the emissions analysis is very sensitive to VMT mix. For example, at high temperatures, a 2.8% change in the heavy duty gas vehicle (HDGV) mix causes about a 10% change in the CO emissions rate, and a 4.8% change in the HDGV mix leads to about a 10% shift in the VOC emissions rate. It is, therefore, important to provide accurate VMT mix values.

Instead of using MOBILE default values, an alternative approach adopted by some metropolitan agencies is to use 24-hour local vehicle classification-counts to determine VMT mix, followed by the application of factors to convert vehicle types in traffic counts to the MOBILE vehicle classes. EPA recommends that local agencies adopt this approach because the MOBILE default values may not be reflective of the local traffic vehicle mix. In this local vehicle count-based approach, the VMT mix is typically stratified by the function classification of roadways to accommodate variations across roadway classes. However, since most counts are conducted only on higher roadway classes (such as interstates and major arterials), there is inadequate information to comprehensively capture variations in VMT mix by roadway class. Values of VMT mix obtained for the higher roadway classes are applied (sometimes after ad hoc adjustments based on judgment) to the lower roadway classes (such as minor arterials, collectors, and local roads.)

There are some other dimensions in estimating VMT mix. Some examples are listed as follows:

- VMT mix can be estimated by functional class using the Highway Performance Monitoring System (HPMS) methodology based on traffic count;
- VMT mix can be estimated for all state owned highways by county;
- VMT mix can be estimated based on fuel consumption records; and
- VMT mix can be estimated for the Cost Responsibility Study based on fuel tax and motor carrier tax records; and so on.

HPMS is a FHWA program, which was introduced in 1978 to strengthen the methods used by the states for collecting, estimating and reporting traffic count data, and to help reduce the effort involved in providing the federal government with necessary traffic data. Based on traffic counts in this program that are distributed over the national highway system, FHWA requires each state to report total state VMT by functional class. The cost of HPMS VMT estimation for the states is significant and because of cost, a sample design for traffic counts is used to develop annual VMT estimates by FHWA functional class and vehicle category for all highways in the state.

Therefore, efforts in improving the accuracy of VMT mix estimation have been made. At a national level, a review of the literature indicates several successful and relatively low-cost approaches for improved VMT mix estimates. Studies in Oregon and Virginia used 24-hour vehicle classification based traffic counts and a mapping approach that improved the seasonal and day-of-week factors used to convert raw counts into VMT mix estimates without requiring the collection of addition data. In addition, some other studies have used simulation models or have modestly increased sample sizes to improve reliability of estimates.

A problem with the state-of-the-art/practice discussed above for VMT mix determination is that they apply aggregate-level values across links in the road network in a region. It was found, in an analysis of VMT mix from 477 different count sites in the U.S., that substantial variations exist in VMT mix across the sites, emphasizing the need for local determination of VMT mix values (rather than using MOBIL default values). The same study also indicates substantial variation in VMT mix even after controlling for roadway class at any given site, underscoring the need to consider explanatory factors other than roadway class in local VMT mix analysis.

Since MOBILE6 is newly released, the practical methods for estimating the other VMT related variables that can exactly meet the needs of MOBILE6, including VMT by Facility, VMT by Hour, Speed VMT, have not been reported till this time. However, practices of general estimation of VMT can be found as mentioned below.

Lee-Gosselin and Richardson (1988) reported a study that looked into the problem of VMT estimation in Canada. They looked at the VMT estimation from different viewpoints and levels, as in this paper (regional VMT and VMT of different road categories, for instance). However, clear mathematical descriptions are not provided by Lee-Gosselin and Richardson. Hoang and Poteat (1980) also applied stratified sampling by stratifying the highway links by volume, area, and facility type. The difference between the two approaches is mainly that Hoang and Poteat calculated required sample sizes for each stratum.

As far as we know, the VMT estimation problem is not fully covered by any standard or guideline concerning traffic counting. The American Association of State Highway and Transportation Officials' Guidelines for Traffic Data Programs ("AASHTO", 1992) provided very little insight into the strategic planning of counting site network, though it gave numerous recommendations of how to carry out counting operations. The same applies to the ASTM Standard Practice for Highway-Traffic Monitoring ("Standard") 1994). All that was basically said is that the gathered data is finally aggregated as the national or regional VMT. This may be sufficient if the total counting site network covers the whole road network well. But there is also a risk that counting sites are distributed to mainly cover important main links and urban areas, and thus the system may give a biased estimate of the total VMT (or at least the less important areas and roads receive less attention and counting effort than they perhaps should).

Räty and Leviäkangäs (1999) showed how the VMT could be estimated by means of stratified probability proportional to size (PPS) cluster sampling. This approach is strategic, showing how the PPS method can be used as a tool to determine the approximate number of counting sites required, rather than operational, which is the next phase of network planning. It calculated the needed total sample size and allocated it optimally to each stratum.

Procedures used by the Texas Transportation Institute (TTI) are documented in developing the Houston-Galveston Nonattainment Counties Mobile Source Emissions Inventories for FY2007 (Dresser *et al.* 2000). The time-of-day VMT and speed estimates for the Houston-Galveston region were developed using the PREPIN2 program. PREPIN2 is one of a series of programs developed by TTI to facilitate the application of EPA's MOBILE5a Hybrid program in estimating mobile source emissions. The PREPIN2 program was developed for use in urban areas that do not have time-of-day assignment and speeds available for air quality analyses. The program inputs a 24-hour assignment and applies the needed seasonal adjustment factors. The time-of-day factors are applied to the seasonally adjusted 24-hour assignment results to estimate the directional time-of-day travel. A simplified version of the HGAC speed model was used to estimate the operational time-of-day speeds for intrazonal trips. These VMT and speeds by link are subsequently input to the IMPSUMA program for the application of MOBILE5a Hybrid emissions rates.

CHAPTER 3

DEVELOPMENT OF MODELS FOR FORECASTING VEHICLE AGE DISTRIBUTION

3.1 Age Distribution and Its Impact to the Emission Estimation of MOBILE

MOBILE's emission factor calculations rely in part on travel fractions for vehicles of each given age and type, which in turn are based on estimates of the registration distribution by age (age 0 - 1 to age 24 - 25^+) for each vehicle types, except motorcycles, for which registration distribution are only provided for the 12^{th} - and - later years of operation (age 0 - 1 to 11 - 12^+).

MOBILE6 users may specify vehicle registration data for 25 vehicle ages for one or more of the 16 composite vehicle types listed in TABLE 1.

TABLE 1 Composite Vehicle Classes for Vehicle Registration Data and Vehicles Miles Traveled Fractions (REG DIST and VMT FRACTIONS commands in MOBILE6)

Number	Abbreviation	Description
1	LDV	Light-Duty Vehicles (Passenger Cars)
2	LDT1	Light Duty Trucks 1 (0-6,000 lbs. GVWR, 3751-5750 lbs. LVW)
3	LDT2	Light Duty Trucks 2 (0-6,000 lbs. GVWR, 0-3750 lbs. LVW)
4	LDT3	Light Duty Trucks 3 (6,001-8,500 lbs. GVWR, 0-3750 lbs. LVW)
5	LDT4	Light Duty Trucks 4 (6,001-8,500 lbs. GVWR, 3751-5750 lbs.
6	HDV2B	Class 2b Heavy Duty Vehicles (8,501-10,000 lbs. GVWR)

7	HDB3	Class 3 Heavy Duty Vehicles (10,001-14,000 lbs. GVWR)
8	HDV4	Class 4 Heavy Duty Vehicles (14,001-16,000 lbs. GVWR)
9	HDV5	Class 5 Heavy Duty Vehicles (16,001-19,500 lbs. GVWR)
10	HDV6	Class 6 Heavy Duty Vehicles (19,501-26,000 lbs. GVWR)
11	HDV7	Class 7 Heavy Duty Vehicles (26,001-33,000 lbs. GVWR)
12	HDV8A	Class 8a Heavy Duty Vehicles (33,001-60,000 lbs. GVWR)
13	HDV8B	Class 8b Heavy Duty Vehicles (>60,000 lbs. GVWR)
14	HDBS	School Buses
15	HDBT	Transit and Urban Buses
16	MC	Motorcycles (All)

Note: This table is copied from Environmental Protection Agency (2001b), where LVW is *loaded vehicle weight* rating, and GVWR is *gross vehicle weight ratings*.

In the input file for MOBILE6, vehicle age fractions are represented by decimals (0.000 through 1.000) for each of the 25 model years and older in the fleet being modeled. MOBILE uses national average annual registration distributions by age, and has provisions allowing the input of alternate data. EPA provides an estimate of the number of vehicles of various ages in operation in the United States as of July 1, 1996 for eighteen GVWR-based vehicle categories, which are listed in TABLE 3-1. So the national annual registration distribution data are based on the analysis of calendar year 1996 registration. Using the default values assumes that the national distribution of vehicles registered by age is the same as the distribution in specific localities. Using national average default values to model specific areas would tend to produce inaccurate emission factors.

EPA encourages local areas to use their local age distributions estimating emission inventories (Cambridge Systematics Inc. and etc. 1996). In the real applications, someone uses the local vehicle registration data for a particular year as input into MOBILE. For example, in developing the Houston-Galveston nonattainment counties gridded mobile source emissions inventories for FY 2007, the 1999 vehicle registration data for the 8 counties were used to run MOBILE5a (Dresser, et al. 2000).

Vehicle age distribution has an important impact on the MOBILE emission factors (HC, CO and NO_x). As discussed in "Sensitivity Analysis of MOBILE6.0" (EPA, 2000,) emissions are affected by changes in the distribution of vehicle ages for a given year. This reflects the deterioration of emissions with vehicle age, which is the main assumption in MOBILE6 emissions calculations. Figure 1 and Figure 2 displays the percent change in CO emissions versus the percent change in the vehicle age fractions for all Vehicles (Source: EPA, 2002, p.20.)

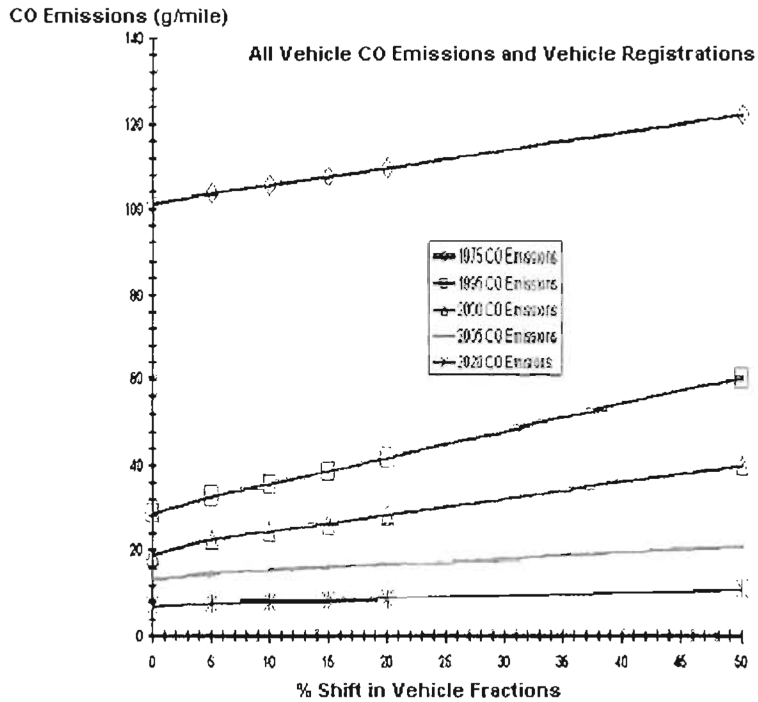


FIGURE 1 All Vehicle CO emissions as a function of the percent change in the fraction of registered vehicles with a given age.

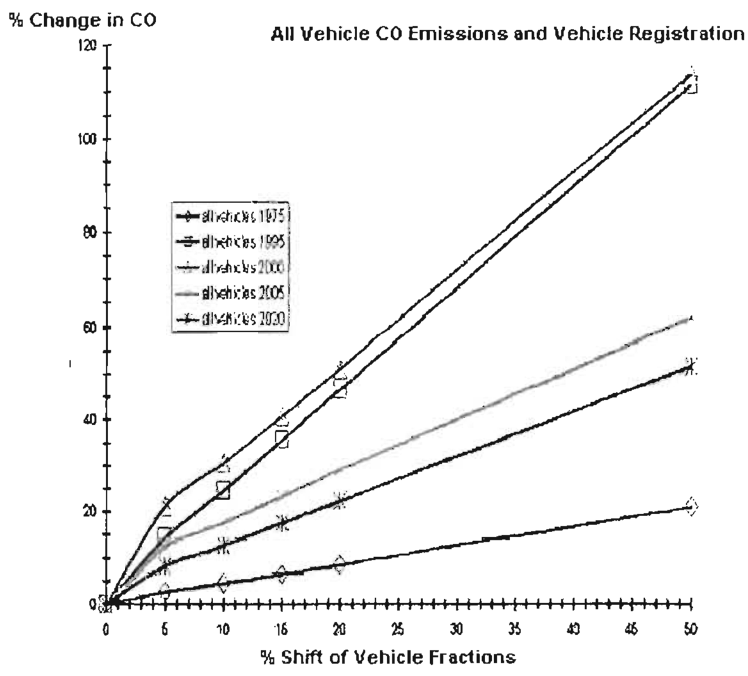


FIGURE 2 All Vehicle CO emissions as a function of the percent change in the fraction of registered vehicles with a given age.

In the above two figures, the percentage is determined relative to the MOBILE6 default registration and the emissions determined with those default vehicle age fractions. The relationships for other types of emissions are similar.

3.2 Model Design for Estimating Vehicle Age Distribution

Vehicle age distribution modeling system is an object in which variables of different kinds interact and produce observable signals (vehicle age distribution), which are usually called *outputs*. Figure 3 is the illustration of this system, where vehicle age distribution, as well as the absolute number of vehicles for a particular vehicle type with a particular age in a certain area, can be regarded as the function of some kinds of *inputs*. These input variables could be either the predictable socioeconomic factors, or the complex unpredictable or immeasurable inputs. The predictable socioeconomic indices may include population, average income, household, population density and etc. If the variables are unpredictable or immeasurable, the chronological series can be used as the input of the function.

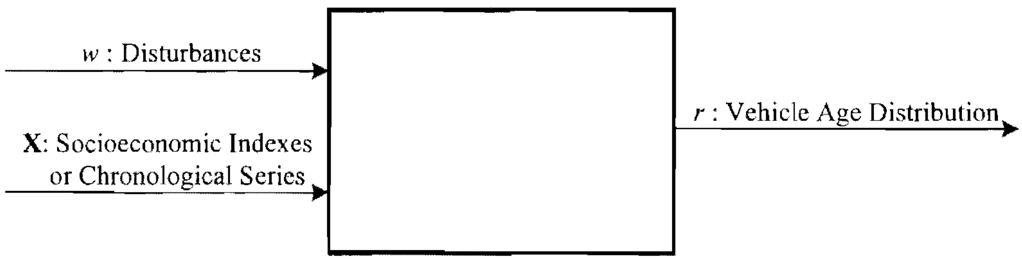


FIGURE 3 Vehicle age distribution modeling system.

According to the theory of system identification, we shall call the assumed relationship among observed input/output variables a *model* of the system (Ljung, 1999.) Sometimes, the model can be constructed from basic physical laws and other well-established relationships. However, a model set whose parameters are basically viewed as vehicles for adjusting the fit to the data and do not reflect physical considerations in the system is called a *black box* (Ljung, 1999.)

Vehicle age distribution is influenced by a lot of factors. The physical relationships among them are not easily identifiably. Therefore, it is natural to think of the *black box* as model set.

The constructed models should contain some parameters that need to be calibrated by the real world collected data. The calibration of parameters can be based on the algorithms like Least Square (Ljung 1999, Crooper and McGillem, 1999). The projection of the age distribution for the target year can be obtained when the input variables for the target year are supplied.

According to whether the age distribution is modeled directly, two types of models are developed for the projection of the future vehicle age distribution. Model Type I (**MT I**) models the number of vehicles for the particular vehicle type in particular age, and then transfers the results to project the future age distribution. Model Type II (**MT II**) models the future age distribution directly. The modeling processes are described next.

MT I:

Suppose v_{kg} is the number of vehicles with type k $(k = 1, 2, \ldots, n_k)$ and age g $(g = 1, 2, \ldots, n_g)$, n_k is the number of total vehicle types, and n_g is the maximum number of vehicle age. Let $\hat{v}_{kg}$ be the estimated value of v_{kg} by a certain model, and then the entire system objective can be represented as:

$$\min \sum_{k=1}^{n_k} \sum_{g=1}^{n_g} \left(v_{kg} - \hat{v}_{kg}\right)^2 \tag{3-1}$$

Since $\left(v_{kg} - \hat{v}_{kg}\right)^2 \geq 0$, so the system objective (3-1) can be decomposed into various sub-system objectives as:

$$\min\left(v_{kg} - \hat{v}_{kg}\right)^2 \qquad \forall k = 1, \ldots n_k, g = 1, \ldots, n_g \tag{3-2}$$

where, each v_{kg} is a function of the vector of inputs $\mathbf{x} = \{x_1, x_2, \ldots, x_{n_x}\}$ (n_x is the total number of inputs).

There are many factors that can affect vehicle age distribution, and the relationship between these factors and age distributions are very complex. Until now no one can build a physical model that can describe this kind of relationship. Since it is very difficult to build a model that can physically represent the relationships between the various inputs (in vector **x**) and the system output v_{kg}, it is reasonable to regard the system as a *black box*. In practical application, it may be necessary to use models that describe the relationships among the system variables in terms of mathematical expressions. From the theory of system identification, the mapping from the input vector **x** to the output v_{kg} can have the following parameterized function form:

$$v_{kg} = a_{kg}^0 + \sum_{i=1}^{n_f} a_{kg}^i f_{kg}^i\left(\mathbf{x} \middle| c_{kg}^{i0}, c_{kg}^{i1}, \ldots, c_{kg}^{in_f}\right) \tag{3-3}$$

where, the parametric matrix $\theta_{kg} = \left[a_{kg}^0, a_{kg}^1, \ldots, a_{kg}^{n_f}; c_{kg}^{10} c_{kg}^{11}, c_{kg}^{12}, \ldots, c_{kg}^{1n_f}; \ldots; c_{kg}^{n_f 0}, c_{kg}^{n_f 1}, c_{kg}^{n_f 2}, \ldots, c_{kg}^{n_f n_f}\right]$ is to be calibrated; f_{kg}^i is regarded as the basic function; and n_f is the total number of the basic functions f_{kg}. The basic function f_{kg}, however, can have different forms. The simplest basic function is the linear one that can be expressed as:

$$f_{kg}^i(\mathbf{x}) = x_i, \qquad \forall\, i = 1, 2, \ldots, n_f = n_x \tag{3-4}$$

which is a linear function of the scalar variable x_i. This kind of relationship is also illustrated in Figure 4(a).

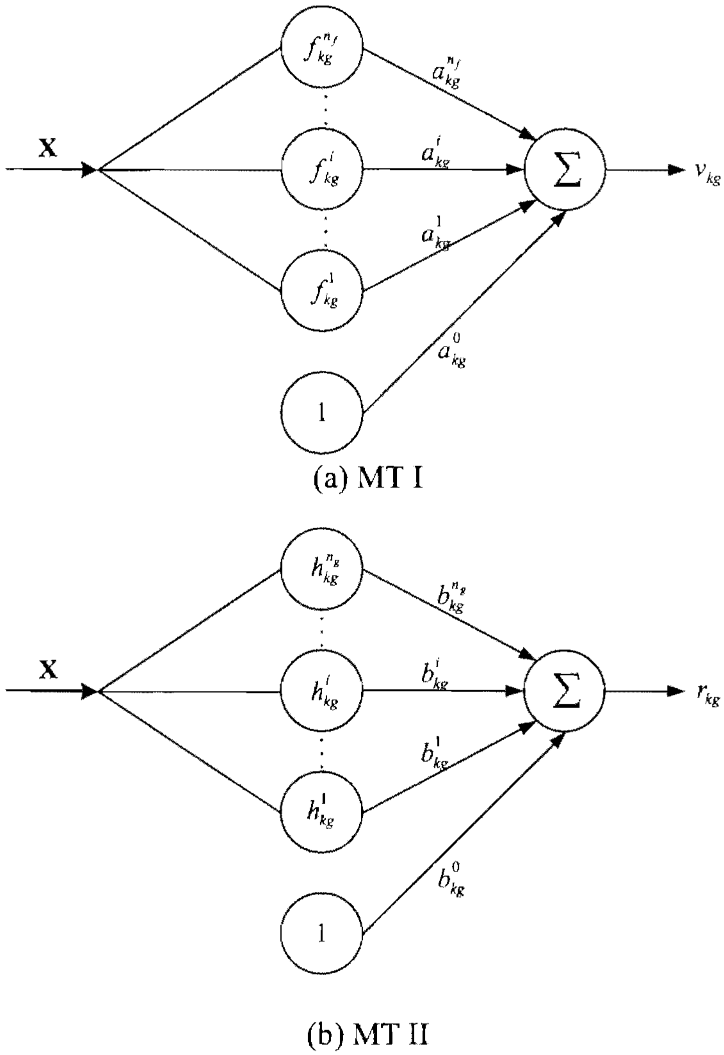


FIGURE 4 Mapping from input X to the output.

The structured model in (3-3) is parameterized with the parameter vector $\theta_{kg} = \left[a_{kg}^{0}, a_{kg}^{1}, \ldots, a_{kg}^{n_f}; c_{kg}^{01}, c_{kg}^{11}, \ldots, c_{kg}^{1n_f}; \ldots; c_{kg}^{n_f 0}, c_{kg}^{n_f 1}, \ldots, c_{kg}^{n_f n_f}\right]$. The search for the best model then becomes a problem of determining or estimating θ_{kg}. Our objective now is to determine a mapping from data sets $Z_v^N = \left(v_{kg}^N, \mathbf{x}^N\right)$ (where, N is the total number of recorded input-output pairs over a time period $1 \le t \le N$), to a series of possible parameters $\hat{\theta}_{kg} = \left[\hat{a}_{kg}^{0}, \hat{a}_{kg}^{1}, \ldots, \hat{a}_{kg}^{n_f}; \hat{c}_{kg}^{10}, \hat{c}_{kg}^{11}, \ldots, \hat{c}_{kg}^{1n_f}; \ldots; \hat{c}_{kg}^{n_f 0}, \hat{c}_{kg}^{n_f 1}, \ldots, \hat{c}_{kg}^{n_f n_f}\right]$, so that the model produces the prediction that is close to the target output. An obvious approach is then to select $\hat{\theta}_{kg}$ so as to fit the calculated values $\hat{v}_{kg}\left(t \middle| \hat{\theta}_{kg}\right)$ as well as possible to the measured inputs by least squares method.

So the *best* value of θ_{kg} is determined from the input-output data set by:

$$\hat{\theta}_{kg} = \arg\min \sum_{t=1}^{N} \left| v_{kg} - \left(\hat{a}_{kg}^{0} + \sum_{i=1}^{n_f} \hat{a}_{kg}^{i} f_{kg}^{i} \left(\mathbf{x} \middle| \hat{c}_{kg}^{i0}, \hat{c}_{kg}^{i1}, \ldots, \hat{c}_{kg}^{in_f} \right) \right) \right|^2 \tag{3-5}$$

The model output, which is the number of the vehicles for type k with age g, will be:

$$\hat{v}_{kg} = \hat{a}_{kg}^{0} + \sum_{i=1}^{n_f} \hat{a}_{kg}^{i} f_{kg}^{i} \left(\mathbf{x} \middle| \hat{c}_{kg}^{i0}, \ldots, \hat{c}_{kg}^{in_f} \right) \tag{3-6}$$

The age distribution r_{kg} can then be calculated by:

$$r_{kg} = \frac{v_{kg}}{\sum_{g=1}^{n_g} v_{kg}} \tag{3-7}$$

MT II:

Model Type II (**MT II**) models the future age distribution r_{kg} directly. Similarly, the entire system objective can be represented as:

$$\min \sum_{k=1}^{n_k} \sum_{g=1}^{n_g} \left(r_{kg} - \hat{r}_{kg} \right)^2 \tag{3-8}$$

$$s.t. \quad \sum_{g=1}^{n_g} r_{kg} = 1 \qquad \forall k = 1, 2, \ldots, n_k \tag{3-9}$$

In (3-8) and (3-9), r_{kg} is the age distribution for vehicle type k with age g, $\hat{r}_{kg}$ is the estimated value of r_{kg} by model, and n_k and n_g are the same as defined before. The constraint $\sum_{g=1}^{n_g} r_{kg} = 1$ is necessary here in order to ensure the sum of the age distribution for a particular vehicle type k is equal to 100%.

Since $\left(r_{kg} - \hat{r}_{kg} \right)^2 \geq 0$, the entire system objective can be decomposed into various sub-system objectives as:

$$\min \left(r_{kg} - \hat{r}_{kg} \right)^2 \qquad \forall k = 1, \ldots n_k, g = 1, \ldots, n_g \tag{3-10}$$

$$s.t. \quad \sum_{g=1}^{n_g} r_{kg} = 1 \qquad \forall k = 1, 2, \ldots, n_k \tag{3-11}$$

where, each r_{kg} is a function of the vector of inputs $\mathbf{x} = \{x_1, x_2, \ldots, x_{n_x}\}$ (n_x is the total number of inputs).

In the same way as for v_{kg}, the mapping from the input vector $\mathbf{x}$ to the output r_{kg} can have the following parameterized function form:

$$r_{kg} = b_{kg}^{0} + \sum_{i=1}^{n_g} b_{kg}^{i} h_{kg}^{i}\left(\mathbf{x} \middle| d_{kg}^{i0}, \ldots, d_{kg}^{in_g}\right) \tag{3-12}$$

where, $\phi_{kg} = \left\lfloor b_{kg}^{0}, b_{kg}^{1}, \ldots, b_{kg}^{n_f}; d_{kg}^{10}, d_{kg}^{11}, \ldots d_{kg}^{1n_g}; \ldots; d_{kg}^{n_g 0}, \ldots d_{kg}^{n_g n_g} \right\rfloor$ is the parametric matrix to be calibrated; h_{kg}^{i} is regarded as the basic function; and n_f is the total number of the basic functions h_{kg}. The basic function h_{kg}, however, can have different forms. The simplest one is the linear one that can be expressed as:

$$h_{kg}^{i}(\mathbf{x}) = x_i \qquad \forall\, i = 1,2,\ldots, n_f = n_x \tag{3-13}$$

which is a linear function of the scalar variable x_i. This kind of relationship is also illustrated in Figure 3-4(b).

The structured model in (3-12) is parameterized with the vector $\phi_{kg} = \left\lfloor b_{kg}^{0}, b_{kg}^{1}, \ldots, b_{kg}^{n_f}; d_{kg}^{10}, d_{kg}^{11}, \ldots d_{kg}^{1n_g}; \ldots; d_{kg}^{n_g 0}, \ldots d_{kg}^{n_g n_g} \right\rfloor$. The search for the best model then becomes a problem of determining or estimating ϕ_{kg}. Our objective now is to determine a mapping from data sets $Z_r^N = \left(r_{kg}^N, \mathbf{x}^N\right)$ (where, N is the total number of recorded input-output pairs over a time period $1 \le t \le N$), to a series of possible parameters $\hat{\phi}_{kg} = \left[\hat{b}_{kg}^{0}, \hat{b}_{kg}^{1}, \ldots, \hat{b}_{kg}^{n_f}; \hat{d}_{kg}^{10}, \hat{d}_{kg}^{11}, \ldots \hat{d}_{kg}^{1n_g}; \ldots; \hat{d}_{kg}^{n_g 0}, \ldots \hat{d}_{kg}^{n_g n_g} \right]$, so that the network produces the prediction that is close to the target output. One of the measurements of closeness may be on a mean square error criterion.

So, the *best* value of θ_{kg} is determined from the data input-output set by:

$$\hat{\phi}_{kg} = \arg\min \sum_{t=1}^{N} \left| v_{kg} - \left(\hat{b}_{kg}^{0} + \sum_{i=1}^{n_f} \hat{b}_{kg}^{i} h_{kg}^{i}\left(\mathbf{x} \middle| \hat{d}_{kg}^{i1}, \ldots, \hat{d}_{kg}^{in_g}\right)\right)\right|^2 \tag{3-14}$$

The model output, that is vehicle age distribution for type k with age g (with no constraint (3-11)), will be:

$$r_{kg}^{e} = \hat{b}_{kg}^{0} + \sum_{i=1}^{n_f} \hat{b}_{kg}^{i} h_{kg}^{i}\left(\mathbf{x} \middle| \hat{d}_{kg}^{i1}, \ldots, \hat{d}_{kg}^{in_g}\right) \tag{3-15}$$

To meet the constraint (3-11), the resulting age distribution r_{kg} can be calculated by:

$$r_{kg} = \frac{r_{kg}^{e}}{\sum_{g=1}^{n_g} r_{kg}^{e}} \tag{3-16}$$

3.3 Model Implementation for Estimating Vehicle Age Distribution

The whole process of the projection of vehicle age distribution includes the calibration of parameters for each model; the examination of the significance test for input indices (if the input indices are predictable); the choice of model types and structures; and the projection of vehicle age for the target year.

Possible data needed for modeling and projection include the socioeconomic indices in the corresponding area in the past years; the age distribution or number of vehicles for all kinds of vehicle types at different vehicle age; socioeconomic indices in the past years for projection; and other necessary background information and user defined requirements.

The socio-economic data may include: population (total; for different age groups…); number of employees, incomes and production of industries (total; agricultural services; construction; manufacturing; transportation and public utilities; and etc.); or even the price of oils and etc. A specific jurisdiction can input as many as the possible socio-economic data they may have. The software MOFAD have the ability to select several most suitable ones to build the model.

Parameter calibration is implemented by the linear square regression approach. The calibration of the parameters includes parameter estimations and interval estimations. The significance test for each index can be conducted by using the result of corresponding parameter estimation and interval estimation. The suitable model type and structure is determined such that the final model meets the requirements of the objective functions in (3-1) and (3-8). The projection of age distribution for the target year can be obtained if all the input socioeconomic indices for the target year are available.

As shown in figure 5, the FORTRAN program with the name MOFAD (**MO**deling and **F**orecasting **A**ge **D**istribution) implements the whole modeling process. A Graphic User Interface (GUI) is developed in Visual Basic language to provide a user friendly platform. Under the GUI shelter, several main specific functional modules are designed to accomplish main functions such as data inputting, model calibration, forecasting, and results outputting. Figure 6 shows the organization map of the program. Three main subroutines include Datain module, MT module, Forecast module. Each module is specified to accomplish a subtask.

The functions of three main modules have been summarized as follows:

Datain Module: is the module used to import all the necessary data into the program, including the socioeconomic index data, vehicle age distribution data, and the control parameters. This module can create an output file called outin.txt, which contains all the input data. This file can be used as a measure of checking if the data have been correctly inputted into the program.

MT Module: is the central module of the program that controls the model selection and calibration work. This module is subdivided into two model types according to the dependent variable format. Each mode type includes five candidate models. Two output files can be created in this module, outde.txt and outsu.txt, which output the detailed and summarized statistical results of model calibration.

Forecast Module: is the module to forecast future year's vehicle age distribution based upon calibrated model. One output file called outst.txt will be created in this module, which contains the estimated vehicle age distribution and can be inputted into MOBILE6 directly..

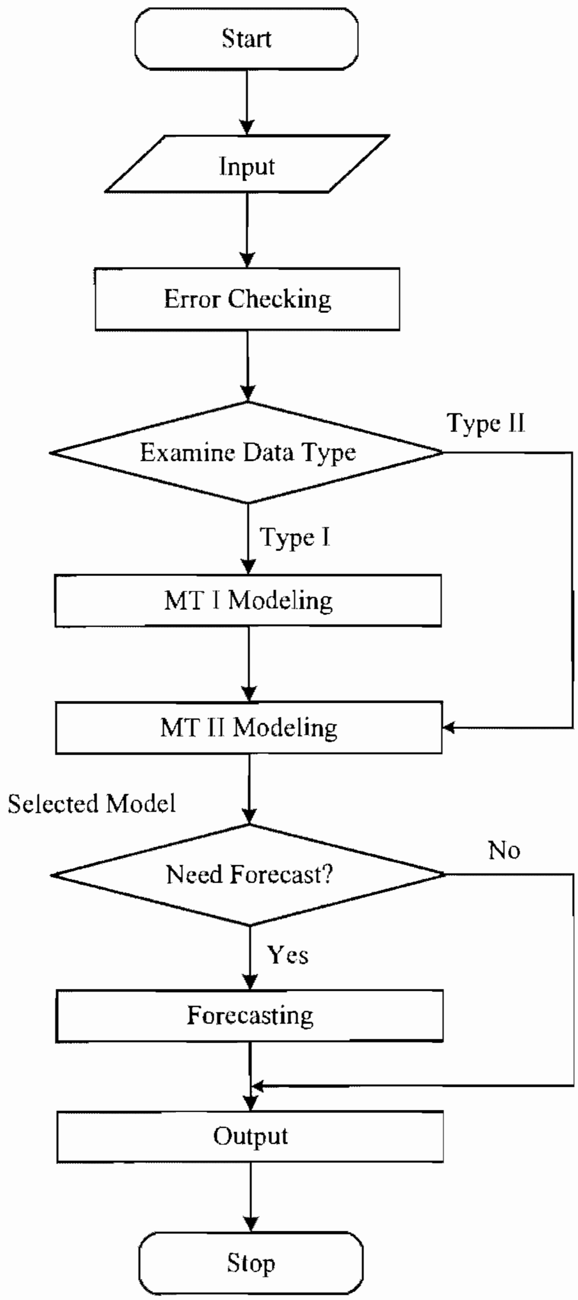


FIGURE 5 Flowchart of the program MOFAD.

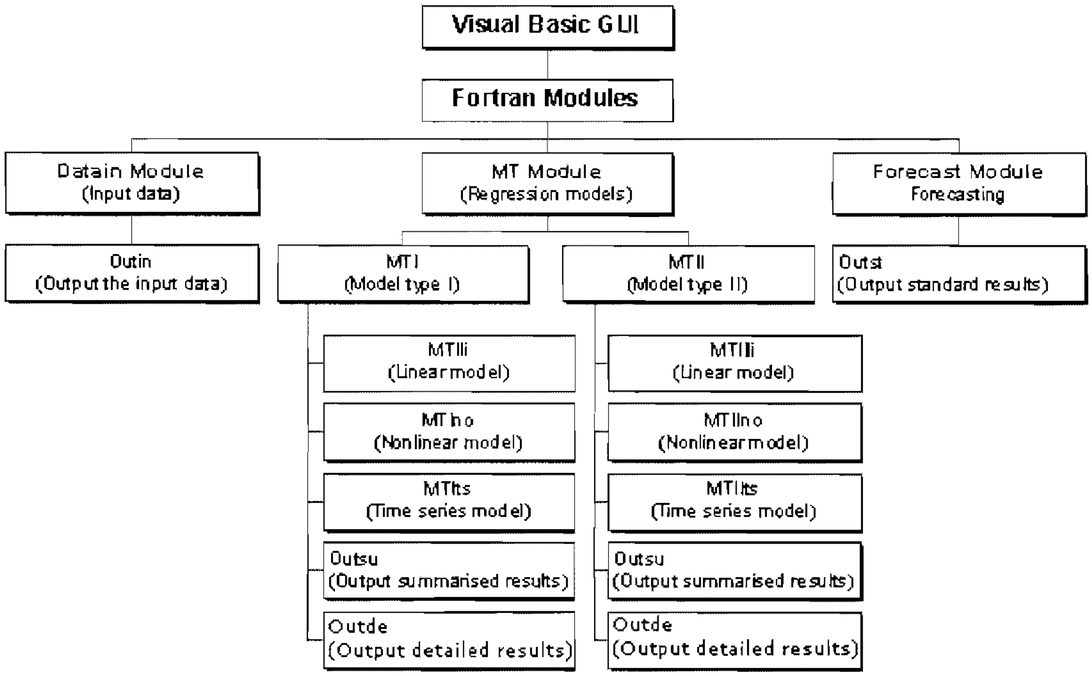


FIGURE 6 Organization map for the subroutines of program MOFAD.

The program can generate 4 types of output files that will meet the various needs of the users. It can provide the detailed modeling information in one of the output file, and give the summarized output in another file. It can also produce the standard output files that can be directly used as one of the input file for MOBILE. The model software MOFAD is during the finalizing stage and will be available for the users soon.

3.4 Model Validation

As described above, MOBILE6 users may specify vehicle registration data at 25 vehicle ages for one or more of the 16 composite vehicle types. In MOFAD, the number of vehicle for every 16 vehicle types at every 25 vehicle ages will be estimated separately. As a result, there are up to 400 models that will be calibrated for each specific county. So, it is difficult to validate a particular single model out of the total 400 models by some certain statistical criteria. Model validation here is more likely be called model verification or confirmation (Flavelle, 1992). More general form of output such as sum and average will be used to test the model group for each vehicle type or vehicle age other than specific model or data. On the other hand, it is valuable to check the whole process of model establishment process to identify how well the model groups perform.

3.4.1 Data Used for Model Validation

To validate the proposed MOFAD model, real data was collected in El Paso and eight counties in HGAC (Houston-Galveston Area Council): Brazoria, Chambers, Fort Bend, Galveston, Harris, Liberty, Montgomery and Waller (Figure 7).

Socioeconomic data were obtained from the website of government information sharing project, the US census bureau, and Bureau of Economic Analysis. Age distribution data were obtained from the Texas Department of Transportation and HGAC, which contain vehicle age registration information from 1994 to 2000, and were used for model calibrations and selections.

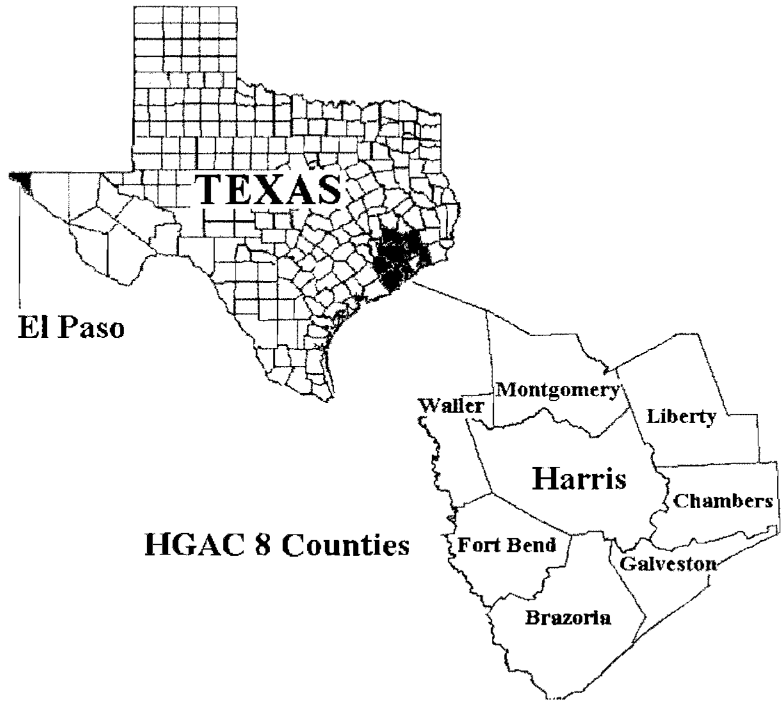


FIGURE 7 Map of the 8 counties in HGAC and El Paso.

Overall 22 socioeconomic indices for each of 8 counties in HGAC and 16 socioeconomic indices for El Paso have been collected. See Table 2 and Table 3 for detailed category of all 22 and 16 socioeconomic items.

TABLE 2 Category of All 22 Socioeconomic Items for HGAC Counties

Item number	Description	Item number	Description
1	Total population	12	Non farm employment
2	Population under age 20	13	Wage salary
3	Population between age 20-65	14	Personal income
4	Population above age 65	15	per capital personal income
5	Construction employment	16	Transportation employment
6	Ag service employment	17	Total employment
7	Farm employment	18	Retail employment

8	Fire employment	**19**	Private employment
9	Government employment	**20**	Whole sale employment
10	Manufactory employment	**21**	Service employment
11	Mining employment	**22**	Oil production

TABLE 3 Category of All 16 Socioeconomic Items for El Paso

Item number	Description	Item number	Description
[1]	Total population	**[9]**	Manufactory employment
[2]	Population under age 20	**[10]**	Mining employment
[3]	Population between age 20-65	**[11]**	Non farm employment
[4]	Population above age 65	**[12]**	Transportation employment
[5]	Construction employment	**[13]**	Total employment
[6]	Agriculture employment	**[14]**	Retail employment
[7]	Fire employment	**[15]**	Whole sale employment
[8]	Government employment	**[16]**	Service employment

3.4.2 Testing Linear, Nonlinear and Time Series Models

For each model, two model types (MTI and MTII) were prepared and five kinds of candidate models were tested. The five candidate models included one linear regression model, three nonlinear models and one pure time series model. The linear regression model has been described in (3-4) and (3-13), while for the time series model the input was the chronological series (*i.e.* the sequence of year) instead of the socioeconomic indices. The three nonlinear models chosen here were all log-linear models listed in the following:

For MT I:
$$v_{kg} = \exp\left(c_{kg}^0 + c_{kg}^1 \log x_1 + \ldots + c_{kg}^n \log x_n\right) \tag{3-17}$$

$$v_{kg} = \exp\left(c_{kg}^0 + c_{kg}^1 x_1 + \ldots + c_{kg}^n x_n\right) \tag{3-18}$$

$$v_{kg} = c_{kg}^0 + c_{kg}^1 \log x_1 + \ldots + c_{kg}^n \log x_n \tag{3-19}$$

For MT II:
$$r_{kg} = \exp\left(d_{kg}^0 + d_{kg}^1 \log x_1 + \ldots + d_{kg}^n \log x_n\right) \tag{3-20}$$

$$r_{kg} = \exp\left(d_{kg}^{0} + d_{kg}^{1} x_1 + \ldots + d_{kg}^{n} x_n\right) \tag{3-21}$$

$$r_{kg} = d_{kg}^{0} + d_{kg}^{1} \log x_1 + \ldots + d_{kg}^{n} \log x_n \tag{3-22}$$

Therefore in running the program for each county, a total of 4000 candidate models (=25 ages * 16 vehicle types * 2 model types * 5 linear or nonlinear models) were to be prepared. The selected model from the 4000 candidate was the one that can meet the requirement of the objective functions (3-2) and (3-10), *i.e.* the one that had the minimum modeling errors.

Table 4 shows the number of different models used in modeling age distribution for 8 HGAC counties. From Table 4 it is shown that for MT I, the selected models came from different five model families (linear model, 3 types of nonlinear models and time series models). Most of them were taken from the linear model and the third nonlinear model (3-19). Only under a few cases the best model for MT I were taken from time series model. For MT II, the results are very interesting. All the selected models were taken from the time series model and the third nonlinear model (3-22), and none were taken from the first nonlinear model (3-20) and the second nonlinear model (3-21). TABLE 5 lists the number of models taken from MT I and taken from MT II. About 41.4% of the final models were taken from MT I, while 58.6% taken from MT II.

TABLE 4 Number of Different Models Used in Modeling Age Distribution for 8 HGAC Counties

	Linear	Nonlinear1	Nonlinear2	Nonlinear3	Time Series	Total
	MT I					
Brazoria	112	69	38	181	0	400
Chambers	15	36	81	268	0	400
Fort Bend	202	30	57	108	3	400
Galveston	158	28	57	156	1	400
Harris	160	52	64	124	0	400
Liberty	153	18	80	149	0	400
Montgomery	264	39	17	80	0	400
Waller	166	37	50	147	0	400
	MT II					
Brazoria	248	0	0	152	0	400
Chambers	225	0	0	175	0	400
Fort Bend	254	0	0	146	0	400
Galveston	182	0	0	215	3	400
Harris	301	0	0	99	0	400
Liberty	238	0	0	161	1	400
Montgomery	272	0	0	128	0	400
Waller	257	0	0	143	0	400

TABLE 5 Number and Percentage of Selected Models from MT I and MT II

	MT I		MT II	
Brazoria	150	37.5%	250	62.5%
Chambers	100	25.0%	300	75.0%
Fort Bend	154	38.5%	246	61.5%
Galveston	186	46.5%	214	53.5%
Harris	164	41.0%	236	59.0%
Liberty	213	53.3%	187	46.8%
Montgomery	178	44.5%	222	55.5%
Waller	179	44.8%	221	55.3%
Average	165.5	41.4%	234.5	58.6%

3.4.3 Goodness-of-Fit with Historical Data

It would be of value to compare the improved method with the other practices. However a simple comparison between the results from the improved method and other practices has no direct meaning. Theoretically speaking, the improved methods considered the impacts of socio-economic factors and provides a family of candidate functions to select, therefore it should result in better estimates.

On the other hand, a comparison of the actual distribution themselves would be valuable. This can be observed from the following discussions on the goodness-of-fit with historical data.

One of the critical criteria of model validation is "the extent to which predictions agree with observations, which can vary from perfect equality (accurate or unbiased) to perfect inequality (inaccurate or biased)". As previously discussed, the model calibration was conducted using the data for El Paso City and the 8 counties in HGAC: Brazoria, Chambers, Fort Bend, Galveston, Harris, Liberty, Montgomery and Waller. It is important to check the fitness of estimated data with real ones for the above city and counties before forecasting. After calibrating the model, data for each historical year from 1994 to 2000 is input into the calibrated model. The estimated age distributions from 1994 to 2000 are compared with the real ones. The relative mean absolute error of fitness has been employed here as the criteria which is shown below:

$$OF = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{\hat{r}_{kg} - r_{kg}}{r_{kg}} \right| \tag{3-23}$$

where OF is the relative error of fitness, $\hat{r}_{kg}$ is the estimated age distribution with type k $(k = 1,2,...,n_k)$ and age g $(g = 1,2,...,n_g)$, n_k is the number of total vehicle types, n_g is the maximum number of vehicle age, and r_{kg} is the historical data.

Figure 8 shows the fitness with historical data for Harris County. All the relative errors for year 1994 to 2000 are within 16%. 1996 and 1997 have the strongest fitness with the surveyed data (7% and 4%).

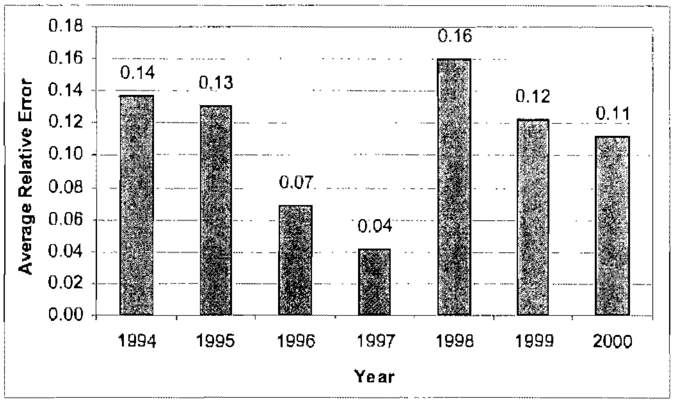


FIGURE 8 Relative error of fitness for Harris County.

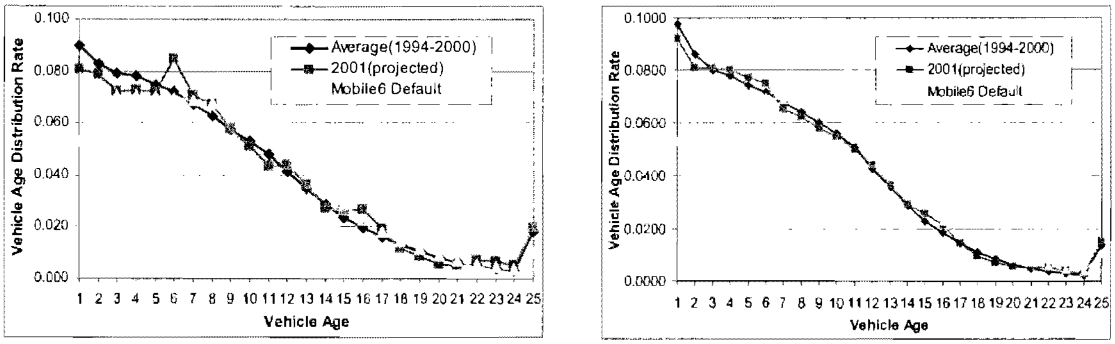
Similar analysis has been implemented for all 8 counties and El Paso City. The average errors for nine areas for the year 1994-2000 have been summarized in Table 6. From Table 6 we can see that the average relative errors for almost all the counties are within 15%, except for Chambers County where the average relative errors are 20%.

TABLE 6 Relative Mean Absolute Error of Fitness for 8 Counties of Houston and El Paso

	1994	1995	1996	1997	1998	1999	2000	Average
Brazoria	0.07	0.16	0.11	0.07	0.11	0.05	0.38	0.14
Chambers	0.20	0.09	0.23	0.25	0.20	0.22	0.21	0.20
Fort Bend	0.21	0.10	0.07	0.06	0.13	0.08	0.10	0.11
Galveston	0.14	0.18	0.13	0.12	0.18	0.13	0.18	0.15
Harris	0.14	0.13	0.07	0.04	0.16	0.12	0.11	0.11
Liberty	0.09	0.07	0.07	0.15	0.13	0.16	0.17	0.12
Montgomery	0.18	0.11	0.12	0.12	0.08	0.17	0.20	0.14
Waller	0.11	0.15	0.15	0.14	0.11	0.06	0.18	0.13
El Paso	0.14	0.19	0.19	0.09	0.14	0.11	0.09	0.14

3.4.4 Predictive Validation

As mentioned above, it is found that MOFAD can perform reliable fit with historical data. The next step attempts to perform the 2001 foresting using the calibrated models. 2001 socioeconomic indices are forecasted and input into the calibrated models to forecast 2001 age distribution. The result is compared with the average historical (1994-2000) and the MOBILE6 default one. The average historical (1994-2000) data came from TxDOT and defined as the average of the vehicle age distributions from year 1994 through 2000. Figure 9 (a, b) presents the result for vehicle type LDV for Harris and Brazoria Counties.



(a) Brazoria **(b) Harris**

FIGURE 9 LDV for Brazoria and Harris County: vehicle age distribution of average (1994-2000), 2001 (Projected) and MOBILE6 default value.

Three curves appear in each of the two figures representing the average historical value, 2001 projected value and the default value in MOBILE6, respectively. Results show that MOBILE6 default values are smaller than local registration rates for younger vehicle age (age < 7), and become bigger for middle age vehicle (age between 8 and 16). The default ones get closer to the local ones for elder vehicles (age > 16). It is interesting that obvious difference exists for one-year old vehicle. MOBILE6 default percentage of vehicles at one year old is smaller than that at nearby vehicle ages while the local one is bigger. Since the vehicle sales year begins in October, MOBILE6 multiplies the estimated age 1 population by 0.75 to account for the fact that approximately 75% of the year's sales will have occurred by July 1^{st} of a given calendar year. It may be the reason why the default value for one-year-old vehicle is smaller than local one.

This result shows that it is important and necessary to collect localized age distribution by vehicle age for 16 composite vehicle types because MOBILE6 default values cannot reflect real-world characteristics of local environment.

3.4.5 Sensitivity Analysis

The objective of the sensitivity analysis is to test the sensitivity for each independent variable to the final age distribution forecasting results. One possible approach of this kind of sensitivity analysis is to give several percentage changes to independent variables, and observe the change of the results obtained from the model. On the other hand, the selection of proper inputs is important to make the model more practical and easier to be implemented.

Sensitivity analyses for 8 HGAC counties and El Paso by changing all 2001 socioeconomic data are implemented. Figure 10 shows the Sum Square of Error (SSE) of all model outputs for each HGAC county with percent changes of 2001 socioeconomic indices ranging from -20% to 20% with an increment of 5%. For most counties, with the increasing changes of independent variables, the resulting SSEs will also change. This means that the model is sensitive to the changes of input variables in most cases.

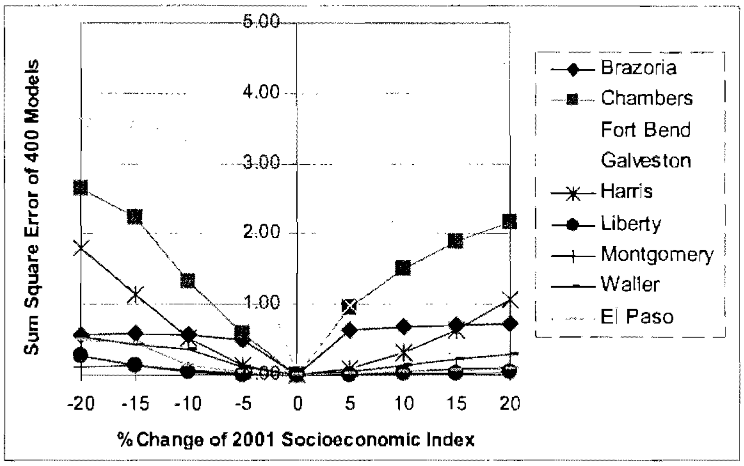


FIGURE 10 SSE changing with % change of 2001 socioeconomic indices for all counties.

The inputs of MOFAD are the socioeconomic indices. However, it is often difficult to collect and forecast all the socioeconomic indices because of the limitation of time and expenses. So it is important to identify which socioeconomic indices play more important roles. This can give users guidance on which indices should be collected.

For our application in HGAC counties and El Paso, a total of 4 independent variables are employed in the final model calibration for each of the overall 400 models which are selected from overall 22 (for HGAC) or 16 socioeconomic indices (for El Paso). As shown in Table 7, all the four most frequently used socioeconomic indices for 8 HGAC counties and El Paso are summarized. The detailed description on which socioeconomic indices each number in Table 7 represent are given in Appendixes B and C.

TABLE 7 Most Frequently Used Four Socioeconomic Items for HGAC Counties and El Paso

County Name	Percent among 400 Models (%)	4 Frequently Selected Socioeconomic indices			
Brazoria	85	6	7	10	19
Chambers	56	8	11	12	19
Fort Bend	58	7	13	20	21
Harris	92	7	8	11	22
Galveston	92	2	10	12	22
Liberty	100	8	14	16	21
Montgomery	76	7	11	14	18
Waller	78.5	9	13	14	21
El Paso	85	[6]	[7]	[10]	[14]

Note: 6 and [6] represent different socioeconomic indices; details are in Table 2 and Table 3.

Due to the fact that high frequently used variables play more important roles to affect the overall age distribution of certain vehicle type, the analysis has been focused on those four most frequently used items for each county. Table 8 shows the average % change of the number of vehicles for all 16 vehicle types at 25 vehicle ages in Harris County when there are 2% changes of input items. In Table 8, MOFAD is very sensitive to the first three items (7, 8, 11) for Harris County, where the percentage changes range from 3.7% to 21.2%, all of which are great than 2%. Figure 11 illustrates the age distribution curves comparisons for LDV of Harris County when changing item 7 (Farm employment, see Appendices B and C for details) of 2001. It is interesting that the curve for increasing and the curve for decreasing are symmetric to the default curve. Therefore, the final age distribution curves are sensitive to the most frequently used socioeconomic items.

TABLE 8 Sensitivity Analyses for All Four Frequently Used Variables of Harris County

Item	% Change of item	% change of vehicle age distribution
7	2	0.209
8		0.037
11		0.090
22		0.011
7	-2	0.212
8		0.037
11		0.089
22		0.011

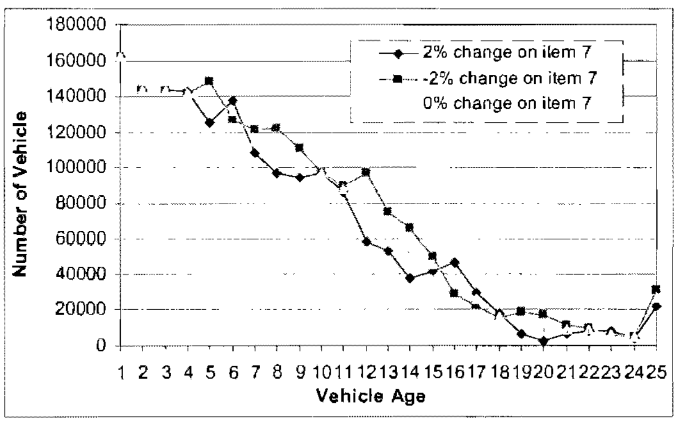


FIGURE 11 Sensitivity analyses with % change of item 7 in Harris County for LDV.

3.4.6 Suboptimal Selection of Independent Variables

A suboptimal selection of independent variables refers to the situation when less number of socioeconomic data than the ideal situation is selected for inputs to MOFAD. The reason for suboptimal selection is that it is often impossible to collect enough socioeconomic data that are all predictable.

In our case study, three more generally and easily forecasted socioeconomic indices: total population, total employment, and personal income are considered as a suboptimal selection. The comparison of effects between the suboptimal selection and the optimal selection is conducted below.

Figure 12 shows one of the results of vehicle type LDV for Brazoria County. It is shown that the suboptimal selection shows smoother curve and much more closer to the original one than the default one. This result also implies that it is possible to get reliable estimation even using less socioeconomic data.

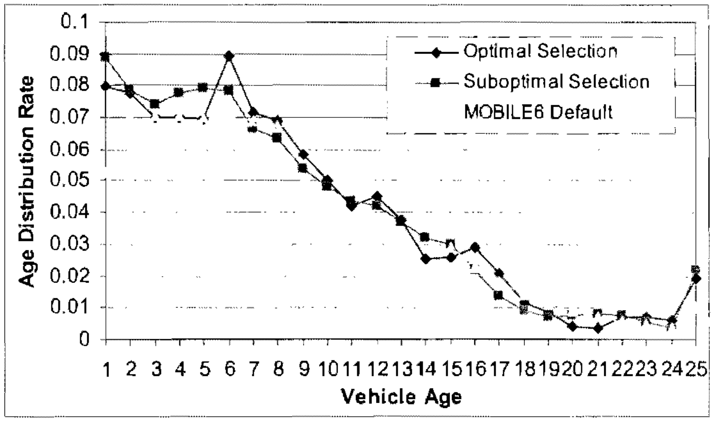


FIGURE 12 2002 estimated age distribution of LDV for Brazoria County.

3.4.7 Analysis of Forecasting Process

When using regression model, it is assumed that the curve derived from regression analysis based upon historical data can represent the trend of dependent variable, so the future year's dependent data can be estimated based on this trend curve. But when there are not enough historical data, this type of forecasting method is probably lack of precision.

For example, in our case study, age distribution models are calibrated using 7 years' historical data for each county. Two different forecasting methods can be applied. The first method is to calculate the future year's age distribution by inputting the target year's socioeconomic indices into the same calibrated model. Due to the limitation of historical data, long-term age distribution forecasting is unlikely to be achieved precisely. The second method is called step-by-step forecasting. In this method, 2001 year's age distribution can be forecasted using the calibrated model derived from 7 years historical data. Then the new 8 years "historical" data that includes 2001 estimated result could be applied to calibrate parameters for new models and forecast results for 2002 year. This procedure can be repeated several times to reach the estimates for mid-term years.

To illustrate the whole procedure, results between these two forecasting methods are compared next. A total of 4-years' age distributions (2001 – 2004) have been estimated. Same comparisons are performed for both the optimal selection and the suboptimal selection.

Figure 13 shows one of the forecasted results of vehicle type LDV for Brazoria County. It is found that even though the second forecasting method takes more complex procedure and is anticipated better confidential result theoretically, there are slight differences between these two methods for suboptimal selection, where only 3 predefined socioeconomic independent indices are used. For both optimal and suboptimal selection of inputs, results of the first method are smaller than those of the second method.

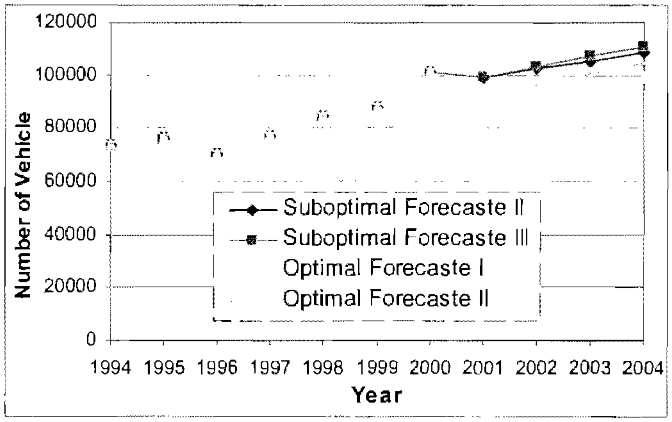


FIGURE 13 Forecasted age distribution of LDV for Brazoria County.

3.5 Impacts to MOBILE6 Results

3.5.1 Estimates of 2001 Emission Factors for 8 HGAC Counties and El Paso

Vehicle age distribution for the 8 counties in HGAC for the year 2001 has been estimated and the forecasted results are input to the MOBILE6. Figure 14 presents the produced three emission factors (VOC, CO and NO_x) by the default age distributions and by the forecasted local ones from MOFAD models for 8 HGAC counties in the year 2001. The x-axis in this figure is purely a categorical classification, while y-axis stands for the emission factors for all vehicles. From the results we can see that there are differences between the two sets of the results, especially for CO. In most of the 8 HGAC counties, the emission factors (especially for CO) are smaller than the ones that are generated by the default age distributions. The only exception is the Chambers County, where the local emission factors are slightly larger than default ones.

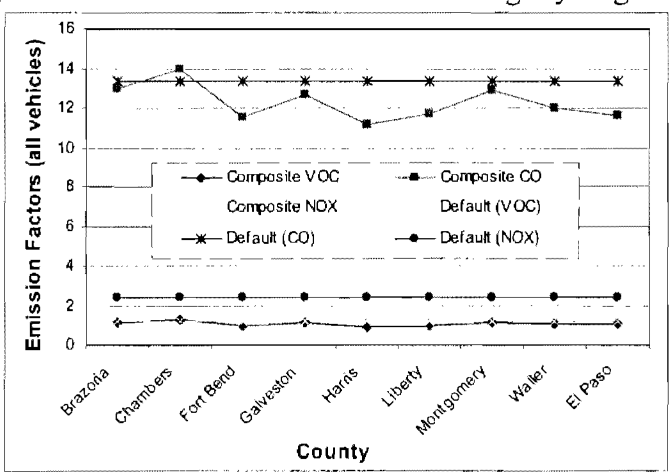
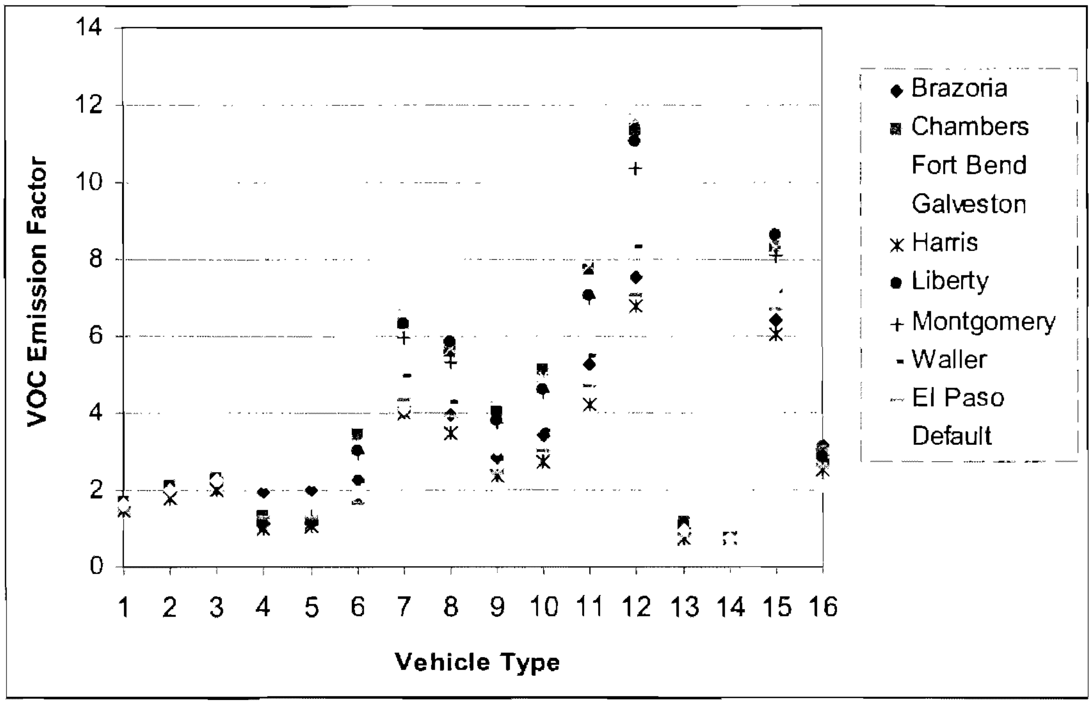
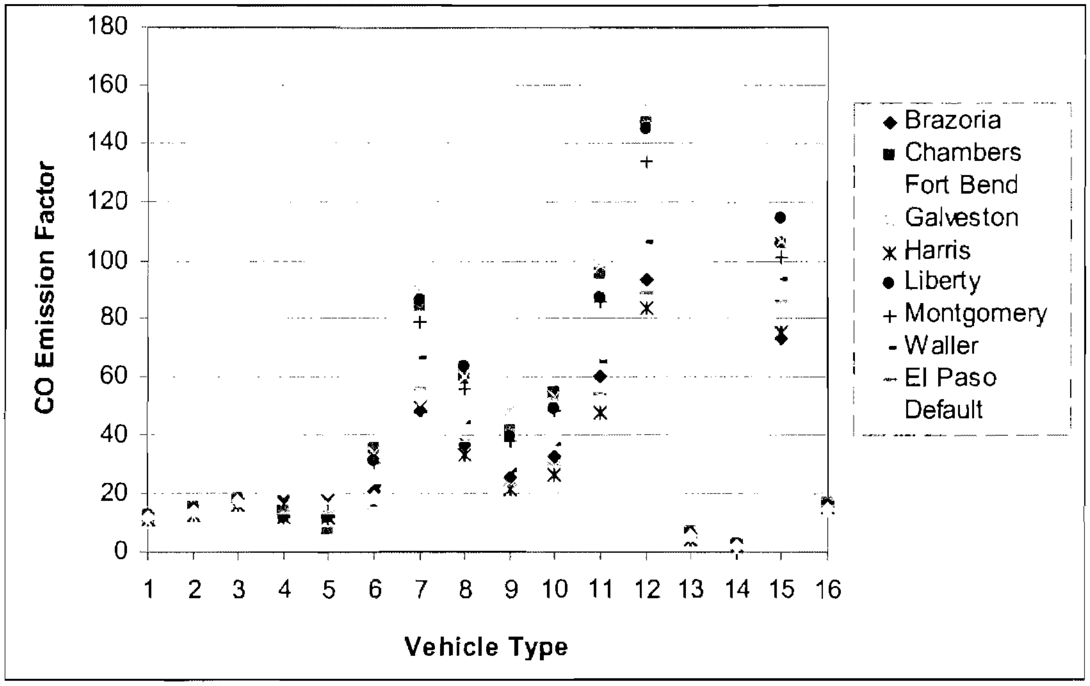


FIGURE 14 Comparison of emission factors by default and forecasted vehicle age distributions for 8 HGAC counties and El Paso in the year 2001.

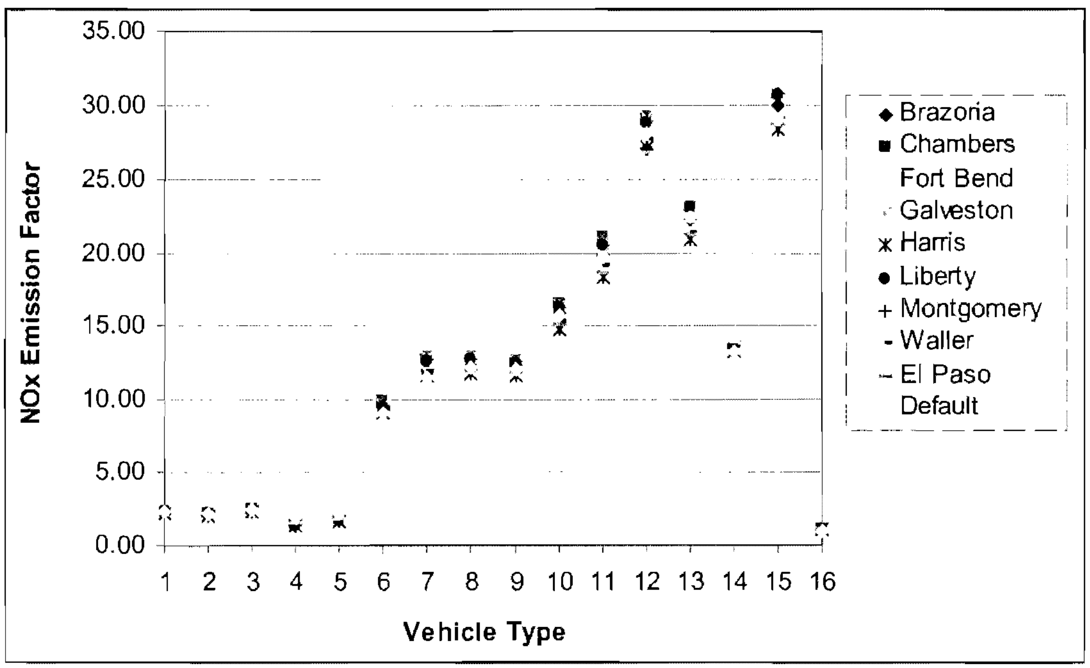
Further emission factors for all 16 composite vehicle types are calculated using 2001 projected age distribution data for 8 HGAC Counties and El Paso. Figure 15 (a, b, c) gives all the results of three emission factors, Composite VOC, Composite CO and Composite NO_x for all the selected areas. In addition, MOBILE6 default value is added to the figures for the comparison purpose. The results show that there are obvious differences on Composite VOC and CO for most of the vehicle types when compared with MOBILE6 default one. For the first three vehicle types, the differences of emission factors are smaller than other vehicle types. For NO_x emission pollutant, the differences for all vehicle types are not as obvious as the other two factors.



(a) VOC



(b) CO



(c) NO_x

FIGURE 15 Comparison of three emission factors by default and forecasted vehicle age distributions for 8 HGAC counties and El Paso in the year 2001.

3.5.2 Impacts of Emission Factors for Suboptimal Selection and Forecasting Process

As discussed earlier, vehicle age distributions from 2001 to 2004 for Brazoria County have been estimated using two different forecasting procedures for both optimal and suboptimal selections of inputs. To analyze the impacts to emission factors for suboptimal selection and forecasting process, all the resulted age distributions are input into MOBILE6. Table 9 shows the comparison results of emission factors from 2001 to 2004 for Brazoria County in terms of different forecasting results of age distributions. The average relative errors of emission factors between optimal and suboptimal selection of socioeconomic indices are only 1.31%; while the average relative errors of emission factors between forecasting method I and II are -1.52%. It is found that little differences among emission factors for all the four selections. Again this proves that the suboptimal selection of input variables is feasible.

TABLE 9 Emission Factors Comparison for Harris County Using Different Age Distribution

Year	Age Distribution Estimation Type	Emission Factors (g/mi)		
		Composite VOC	Composite CO	Composite NO_x
2001	Suboptimal I	0.909 (17%)	10.859 (12%)	2.248 (2%)
	Suboptimal II	0.909 (17%)	10.859 (12%)	2.248 (2%)
	Optimal I	0.940 (14%)	11.171 (9%)	2.241 (3%)
	Optimal II	0.940 (14%)	11.171 (9%)	2.241 (3%)
	MOBILE6 Default	1.097	12.287	2.300
2002	Suboptimal I	0.834 (18%)	10.339 (11%)	2.091 (3%)
	Suboptimal II	0.834 (18%)	10.343 (11%)	2.089 (3%)
	Optimal I	0.858 (15%)	10.566 (9%)	2.116 (2%)
	Optimal II	0.834 (18%)	10.343 (11%)	2.089 (3%)
	MOBILE6 Default	1.015	11.639	2.150
2003	Suboptimal I	0.768 (18%)	9.673 (11%)	1.889 (4%)
	Suboptimal II	0.768 (18%)	9.675 (11%)	1.889 (4%)
	Optimal I	0.785 (16%)	9.851 (9%)	1.939 (2%)
	Optimal II	0.768 (18%)	9.675 (11%)	1.889 (4%)
	MOBILE6 Default	0.937	10.845	1.974
2004	Suboptimal I	0.661 (9%)	7.994 (12%)	1.662 (5%)
	Suboptimal II	0.66 (10%)	8.002 (12%)	1.664 (5%)
	Optimal I	0.671 (8%)	8.097 (11%)	1.722 (2%)
	Optimal II	0.660 (10%)	8.002 (12%)	1.664 (5%)
	MOBILE6 Default	0.820	9.087	1.755

Note: The number in bracket represents relative error compared with MOBILE6 default.

3.6 Summary on Selection of Independent Variables

Socio-economic developments are fundamental factors of transportation characteristics which include vehicle age distribution. Some vehicle age distribution models use variables like vehicle ownership, new vehicle registrations, etc. These variables are eventually influenced by socio-economic factors and are secondary level variables.

MOFAD uses socio-economic indices directly as the independent variables. The socio-economic indices that are listed in Table 2 and Table 3 may not be fully suitable to all the real applications. However, the program will select the "best ones" as the final inputs to the model based on the predefined criteria.

As illustrated in section 3.4.6, the suboptimal selection of independent variables is also necessary to identify the needed socio-economic indices. In the case study, three predictable socioeconomic indices are selected as independent variables. They are total population, total employment, and personal income. The comparison results in Figure 12 and Table 9 show that both the estimated vehicle age distribution and the emission impacts are very close between the suboptimal selection and the optimal selection.

As a whole, MOFAD provides flexibility for the selection of socio-economic indices in real application. The users may provide multiple candidate socio-economic indices. Table 2 and Table 3 are just examples in the case study. The program will help to select some (e.g., four) of them as the final ones based on the optimal selection procedure.

However, the users may also directly provide several easily predicted socio-economic indices (e.g. population, average income) to the model and skip the procedure of optimal selection.

CHAPTER 4

ADJUSTING VEHICLE MILEAGE ACCUMULATION RATES BASED ON SMALL SAMPLE SURVEY

4.1 Impact of Mileage Accumulation Rates on Emission Factors

Mileage accumulation rates are mainly used in the MOBILE6 model for weighting the vehicle miles traveled (VMT) by vehicle age, and for calculating the total mileage accumulation by vehicle age, which is used to estimate the emission factors for each vehicle age taking into account the "deterioration" of air pollution control devices.

The vehicle mileage accumulation rates have considerable impacts on three emission factors (HC, CO and NO_x). Figure 16 and 17 show the plots of the percentage changes of emission factors with the percentage change of vehicle mileage accumulation rates for vehicle types HDDV2B and LDGV. Emission factors were all generated by MOBILE6, while mileage accumulation rates were changed from -50% to 50% (with an increment of 10%) around the default ones. It is shown in these two figures that the increase of the percentage of vehicle mileage accumulation rate will result in significant changes of emission factors. For example for vehicle type HDDV2B in Figure 16, all emission factors are increasingly proportional to mileage accumulation rates including VOC, CO and NO_x. However, it is interesting to note that for vehicle type LDGV in Figure 17 and for most of the other vehicle types, not all the emission factors are always increasingly proportional to mileage accumulation rates. For those vehicle types, VOC will increase when mileage accumulation rates are far less than the default values. Nevertheless, regardless of the kind of trends between the change of emission factors and the change of mileage accumulation rates, the emission factors are always sensitive to the change of mileage accumulation rates for all types of vehicles. The impacts of mileage accumulation rates on emission factors are very high.

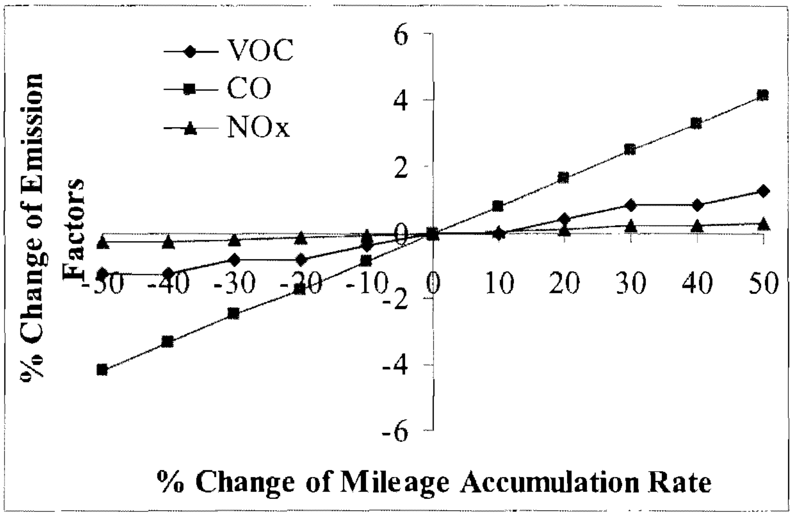


FIGURE 16 Percentage changes of emission factors with percentage changes of mileage accumulation rates for HDDV2B.

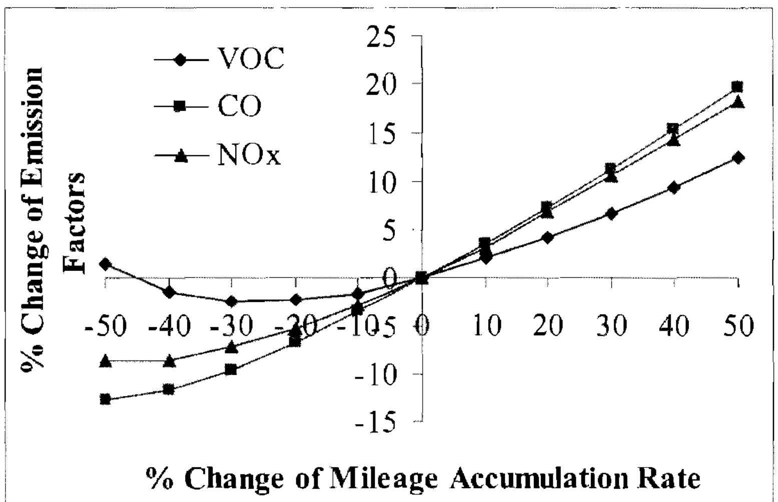


FIGURE 17 Percentage changes of emission factors with percentage changes of mileage accumulation rate for LDGV.

4.2 Algorithm Description

4.2.1 Basic Ideas

Practically for most local areas, it is very hard to conduct a large-scale full-sized survey on mileage accumulation rates due to the survey cost. Also, inspection and maintenance (I/M) data may not be always available for kinds of reasons. However, as stated before, using the default values directly, which differs from the local ones, may cause inaccurate estimates of emission factors.

A realistic approach to solving the above problem is to conduct a small sample survey in the concerned area, and then incorporate the collected information into MOBILE6 default values. The following is the description of the adjusting algorithm.

Suppose in a small sized survey in the concerned area, the average mileage accumulated in the past year for vehicle type s and age g (where, $g = 1, 2, ..., n_g$ and n_g is the total number of vehicle age required by MOBILE6) is: u_{sg}. Let $\bar{u}_{mg}$ be the default value of vehicle mileage accumulation rate with vehicle type m and age g, where m is the vehicle type in MOBILE6, $m = 1, 2, ..., M$; M is the total number of vehicle types defined in MOBILE6.

It is a practical fact that it is rather difficult in some juristic areas to obtain the mileages for all detailed 28 vehicle types as required in MOBILE6. In addition, it is also not guaranteed that the divisions of vehicle types in survey conducting or data collecting processes are exactly the same as those required by MOBILE6. For example in the case of small sample survey, it is difficult for the surveyors to distinguish all the 28 vehicle types since some of the differences are not so obvious from outside. The realistic way is to classify the surveyed vehicle types in the commonly recognized forms, such as car, sport utility vehicle (SUV), truck and etc. This means that vehicle type s from the survey may not be consistent with any vehicle type defined in MOBILE6. So the difference between vehicle types should be considered in developing the practical algorithm for adjusting mileage accumulation rates.

In view of this, let's set up the matching groups converting surveyed vehicle types to MOBILE6 vehicle types. Suppose there are a total of P groups that can link the surveyed vehicle types and the MOBILE6 vehicle types. The group number p is listed in the first column ($p = 1, 2, ..., P$) in TABLE1. The second column lists the surveyed vehicle types, where s_p^i is the $i^{\text{-th}}$ surveyed vehicle type in group p, $i = 1, 2, ..., n_{sp}$ and n_{sp} is the total number of surveyed vehicle types in group p. The vehicle types in the entire second column $s_1^1, s_1^2, ..., s_1^{n_{s1}}$, $s_2^1, s_2^2, ..., s_2^{n_{s2}}$, ... $s_P^1, s_P^2, ..., s_P^{n_{sP}}$ consist of all the surveyed vehicle types. The third column is the MOBILE6 vehicle types, where m_p^j is the $j^{\text{-th}}$ MOBILE6 vehicle type in the $p^{\text{-th}}$ group, where $j = 1, 2, ..., n_{mp}$, and n_{mp} is the total number of MOBILE6 vehicle types in the $p^{\text{-th}}$ group. The total combination $\mathbf{K} = \{k\}$ of the vehicle types in the third column $m_1^1, m_1^2, ..., m_1^{n_{m1}}$, $m_2^1, m_2^2, ..., m_2^{n_{m2}}$, ... $m_P^1, m_P^2, ..., m_P^{n_{mP}}$ should cover all the 28 vehicle types for MOBILE6.

After survey, sampled mileage accumulation rates over different surveyed vehicle types can be obtained. To better reflecting the difference between the local mileage accumulation rates and the MOBILE6 default values, the ratio of surveyed mileage accumulation rate over the corresponding default one is chosen as the analytical basis. This type of ratio is called *Variation Ratio* (VR), which is defined as

$$r_{s_p^i,g}(l) = \frac{u_{s_p^i g}(l)}{\bar{u}_{s_p^i,g}}, \quad \forall l = 1,\ 2,\ ...\ L_{s_p^i,g} \tag{4-1}$$

In equation (4-1), $r_{s_p^i,g}(l)$ is the Variation Ratio (VR) for vehicle type s_p^i, age g and for sample l. $L_{s_p^i,g}$ is the total sample size for the corresponding surveyed vehicle type. $u_{s_p^i,g}(l)$ is the surveyed mileage accumulation rate for sample l. $\bar{u}_{s_p^i,g}$ is the default value for the surveyed vehicle type s_p^i and age g, which can be converted from the default values for MOBILE6 vehicle types by using the following formula

$$\bar{u}_{s_p^i,g} = \frac{1}{m_p}\sum_{j=1}^{m_p} u_{m_p^j,g} \tag{4-2}$$

$\bar{u}_{m_p^j,g}$ is the default value of vehicle mileage accumulation rate with MOBILE6 vehicle type m_p^j and age g. m_p^j and s_p^i are all members in group p in Table 10.

In the real case, the sample size may not be exactly the same for each different surveyed vehicle type. Furthermore, the sample errors may also not be the same.

The basic idea of the algorithm for estimating the mileage accumulation rate in a local area is to analyze the mean and confidence interval of surveyed mileage accumulation rate for different surveyed vehicle types, and then to transfer the results into the MOBILE6 vehicle types. The next part is the detailed description of this process.

TABLE 10 Matching Groups Relating Surveyed Vehicle Types to MOBILE6 Vehicle Types

Group Number	Surveyed Vehicle Type	MOBILE6 Vehicle Type
1	$s_1^1, s_1^2, \ldots, s_1^{n_{s1}}$	$m_1^1, m_1^2, \ldots, m_1^{n_{m1}}$
2	$s_2^1, s_2^2, \ldots, s_2^{n_{s2}}$	$m_2^1, m_2^2, \ldots, m_2^{n_{m2}}$
...	...	...
p	$s_p^1, s_p^2, \ldots, s_p^{n_{sp}}$	$m_p^1, m_p^2, \ldots, m_p^{n_{mp}}$
...	...	...
P	$s_P^1, s_P^2, \ldots, s_P^{n_{sP}}$	$m_P^1, m_P^2, \ldots, m_P^{n_{mP}}$

Note: The correspondence between the surveyed and MOBILE6 vehicle types for case study is listed in the later table (Table 11).

4.2.2 Detailed Description

Suppose for surveyed vehicle type s_p^i, the total number of sample size is $L_{s_p^i,g}$. Then, the sampled Variance Ratio (VR) for this vehicle type $\bar{r}_{s_p^i,g}$ is

$$\bar{r}_{s_p^i,g} = \frac{1}{L_{s_p^i,g}} \sum_{l=1}^{L_{s_p^i,g}} r_{s_p^i,g}(l) \tag{4-3}$$

Equation (4-3) is a point estimation of VR for vehicle type s_p^i and age g. Such estimates, however, do not convey information on the degree of accuracy of these estimates. Consideration on the variability of the data in the sampling process is necessary. Theoretically, variability is calculated by variance $\sigma_{S_p^i,g}$, sample size, confidence intervals and etc. Conversely, variances of the samples are normally unknown in the real cases. The confidence interval with unknown variance can be estimated based on the *sample standard deviation* $S_{s_p^i,g}$, which can be expressed as (Ang and Tang, 1975)

$$\left\langle \bar{r}_{s_p^i,g} \right\rangle_{1-\alpha} = \left[\bar{r}_{s_p^i,g} - t_{\alpha/2, L_{s_p^i,g}-1} \frac{S_{s_p^i}}{\sqrt{L_{s_p^i,g}}} ; \bar{r}_{s_p^i,g} + t_{\alpha/2, L_{s_p^i,g}-1} \frac{S_{s_p^i}}{\sqrt{L_{s_p^i,g}}} \right] \tag{4-4}$$

where, $\bar{r}_{s_p^i,g}$ and $S_{s_p^i,g}$ are sample mean and sample standard deviation; $L_{s_p^i,g}$ is the sample size, and $(1-\alpha)$ is the specified confidence level; $\mp t_{\alpha/2, L_{s_p^i,g}-1}$ denotes the percentile of the t-distribution variate T at the cumulative probability level $\alpha/2$ and $(1-\alpha/2)$, respectively.

The statistical meaning of equation (4-4) is that the real mileage accumulation rate, or the population mean of the samples, is located in the confidence interval $\left\langle \bar{r}_{s_p^i,g} \right\rangle_{1-\alpha}$ with the probability of $(1-\alpha)$. Therefore, the possible estimated Maximum Relative Error (MRE) with the confidence of $(1-\alpha)$ is

$$\mathrm{MRE}_{s_p^i,g}^{1-\alpha} = t_{\alpha/2,\left(L_{s_p^i,g}-1\right)} \frac{S_{s_p^i}}{\sqrt{L_{s_p^i,g}}} \Big/ \bar{r}_{s_p^i,g} \cdot 100\% \tag{4-5}$$

Evidently, the smaller sample standard deviation or large sample size would increase the accuracy of the sample as the estimator of the population mean. Equation (5) requires the large sample size and small standard deviation. In the case of small sample survey, the sample size for a particular vehicle type s_p^i at a particular age g is normally too small, which will cause inaccuracy estimation. Then again, the sample size as well as the corresponding estimation accuracy will increase if the sample is aggregated along vehicle age g. This assumes that the Variation Ratio (VR) $\bar{r}_{s_p^i,g}$ remains constant along vehicle ages for a particular surveyed vehicle type. This means

$$\bar{r}_{s_p^i,g} = \bar{r}_{s_p^i} = \frac{1}{G} \sum_{g=1}^{g} \left(\frac{1}{L_{s_p^i,g}} \sum_{l=1}^{L_{s_p^i,g}} r_{s_p^i,g}(l) \right) \qquad \forall g = 1,\ 2,\ \ldots,\ G \tag{4-6}$$

where, $\bar{r}_{s_p^i}$ is the Variation Ratio (VR) for surveyed vehicle type s_p^i, G is the total vehicle age.

The sample size covered by equation (4-6) will be $L_{s_p^i}$ rather than $L_{s_p^i,g}$, the relationship between them is

$$L_{s_p^i} = \frac{1}{G}\sum_{g=1}^{G} L_{s_p^i,g} \tag{4-7}$$

Therefore, the possible estimated Maximum Relative Error (MRE) with the confidence of $(1-\alpha)$ in this case is

$$\mathrm{MRE}_{s_p^i}^{1-\alpha} = t_{\alpha/2,\left(L_{s_p^i}-1\right)} \frac{S_{s_p^i}}{\sqrt{L_{s_p^i}}} \Big/ \bar{r}_{s_p^i} \cdot 100\% \tag{4-8}$$

$\mathrm{MRE}_{s_p^i}^{1-\alpha}$ should be much smaller than $\mathrm{MRE}_{s_p^i,g}^{1-\alpha}$ since $L_{s_p^i}$ in equation (4-8) is much larger than $L_{s_p^i,g}$ in equation (4-5). This increases the accuracy of estimation.

The minimum sample size on a particular surveyed vehicle type can be required by the following equation in case both the sample mean and sample variance are known with a certain confidence

$$L_{s_p^i} > \mathrm{int}\left(t_{\alpha/2,\left(L_{s_p^i}-1\right)} \cdot \frac{S_{s_p^i}}{\mathrm{MRE}_{s_p^i}^{1-\alpha}} \Big/ \bar{r}_{s_p^i}\right)^2 \tag{4-9}$$

where, $\mathrm{int}(\cdot)$ means taking integer number.

In equation (4-9), $\mathrm{MRE}_{s_p^i}^{1-\alpha}$ is a pre-required maximum relative error with the confidence $(1-\alpha)$. $t_{\alpha/2,L_{s_p^i,g}-1}$ can be firstly chosen like $L_{s_p^i} = \infty$ to get an initial $L_{s_p^i}$. Then, based on this initial $L_{s_p^i}$, a new $t_{\alpha/2,L_{s_p^i,g}-1}$ can be gotten for a better $L_{s_p^i}$.

After getting the mean, variance and maximum relative errors for Variation Ratios (VR) of mileage accumulation rates on all surveyed vehicle types, these results will be easily transferred to the corresponding MOBILE6 vehicle types.

$$\bar{r}_{m_p^j} = \frac{1}{n_{sp}}\sum_{i=1}^{n_{sp}}\left[\frac{1}{G}\sum_{g=1}^{g}\left(\frac{1}{L_{s_p^i,g}}\sum_{l=1}^{L_{s_p^i,g}} r_{s_p^i,g}(l)\right)\right] \qquad \forall j = 1,\ 2,\ \ldots\ n_{mp} \tag{4-10}$$

In equation (4-10), $\bar{r}_{m_p^j}$ and $\bar{r}_{s_p^i,g}$ are in the same matching group as listed in Table1. For those MOBILE6 vehicle types, where no matching surveyed vehicle type exists, their Variation Ratios (VR) $\bar{r}_{m_p^j}$ can be set as the total average for all the surveyed vehicle types.

$$\bar{r}_{m_p^j} = \frac{1}{P_s}\sum_{p=1}^{P_s}\left[\frac{1}{n_{sp}}\sum_{i=1}^{n_{sp}}\left[\frac{1}{G}\sum_{g=1}^{g}\left(\frac{1}{L_{s_p^i,g}}\sum_{l=1}^{L_{s_p^i,g}} r_{s_p^i,g}(l)\right)\right]\right] \qquad \forall j = 1,\ 2,\ \ldots\ n_{mp} \tag{4-11}$$

In equation (4-11), P_s is the number of the matching groups where surveys are conducted.

With the help of the above stages, means and variances of VR for all the MOBILE6 vehicle types can be obtained.

Therefore, the estimated vehicle mileage accumulation rates $\hat{u}_{m_p^j}$ for all MOBILE6 vehicle types can be obtained by multiplying the default one $\bar{u}_{m_p^j}$ with the corresponding VR.

$$\hat{u}_{m_p^j} = \bar{u}_{m_p^j} \cdot \bar{r}_{m_p^j} \tag{4-12}$$

For the purpose of getting a general picture on the difference between the local vehicle mileage accumulation rates and the default ones, this is probably one of the feasible estimation approaches in case the sample size is not sufficient while considerations are focused on both particular vehicle type and particular age.

4.2.3 Steps of Whole Algorithm

Figure 18 illustrates the whole modeling process. In summary, the basic process of the algorithm can be divided into the following five steps.

Step 1: Set-up the matching table for surveyed vehicle types and MOBILE6 vehicle type;

Step 2: Obtain the Variation Ratio for all sampled vehicle types by Equation (4-1)-(4-2);

Step 3: Calculate the mean, confidence interval and MRE for each surveyed vehicle types by Equation (4-3)-(4-5);

Step 4: Convert the results in Step 3 into MOBILE6 vehicle types;

Step 5: Calculate the final estimates of vehicle mileage accumulation rates by Equation (4-12).

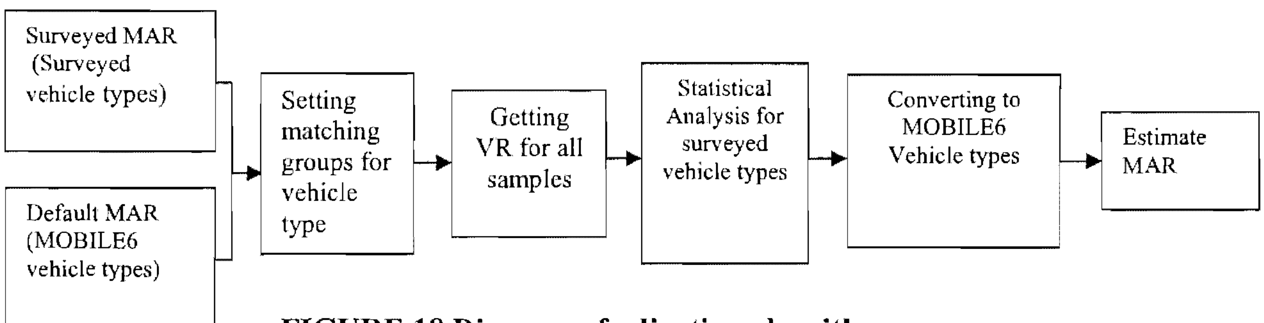


FIGURE 18 Diagram of adjusting algorithm process.

4.3 CASE STUDIES IN HOUSTON AND EL PASO

To test and validate the proposed algorithm, case studies were conducted in two Texas areas: Harris County of Houston and El Paso. Houston is a big city with the population of 1.95 million and an area of 602 square miles and Harris County is the center and most urbanized area in Houston; while El Paso is a relatively medium one with the population of 0.56 million and an area of 250 square miles including both El Paso city and El Paso county (US Census Bureau, 2001). Sample surveys on mileage accumulation by vehicle types were conducted, and the corresponding adjusting factors together with the adjusted mileage accumulation rates were calculated afterwards.

4.3.1 Survey Process and Data Preparation

This survey is to investigate selected vehicle users based on the vehicle classes and vehicle ages regarding the annual total mileage traveled. The vehicle types surveyed include car, SUV (Sport Utility Vehicle), van and truck for both Houston and El Paso areas. The reason to use the simplified vehicle types in the survey is that few survey users were able to determine which MOBILE6 vehicle types their vehicles belong to. The vehicle make year, model and vehicle weight were requested. More importantly, vehicle mileage driven in the year 2000 and the total mileage on the odometer were recorded. The participants estimated the mileage for the year 2000. Since the survey was conducted in the spring of 2001, their estimated mileages should be trustable. Simultaneously, the background information such as the vehicle owners' age group, ethnic group, sex, number of household members, average household income, residential area and etc. were also collected for reference. Since there was no restriction to the socioeconomic background of the participants, their socioeconomic properties should be randomly distributed.

Surveys were conducted in different locations of both areas, such as Department of Public Safety, license renewed offices, inspection and maintenance stores, oil stations, public parking lots, and etc. This sample survey was conducted in the spring of 2001 in the both areas simultaneously.

During the survey, a total of 932 survey forms have been returned (805 in Houston area and 127 in El Paso area), with 1255 vehicles surveyed (1076 in Houston and 179 in El Paso). After eliminating the wrong records (for example, the 2000 mileage was less than the total mileage) and restricting Houston samples for Harris County only, 935 samples (798 in Houston and 137 in El Paso) were remained.

Since there are 100 combinations (4 vehicle types × 25 ages) if both surveyed vehicle types and ages are considered, the sample sizes in this survey are not sufficient to guarantee the survey accuracy. Therefore in this algorithm samples were aggregated in terms of surveyed vehicle types only in order to increase the sample size as well as the survey accuracy.

4.3.2 Corrected Mileage Accumulation Rate for Houston and El Paso

As the first step of this algorithm, the matching table for surveyed vehicle types and MOBILE6 vehicle types was set up. In the application in Harris County and El Paso, four vehicle types were surveyed: Car, SUV, Van and Truck. This means that in this case the total surveyed vehicle type n_s is: 4, where $s = 1, 2, 3, 4$ representing Car, SUV, Van and Truck, respectively. According to the requirement by MOBIILE6, a total of 28 vehicle types are needed. The matching table linking the surveyed vehicle types and the MOBILE6 vehicle types are displayed in Table 11.

TABLE 11 Matching Groups for Converting Vehicle Types for Survey to That for MOBILE6

Group Number	Vehicle Types in Survey	Vehicle Types in MOBILE6
1	Car	LDGV, LDDV
2	SUV	LDGT1

3	SUV, Van	LDGT2, LDGT3, LDGT4, LDDT12, LDDT34
4	Truck	HDGV2b, HDGV3, HDGV4, HDGV5, HDGV6, HDGV7, HDDV2b, HDDV3, HDDV4, HDDV5, HDDV6, HDDV7
5	None	HDGV8a, HDGV8b, HDDV8a, HDDV8b, HDGB, HDDBT, HDDBS, MC

Then, VRs (the Variation Ratios) for all sampled vehicles were obtained. These VRs were sorted according to vehicle types. Based on the equations in the previous sections, the sample means, sample standard deviations and other statistical results for different surveyed vehicle types were calculated, which were listed in Table 12 (for Harris County of Houston) and Table 13 (for El Paso).

TABLE 12 Results of Statistical Analysis for Ratios in Harris County

	Average	Std	Sample Size	90% Confidence Interval				95% Confidence Interval			
				Half Interval	Left Margin	Right Margin	MRE (%)	Half Interval	Left Margin	Right Margin	MRE (%)
Car	1.422	1.164	426	0.093	1.329	1.515	6.53	0.135	1.287	1.557	9.50
Truck	1.382	1.381	231	0.135	1.247	1.517	9.76	0.161	1.221	1.542	11.63
SUV	1.149	0.738	81	0.135	1.015	1.284	11.73	0.161	0.989	1.310	13.98
Van	1.214	1.071	60	0.228	0.987	1.442	18.74	0.271	0.943	1.485	22.32
SUV, Van	1.177	0.892	141	0.124	1.054	1.301	10.50	0.147	1.030	1.324	12.51
All Type	1.367	1.192	798	0.069	1.298	1.436	5.08	0.083	1.284	1.450	6.05

Note: MRE means the possible Maximum Relative Error

TABLE 13 Results of Statistical Analysis for Ratios in El Paso

	Average	Std	Sample Size	90% Confidence Interval				95% Confidence Interval			
				Half Interval	Left Margin	Right Margin	MRE (%)	Half Interval	Left Margin	Right Margin	MRE (%)
Car	0.690	0.272	93	0.046	0.643	0.736	6.73	0.140	0.549	0.830	20.32
Truck	0.476	0.261	15	0.090	0.386	0.565	18.88	0.107	0.369	0.583	22.49
SUV	0.615	0.244	20	0.090	0.525	0.704	14.61	0.107	0.508	0.722	17.40

Van	0.703	0.267	9	0.147	0.556	0.849	20.86	0.175	0.528	0.877	24.86
SUV, Van	0.642	0.250	29	0.076	0.566	0.718	11.90	0.091	0.551	0.733	14.18
All Type	0.656	0.273	137	0.038	0.618	0.694	5.84	0.046	0.610	0.702	6.96

Note: MRE means the possible Maximum Relative Error

In Table 12 and Table 13, estimation intervals for both 90% and 95% confidence were given together with the possible Maximum Relative Errors. From the two tables it is shown that the possible Maximum Relative Errors (MREs) are not too big for most of the surveyed vehicle types. MREs will increase if the confidence probability increases. For some vehicle types, the sample sizes of which are not so big, the corresponding MREs are bigger than the others. However, MREs of VRs for all surveyed vehicle types are relatively small (5.08% for Harris County and 5.84% for El Paso, respectively), which means the total average of VRs for the two areas are acceptable. Therefore, the survey results in a whole are acceptable.

So should the result for a specific surveyed vehicle type be accepted or not? It depends on the criteria the user sets for them. For example, results for almost all the vehicle types in Table 12 and Table 13 can be directly accepted if the user sets the criteria as 20% under 90% confidence probability or 25% under confidence probability.

For the case when MRE of VR for a particular surveyed vehicle type cannot meet the user's criteria, the best way is to make a supplementary survey to increase the sample size and reduce its variance and MRE. The minimum sample size can be determined according to equation (4-9). On the other hand, the sample mean of VR for all vehicle types can also be applied to that for a particular surveyed vehicle type in case its MRE cannot meet the user's criteria. Nevertheless, that later way can only be used when there are indeed difficulties to make such a supplementary survey.

Since the example is only for the illustration of the proposed examples, the entire results in Table 12 and Table 13 were accepted for the further demonstration.

After obtaining the estimates of VRs for the surveyed vehicle types, the estimates of VRs for all the MOBILE6 vehicle types can therefore be transferred based on the matching table (Table 11). For example in the first row of Table 11, Car in surveyed vehicle type corresponds to LDGV and LDDV in MOBILE6 vehicle type. So VRs for LDGV and LDDV were taken from VR for Car. It should be noted that in group 3 in Table 11, both SUV and Van correspond to five MOBILE6 vehicle types. VRs for these five MOBILE6 vehicle types were taken from the sample mean based on samples for both SUV and Van, which is different from the simple average of VR for SUV and for Van, respectively.

For those MOBIEL6 vehicle types, the matching MOBILE6 vehicle types cannot be located in Table 11, and the total sample mean of VR for all surveyed vehicle types was used for their VRs. Although this may not be exactly true, it is believed that the total average VR for all surveyed vehicle types is still closer to the real ones than the default values.

Table 14 lists all the estimates of VRs for all MOBILE6 vehicle types. It is shown that VRs in Harris County of Houston are bigger than one while those for El Paso are smaller than one.

TABLE 14 Variance Ratios (VRs) of Mileage Accumulation for Harris and El Paso

(a) Harris County

VT	1	2	3	4	5	6	7
VR	1.42	1.15	1.18	1.18	1.18	1.38	1.38
VT	**8**	**9**	**10**	**11**	**12**	**13**	**14**
VR	1.38	1.38	1.38	1.38	1.37	1.37	1.42
VT	**15**	**16**	**17**	**18**	**19**	**20**	**21**
VR	1.18	1.38	1.38	1.38	1.38	1.38	1.38
VT	**22**	**23**	**24**	**25**	**26**	**27**	**28**
VR	1.37	1.37	1.37	1.37	1.37	1.37	1.18

VT: vehicle type

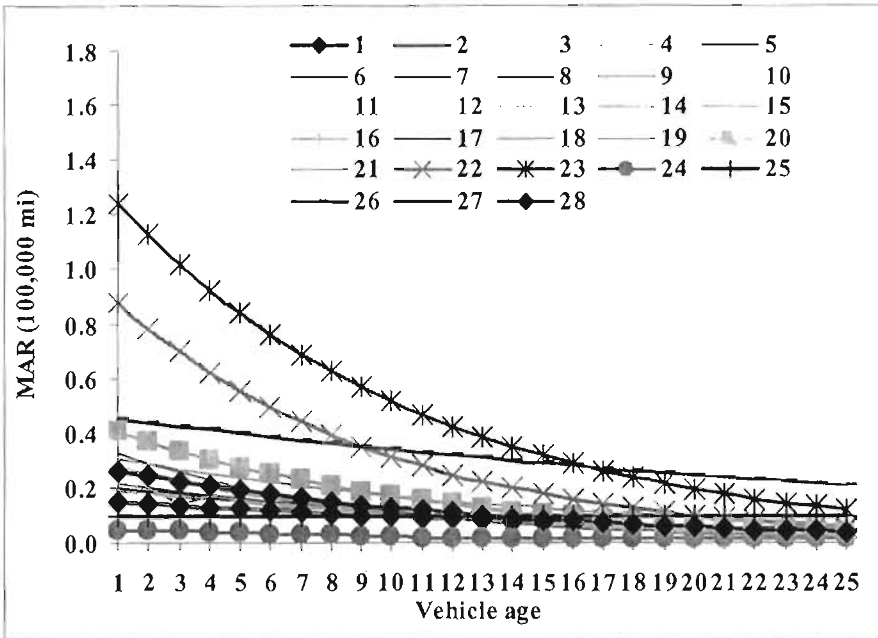
(b) El Paso

VT	1	2	3	4	5	6	7
VR	0.69	0.61	0.64	0.64	0.64	0.48	0.48
VT	**8**	**9**	**10**	**11**	**12**	**13**	**14**
VR	0.48	0.48	0.48	0.48	0.66	0.66	0.69
VT	**15**	**16**	**17**	**18**	**19**	**20**	**21**
VR	0.64	0.48	0.48	0.48	0.48	0.48	0.48
VT	**22**	**23**	**24**	**25**	**26**	**27**	**28**
VR	0.66	0.66	0.66	0.66	0.66	0.66	0.64

VT: vehicle type

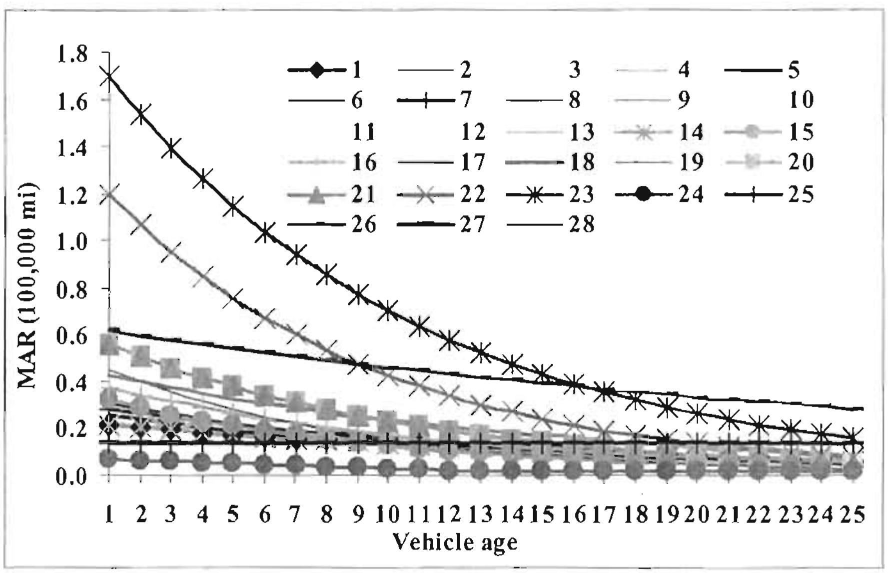
The final estimation of mileage accumulation rates for all vehicle types was very simple when all VRs (Variance Ratios) were obtained. The only thing to do was to multiply the default values with the corresponding Variance Ratios.

Figure 4 shows the nationwide mileage accumulation rates for all vehicle types. Each curve in the figure represents each vehicle type. The corrected local mileage accumulation rates for Harris and El Paso areas are plotted in Figure 5 and Figure 6.



Note: Each line represents one vehicle type.

FIGURE 19 Default nationwide mileage accumulation rate.



Note: Each line represents one vehicle type.

FIGURE 20 Corrected local mileage accumulation rate for Harris County.

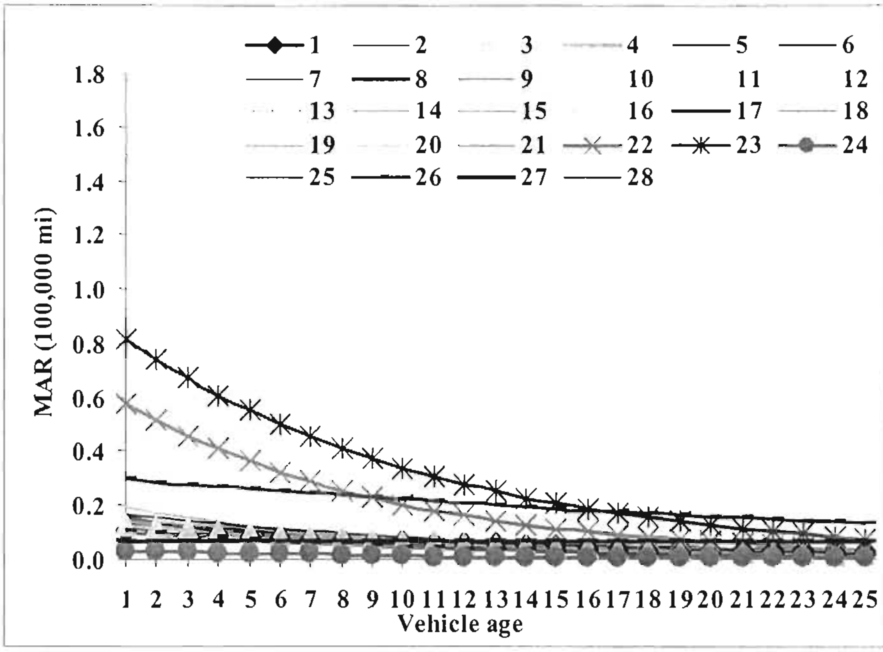


FIGURE 21 Corrected local mileage accumulation rate for El Paso area.

From Figure 20 and Figure 21, it is easy to see that the corrected local mileage accumulation rates in Harris County are much higher than the default ones, while the corrected local mileage accumulation rates in El Paso are much lower than the default ones. By calculation, the average Variance Ratios for Harris County of Houston and El Paso are 1.34 and 0.58 (mathematical mean), respectively. These results make sense since Harris is a big urbanized area with large population lived in or nearby, while El Paso is a relatively smaller one. People in Harris should travel more than the average U.S. cities, while people in El Paso may travel less than the average.

4.3.3 Impacts on Emission Factors

As shown above, vehicle mileage accumulation rates have significant impacts on the estimates of emission factors by MOBILE6. The adjusted mileage accumulation rates in both Harris County and El Paso areas were input into MOBILE6 and the estimates of emission factors were compared with those by inputting the default mileage accumulation rates. The comparison results are presented in Table 6 and Figure 7. In Table 6 it is shown that the emission factors vary a lot among Harris County and El Paso. According to the relative differences shown in Figure 7, the emission factor CO in Harris is 9.8% higher than the default, while El Paso is 9.5% lower than the default one. For NO_x, it shows similar characteristic since the relative differences with the default one are 4.0% and -3.5%, respectively. The result for VOC is the same. The relative error for VOC in Harris is 6.4% higher than the default while that in El Paso is 4.0% lower than default one. In a whole, emission in Harris County is higher than the estimates based on the default mileage accumulation rates, while emission in El Paso is lower than the estimates based on the default mileage accumulation rates.

TABLE 15 Emission Factors in Harris and El Paso by Corrected MAR compared with Default Ones

Emission Factors	Default	El Paso	Harris
VOC	0.673	0.646	0.716
CO	13.39	12.116	14.703
NO_x	2.421	2.336	2.519

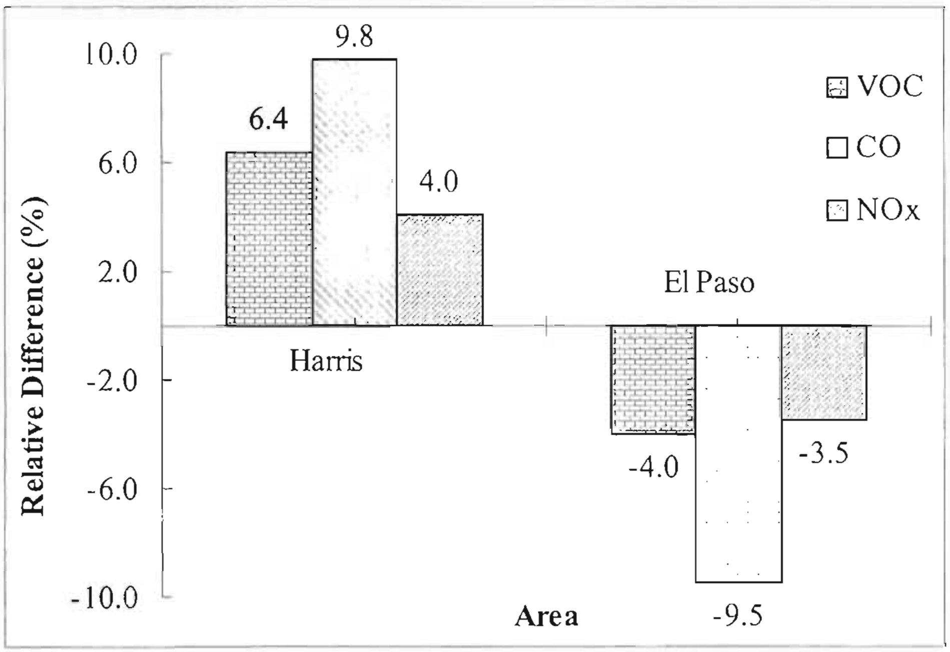


FIGURE 22 Relative differences of emission factors when compared adjusted mileage accumulation rates with MOBILE6 default values for Harris and El Paso.

CHAPTER 5

COLLECTING INFORMATION ON ESTIMATION OF VMT RELATED VARIABLES

5.1 Requirements in MOBILE5 &6 on VMT Related Variables

5.1.1 Definition of VMT and VMT mix

VMT (vehicle miles traveled or vehicle miles of travel) is a unit to measure vehicle travel made by a vehicle, such as an automobile, van, pickup truck, or motorcycle. Each mile traveled is counted as one vehicle mile regardless of the number of persons in the vehicle.

The vehicle mile traveled (VMT) mix specifies the fraction of total highway VMT that is accumulated by each of the different vehicle types.

VMT & mix are important travel indexes in the emission estimation model MOBILE. Emissions analysis is very sensitive to VMT mix. For example, for MOBILE5 at high temperature, a 2.8% change in HDGV mix causes about a 10% change in the CO rate; at high temperature, a 4.8% change in HDGV mix leads to about a 10% shift in the VOC rate.

5.1.2 Requirement in MOBILE5 on VMT mix

From section 5.1.2 to section 5.1.3, the requirements in MOBILE, which come from MOBILE User's Guide (Environment Protection Agency, 1994, p. 2-22; Environment Protection Agency, 2000, Chapter 5; Environment Protection Agency, 2001a, Chapter 5), will be summarized. The VMT mix is used in MOBILE5 only to calculate the composite (all vehicles, or fleetwide) emission factor for a given scenario on the basis of the eight vehicle class-specific emission factors.

In MOBIEL5, the users can choose to use the MOBILE5 national VMT mix (VMFLAG=1), the input of one alternate VMT mix (in One-time Data) for use in all scenarios of a given MOBILE5 run (VMFLAG=3), or the input of a different alternate VMT mix (in Scenario data) for each scenario (VMFLAG=2).

In MOBILE5, VMT mix is the fraction of total highway VMT that is accumulated by each of the 8 vehicle types. Each VMT mix supplied as input must consist of a set of eight fractional values, representing the fraction of total highway VMT accumulated by each of the eight vehicle types. All values must be between zero and one, and the eight values must sum to 1.0 (MOBILE5 produces an error message and does not execute the run if these constrains are not met).

The format of the VMT mix record(s) is **8F4.3**. The values correspond to the eight vehicle types in this order: LDGV, LDGT1, LDGT2, HDGV, LDDV, LDDT, HDDV, and MC. An example of a VMT mix record specifying that 65% of all VMT is accumulated by LDGVs and that each of the other seven vehicle types accounts for 5% of all VMT is shown below. Note that this format does not include leading zeros or blanks between the individual values.

.650.050.050.050.050.050.050.050

5.1.3 New Version of MOBILE and Changes on VMT and VMT mix

The present version of MOBILE used in Texas and other states is MOBILE5. However, the new version MOBILE6 will be fully released soon. In the summer of 2001, the trial version of MOBILE6 has already been sent to all the states' DOT. Therefore, in estimating the VMT mix, it is practical to consider all the requirements in the new version (MOBILE6), instead of only in the old version (MOBILE5) as required in the project proposal.

In MOBILE5 environment, VMT information was used outside of the MOBILE5. VMT information is not needed to run the model. VMT was used to estimate emission inventory. In MOBILE6 environment, local VMT data or the national default is required when to model local conditions.

In MOBLIE6, there are many changes on VMT related functions compared with MOBILE5. Table 16 lists the names and functions of VMT related commands and their corresponding functions. Table 17 lists the differences of the VMT related commands and functions between MOBILE5 and MOBILE6.

TABLE 16 Name and Functions of VMT Related Commands in MOBILE6

Command and Name	Command and Function
VMT FRACTIONS	Allows user to apply alternate VMT factions by each of 16 combined vehicle types
VMT BY FACILITY	Allows user to supply alternate VMT distributions by facility type that override M6 defaults for each scenario. 4 Road Types * 24 Hours = 96 VMT Fractions (freeway, arterial, local and ramp)*(6 AM ~5 AM)

VMT BY HOUR	Allows user to apply alternate hourly distributions of VMT that override M6 defaults for each scenario. 24 Hours; all facility types (24 values must add to 1)
SPEED VMT	Allows user to enter VMT distribution across 14 preselected speed ranges for each of the 24 hours of the day for each scenario.

TABLE 17 VMT Related MOBILE6 Commands and the Difference between MOBILE5 and MOBILE6

M6 Command	Difference Between M5 and M6
VMT FRACTIONS	In M5, only 8 VMT fractions needed instead of the 28 fractions needed for M6.
VMT BY FACILITY	New features, no precedent in M5.
VMT BY HOUR	New features, no precedent in M5.
SPEED VMT	In M5, a single average speed could be specified for all or for 8 individual vehicle types. M6 requires speed distributions for each hour.

From Table 16 and Table 17, we can see that there are many new features in MOBILE6. The format and part of the default VMT related variables in MOBILE6 are listed in Appendix E, F, G, and H.

5.1.4 Converting of MOBILE5 Vehicle Classes into MOBILE6 Vehicle Classes

Chapter 5 of the User's Guide of MOBILE6 (Environment Protection Agency, 2001a) mentions the conversion of MOBILE5 vehicle classes into MOBILE6 vehicle classes. The MOBILE5 the emission factor model requires the VMT split by eight vehicle classes. The vehicle classes are based on the size and weight of vehicles as well as the type of fuel used. The eight vehicle classes are: light-duty gasoline vehicle (LDGV), light-duty gasoline truck type 1 (LDGT1), light-duty gasoline truck type 2 (LDGV2), heavy duty gasoline vehicle (HDGV), light duty diesel vehicle (LDDV), light duty diesel truck (LDDT), heavy duty diesel vehicle (HDDV), and motorcycle (MC).

So MOBILE5 accounted for only eight vehicle classes, but MOBILE6 has greatly expanded the number of individual vehicle classes to 28 as listed in Appendix I. In some contexts, MOBILE6 input is provided in terms of 16 combined vehicle classes as listed in Figure 1. The difference between the 28 vehicle classes and the 16 vehicle classes is that the 28 vehicle classes divide vehicle types also according to whether the vehicle use gasoline or diesel, while the 16 vehicle classes do not have this kind of division. In some cases, aggregated user-supplied MOBILE5 data will be used for each of the vehicle classes in MOBILE6. In other cases, such as distributions, the MOBILE5 values must be further split by new vehicle classes for use in MOBILE6.

Because of the unequal growth that occurs in various vehicle classes, the VMT distribution by vehicle class becomes a function of calendar year. MOBILE5 allows the user to enter eight VMT values, corresponding to the eight vehicle classes represented in the MOBILE5 output. MOBILE6 allows the user to enter 16 VMT values by combined vehicle classes.

Whereas MOBILE5 allowed the user to enter separate VMT for diesel-and gasoline-fueled vehicle classes, MOBILE6 requires that VMT by vehicle class be supplied in terms of the 16 combined gasoline and diesel-fuel categories. In MOBILE6, the VMT by vehicle class is split internally – accounting for the diesel sales fractions and annual mileage accumulation rates – in order to ensure that all of the fleet descriptions and activity values are consistent with one another. The first step in converting MOBILE5 to MOBILE6 VMT fractions is to combine the VMT fractions for gasoline and diesel categories into five composite gasoline/diesel groupings:

- LDV Group = LDGV + LDDV
- LDT Group 1 = LDGT1 + LDDT
- LDT Group 2 = LDGT2
- HDV Group = HDGV + HDDV
- MC Group = MC

The sum of the VMT fractions from the five groups should still equal to 1. These fractions are then adjusted using factors calculated from the default distributions of VMT from MOBILE6 for the appropriate calendar year. When the adjustments are completed properly, the sum of the 16 MOBILE6 VMT fractions will be 1.

TABLE 18 Converting MOBILE5 Vehicle Classes into MOBILE6 Vehicle Classes

16 Combine MOBILE6 Vehicle Classes	**VMT Fraction Calculation**
LDV	LDV Group
LDT1	LDT Group 1 * A
LDT2	LDT Group 1 * B
LDT3	LDT Group 2 * C
LDT4	LDT Group 2 * D
HDV2b	HDV Group * E
HDV3	HDV Group * F
HDV4	HDV Group * G
HDV5	HDV Group * H
HDV6	HDV Group * I
HDV7	HDV Group * J
HDV8a	HDV Group * K
HDV8b	HDV Group * L
HDBS	HDV Group * M
HDBT	HDV Group * N
MC	MC Group

Source: Environment Protection Agency, 2001a, P. 147.

The values A through N are taken for the appropriate calendar year. They are calculated from the default MOBILE6 VMT fractions for that calendar year. The terms A and B, C and D, and E through N should each add up to 1. The resulting 16 VMT fractions are supplied to MOBILE6 using the VMT FRACTIONS command.

5.2 Information on Estimation of VMT Related Variables

5.2.1 Sources of Information Collected

Information collected includes the reports from US EPA (United States Environmental Protection Agency), papers and reports from relevant journals, conference proceedings and government websites. A survey through e-mail was conducted to obtain more information on what kind of methodologies are used by the other states.

5.2.1.1 EPA Documents

EPA provides a guidance to assist users of the MOBILE6 highway vehicle emission factor model in the preparation of traffic activity inputs. It offers the recommendations on how to develop national wide distributions of vehicle miles traveled (VMT) by time of day, facility type and average speed.

5.2.1.2 Other Literatures

Other literatures include the reports from other government agencies, papers on journals, conference proceedings, etc. From these literatures the information of VMT estimation in the following states are obtained: Oregon, California, Florida, Idaho, Washington and Wisconsin.

5.2.1.3 Survey by E-mail

To better obtain the current practice of the VMT & mix estimation approaches in the other states and agencies, a survey by e-mail was conducted. The persons surveyed are those who attended the *TRB Technical Meeting and Workshop - Impacts of Recent Transportation Air Quality Modeling Improvements: Emphasis on MOBILE6 and EMFAC2000*, held on June 3-5, 2001 in Irvine California. The workshop was sponsored by the Transportation Research Board's Transportation/Air Quality Committee (A1F03), and addressed the new EPA and California mobile source emission factor models and their use in the transportation community. It attracted national-wide persons who apply, develop, or use the result of mobile source emissions models, or are involved in regional transportation and air quality planning. In this e-mail survey, 4 questions were designed as listed in the following:

1. What's the current approach they are using in estimating VMT mix for MOBILE5 input?
2. What kinds of approaches they are going to use when MOBILE6 is released?
3. Do they have any project/planning in setting up new algorithm(s) in estimating the VMT related parameters for MOBILE6?
4. Any other information on this matter they may provide.

The e-mail was sent out on Oct. 10, 2001 and Oct. 16, 2001, and the replies were received soon. A total of 133 persons were surveyed with 15 responded. For some important valuable replies, the follow-up e-mails were sent to the relevant persons to get more specific detailed information. Appendix J lists the names of persons and agencies/companies that responded the survey through e-mail. The responded ones include those from FHWA, different states (Colorado, Georgia, Florida, New York State, California); some regional councils and national laboratories (North Center TX Council of Governments, Wasatch Front Regional Council; Lawrence Berkeley National Laboratory); and some companies (Cambridge Systematics, Inc., Stan COG, ENVIRON International Corp).

5.2.2 Current Methodologies on estimation of VMT Related Variables

According to the information collected, there are several methodologies on estimation of VMT related variables. EPA gives a guidance involving the development and application of methods to estimate detailed national wide VMT related variables. The results serve as the national default values. It uses the traffic count data and the travel demand model to estimate the VMT related variables for five selected urban areas and estimated national time-of-day and speed distributions of urban VMT derived by extrapolation of results for four of the five selected urban areas.

Bhat and Nair (2000) formulate and estimate a fractional split model that determines the VMT mix ratio as a function of several informative variables, including physical attributes of links, the operating characteristics of links, aggregate area type characterizations of the traffic survey zone in which the link lies, and the land use attributes of the zone. This model is currently being embedded within a GIS platform to predict the VMT mix on all links of the Dallas Fort Worth metropolitan region.

For the practices in the other states, some use the MOBILE defaults, some use the HPMS traffic count data, some estimate according to the percentage of vehicles registered within the state, some use the fuel consumption based finance method, the policy procedure, etc.

5.2.3 Guidance by EPA

The EPA report *EPA420-P-99-006* (entitled as "*Development of Methodology for Estimating VMT Weighting by Facility Type*") concludes the results of work conducted for the development and application of methods to estimate certain aspects of on-road vehicle activity. In particular, this work was designed to estimate VMT on different classes of roadways by time of day and speed, and to investigate other vehicle activity characteristics. The materials in this section were summarized from that report (Environmental Protection Agency, 1999a.)

Two methods are developed for development of VMT distributions by facility class and speed. The first one works directly using vehicle count data. The second requires processing of regional travel demand model outputs. These two methods use data which are most capable to be obtained by local and state agencies, and neither method relies on databases of observed speeds. In these methods, speeds are estimated using facility characteristics and level of traffic congestion. Actual speed data can and should address the efforts of local characteristics that influence driver behavior and speeds, such as roadway lay-out (curves, hills, visibility, and distances between intersections) and signal coordination.

5.2.3.1 Method 1 – Working with traffic count data

It is relatively straightforward to estimate total VMT from vehicle count databases, although as noted later in this section, there are a number of ways in which biases can enter the calculation. Most regions use similar methods to estimate total VMT by functional class. Area type is available and used in many regions. The VMT estimation procedure is:

1. Calculate the sum of counts in each functional class (by area type if possible)
2. Determine the sample size in each functional class (the number of counters)
3. Determine the average volume by dividing total count by sample size
4. Obtain miles of facility in each class (available from DOT or GIS databases)
5. Calculate VMT by class as average volume multiplied by the number of miles of facility

Several key issues are immediately apparent if the VMT estimates are intended to be used in emission calculations. First, the classification of roadways must match with the four functional classes used in MOBILE6. Thus data for major and minor arterials and collectors may need to be merged into the MOBILE "arterial" class. The MOBILE "freeway" class might include data reported for "interstate" and "expressway" classes as well.

Frequently, counts will not be available for ramps. In this case, ramp VMT can be estimated as a fraction of freeway VMT, possibly by area type, based upon VMT estimates from a regional travel demand model. Rapid acceleration events on on-ramps can have significant contributions to total emissions, so realistic estimation of ramp VMT is important.

Common problems with count data include biases arising from the selection of roadways that are sampled or from idiosyncrasies of the counting device. For example, area using road tube counters may have undercounts on multilane facilities, especially during peak traffic periods. It occurs when two cars crossing the tube at the same time. (On freeways, this problem can be corrected by switching to ramp on/off counts). Also, sometimes data are combined without correcting underlying differences in the collection method.

Another problem that can occur is that there are too little count data for a particular facility type (or facility/area type combination). In these cases, one can combine two similar classes or extrapolate data from another similar class. The overall result, however, is an increase in the associated uncertainty of these estimates.

Addressing the speed dependence of emission rates in MOBILE6 requires that VMT for arterials and freeways be further disaggregated by either speed or LOS. Since characterizing traffic behavior using speed estimates provides better precision and sensitivity than would the relatively coarse LOS classes do, they focus on deriving speed distributions rather than LOS.

There are generally two methods available for estimating speeds. The first uses procedures from the Highway Capacity Manual (HCM). The second uses volume/capacity relationships expressed in the Bureau of Public Roads (BPR) curves (or modified BPR curves). The accuracy of both methods falls substantially when applied to arterials, due to the complications caused by controls (signalization).

5.2.3.2 Method 2 – Working with Travel Demand Models

Travel demand models (TDMs) provide another source of estimates of vehicle activity by function class, time of day, and speed. The modeling process assigns trips (defined by an origin and a destination within the roadway network) to roadway segments. To the extent that model inputs capture all trips within a region, TDMs provide comprehensive regional VMT estimates and avoid the uncertainties associated with extrapolation of traffic volumes from count data at selected locations. They provide less detail, however, regarding volume fluctuation by time of day, vehicle type, and speeds than what can be obtained from measurements, except to the extent that available data are used to provide such detail in model output.

Because of the difficulties that can arise in achieving both accurate assignments and accurate speeds in TDMs, it may be preferable to calculate speed externally. Post-processing software is available that uses HCM procedures and BPR curves to calculate hourly congested speeds and produce summaries of regional VMT distributions. The general speed post-processing algorithm operates on hourly link volumes is (even if the TDM outputs are multiple hour or daily assignments) as follows:

1. Distribute link-level volumes by hour of day using user-provided or default temporal distributions (usually from count data sets).
2. Calculate hourly VMT by multiplying link distance by hourly volume.
3. Calculate the v/c ratio using either link-specific capacities or lookup tables.
4. Apply the BPR curve, using link-specific free flow speeds or lookup tables, to arrive at hourly congested speeds.

There are several areas in which TDMs may fail to provide comprehensive VMT estimates. These relate to both the preparation of inputs used in modeling and in the level of detail incorporated in trip and network inputs.

Information on travel by vehicle class is typically not available directly in TDMs. However, as TDMs focus primarily on travel by individuals rather than goods movement, this approach provides little value for identifying medium and heavy truck activity. Goods movement models are under development, but at present, simple adjustment factors are more commonly used to estimate incremental freight-related VMT to be added to modeled volumes. Time of day, day of week, and seasonal variation of freight travel should be evaluated separately, based on local data.

5.2.3.3 Development of National Default VMT and Speed Distributions

Vehicle activity estimates derived from both traffic counts and travel demand models were used to develop distributions of VMT by functional class, speed, and time of day for five urban areas. The data were merged to the four functional classes in MOBILE6: freeways, arterials, local roads, and ramps. The five example urban areas were: Chicago, IL, Houston, TX, Charlotte, NC, Ada County ID (Boise region), and New York, NY.

Results for Chicago, Houston, and Boise were obtained using travel demand model outputs and the Caltrans Direct Travel Impact Model (DTIM2). Results for Charlotte and New York were based on traffic count data and a FORTRAN program developed for this purpose. Both methods produce hourly speed estimates based on the level of congestion (ratio of volume to capacity), roadway type, and free flow speed. In addition to these five areas, VMT and speed statistics by functional class were also obtained for three additional cities from chase car data collected by EPA and CARB, (Sierra Research, 1997). These cities were: Baltimore, MD; Spokane WA; and Los Angeles CA.

To develop national default distributions, the area-specific results are extrapolated, using the assumption that the cities for which distributions are available can be used as surrogates or prototypes for other urban areas. The distributions for these eight areas, along with Highway

Performance Monitoring System VMT data (HPMS, 1995), provided a basis for calculating a national default VMT weighting. Although the data from all eight cities are summarized, it was not possible to use data for all cities in developing national defaults because of insufficient data to determine both functional class and temporal dependence of volume and speed.

In order to develop estimates of national class, time of day, and speed, the characteristics identified for the four cities (Chicago, IL; Houston, TX; Charlotte, NC; and New York NY) for which hourly speeds could be obtained were assigned to urban area throughout the country. Urbanized area 1995 daily VMT by functional class were obtained from HPMS (1995). A "best-fit" procedure was used to select which of the four cities' characteristic temporal and speed profiles would be assigned to each urban area.

HPMS interstate and freeway/expressway classes were combined, as were arterial and collector classes to provide VMT values corresponding to the MOBILE6 functional classes. Ramp VMT was assumed to be 8.7 percent of freeway VMT. Fractional VMT for the four functional classes was then calculated for each urban area.

The temporal variation and speed distributions of VMT by functional class for Chicago, Houston, Charlotte, or New York were assigned to each HPMS urban area based on which had a functional class VMT distribution that was most similar. Similarity was determined by a "distance" calculation based on the sum of squares of the differences between fractional VMT for each functional class. The sum of HPMS functional class VMT values for all urban areas assigned to a particular prototype city was determined and was used as the prototype city's weight in calculating national VMT distributions. The following equation was used to calculate "distances" between the prototype cities and HPMS urban areas (Environmental Protection Agency, 1999a, p. 3-5):

"Distance" from HPMS urban area to prototype city

$$=((\text{fracVMT\_freeway}_{HPMS})-(\text{fracVMT\_freeway}_{\text{proto-hpms}}))^2$$

$$+((\text{fracVMT\_art/col}_{HPMS})-(\text{fracVMT\_art/col}_{\text{proto-hpms}}))^2$$

$$+((\text{fracVMT\_local}_{HPMS})-(\text{fracVMT\_local}_{\text{proto-hpms}}))^2$$

$$+((\text{fracVMT\_ramp}_{HPMS})-(\text{fracVMT\_ramp}_{\text{proto-hpms}}))^2$$

The assignment of HPMS functional class VMT to the four prototype cities is shown in Table 19. Approximately, 50 percent of total VMT occurs on arterial and collectors, 34 percent on freeways, and 13 percent on local roads. Ramp VMT is estimated as a percentage of freeway VMT. HPMS data include VMT accumulated by all vehicle types. National summary data from HPMS (HPMS, 1995) show approximately 7.8 percent of urban interstate VMT to be accumulated by buses, combination trucks, and single unit 6-tire or more trucks, and approximately 4.1 percent of other urban VMT to be attributable to these classes.

TABLE 19 Total HPMS VMT Assigned to Each Prototype City (in Thousands)

	Freeways	Arterials & Collectors	Locals	Ramps	Total
Charlotte	87631	127404	72689	7623	295348

Chicago	291757	749362	165148	25382	1231650
Houston	395167	358956	107253	34379	895756
New York	504841	626451	142653	43921	1317866
Total	1279396	1862173	487743	111307	3740620

For emission calculations using MOBILE6, both the freeway and arterial/collector functional classes are speed dependent, and default values for temporal distribution of travel may be needed to estimate congestion and speeds in urban areas. In addition, distribution of vehicle activity by time of day for all facility types is obviously needed for the preparation of hourly emission estimates, and also if diurnal temperature variations are to be used in estimating emissions. Table 20 shows the hourly distributions, using the assigned HPMS VMT values as a weighted average of the four prototype city distributions, using the assigned HPMS VMT values as weights. Since no hourly ramp data were available for any of the cities, it is reasonable to assume that hourly ramp distributions are similar to those for freeways. The distributions can be used in conjunction with the methods to estimate hourly VMT and speed distributions based on daily traffic volumes from either travel demand models or traffic count data. For national urban emissions estimation, the national VMT totals by facility type can be multiplied by the corresponding hourly fractions to obtain hourly VMT by facility type.

TABLE 20 Hourly Distribution of National VMT by Functional Class

Hour	Freeways	Arterials & Collectors	Locals
1	0.0135	0.0091	0.0098
2	0.0112	0.0070	0.0076
3	0.0108	0.0064	0.0068
4	0.0108	0.0063	0.0066
5	0.0130	0.0079	0.0081
6	0.0227	0.0162	0.0159
7	0.0652	0.0523	0.0509
8	0.0744	0.0739	0.0733
9	0.0648	0.0655	0.0679
10	0.0566	0.0549	0.0548
11	0.0546	0.0540	0.0526
12	0.0567	0.0595	0.0577
13	0.0576	0.0631	0.0614
14	0.0557	0.0580	0.0573
15	0.0584	0.0608	0.0603
16	0.0594	0.0662	0.0653
17	0.0750	0.0790	0.0804
18	0.0666	0.0764	0.0782
19	0.0432	0.0541	0.0542
20	0.0352	0.0411	0.0407

21	0.0296	0.0315	0.0313
22	0.0264	0.0263	0.0264
23	0.0216	0.0179	0.0187
24	0.0171	0.0126	0.0136

5.2.4 Fractional Split Model

Bhat and Nair (2000) proposes and implements a fractional split model that predicts the VMT mix on links as a function of the functional roadway classification of the link, the physical attributes of the link, the operating conditions on the link, and the attributes of the traffic analysis zone in which the link lies.

Several data sources are used in the analysis. These include: a) vehicle classification counts conducted in the Dallas-Fort Worth area by the Texas Department of Transportation's (TxDOT) Regional Planning Organization (R.P.O.) and the Division 10 of TxDOT, b) 1996 GIS-based road network file for the Dallas-Fort Worth area, c) Zonal level land use characteristics file of the Dallas-Fort Worth area, and d) 1996 GIS-based Dallas-Fort Worth zonal coverage file. The latter three data files were obtained from the North Central Texas Council of Governments (NCTCOG).

The model results can be applied in forecasting mode to determine the VMT mix in the six vehicle types: autos, PUVs, SUVs, trucks, buses, and motorcycles/two wheelers. The model-predicted VMT mix in the six vehicle types has to be converted into the eight-class EPA vehicle classification for input into the MOBILE5 emissions factor model. However, variations in VMT mix across different times of the day are not captured in the model. And seasonal variations in VMT mix are also not incorporated in the model. Since the fractional model is for getting VMT mix as the input of MOBILE5, it is not well ready for getting the VMT related inputs of MOBILE6.

5.2.5 HGAC practice: 24 hour assignment

The time-of-day VMT and speed estimates for the Houston-Galveston region were developed using a program called PREPIN2. PREPIN2 is one of a series of programs developed by TTI to facilitate the application of EPA's MOBILE5a Hybrid program in estimating mobile source emissions. The PREPIN2 program was developed for use in urban areas that do not have time-of-day assignments and speeds available for air quality analyses. The program inputs a 24-hour assignment and applies the needed seasonal adjustment factors. The time-of-day factors are applied to the seasonally adjusted 24-hour assignment results to estimate the directional time-of-day travel. The HGAC speed models are used to estimate the operational time-of-day speeds by direction on the links. Special intra-zonal links are defined and the VMT and speeds for intra-zonal trips are estimated. These VMT and speeds by link are subsequently input to a program called IMPSUMA for the application of MOBILE5a Hybrid emissions rates.

For the development of girded emissions, the HGAC 24-hour assignment was used as input to the PREPIN2 program. For a given application, 24 applications of PREPIN2 are run to estimate the directional VMT and speeds for each of the 24 one-hour time periods comprising the 24-hour period.

The primary output of PREPIN2 is a data set for the subject time period containing two records for each link. One record specifying the estimated time-of-day VMT and speed in the peak, or principal, direction and the second record specifying the estimated VMT and speed in the opposite direction. This data set is subsequently input to the IMPSUMA program, which applies the MOBILE5a Hybrid emissions' rates to estimate the mobile source emissions for each link. The program VMTSUM calculates the VMT by time period for input into IMPSUMA to incorporate the diurnal emissions into the appropriate time period. Finally, a program SUMALL combines the time-of-day emissions estimates to obtain 24-hour girded emissions.

5.2.6 Practices in other states

From the information collected, there appear to be several general approaches taken by other states in developing the VMT distribution:

- MOBILE defaults are used. The "default" VMT distribution in MOBILE is not actually fixed, but is a function of the user-input registration (age) distribution and the MOBILE default mileage accumulation rates by vehicle type. Georgia and Massachusetts use this approach. California's approach is similar in that state-specific registration distribution data and mileage accumulation rates (from I/M data) are used to produce a VMT mix.
- HPMS data are used to obtain light-duty vs. heavy-duty VMT percentages. EPA data (contained in the guidance document Use of Locality-Specific Transportation Data for the Development of Mobile Source Emission Inventories, and consistent with MOBILE defaults) are then used to allocate the HPMS data to MOBILE vehicle classes. Connecticut and Texas have taken this approach. Georgia has also explored the use of HPMS data and found that it gave them roughly five percent lower emissions compared to the use of MOBILE defaults. Maryland used "old state highway count" data to allocate light vs. heavy duty VMT.
- State vehicle registration data are used to develop all categories; i.e., VMT is split according to the percentage of vehicles registered within the state. New York and Delaware indicated that they used this approach. They acknowledged that this does not reflect the fact that mileage accumulation rates between heavy-duty and light-duty vehicles may differ. The state analyzes the vehicle registration database to count the number of vehicles registered by MOBILE5 vehicle type (8 categories). This provides information for the vehicle age distribution data input to MOBILE5. Then they assume that the VMT in the state is proportional to the % of vehicles registered by vehicle type.
- Other approaches such as the fuel consumption based Finance method, the policy procedure, and etc. are also used in some of the states.

The following are descriptions of the VMT estimation in some states including Colorado State, Oregon State and Wisconsin State.

5.2.6.1 Colorado Practice

Colorado has been using local VMT mix information that was collected (actual on-board counts) in Denver in the late 1980's. Several small scale counting efforts during the 90's has confirmed that this late 80's information remains relatively representative of the distribution of VMT over the fleet. Over the next six months, this information will be updated with a new study that will take place in the major metropolitan areas of the state. They are doing "cluster counting" - a system of counting vehicles at intersections multiple times and a multiple locations within the intersection. This system was devised by their Department of Transportation for VMT counting needs. They believe it will be an appropriate methodology for the needs to update VMT by MOBILE6 model vehicle types. Until this new information is available, they will probably use the Mobile6 default VMT mix distributions.

5.2.6.2 Oregon Practice

In July 2000, David Evans and Associates, Inc. (EDA) and Cambridge Systematics, Inc. (CS), consultants, evaluated the Oregon Department of Transportation's (ODOT's) existing procedures for estimating statewide VMT and to bring each of these procedures into closer alignment, with the possibility of identifying a single, effective method for estimating statewide VMT.

Three different statewide VMT estimation procedures have been developed and are utilized by ODOT for different purposes: Traffic data, Finance and Policy.

- Traffic Data procedure. The ODOT Transportation Data Section has two methods for estimating VMT based on traffic count data. The first method is the Highway Performance Monitoring System (HPMS) developed by the Federal Highway Administration (FHWA). The second method is based on the Mileage Control File (MCF) database, which provides VMT estimates for all highways on the State Highway System (SHS) as part of Oregon's Traffic Monitoring System (TMS) for Highways. ODOT did not use the MCF to estimate statewide VMT. The combination of the HPMA and MCF methods used by ODOT were referenced as the "Traffic Data procedure".
- The ODOT Financial and Economic Analysis Section estimates VMT based on fuel consumption records. This procedure is cited as the "Finance method".
- The ODOT Policy Section has historically developed statewide VMT estimates for Oregon's Highway Cost Allocation Studies (HCAS). These estimates are primarily based on developing an accurate factor for expanding SHS VMT into statewide VMT. This ODOT procedure is cited as the "Policy method".

Table 21 provides a summary of key advantages (Pros) and disadvantages (Cons) of each of the three existing statewide VMT estimation procedures used by ODOT (David Evans and Associates, 2000.) These pros and cons were identified through a coordinated effort between the consultants and ODOT.

TABLE 21 Key Advantages (Pros) and Disadvantages (Cons) of Existing ODOT Statewide VMT Estimation Procedures

Traffic Data Procedure	Finance Procedure	Policy Procedure
Pros • Used in FHWA's Highway Statistics Report and other national publications • Based on actual traffic data • Allows consistent comparison between states • HPMS is only method that procedures data by roadway functional class and area type • HPMS has info. For non-SHS • Used for other purposes in addition to statewide VMT	*Pros* • Consistent with ODOT revenue estimates • Requires relatively few data inputs • Heavy vehicle VMT based on actual reported mileage • Effective method for long-range forecasts and consistent with forecast revenues	*Pros* • Same SHS VMT estimate as MCF • Provides info. by jurisdictional class and estimated and projected VMT by vehicle type
Cons • Complex, data-intensive and resource-intensive methods • Counts are only taken once every three years • HPMS provides limited data for Rural Minor Collectors, and Urban and Rural Local roads	*Cons* • Does not allow for consistent comparison between states • Does not procedure data by roadway functional class, jurisdictional class, or vehicle type • Fuel economy for medium-heavy vehicles based on 1992 data • Relies on several assumptions and data collected from other agencies	*Cons* • Does not allow for consistent comparison between states • Dependent upon increasingly outdated data • Continued use of fitted statewide VMT to SHS VMT ratio will lead to decreased SHS % of statewide VMT • Does not provide info. by roadway functional class

5.2.6.3 Wisconsin Practice

WisDOT develops estimates of statewide VMT based on three independent approaches:

- A fuel-based approach that provides a direct estimate of statewide VMT based on gasoline and diesel fuel consumption in the state multiplied by auto and truck fleet fuel efficiency (MPG) estimates. WisDOT uses the statewide VMT total from this fuel-based method as their control total.
- A traffic count-based method that uses the traffic count information available from automatic traffic recorders (ATRs) located around the state to estimate the percent change in AADT weighted by functional classification (except for Locals and Rural Minor Collectors). WisDOT currently has a total of 146 ATRs located throughout the state, including nine on non-state highways. Of these 146 ATRs, WisDOT typically ends up with complete data (without the effects of highway construction or detours or equipment-related problems) from approximately 100 of the ATRs to get a percent change comparison over two years. They compare the changes in AADT levels at these ATRs, summed at an aggregate functional class (at least 10 and hopefully 30 or more), weight the functional classification levels by the proportion of VMT they carry (from the previous year's HPMS results), and arrive at a statewide weighted percent change estimate from the previous year. Since WisDOT has very few ATRs located on the lower functionally classified highways, however, they have little information about VMT changes on local roads.
- A second count-based approach uses the annual change shown for the interstate, freeway, and arterial and collector highways in the state from the annual HPMS VMT estimates. Since they take 48-hour coverage traffic counts on virtually every segment of state highway on a three-year cycle, WisDOT uses the HPMS Universe rather than the HPMS Sample Segments. With only one-third of the counts current, however, the other two-thirds get growth factored up to the current year.

WisDOT estimates total VMT for all Rural Minor Collectors (also not required for HPMS) directly from the local roads files that contain AADT estimates for each segment.

CHAPTER 6

IMPROVEMENT TO ESTIMATION OF VMT RELATED VARIABLES

In Chapter 5, the information collection, methodology evaluation and traffic count collection for estimation of VMT related variables were summarized. This chapter emphasis on improvements of estimation of VMT related variables.

6.1 Methodology Description

Current Traffic Count Method simply extends the traffic on the count station onto other links without any consideration of the link attributes. The fractional split model only considers link-attributes by setting up relationships between link-attributes and VMT mix, with no consideration of useful traffic counts. The improved methodology strives to set up the relationships between link volume and the count data as well as their link attributes. The disaggregating of volume in hour of day and speed will follow the EPA Traffic Count Method since it is a useful and the only method that can disaggregate volume according to the requirements by MOBILE6.

6.1.1 Volume Estimation

The important improvement to the estimation of VMT related variables is the estimation of link traffic volume. Suppose $c_{kt}^{i_y}$ is the volume for vehicle type t at count station i_y, where $i_y = 1, 2, ..., I_y$, I_y is the total number of count stations on road type y, y is the road type, t is the vehicle type.

On link k, volume for vehicle type t with road type y can be estimated from any of the count station i_y located on the same road type (*i.e.* type y):

$$v_{kt}^{i_y} = c_{kt}^{i_y} \cdot e^{\alpha_{kt}^0 + \sum_{j=1}^{n_x} \alpha_{kt}^j \cdot x_{kt}^j} \tag{6-1}$$

where, α_{kt}^j $(\forall j = 0, 1, 2, \dots n_x)$ is the unknown coefficient to be calibrated from simulation results; n_x is the total number of link attribute types to be included in the estimation process; x_{kt}^j is the j-th link attributes.

To calibrate the coefficients α_{kt}^j $(\forall j = 0, 1, 2, \dots n_x)$, Equation (6-1) can be transferred into the following form:

$$\ln\left(v_{kt}^{i_y}\right) - \ln\left(c_{kt}^{i_y}\right) = \left[1, x_{kt}^1, \dots, x_{kt}^{n_x}\right] \cdot \left[\alpha_{kt}^0, \alpha_{kt}^1, \dots, \alpha_{kt}^{n_x}\right]^{\mathrm{T}} \tag{6-2}$$

Let $\mathbf{Y}_{kt}^{i_y} = \ln\left(v_{kt}^{i_y}\right) - \ln\left(c_{kt}^{i_y}\right)$, $\mathbf{X}_{kt}^{i_y} = \left[1, x_{kt}^1, \dots, x_{kt}^{n_x}\right]$, and $\mathbf{A}_{kt}^{i_y} = \left[\alpha_{kt}^0, \alpha_{kt}^1, \dots, \alpha_{kt}^{n_x}\right]^{\mathrm{T}}$, then the following simpler form can be obtained:

$$\mathbf{Y}_{kt}^{i_y} = \mathbf{X}_{kt}^{i_y} \cdot \mathbf{A}_{kt}^{i_y} \tag{6-3}$$

The calibrated coefficient matrix $\hat{\mathbf{A}}_{kt}^{i_y} = \left[\hat{\alpha}_{kt}^0, \hat{\alpha}_{kt}^1, \dots, \hat{\alpha}_{kt}^{n_x}\right]^{\mathrm{T}}$ in Equation (6-3) can then be obtained by a multivariate regression analysis using any standard routine.

After the calibration process, the volume on link k for vehicle type t can be estimated based on the count data from link i_y:

$$\hat{v}_{kt}^{i_y} = c_{kt}^{i_y} \cdot e^{\left[1, x_{kt}^1, \dots, x_{kt}^{n_x}\right]\left[\hat{\alpha}_{kt}^0, \hat{\alpha}_{kt}^1, \dots, \hat{\alpha}_{kt}^{n_x}\right]^{\mathrm{T}}} \tag{6-4}$$

Since there are I_y count stations, a total of I_y estimated values for the same link volume can be obtained. The final estimated volume $\hat{v}_{kt}$ could then be estimated as an average of all these I_y estimations:

$$\hat{v}_{kt} = \frac{1}{I_y} \sum_{i_y=1}^{I_y} \hat{v}_{kt}^{i_y} \tag{6-5}$$

In Equation (6-1), the link attributes that may be included in the estimation process can be link width, link length, link speed, land use types, urbanized types, etc. All of these link attributes should be quantified beforehand for the convenience of calibration and calculation. The more link attributes that are related to VMT information, the more precision can be obtained during the estimation process.

6.1.2 Volume and VMT Disaggregating

After obtaining the estimated traffic volume for all the links, the link volume should be disaggregated according to the hours of day so that hourly VMT can be obtained. By applying the BPR curve, the speed VMT can be estimated as well.

These disaggregating processes of the methodology are conceptually straightforward, although their calculations might be relatively complex. To obtain the hourly VMT, the distribution of link-level volumes by hour of day should be prepared by using the user-provided distribution. If the user does not provide this kind of information, or the user can only provide this kind of distribution for some particular links (e.g. only for some freeway, or some arterial road), the default temporal distribution can be applied to the links, where local distributions are missing. By multiplying the link distance with the hourly volume, the hourly link VMT can then be obtained directly.

In order to get the hourly speeds for the use of obtaining speed VMT, the v/c ratios need to be calculated beforehand in case the BPR (Bureau of Public Roads) curve (or modified BPR curve) is to be applied. The reason why BPR is preferred is that BPR is not data intensive while HCM approach needs more local information.

So the hourly-congested speeds can be achieved by applying the BPR curve, the standard form of which is:

$$s_k = s_{kf} \Big/ \left(1 + a_k \left(v_k / c_k\right)^{b_k}\right) \tag{6-6}$$

where: s_k is a predicted mean speed on link k, s_{kf} is the free-flow speed on link k, v_k is the hourly volume on link k, c_k is the practical capacity on link k, and a_k and b_k are parameters related to the local traffic flow characteristics. It is suggested in EPA guidance that for signalized facilities (arterials, collector, and local), the parameter a_k can be chosen as 0.05, and for unsignalized facilities (freeways, highways, and expressways), the parameter a_k can be chosen as 0.20. Under the both situations, the parameter b_k can be chosen as 10.

Free-flow speed is defined as the space mean speed of traffic when volumes are so light that they have negligible effect on speed and is estimated to be 1.115 times the speed at capacity. Dowling *et al* (1994) set up relationships for free-flow speed s_{kf} (in mph) on link k as follows:

$$s_{kf} = \begin{cases} 0.88 \cdot s_{kp} + 14, & \forall s_{kp} > 50 \\ 0.79 \cdot s_{kp} + 12, & \forall s_{kp} < 50 \end{cases} \tag{6-7}$$

Practical capacity is defined as 80% of maximum capacity. The v/c ratios can be calculated by using the estimated link hourly volumes and link-specific capacities. The effects of signal control will reduce the accuracy of the speed - v/c relationships for arterial and local roads. Nevertheless, the BPR curves may still be a practical approach for estimating arterial speeds, unless local data on control parameters by facility and area type are available to at least construct look-up tables.

After having obtained the hourly distributions of VMT by speed for all types of facilities, it is easy to get the other VMT related variables required in MOBILE6. Therefore, VMT BY FACILITY; VMT MIX; VMT by HOUR and VMT BY SPEED can all be generated.

6.2 Implementation Procedure

The improved approach to estimate VMT related variables for MOBILE6 can be summarized into the following six steps:

Step 1: Estimate the volume at the links where no traffic counts are collected;

Step 2: Distribute link-level volumes by hour of day using user-provided or default temporal distributions (usually from count data sets);

Step 3: Calculate hourly VMT by multiplying link distance by hourly volume;

Step 4: Calculate the v/c ratio using either link-specific capacities or lookup tables;

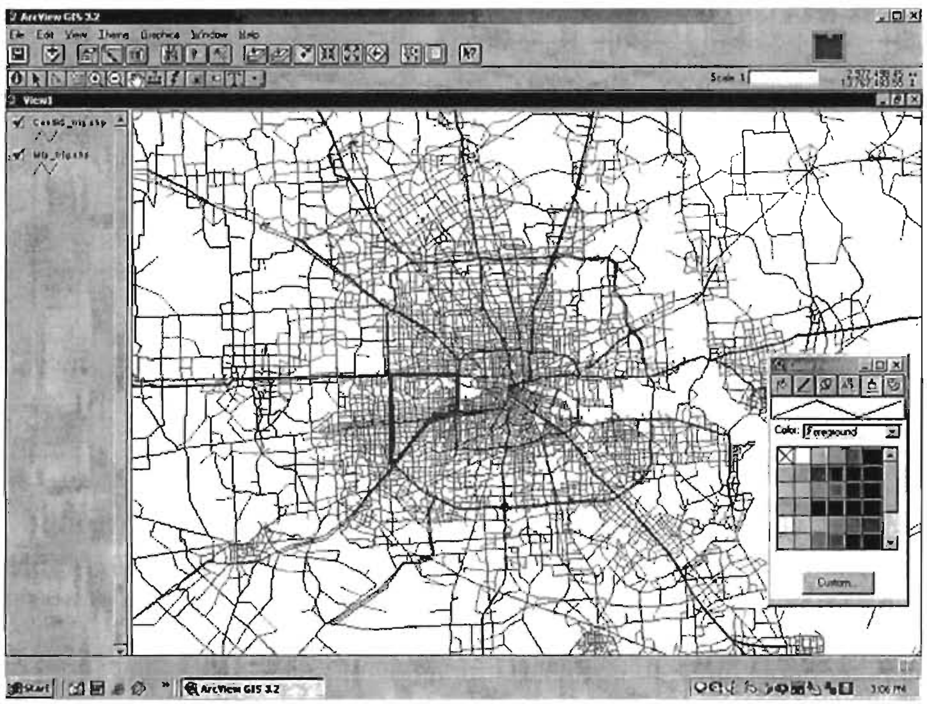
Step 5: Apply the BPR curve, using link-specific free flow speeds or lookup tables, to arrive at hourly congested speeds; and

Step 6: Obtain all VMT related variables required by MOBILE6.

The above procedures are similar to that for Traffic Count Method except the important new features in Step 1, where the link traffic volume is estimated by Equation (6-1) instead of the simple extension of the count volume from the station.

6.3 Case Study in Southwest Houston

To illustrate the proposed improvement, a case study was conducted in the southwest Houston, Texas, which is shown in Figure 23. The selected sub-network contains 276 links with 34 freeway links, 110 arterial links and 132 local street links. While there is no ramp information available, which is also the case for most of the local situations, the final ramp VMT can be a portion of the estimated freeway VMT. The EPA guidance suggests this portion as 8.7%.



Note: illustrates the area for case study

FIGURE 23 GIS network for Houston area in the format of ArcView.

6.3.1 Information Collected

Information needed for illustration of the improved methodology include the link traffic count data and the link attributes information. The link traffic count data were collected from the 1996 traffic map for Harris County from TxDOT, while the link attributes information were based on the descriptions embedded in the GIS network data regarding the 1996 Houston GIS data from Houston – Galveston Area Council.

The 1996 traffic map for Harris County was prepared by the Transportation Planning and Programming Division of Texas Department of Transportation, in cooperation with the US Department of Transportation, Federal Highway Administration. All counts on the map were 24 hour axle counts divided by 2. They were generally made between the dates of September 9, 1996 and December 16, 1996 with the exception of a few ties in counts between bordering counties from other TxDOT districts.

The 1996 Houston GIS network data were stored in the format of ArcView. It contains both the network itself and its corresponding network information. Figure 23 shows the GIS network of whole Houston area in the format of ArcView, where different facility types can be displayed in different colors. In the database, there is the information for the link attributes, such as: number of lanes, link width, link length, mean speed, night speed limit, land use type, etc. This information can be used directly for the calibration and estimation.

6.3.2 Link Volume and VMT Estimation

The link volume estimation started from the calibration of Equation (6-1) or (6-3). Based on the information available, a total of 4 link attributes were selected in this case study, namely link length, mean speed, night speed limit, and land use type. The selection of link attributes in other areas may not follow this. However, the selected ones should be sensitive to the link volume as well as the resulting VMT estimation. According to the information collected, there are 6 land use types, which were quantified into digits 1 to 6 for the convenience of later calibration. Except the land use type, all the other link attributes have already been quantified when the data were archived.

Among the 276 links, 12 freeway links, 15 arterial links and 15 local street links were assigned the count data. This means that it is supposed that the volumes in these 42 links were treated as real data while the rest 234 links were left blank, which might be estimated by the proposed methodology.

The calibrations of coefficients were conducted among the "known" count data, i.e. the 42 links where count data were available. Following the regular regression process, the calibrated coefficients in the Equation (6-1) or (6-3) are listed in Table 22. In Table 22, the calibrated coefficients for freeway, arterial, and local streets are different, meaning that the different relationships existed inside.

TABLE 22 Coefficients for Volume Estimation Model Calibrated by Southwest Houston Real Data

Coefficients	Facility Types		
	Freeway	Arterial	Local Street
α_0	-8.6×10^{0}	-9.2×10^{-1}	2.7×10^{0}
α_1	-3.6×10^{-6}	1.8029×10^{-5}	-5.8×10^{-5}
α_2	-3.0×10^{-4}	-4.2×10^{-2}	5.6×10^{-2}
α_3	1.5×10^{-1}	4.8×10^{-2}	-1.4×10^{-1}
α_4	-1.0×10^{-2}	-3.9×10^{-3}	-3.5×10^{-2}

The estimation of traffic volume can be realized based on Equations (6-4) and (6-5). Equation (6-5) is necessary since there were more than one link with traffic count data for each of the three facility types. The average to the different estimations on the same link was the final estimated link volume.

The link VMT could then be obtained by multiplying the estimated link volume with its corresponding link length. By aggregating the total VMT subjected to the same type of facility, the total VMT for this facility would easily be obtained.

To validate the proposed improvement, a total of 4 scenarios were conducted to compare the estimation results. Scenario 1 calculated the real VMT on all links, which is treated as an ideal one. Scenario 2 estimated the VMT on each links based on the EPA Traffic Count Method. The reason why this method was selected is that this is a good method and has been recognized by EPA and public. Scenario 3 used the proposed improvement to estimate link VMT. In Scenario 3, link VMT was estimated by the calibrated coefficients and link volume estimation equations. Considering the fact that in some local areas, the traffic count data are not available for some facility types such as local streets, Scenario 4 was set up. Scenario 4 assumes that only freeway count data were available and only coefficients for freeway link volume estimation were calibrated. The link volumes for other facility types must also use the calibrated coefficients for freeway.

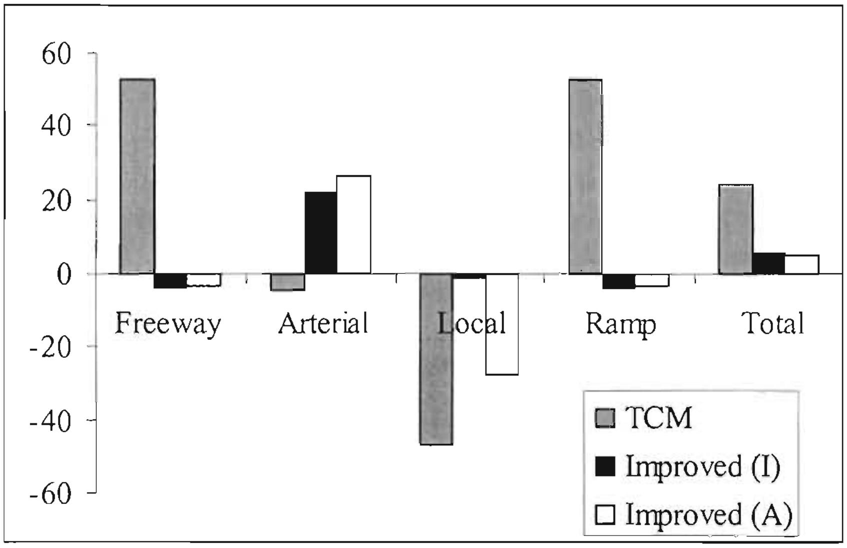
Table 23 lists the total VMT estimates by all facility types based on different scenarios in southwest Houston area. In Table 23, the four scenarios were marked as "Ideal", "TCM", "Improved (I)" and "Improved (A)" for better understanding the meanings. It is shown that the estimates of Total VMT based on both Scenario 3 and 4 (i.e. both use improved methods) were closer to the ideal one than the estimates on Scenario 2 (TCM method).

TABLE 23 Comparison of Total VMT by Facility Types Based on Different Methods in Southwest Houston Area

VMT	Freeway	Arterial	Local	Ramp	Total
Ideal	1.15×10^{10}	7.04×10^{9}	1.85×10^{9}	1.00×10^{9}	2.14×10^{10}
TCM	1.76×10^{10}	6.73×10^{9}	9.84×10^{8}	1.53×10^{9}	2.68×10^{10}
Improved (I)	1.11×10^{10}	8.59×10^{9}	1.83×10^{9}	9.66×10^{8}	2.25×10^{10}
Improved (A)	1.12×10^{10}	8.88×10^{9}	1.34×10^{9}	9.74×10^{8}	2.24×10^{10}

Note: Improved (I) means that the improved model calibrated each facility type independently
Improved (A) means that the calibrated model for freeway was used to links of all facility types

To better compare the results on the 4 scenarios, the relative errors of the estimates on each scenario (except scenario 1) were calculated and listed in Figure 24. In Figure 24, it is obvious that TCM resulted in the largest relative errors among all the scenarios, especially for freeway, ramp and total VMT.



Note: Improved (I) means that the improved model calibrated each facility type independently; Improved (A) means that the calibrated model for freeway was used to links of all facility types

FIGURE 24 Comparison of relative errors (%) of total VMT estimation by facility types based on different methods in Southwest Houston area.

Based on the above estimation, the VMT fractions on all facility types can be calculated, and the results were listed in Table 24. Again, the VMT fractions on facility types by improved two methods were closer to the ideal one.

TABLE 24 Comparison of VMT Split by Facility Types Based on Different Methods in Southwest Houston Area

VMT	Freeway	Arterial	Local	Ramp	Total
Ideal	0.539	0.329	0.086	0.047	1.000
TCM	0.656	0.251	0.037	0.057	1.000
Improved (I)	0.493	0.382	0.082	0.043	1.000
Improved (A)	0.499	0.397	0.061	0.043	1.000

Note: Improved (I) means that the improved model calibrates each facility type independently; Improved (A) means that the calibrated model for freeway was used to the links of all facility types.

After estimating the link volume and VMT by facility type, the next parts just followed the steps in the section on "Implementation Procedure," which is similar to the process of Traffic Count Method. After disaggregating facility VMT into hours and speed distribution, all the MOBILE6 required parameters were obtained.

6.3.3 Impact on Emission Estimation

The final estimated local VMT related variables were input into MOBILE6 so that the impacts on emission estimations could be obtained. Table 25 lists the estimates for three emission factors VOC, CO and NO_x based on different scenarios. Figure 25 illustrates the relative errors for the different methods. From both Table 25 and Figure 25 it is shown that the proposed improvements have better estimation on the emission factors. They are better than both the nationwide default one and TCM estimation. It is interesting to note that all the three emission factors for the both improved methods are smaller than the real one although all the relative errors are relatively small comparing with the other two methods. For default values, CO and NO_x are much smaller than the real one, while VOC is bigger. For TCM, CO and NO_x are much bigger than real ones while VOC is much smaller.

TABLE 25 Emission Factors Estimates Based on Different Methods of VMT Estimation in Southwest Houston Areas

	Ideal	TCM	Improved (I)	Improved (A)
VOC	0.467	0.448	0.459	0.454
CO	5.259	5.459	5.304	5.309
NO_x	0.904	0.963	0.927	0.927

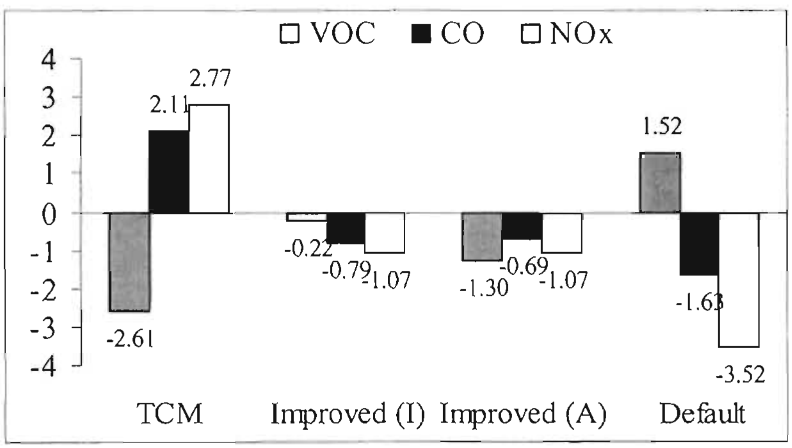


FIGURE 25 Relative errors (%) for estimates of emission factors based on different scenarios.

CHAPTER 7

CONCLUSION

In this research report, the state-of-the-art and the state-of-the-practice related to the project were conducted and a large amount of literatures were reviewed. Data collecting and modeling for vehicle registrations and mileage accumulation were made. The collection of information on VMT mix estimation was also carried out to form the basis for the improvements of VMT mix estimation methodologies.

As for modeling the vehicle age distribution, two model types were used; each of which contains the linear model, nonlinear model and time series model. The age distribution forecasting model MOFAD has been tested for 8 HGAC counties and in El Paso. The performance of the model was validated in several different aspects.

Goodness-of-fit has been performed to check the extent to which predictions agree with observed data. It has been found that, for most counties except for Chambers County, the average relative errors are relatively small. Based upon the fact that there are only 7 years data for modeling, the results are reasonable. More reliable model is anticipated in case data for more years is available. Sensitivity analysis was performed to identify the importance of parameters and independent variables. For each test area, four most frequently used socioeconomic indices were summarized. Suboptimal selections of three predictable socioeconomic indices were used for modeling and forecasting. Comparisons between suboptimal and optimal selection have been performed. Analytical results show that MOFAD can still work well when suboptimal selected socioeconomic indices were input.

Furthermore, impacts of changes of MOFAD on MOBILE6 outputs were also analyzed. For most of the vehicle types, the emission factors change obviously when MOFAD results were used to substitute the default age distribution. However, the impacts of emission factors for different selections of independent variables (optimal selection and suboptimal selection), and of forecasting methods (method I and II) are not so significant.

It should be noted that the prediction of age distribution by the proposed model contains more information, including socioeconomic indexes and local distribution in the recent years. It is not based on the simple expending of the current trends of vehicle age distribution. The basic idea is to build the relationship between the socioeconomic indexes and vehicle age distribution. In the future years, vehicle age distribution can be properly predicted providing the socioeconomic indexes are provided. This is the only modeling effort of this type of problem till now.

The modeling of the correcting process for mileage accumulation was developed mathematically in this report. The adjusting algorithm is developed for obtaining local mileage accumulation rates based on small sample survey. In the case of small sample survey, although the individual survey results cannot be directly used as the local mileage accumulation rates, the entire survey's result is valuable and contains information that can be used to estimate the local mileage accumulation rates. The proposed algorithm makes full use of both the local survey results and the nationwide default ones. It is a practical and feasible way for some of the local juristic areas although it may not be the optimal and unique one.

Case studies in Houston and El Paso areas illustrate the whole operation process and the impacts on the estimates of emission factors were shown. From the results, the real mileage accumulation in Houston area is 1.85 times higher than the national-wide default value, and in El Paso it is 0.55 times than the default value.

According to the information collected, there are several methodologies on VMT mix estimation till now. EPA gives a guidance involving the development and application of methods to estimate detailed national wide VMT related variables. The results serve as the national default values. Bhat and Nair (2000) formulate and estimate a fractional split model. In develop the methodology used for the Houston-Galveston Nonttainment Counties gridded mobile source emissions inventories for FY 2007, the 24-hour traffic assignment are used in the analysis to obtain the VMT mix, which can be used as the input of MOBILE5. For the practices in the other states, some use the MOBILE defaults, some use the HPMS traffic count data, some estimate according to the percentage of vehicles registered within the state, some use the fuel consumption based finance method, the policy procedure, and etc.

In this research, the improvements to the VMT estimation were proposed which considered both the link attributes and the traffic count information. The proposed model for volume estimation is easy to be calibrated. Case study and model calibration in southwest Houston show that the improved approach is better than both the EPA Traffic Count Method and the nationwide MOBILE6 defaults in terms of the estimation of both VMT and emission factors. In order to apply the improved approach to the real-world networks, more calibrations and validations under various environments are necessary, which will be the next step of this research.

REFERENCES

AASHTO, "Guidelines for traffic data programs," AASHTO, Washington, D.C., 1992.

Albright D., "A Quick Cluster Control Method: Permanent Control Station Cluster Analysis in Average Daily Traffic Calculations," *Transportation Research Record* 1134, pp. 57-64.

Albright D., "History of Estimating and Evaluating Annual Traffic Volume Statistics," *Transportation Research Record* 1305, pp. 103-107.

Ang H.-S. and W. H. Tang. *Probability Concepts in Engineering Planning and Design*. John Wiley & Sons, Inc. 1975.

ASTM, "Standard Practice for Highway-traffic Monitoring," E. 1442. ASTM, West Conshohocken, Pa., 1994.

Bester C.J., J.D. De, and B. Jouber, "Classifying a Rural Road Network for Traffic Counting," *Transportation Research Record* 1090, pp. 22-26.

Bhat C.R. and H.S. Nair, "VMT Mix Modeling For Mobile Source Emissions Forecasting: Formulation And Empirical Application," Presented at *79th TRB Annual Meeting*. Washington, D.C., January 2000.

Bureau of Economic Analysis, United States Department of Commerce. http://www.bea.doc.gov. Accessed February 20, 2001.

Cambridge Systematics Inc., Eastern Research Group, and Radian Corporation, "Volume IV: Chapter 2, Use of Locality-Specific transportation Data for the Development Mobile Source Emission Inventories," Mobile Source Committee, Emission Inventory Improvement Program. Sept. 1996. Website: *http://www.epa.gov/ttnchie1/eiip/techreport/volume04/iv02.pdf* (Accessed Date: Nov. 08, 2001).

Center for Urban Transportation Studies, University of Wisconsin-Milwaukee, "Guidebook on Statewide Travel Forecasting," March 1999.

Chatterjee, A. et al. *Improving Transportation Data for Mobile Source Emission Estimates*. National Cooperative Highway Research Program Report 394, Transportation Research Board, 1997.

Chin S.-M., D.L. Greene, J. Hopson, H.L. Hwang, and B. Thompson, "Toward National Indicators of VMT and Congestion Based on Real-time Traffic Data," *http://micron4.ed.ornl.gov*, 1999.

Chung J.-H., K. Viswanathan, and K. G. Goulias, "Design of an Automatic Comprehensive Traffic Data Management System for Pennsylvania," presentation at *the 77th Annual Meeting of Transportation Research Board* (Washington D.C., January 1998).

Cochran, M., "Sampling Techniques," 3rd ed., Wiley, New York, 1977.

Crooper, G. R., and C. D. McGillem, "Probabilistic Methods of Signal and System Analysis," 3rd Edition, Oxford University Press, Inc. New York, USA, 1999.

David Evans and Associates, Inc. and Cambridge Systematics, Inc., 2000, "Oregon Department of Transportation Statewide Vehicle Miles of travel (VMT) Study," Prepared for Oregon Department of transportation (ODOT). July 2000.

Davis G. A., "Accuracy of Estimates of Mean Daily Traffic – A Review," *Transportation Research Record* 1593, pp. 12-16.

Davis G.A. and Y. Guan, "Bayesian Assignment of Coverage Count Locations to Factor Groups and Estimation of Mean Daily Traffic," *Transportation Research Record* 1542, pp. 30-37.

Direnzo J. N. Serianni and L. King, "Improving the Productivity of the Pennsylvania Transportation Department's Traffic Count Program," *Transportation Research Record* 1050, pp. 13-20.

Dowling. R. *Technical Memorandum 8-1 User's Guide for Speed Processor*. Prepared for the California Department of Transportation, Dowling Associates, June 14, 1994.

Dresser G.B.; T. Qu and C. E. Bell, "Gridded Mobile Source Emissions Estimates for the Houston-Galveston Nonattainment Counties to Support the Attainment Demonstration SIP," *TTI Study* No.: 402011-23, Texas Transportation Institute. December 2000.

Dresser, G. B., and C. E. Bell, "Development of Gridded Mobile Source Emissions Estimates For The Houston Galveston Nonattainment Counties," Report No. TRNCC/TX-98-402001, Texas Transportation Institute, Dec. 1998.

Environment Protection Agency. *User's Guide to MOBILE5 (Mobile Source Emission Factor Model)*. EPA-AA-AQAB-94-01. United States Environment Protection Agency, May, 1994.

Environment Protection Agency. *Update of Fleet Characterization Data for Use in MOBILE6 – Final Report*. EPA420-P-98-016. United States Environment Protection Agency, June. 1998. http://www.epa.gov/oms/m6.htm. Accessed July 18, 2001.

Environmental Protection Agency (1992), "Section 187 VMT Forecasting and Tracking Guidance," *http://www.bts.gov/ntl/DOCS/vmt.html*, January 1992.

Environmental Protection Agency (1998), "Update of Fleet Characterization Data for Use In MOBILE6 – Final Report," United States Environmental Protection Agency, Office of Mobile Source, Publication No. EPA420-P-98-016, June 1998.

Environmental Protection Agency (1999a), "Development of Methodology for Estimating VMT Weighting by Facility Type, Draft," Report EPA420-P-99-006, M6. SPD. 003. February 1999.

Environmental Protection Agency (1999b), "Guidance for the Development of Facility Type VMT and Speed Distributions, Draft," Report EPA420-P-99-004, M6. SPD. 004. February 1999.

Environmental Protection Agency (1999c), "Analysis of Emissions Deterioration Using Ohio and Wisconsin IM240 Data," United States Environmental Protection Agency, Office of Mobile Source, Publication No. EPA420-P-99-013, March 1999.

Environmental Protection Agency (1999d), "Fleet Characterization Data for MOBILE6: Development and Use of Age Distributions, Average Annual Mileage Accumulation Rates and Projected Vehicle Counts for Use in MOBILE6," United States Environmental Protection Agency, Office of Mobile Source, Publication No. EPA420-P-99-011, April 1999.

Environmental Protection Agency (2000), "User's Guide to MOBILE6: Mobile Source Emission Factor Model, DRAFT," Report EPA420-P-00-005. November 2000.

Environmental Protection Agency (2001a), "Draft User's Guide to MOBILE6," EPA420-D-01-002a. United States Environmental Protection Agency, Feb. 2001. http://www.epa.gov/oms/m6.htm (Accessed July 18, 2001).

Environmental Protection Agency (2001b), "Fleet Characterization Data for MOBILE6: Development and Use of Age Distributions, Average Annual Mileage Accumulation Rates and Projected Vehicle Counts for Use in MOBILE6," EPA420-P-99-011. United States Environmental Protection Agency, Apr. 1999. http://www.epa.gov/oms/m6.htm (Accessed July 18, 2001).

Environmental Protection Agency (2001c), "Update of Fleet Characterization Data for Use in MOBILE6 – Final Report," EPA420-P-98-016. United States Environmental Protection Agency, June. 1998. http://www.epa.gov/oms/m6.htm (Accessed July 18, 2001).

Erlbaum N.S., "Improving Air Quality Models in New York State: Utility of the 1995 Nationwide Personal Transportation Survey," May 13, 1999.

Estimating vehicle miles of travel by using random sampling techniques. Transportation Research record 779, Transportation Research Board, Washington, D. C., 6-10, 1980.

Faghri A. and J. Hua, "Roadway Seasonal Classification Using Neural Networks," *Journal of Computing in Civil Engineering*, Vol. 9, No. 3, July 1995, pp. 209-215.

Faghri A. and P. Chakroborty, "Development and Evaluation of a Statistically Reliable Traffic Counting program," *Transportation Planning and Technology*, 1994, Vol. 18, pp. 223-237.

Federal Highway Administration, U.S. Department of Transportation, "Highway Performance Monitoring System (NPMS) Field Manual," Federal Highway Administration, U.S. Department of Transportation, http://www.fhwa.dot.gov/legsregs/directives/orders/m56001b.htm.

Federal Highway Administration, U.S. Department of Transportation, *Traffic Monitoring Guide,* Third Edition, February 1995.

Federal Highway Administration, U.S. Department of Transportation, "Participant's Notebook," *Workshop on transportation air quality analysis*, FHWA-141-94-011, September 1994.

Flavelle, P. A. Quantitative Measure of Model Validation and its Potential Use for Regulatory Purpose. *Advances in Water Resources*, Vol. 15, No.1, January 1992, pp. 5-13.

Hallenbeck M.E., "Development of an Integrated Statewide Traffic-Monitoring System," *Transportation Research Record* 1050, pp. 5-12.

Heiken J. G. and et al. Methodology for Gathering Locality-Specific Emission Inventory Data. EVIRON Corporation and Sonoma Technology, Inc. Prepared for US Environmental Protection Agency, June 19, 1996.

Hoang, L.T., and V. P. Poteat, "Estimating vehicle miles of travel by using random sampling techniques," *Transportation Research Record* 779, Transportation Research Board, Washington, D. C., 6-10, 1980.

HPMS, "Urbanized Area Summaries: Mileage and Daily Vehicle-Miles of Travel (TABLE HM-71)," and "Highway Usage and Performance: Vehicle Miles of travel by Highway Category, Vehicle Type, and Related Data (TABLE VM-1)," Internet address: *http://www/fhwa.dot.gov/ohim/1994/section5.htm*

Hu P., "ITS as An Alternative Data Resource: A Demonstration Project with Florida and New York Data," Oak Ridge National Laboratory, Presented at *the 9th Annual Meeting of ITS America* (Washington, D.C. April 21, 1999).

Hu P.S. S.C. Davis, and R.L. Schmoyer, "Registrations and Vehicle Miles of Travel of Light-Duty Vehicles," 1985-1995, ORNL-6936.

Information Services, Oregon State University. *Government Information Sharing Project.* http://govinfo.kerr.orst.edu/index.html. Accessed Feb. 20, 2001

Jaraiedi A., "Statistical Procedures to Determine Seasonal Factors for Traffic Volume Monitoring in West Virginia, Final Report to West Virginia Department of Highway,) 1997.

Jones, S.L. Jr. and E.D. Arnold, Jr., "Investigation of Traffic Count Procedures on Unpaved Roads-VTRC, Final report," Virginia Transportation Research Council, Charlottesville, Virginia, VTRC 95-R19, April 1995. *http://www.bts.gov/ntl/DOCS/433VTRC.html.*

Lee-Gosselin M.E.H., and A.J. Richardson, "An approach to developing comparable national estimate of vehicle travel using traffic count data," Proc., *16th PTRC Transport and Planning Annual Meeting – Seminar D.* 1988.

Lingras P., "Classifying Highways: Hierarchical Grouping Verse Kohonen Neural Networks," *Journal of Transportation Engineering*, Vol. 121, No. 4, July/August, 1995, pp. 364-368.

Ljung, L., "System Identification: Theory for the User," 2nd Edition, Prentice-Hall, Inc., Upper Saddle River, New Jersey, USA, 1999.

Memmott J.L., "Adequacy of the Sample Size and Accuracy of the Highway Performance Monitoring System for Use in Texas," *Transportation Research Record* 1134, pp. 32-39.

Miller, T. L., W.T. Davis, G.D. Reed, P. Doraiswamy, A. Tang, and P. Sanhueza, "Corrections to Mileage Accumulation Rates for Older Vehicles and The Effect on Air Pollution emissions," Department of Civil and Environmental engineering, presentation at *the 80th Annual Meeting of Transportation Research Board* (Washington D.C., January 2001).

Mohamad D., K.C. Sinha, T. Kuczek, and C.F. Scholer, "An Annual Average Daily Traffic Prediction Model for County Roads," presentation at *the 77th Annual Meeting of Transportation Research Board* (Washington D.C., January 1998).

Peyrebrune H.L., "Strategic Reassessment of the Highway Performance Monitoring System," *FHWA*, Publication Bo. FHWA-PL-98-011, September 1997.

Pearson D.F., S.S. Frankel. Improved Forecasts of Vehicle Size and Age Distribution and the Impact on Energy Consumption and Vehicle Emissions, SWUTC/93/60030-1. Project No. 460030. Southwest Region University Transportation Center, College Station, TX. December 1993.

Qiao, F., L. Yu, G. Li, and X. Wang. *Modeling Vehicle Age Distribution for Air Quality Analysis.* Preprints of *the 81st Annual Meeting of Transportation Research Board* (CD-ROM), Washington D.C., January 2002.

Räty P. and P. Leviakangas, "Estimating Vehicle Kilometers of Travel Using PPS Sampling Method," *Journal of Transportation Engineering*, Vol. 125, No. 1, January/February, 1999.

Ritchie S.G. and M.E. Hallenbeck, "Evaluation of a Statewide Highway Data Collection Program," *Transportation Research Record* 1090, pp. 27-35.

Ritchie, S.G., "A Statistical Approach to Statewide Traffic Counting," *Transportation Research Record* 1090, pp. 14-21

Schmoyer R. P.S. Hu, and P. Swank, "Analysis of Vehicle Classification and Truck Weight Data of the New England States, Final Report," Oak Ridge National Laboratory for U.S. Department of Energy, 1996.

Sharma S.C. and A. Werner, "Improved Method of Grouping Provincewide Permanent Traffic Counters," *Transportation Research Record* 815, 12-17.

Sharma S.C. and R.R. Alllpuram, "Duration and Frequency of Seasonal Traffic Counts," *Journal of Transportation Engineering*, Vol. 119, No.3, May/June, 1993, pp. 344-359.

Sharma S.C. and Y. Leng, "Seasonal Traffic Counts for a Precise Estimation of AADT," *ITE Journal*, September 1994, pp. 21-28.

Sharma S.C., B.M. Gulati, and S.N. Rizak, "Statewide Traffic Volume Studies and Precision of AADT Estimates," *Journal of Transportation Engineering*, Vol. 122, No. 6, November/December, 1996, pp. 430-439.

Sharma S.C., G.X. Liu, and S. Thomas, "Sources of Error in Estimating Truck Traffic from Automatic Vehicle Classification Data," *Journal of Transportation and Statistics*, October 1998.

Sharma S.C., P. Lingras, F. Xu, and G.X. Liu, "Neural Networks as an Alternative to the Traditional Factor Approach of AADT Estimation from Traffic Counts," Presented at *78th Transportation Research Board* (Washington D.C., January 1999).

Sierra Research, "Development of Speed Correction Cycles," Prepared for U.S. EPS by Sierra Research, Sacramento, California. Report No. SR97-04-01. April 30, 1997.

Skabardonis, A., and R. Dowling. Improved Speed –Flow Relationships for Planning Applications. In *Transportation Research Record 1572*, TRB, National Research Council, Washington, D.C., 1997, pp. 18-23.

Systems Applications International, Inc. *Development of Methodology for Estimating VMT Weighting by Facility Type*. Report EPA420-R-01-009, M6. SPD. 003. Environment Protection Agency. April 2001.

Systems Applications International, Inc. *Guidance for the Development of Facility Type VMT and Speed Distributions, Draft*. Report EPA420-P-99-004, M6. SPD. 004. Environment Protection Agency. February 1999.

Texas Transportation Institute, "Development of Gridded Mobile Source Emissions Estimates for the Houston-Galveston Nonattainment Counties FY2007 in Support of the Coast Project," Report No. TNRCC/TX-98-402001, 1998.

Tracie R. Jackson. *Fleet Characterization Data for MOBILE6: Development and Use of Age Distributions, Average Annual Mileage Accumulation Rates, and Projected Vehicle Counts for Use in MOBILE6 (M6.FLT.007)*. EPA420-R-01-047. U.S. Environmental Protection Agency, September 2001. http://www.epa.gov/otaq/models/mobile6/r01047.pdf. Accessed July 06, 2002.

U.S. Census Bureau, United States Department of Commerce. http://www.census.gov. Accessed February 20, 2001.

U.S. Environmental Protection Agency. *Frequently Asked Questions on MOBILE6.* January 16, 2002

U.S. Environmental Protection Agency. *User's Guide to MOBILE6.0: Mobile Source Emission Factor Model.* January 2002

United States Environmental Protection Agency. *Analysis of Emissions Deterioration Using Ohio and Wisconsin IM240 Data*, United States Environmental Protection Agency, Office of Mobile Source, Publication No. EPA420-P-99-013, March 1999.

US Census Bureau, *Publications*, United States Department of Commerce. *http://www.census.gov* (Accessed Feb. 20, 2001).

US Census Bureau, United States Department of Commerce. http://www.census.gov. Accessed July 06, 2002.

Weinblatt, H., "Using Seasonal and Day-of-Week Factoring to Improve Estimates of Truck Vehicle Miles Traveled," *Transportation Research Record* 1522, pp.1-8.

Xia Q., F. Zhao, L.D. Shen, D. Ospina, and Z. Chen, "Estimation of Annual Average Daily Traffic for Non-State Roads in a Florida County," Presented at *78th Transportation Research Board* (Washington D.C., January 1999).

Yu, L., F. Qiao, G. Li, and X. Wang. *Forecasting Traffic Characteristics for Air Quality Analysis*. Interim Report No. TxDOT 4142-8. Texas Department of Transportation, March 2002.

Appendix A

Annual Mileage Accumulation Curve Fit Equations

Vehicle Class	Equation
LDGV	$y = 15684e^{-0.0506x}$
LDDV	$y = 15684e^{-0.0506x}$
LDGT1	$y = 17.472x^2 - 1163.7x + 20642$
LDGT2	$y = 22905e^{-0.0712x}$
LDDT1	$y = 30028e^{-0.104x}$
LDDT2	$y = 28231e^{-0.0808x}$
HDGV (2B-3)	$y = 21250e^{-0.0618x}$
HDGV (4-8)	$y = 23243e^{-0.0829x}$
HDGSB	$y = 9939$
HDGTB	$y = 38654e^{-0.0958x}$
HDDV (2B)	$y = 29657e^{-0.0888x}$
HDDV (3)	$y = 37008e^{-0.1222x}$

HDDV (4-5)	$y = 32625e^{-0.0656x}$
HDDV (6-7)	$y = 44883e^{-0.0983x}$
HDDV (8A)	$y = 98554e^{-0.1153x}$
HDDV (8B)	$y = 137024e^{-0.0982x}$
HDDSB	$y = 9939$
HDDTB	$y = 46659e^{-0.0324x}$

X=model year – 1990

Y=Annual mileage (miles)

Appendix B

Average Annual Mileage Accumulation (Curve Fit Data)

U.S. Levels
(12 months estimate)

	LDV		**LDGT**		**LDDT**		**HDGV**		**HDGB**	
Vehicle Age	**LDGV**	**LDDV**	**LDGT** 0-6000	**LDGT** 6001-8500	**LDDT** 0-6000	**LDDT** 6001-8500	**2B-3** 8501-14000	**4-8** >14000	**S.BUS ANY WGT.**	**T.BUS ANY WGT.**
1	14910	14910	19496	21331	27059	26040	19977	21394	(a)	35123
2	14174	14174	18284	19565	24384	24018	18779	19692		31914
3	13475	13475	17308	18500	21973	22154	17654	18125		28999
4	12810	12810	16267	17228	19801	20434	16596	16683		26350
5	12178	12178	15260	16044	17843	18848	15601	15356		23942
6	11577	11577	14289	14942	16079	17385	14666	14134		21755
7	11006	11006	13352	13915	14490	16036	13787	13010		19768
8	10463	10463	12451	12959	13057	14791	12961	11975		17962
9	9947	9947	11584	12068	11766	13643	12184	11022		16321
10	9456	9456	10752	11239	10603	12584	11454	10145		14830
11	8989	8989	9955	10466	9555	11607	10768	9338		13475
12	8546	8546	9194	9747	8610	10706	10122	8595		12244
13	8124	8124	8467	9077	7759	9875	9516	7911		11126
14	7723	7723	7775	8453	6992	9109	8946	7282		10109
15	7342	7342	7118	7872	6301	8402	8409	6703		9186
16	6980	6980	6496	7331	5678	7749	7905	6169		8347
17	6636	6636	5909	6827	5116	7148	7432	5679		7584
18	6308	6308	5356	6358	4610	6593	6986	5227		6891
19	5997	5997	4839	5921	4155	6081	6568	4811		6262
20	5701	5701	4357	5514	3744	5909	6174	4428		5690
21	5420	5420	3909	5135	3374	5174	5804	4076		5170
22	5152	5152	3497	4782	3040	4772	5456	3752		4698

23	4898	4898	3120	4454	2740	4402	5129	3453		4268
24	4656	4656	2777	4148	2469	4060	4822	3178		3879
25	4427	4427	2470	3863	2225	3745	4533	2926		3524
26	4208	4208	2197	3597	2005	3454	4261	2693		3202
27	4001	4001	1959	3350	1807	3186	4006	2479		2910
28	3803	3803	1756	3120	1628	2939	3766	2281		2644
29	3616	3616	1589	2905	1467	2711	3540	2100		2402
30	3437	3437	1456	2706	1322	2500	3328	1933		2183

LDV Light duty vehicle

LDGV Light duty gasoline vehicle

LDDV Light duty diesel vehicle

LDGT Light duty gasoline truck

(a) Average school bus mileage for all ages = 9,939

LDDT Light duty diesel truck

HDGV Heavy duty gasoline vehicle

HDGV Heavy duty gasoline vehicle

HDGB Heavy duty gasoline bus

Annual Mileage Accumulation (Curve Fit Data)

(12 months estimate)

(Continued)

U.S. Levels

Vehicle Age	HDDV						HDDB	
	2B 8501-10000	**3** 14001-19500	**4-5** 14001-19500	**6-7** 19501-33000	**8A** 33001-60000	**8B** >60000	**S.Bus** Any WGT.	**T. Bus** Any WGT.
1	27137	32751	30653	40681	87821	124208	(a)	45171
2	24831	28984	28622	36872	78257	112590		43731
3	22721	25650	26805	33420	69735	102060		42337
4	20791	22699	25103	30291	62141	92514		40987
5	19024	20088	23509	27455	55374	83861		39681
6	17407	17778	22016	24885	49343	76017		38416
7	15928	15733	20618	22555	43970	68907		37191
8	14575	13923	19309	20443	39181	62462		36005
9	13336	12321	18083	18529	34915	56620		34857
10	12203	10904	16935	16795	31112	51324		33746
11	11166	9650	15860	15222	27724	46523		32670
12	10217	8540	14853	13797	24705	42172		31629
13	9349	7557	13910	12505	22015	38228		30620
14	8555	6688	13026	11335	19617	34652		29644
15	7828	5919	12199	10273	17481	31411		28699
16	7163	5238	11425	9312	15577	28473		27784
17	6554	4635	10699	8440	13881	25810		26898
18	5997	4102	10020	7650	12369	23396		26041
19	5488	3630	9384	6933	11022	21208		25211
20	5021	3213	8788	6284	9822	19224		24407
21	4995	2843	8230	5696	8752	17426		23629
22	4204	2516	7707	5163	7799	15796		22875

23	3847	2227	7218	4679	6950	14319		22146
24	3520	1971	6760	4241	6193	12979		21440
25	3221	1744	6331	3844	5518	11765		20757
26	2947	1543	5929	3484	4918	10665		20095
27	2697	1366	5552	3158	4382	9667		19454
28	2468	1209	5200	2862	3905	8763		18834
29	2258	1070	4869	2594	3480	7944		18234
30	2066	947	4560	2352	3101	7201		17652

HDDV Heavy duty diesel vehicle

HDDB Heavy duty diesel bus

(a) Average school bus mileage for all ages = 9,939

Appendix C

Motorcycle Age Distribution and Mileage Accumulation Rates for Use in MOBILE6

Age	Registration Distribution	Mileage Accumulation Rates
1	0.144	4,786
2	0.168	4,475
3	0.135	4,164
4	0.109	3,853
5	0.088	3,543
6	0.07	3,232
7	0.056	2,921
8	0.045	2,611
9	0.036	2,300
10	0.029	1,989
11	0.023	1,678
12+	0.097	1,368

Note: Motorcycle vehicle count is 4,219,000 for all years, pre-1982 through 2050.
Source: 1987 Motorcycle Statistical Annual, Motorcycle Industry Council, Inc.

Appendix D

Vehicle Mileage Survey Form

Circle the correct multiple choice answer that applies in Part I. and Part II. Fill in the TABLE Chart in Part III.

I. Background Information:

1.What age group do you fall under?
a. 17 yr. and below b. 18-24 yr. c. 25-31 yr. d. 32-38 yr.
e. 39-45 yr. f. 46-52 yr. g. 52 yr. and above

2. What is your ethnic group?
a. Hispanic b. Caucasian c. African- American d. Pacific- Asian e. Native American
f. Other

3. What is your sex?
a. Male b. Female

II. Household Information:

4. How many members are in your household?
a. 1-3 b. 4-6 c. 7-9 d. 10 and above

5. What is the average household income? (Optional)
a. 16,000 or below b. 17,000- 24,000 c. 25- 32,000 d. 33,00- 40,000 e. 41,000- 48,000
f. 49,000- 55,000 g. 55,000 and above

6. What area of the city do you reside at?
a. North b. Northeast c. South d. Southeast e. Southwest f. Downtown- Central
g. Other (specify):

III. Please fill in the TABLE Chart for all the vehicles you own.

Number Of Vehicles	Vehicle Type (Car, Van, Truck, etc)	Make and Model / Vehicle Weight	Year of Make	Number of Miles driven in Yr. 2000	Total Mileage on the Odometer	County You Reside
1						
2						
3						

Appendix E

Default VMT mix in MOBILE6

```
*  Last change:  MC    6 Dec 2000    2:17 pm
*
* VMT is not read as an external file.  The following input is an
* example of the VMT input label followed by values representing the
* 2010 calendar year values for VMT mix.
*
* The sixteen values represent the
* distribution of all vehicle miles traveled (VMT) by each of 16 vehicle
* classes. These numbers are read in a "free" format.  This file
* contains a recommended form.
*
* The sixteen vehicle classes are:
*
*  1  LDV    Light-Duty Vehicles (Passenger Cars)
*  2  LDT1   Light-Duty Trucks 1 (0-6,000 lbs. GVWR, 0-3750 lbs. LVW)
*  3  LDT2   Light Duty Trucks 2 (0-6,001 lbs. GVWR, 3751-5750 lbs. LVW)
*  4  LDT3   Light Duty Trucks 3 (6,001-8500 lbs. GVWR, 0-3750 lbs. LVW)
*  5  LDT4   Light Duty Trucks 4 (6,001-8500 lbs. GVWR, 3751-5750 lbs. LVW)
*  6  HDV2B  Class 2b Heavy Duty Vehicles (8501-10,000 lbs. GVWR)
*  7  HDV3   Class 3 Heavy Duty Vehicles (10,001-14,000 lbs. GVWR)
*  8  HDV4   Class 4 Heavy Duty Vehicles (14,001-16,000 lbs. GVWR)
*  9  HDV5   Class 5 Heavy Duty Vehicles (16,001-19,500 lbs. GVWR)
* 10  HDV6   Class 6 Heavy Duty Vehicles (19,501-26,000 lbs. GVWR)
* 11  HDV7   Class 7 Heavy Duty Vehicles (26,001-33,000 lbs. GVWR)
* 12  HDV8A  Class 8a Heavy Duty Vehicles (33,001-60,000 lbs. GVWR)
* 13  HDV8B  Class 8b Heavy Duty Vehicles (>60,000 lbs. GVWR)
* 14  HDBS   School Busses
* 15  HDBT   Transit and Urban Busses
* 16  MC     Motorcycles (All)
*
* All values must be less than or equal to 1 and greater than or equal to
* zero.  The sum of all 16 values must be exactly equal to 1, otherwise
```

```
* the model will normalize the values so that they equal 1. All 16 values
* must be entered each time the VMT MIX label is used.
*
* The default value for VMT mix varies by calendar year, based on the
* value of vehicle counts in the model. Vehicle count by calendar year
* is not a user input. The following values are for the 2010 calendar
* year.
*
VMT MIX          :
0.354 0.089 0.297 0.092 0.041 0.040 0.004 0.003
0.002 0.008 0.010 0.012 0.040 0.002 0.001 0.005
```

Appendix F

Default VMT BY HOUR in MOBILE6

```
VMT BY HOUR
*
* Fraction of all vehicle miles traveled by hour of the day.
* First hour is 6 a.m.
*
    0.0569 0.0740 0.0655 0.0555 0.0540 0.0582
    0.0608 0.0571 0.0598 0.0636 0.0777 0.0730
    0.0501 0.0389 0.0308 0.0264 0.0194 0.0144
    0.0108 0.0086 0.0081 0.0080 0.0098 0.0186
```

Appendix G

Default VMT BY FACILITY in MOBILE6

```
VMT BY FACILITY
* VMT fractions are listed for 28 vehicle classes.
* For each class, 24 sets of values represent the hours of the day.
* For each class and hour, 4 values represent the VMT distribution on
* freeway, arterial, local and ramps--in that order.
 1 0.392 0.457 0.117 0.034
   0.344 0.497 0.129 0.030
   0.338 0.497 0.135 0.029
   0.349 0.492 0.129 0.030
   0.346 0.497 0.127 0.030
   0.333 0.509 0.129 0.029
   0.324 0.516 0.132 0.028
   0.334 0.506 0.131 0.029
   0.334 0.506 0.131 0.029
   0.320 0.519 0.134 0.028
   0.330 0.506 0.135 0.029
   0.312 0.521 0.140 0.027
   0.295 0.538 0.141 0.026
   0.310 0.527 0.137 0.027
   0.329 0.510 0.133 0.029
   0.343 0.497 0.131 0.030
   0.381 0.460 0.126 0.033
   0.405 0.437 0.123 0.035
   0.426 0.418 0.118 0.037
```

	0.443	0.403	0.115	0.039
	0.457	0.394	0.110	0.040
	0.461	0.391	0.107	0.040
	0.453	0.400	0.108	0.039
	0.418	0.434	0.112	0.036
2	0.392	0.457	0.117	0.034
	0.344	0.497	0.129	0.030
	0.338	0.497	0.135	0.029
	0.349	0.492	0.129	0.030
	0.346	0.497	0.127	0.030
...				
28	0.392	0.457	0.117	0.034
	0.344	0.497	0.129	0.030
	0.338	0.497	0.135	0.029
	0.349	0.492	0.129	0.030
	0.346	0.497	0.127	0.030
	0.333	0.509	0.129	0.029
	0.324	0.516	0.132	0.028
	0.334	0.506	0.131	0.029
	0.334	0.506	0.131	0.029
	0.320	0.519	0.134	0.028
	0.330	0.506	0.135	0.029
	0.312	0.521	0.140	0.027
	0.295	0.538	0.141	0.026
	0.310	0.527	0.137	0.027
	0.329	0.510	0.133	0.029
	0.343	0.497	0.131	0.030
	0.381	0.460	0.126	0.033
	0.405	0.437	0.123	0.035
	0.426	0.418	0.118	0.037
	0.443	0.403	0.115	0.039
	0.457	0.394	0.110	0.040
	0.461	0.391	0.107	0.040
	0.453	0.400	0.108	0.039
	0.418	0.434	0.112	0.036

Appendix H

Default VMT BY FACILITY in MOBILE6

```
SPEED VMT
1 1  0.0083 0.0272 0.0210 0.0224 0.0217 0.0381 0.0344 0.0536 0.0614 0.0700 0.2507 0.1150 0.2550 0.0212
1 2  0.0260 0.0066 0.0076 0.0156 0.0282 0.0326 0.0344 0.0361 0.0360 0.0435 0.2453 0.1729 0.3023 0.0129
1 3  0.0259 0.0033 0.0064 0.0057 0.0126 0.0281 0.0342 0.0349 0.0407 0.0369 0.2181 0.1066 0.4399 0.0127
1 4  0.0145 0.0096 0.0021 0.0022 0.0041 0.0166 0.0232 0.0373 0.0418 0.0449 0.2248 0.1190 0.4422 0.0177
1 5  0.0083 0.0086 0.0052 0.0032 0.0040 0.0163 0.0232 0.0364 0.0375 0.0420 0.2352 0.1170 0.4454 0.0177
1 6  0.0072 0.0034 0.0042 0.0098 0.0121 0.0244 0.0289 0.0327 0.0401 0.0392 0.2294 0.1011 0.4538 0.0137
1 7  0.0103 0.0023 0.0064 0.0087 0.0147 0.0281 0.0335 0.0328 0.0345 0.0354 0.2294 0.0964 0.4547 0.0128
1 8  0.0083 0.0075 0.0052 0.0043 0.0054 0.0182 0.0257 0.0381 0.0380 0.0421 0.2258 0.1118 0.4512 0.0184
1 9  0.0113 0.0065 0.0052 0.0023 0.0039 0.0206 0.0279 0.0358 0.0383 0.0517 0.2147 0.1151 0.4484 0.0183
1 10 0.0155 0.0075 0.0034 0.0042 0.0081 0.0272 0.0324 0.0363 0.0315 0.0390 0.2124 0.0644 0.5000 0.0181
1 11 0.0156 0.0411 0.0225 0.0199 0.0284 0.0316 0.0500 0.0488 0.0446 0.0555 0.2223 0.1092 0.2957 0.0148
1 12 0.0186 0.0113 0.0046 0.0110 0.0183 0.0261 0.0488 0.0383 0.0314 0.0534 0.2235 0.1237 0.3736 0.0174
1 13 0.0176 0.0064 0.0010 0.0024 0.0034 0.0155 0.0191 0.0315 0.0357 0.0515 0.2134 0.0674 0.5178 0.0173
1 14 0.0135 0.0043 0.0031 0.0010 0.0012 0.0094 0.0177 0.0258 0.0264 0.0550 0.2060 0.0980 0.5209 0.0177
1 15 0.0094 0.0031 0.0025 0.0007 0.0012 0.0069 0.0166 0.0216 0.0257 0.0476 0.2169 0.1048 0.5228 0.0202
1 16 0.0054 0.0018 0.0018 0.0004 0.0011 0.0045 0.0155 0.0175 0.0250 0.0401 0.2277 0.1117 0.5246 0.0229
1 17 0.0027 0.0010 0.0014 0.0002 0.0011 0.0028 0.0147 0.0147 0.0245 0.0352 0.2350 0.1162 0.5259 0.0246
1 18 0.0013 0.0006 0.0012 0.0001 0.0011 0.0020 0.0144 0.0133 0.0242 0.0327 0.2386 0.1185 0.5265 0.0255
1 19 0.0000 0.0001 0.0010 0.0000 0.0011 0.0012 0.0140 0.0119 0.0240 0.0302 0.2422 0.1208 0.5271 0.0264
1 20 0.0000 0.0013 0.0000 0.0000 0.0000 0.0010 0.0115 0.0097 0.0200 0.0241 0.2450 0.1285 0.5271 0.0318
1 21 0.0000 0.0003 0.0010 0.0000 0.0000 0.0008 0.0103 0.0086 0.0181 0.0206 0.2464 0.1321 0.5271 0.0347
1 22 0.0000 0.0013 0.0000 0.0000 0.0000 0.0008 0.0107 0.0081 0.0170 0.0199 0.2451 0.1341 0.5271 0.0359
1 23 0.0021 0.0003 0.0000 0.0010 0.0000 0.0010 0.0118 0.0100 0.0205 0.0224 0.2452 0.1274 0.5271 0.0312
1 24 0.0031 0.0003 0.0000 0.0010 0.0001 0.0011 0.0134 0.0124 0.0240 0.0267 0.2404 0.1226 0.5271 0.0278
2 1  0.0004 0.0052 0.0061 0.0053 0.0158 0.0854 0.3210 0.1382 0.2804 0.0595 0.0628 0.0103 0.0095 0.0001
2 2  0.0036 0.0029 0.0059 0.0234 0.0735 0.1114 0.2842 0.0950 0.2633 0.0396 0.0698 0.0107 0.0169 0.0000
2 3  0.0033 0.0021 0.0032 0.0085 0.0436 0.1130 0.2914 0.1076 0.2835 0.0424 0.0719 0.0091 0.0204 0.0000
2 4  0.0030 0.0015 0.0011 0.0015 0.0183 0.1001 0.2910 0.1246 0.3013 0.0535 0.0743 0.0094 0.0204 0.0000
```

```
2 5  0.0030 0.0014 0.0005 0.0017 0.0181 0.1008 0.2898 0.1246 0.3015 0.0537 0.0751 0.0094 0.0204 0.0000
2 6  0.0034 0.0017 0.0021 0.0049 0.0344 0.1091 0.2894 0.1125 0.2932 0.0460 0.0735 0.0093 0.0205 0.0000
2 7  0.0040 0.0021 0.0027 0.0078 0.0427 0.1134 0.2857 0.1083 0.2886 0.0427 0.0724 0.0091 0.0205 0.0000
2 8  0.0038 0.0025 0.0020 0.0022 0.0216 0.1034 0.2834 0.1243 0.3020 0.0515 0.0736 0.0094 0.0203 0.0000
2 9  0.0041 0.0024 0.0020 0.0034 0.0249 0.1049 0.2844 0.1215 0.2986 0.0489 0.0751 0.0093 0.0205 0.0000
2 10 0.0052 0.0027 0.0032 0.0085 0.0450 0.1151 0.2822 0.1024 0.2835 0.0419 0.0777 0.0096 0.0230 0.0000
2 11 0.0049 0.0165 0.0087 0.0224 0.0652 0.1222 0.2809 0.0959 0.2557 0.0405 0.0651 0.0095 0.0125 0.0000
2 12 0.0055 0.0071 0.0082 0.0219 0.0675 0.1169 0.2771 0.0915 0.2637 0.0394 0.0712 0.0106 0.0194 0.0000
2 13 0.0043 0.0024 0.0016 0.0038 0.0255 0.1005 0.2849 0.1205 0.2996 0.0497 0.0761 0.0100 0.0211 0.0000
2 14 0.0038 0.0021 0.0018 0.0015 0.0115 0.0734 0.2923 0.1219 0.3170 0.0641 0.0794 0.0100 0.0211 0.0001
2 15 0.0037 0.0017 0.0012 0.0019 0.0103 0.0558 0.3040 0.1067 0.3309 0.0702 0.0824 0.0100 0.0211 0.0001
2 16 0.0036 0.0018 0.0009 0.0012 0.0109 0.0530 0.3056 0.1064 0.3320 0.0707 0.0827 0.0100 0.0211 0.0001
2 17 0.0034 0.0009 0.0007 0.0015 0.0104 0.0531 0.3065 0.1064 0.3325 0.0706 0.0829 0.0100 0.0211 0.0000
2 18 0.0030 0.0013 0.0016 0.0018 0.0103 0.0528 0.3057 0.1061 0.3327 0.0704 0.0831 0.0100 0.0211 0.0001
2 19 0.0000 0.0000 0.0000 0.0003 0.0087 0.0502 0.3303 0.1054 0.3306 0.0699 0.0733 0.0100 0.0211 0.0002
2 20 0.0001 0.0000 0.0000 0.0000 0.0082 0.0496 0.3302 0.1057 0.3293 0.0696 0.0757 0.0101 0.0211 0.0004
2 21 0.0000 0.0000 0.0000 0.0000 0.0081 0.0491 0.3306 0.1060 0.3298 0.0693 0.0755 0.0101 0.0211 0.0004
2 22 0.0000 0.0000 0.0000 0.0000 0.0077 0.0489 0.3291 0.1060 0.3316 0.0692 0.0758 0.0101 0.0211 0.0005
2 23 0.0000 0.0000 0.0000 0.0000 0.0082 0.0497 0.3286 0.1056 0.3311 0.0697 0.0756 0.0101 0.0211 0.0003
2 24 0.0000 0.0000 0.0000 0.0000 0.0085 0.0502 0.3271 0.1054 0.3324 0.0699 0.0752 0.0100 0.0211 0.0002
*
* Comments are not allowed before the end of the data!
*
* Fraction of vehicle miles traveled within an hour within an average speed bins by hour of the day.
* The first hour is 6 a.m.
*
* Freeways
* Hr  2.5   5.0  10.0  15.0  20.0  25.0  30.0  35.0  40.0  45.0  50.0  55.0  60.0  65.0+
*
* Arterial and Collector Roadways
* Hr  2.5   5.0  10.0  15.0  20.0  25.0  30.0  35.0  40.0  45.0  50.0  55.0  60.0  65.0+
```

Appendix I

Composite MOBILE6 Vehicle Classifications (STARTS PER DAY Command)

Number	Abbreviation	Description
1	LDGV	Light-duty Gasoline Vehicles (Passenger Cars)
2	LDGT1	Light-duty Gasoline Tracks 1 (0-6,000lbs. GVWR, 0-3750lbs.LVW)
3	LDGT2	Light-duty Gasoline Tracks 2 (0-6,001lbs. GVWR, 3751-5750lbs.LVW)
4	LDGT3	Light-duty Gasoline Tracks 3 (6,001-8500lbs. GVWR, 0-3750lbs.LVW)
5	LDGT4	Light-duty Gasoline Tracks 4 (6,001-8500lbs. GVWR, 3751-5750lbs.LVW)
6	HDGV2B	Class 2b Heavy Duty Gasoline Vehicles (8501-10,000lbs. GVWR)
7	HDGV3	Class 3 Heavy Duty Gasoline Vehicles (10,001-14,000lbs. GVWR)
8	HDGV4	Class 4 Heavy Duty Gasoline Vehicles (14,001-16,000lbs. GVWR)
9	HDGV5	Class 5 Heavy Duty Gasoline Vehicles (16,001-19,500lbs. GVWR)
10	HDGV6	Class 6 Heavy Duty Gasoline Vehicles (19,501-26,000lbs. GVWR)
11	HDGV7	Class 7 Heavy Duty Gasoline Vehicles (26,001-33,000lbs. GVWR)

12	HDGV8A	Class 8a Heavy Duty Gasoline Vehicles (33,000-60,000lbs. GVWR)
13	HDGV8B	Class 8b Heavy Duty Gasoline Vehicles (>60,000lbs. GVWR)
14	LDDV	Light Duty Diesel Vehicles (Passenger cars)
15	LDDT12	Light Duty Diesel trucks 1 and 2 (0-6,000lbs. GVWR)
16	HDDV2B	Class 2b Heavy Duty Diesel Vehicles (8501-10,000lbs. GVWR)
17	HDDV3	Class 3 Heavy Duty Diesel Vehicles (10,001-14,000lbs. GVWR)
18	HDDV4	Class 4 Heavy Duty Diesel Vehicles (14,001-16,000lbs. GVWR)
19	HDDV5	Class 5 Heavy Duty Diesel Vehicles (16,001-19,500lbs. GVWR)
20	HDDV6	Class 6 Heavy Duty Diesel Vehicles (19,501-26,000lbs. GVWR)
21	HDDV7	Class 7 Heavy Duty Diesel Vehicles (26,001-33,000lbs. GVWR)
22	HDDV8B	Class 8a Heavy Duty Diesel Vehicles (33,000-60,000lbs. GVWR)
23	HDDV8B	Class 8b Heavy Duty Diesel Vehicles (>60,000lbs. GVWR)
24	MC	Motorcycles (Gasoline)
25	HDGB	Gasoline Buses (School, Transit and Urban)
26	HDDBT	Diesel Transit and Urban Buses
27	HDDBS	Diesel School Buses
28	LDDT34	Light Duty Truck 3 and 4 (6,001-8500lbs.GVWR)

Appendix J

Names of Persons and Agencies/Companies for Responded E-mail Survey on VMT Approaches

Name of Persons	Name of Agencies/Companies
Alison K. Pollack	ENVIRON International Corp.
Andrew Edwards	Air Quality Specialist Southern Resource Center - FHWA
Barbara MacRae	Colorado Department of Public Health and Environment Air Pollution Control Division
Christopher Porter	Cambridge Systematics, Inc.
Dawn Wills	Transportation Planner North Center TX Council of Governments
JIM Dileo	Colorado Department of Public Health and Environment Air Pollution Control Division
Jonathan Morton	Georgia Department of Natural Resources
Joon Byun	Air Quality Modeling Specialist Eastern Resource Center, FHWA
Kevin N. Black	FHWA

Kip Billings	Wasatch Front Regional Council
Lark Downs	Stan COG
Richard McElveen	Florida Department of Environmental Protection
Tom Wenzel	Lawrence Berkeley National Laboratory
Walter Pienta	NYS Department of Environmental Conservation
Wayne Luney	California DOT