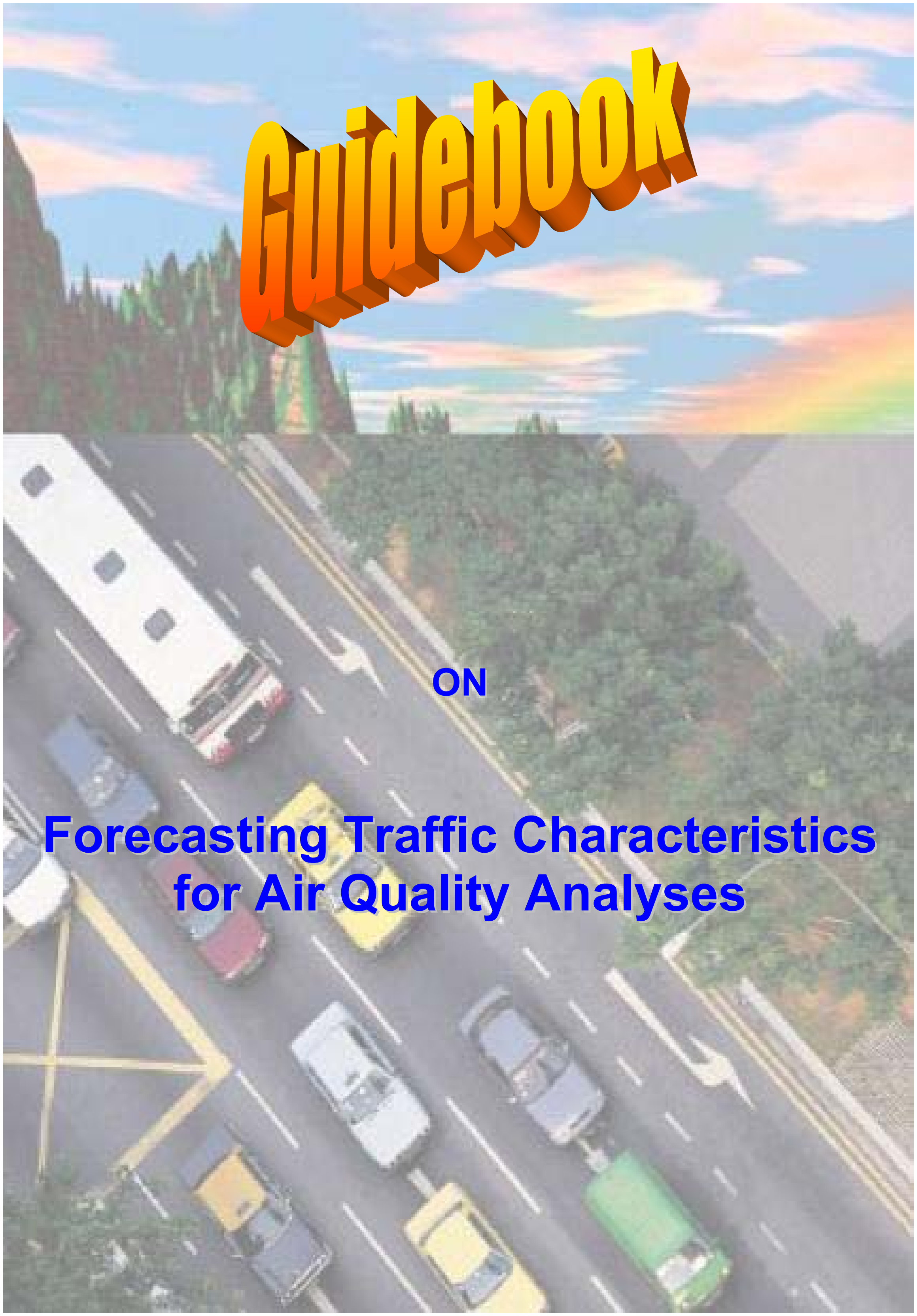
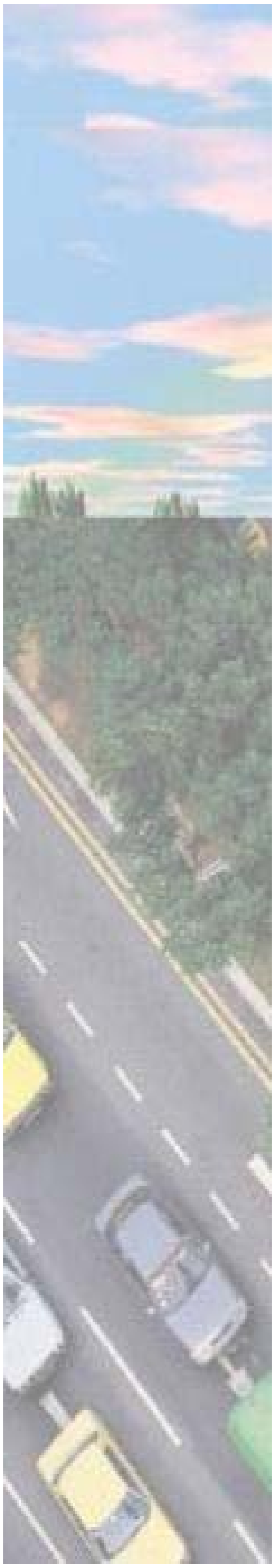


Guidebook

ON

Forecasting Traffic Characteristics for Air Quality Analyses



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.

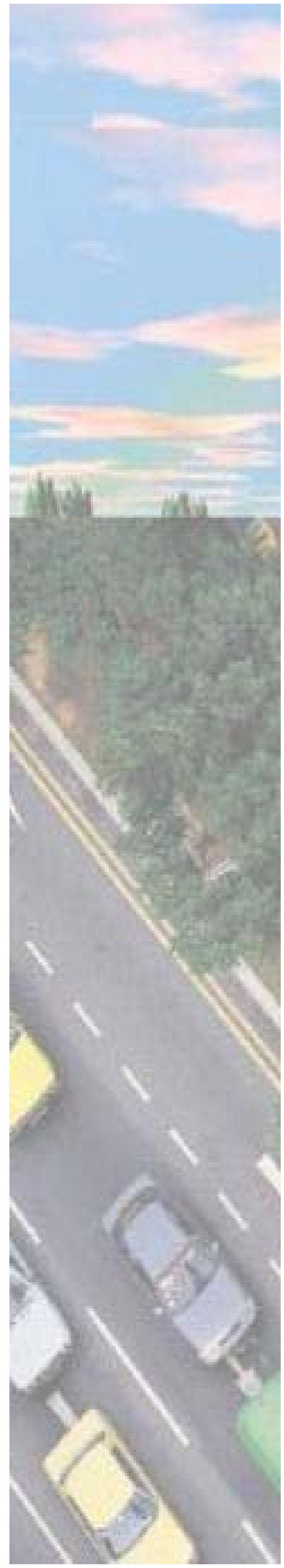
The United States Governments and the states of Texas do not endorse products or manufacturers. Trade or manufactures' names appear herein solely because they are considered essential to the object of this report.

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 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.

Guidebook on Forecasting Traffic Characteristics for Air Quality Analyses



By

Lei Yu, Ph.D., P.E.,

Fengxiang Qiao, Ph.D., and Guangchun Li

Texas Southern University

And

Hong Oey, Ph. D., P.E.

University of Texas at El Paso

Project Number 0-4142

Research Project Title: Forecasting Traffic Characteristics for Air Quality Analysis

Sponsored by the

Texas Department of Transportation

In Cooperation with the

U.S. Department of Transportation of Transportation

Federal Highway Administration

August 2002

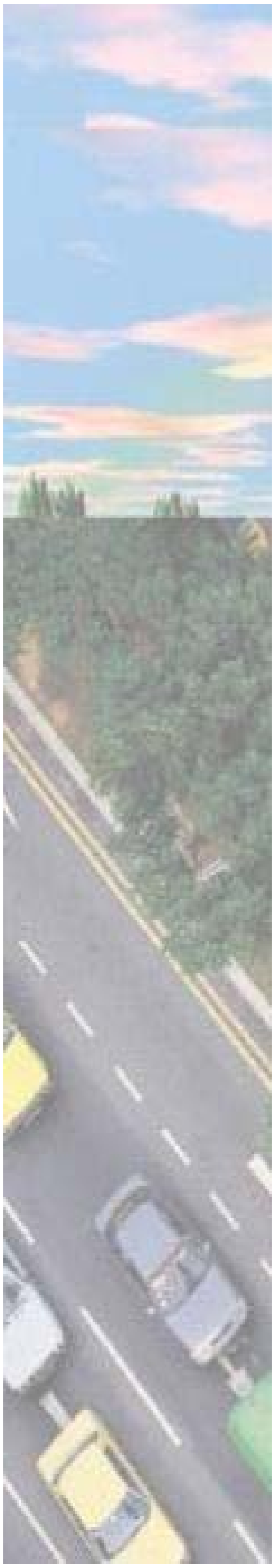
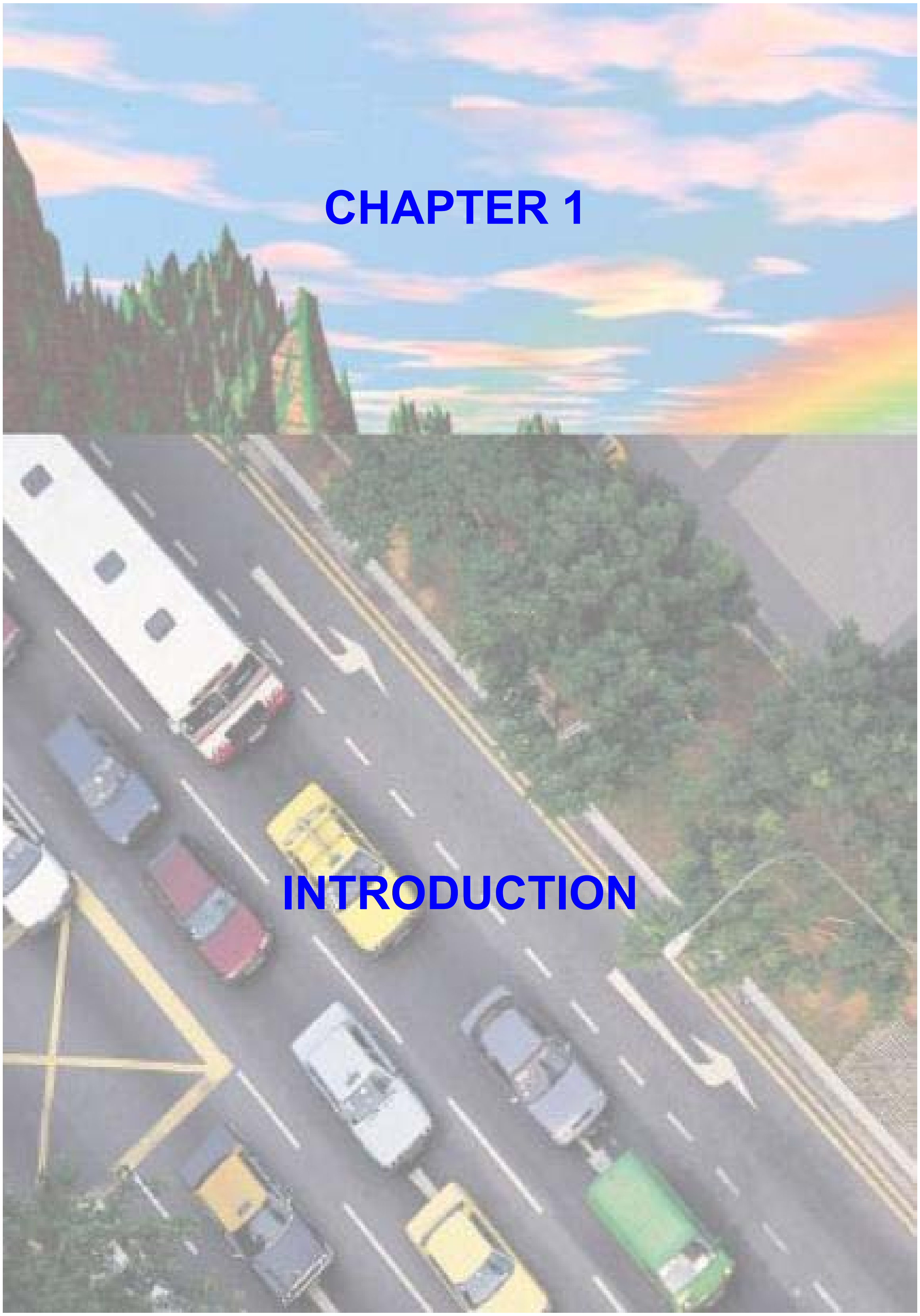
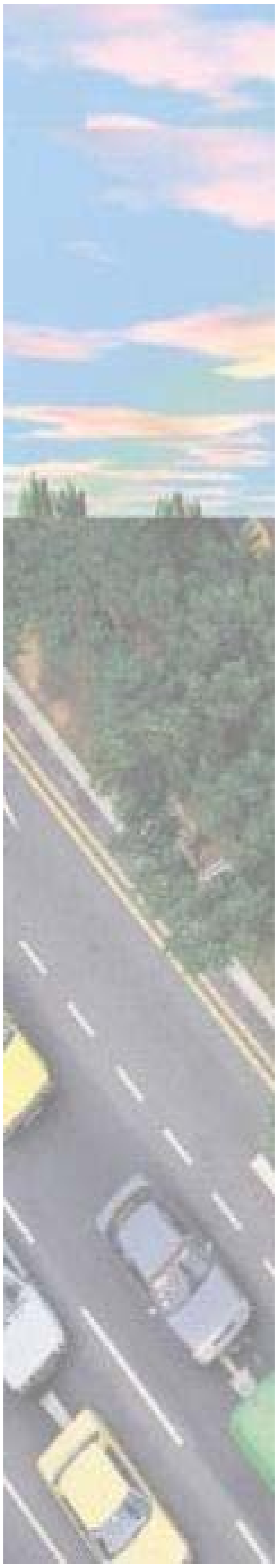


TABLE OF CONTENTS

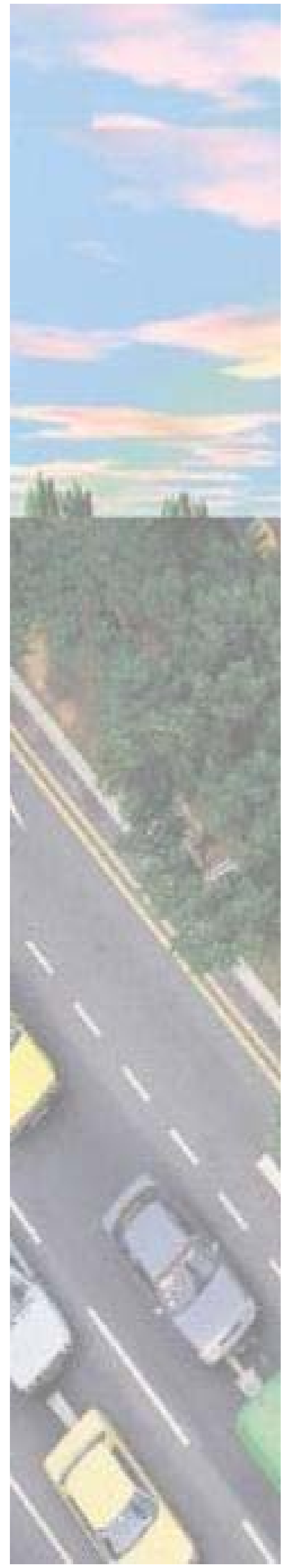
CHAPTER 1



INTRODUCTION



VEHICLE EMISSION ESTIMATION MODEL



Environmental Protection Agency (EPA) requires that on-road mobile source emissions such as carbon monoxide (CO), volatile organic compounds (VOCs), oxides of nitrogen (NO_x), particular matter (PM), etc. do not exceed the motor vehicle emission budgets estimated in the Rate of Progress (ROP) and Attainment Demonstration State Implementation Plan (SIP). The only model that was approved by EPA for use by all states except California for the conformity determination analysis is MOBILE.

MOBILE is used to calculate current and future emission inventories of these emissions. These inventories are used to make decisions about air pollution policy at the local, state and national levels. Inventories based on MOBILE are also used to meet the federal Clean Air Act's State Implementation Plan (SIP) and transportation conformity requirements, and are sometimes used to meet requirements of the National Environmental Protection Act (NEPA).

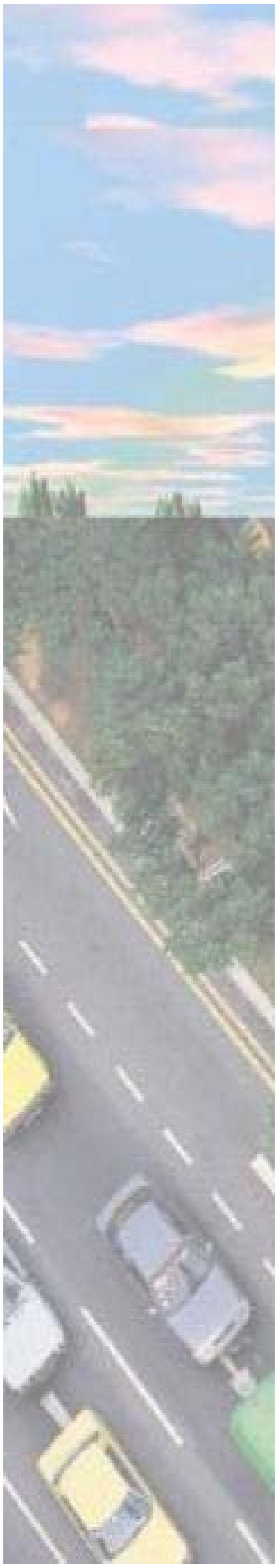
MOBILE model was first developed in 1978. Since that time, it has been updated many times to reflect the growing understanding of vehicle emissions, and to cover new emissions regulations and modeling needs. Although some updates were made in 1996 with the release of MOBILE5b, MOBILE6, released in 2002, is the first major revision to MOBILE since MOBILE5a.

IMPACT OF TRAFFIC CHARACTERISTICS TO EMISSION FACTORS

Mobile source emission related travel indicators are crucial parts in using MOBILE. The related traffic indicators include the vehicle age distribution, mileage accumulation rates by vehicle type, vehicle miles traveled (VMT) related variables, etc.

Traffic characteristics have considerable impacts on emission factors. By changing vehicle age distribution, mileage accumulation rates or VMT (& mix), the emission factors estimated by MOBILE will vary considerably. For example, for MOBILE5 at high temperature, a 2.8% change in HDGV mix causes about a 10% change in the CO rate and a 4.8% change in HDGV mix leads to about a 10% shift in the VOC rate. Therefore, it is very important to obtain the reliable estimates of the traffic characteristics.

MOBILE includes default values for a wide range of conditions that affect emissions. These default values are designed to represent "national average" input data. However, variations in roadway network characteristics between different areas are big enough to justify the use of locally developed travel indicators.



Overview of Guide

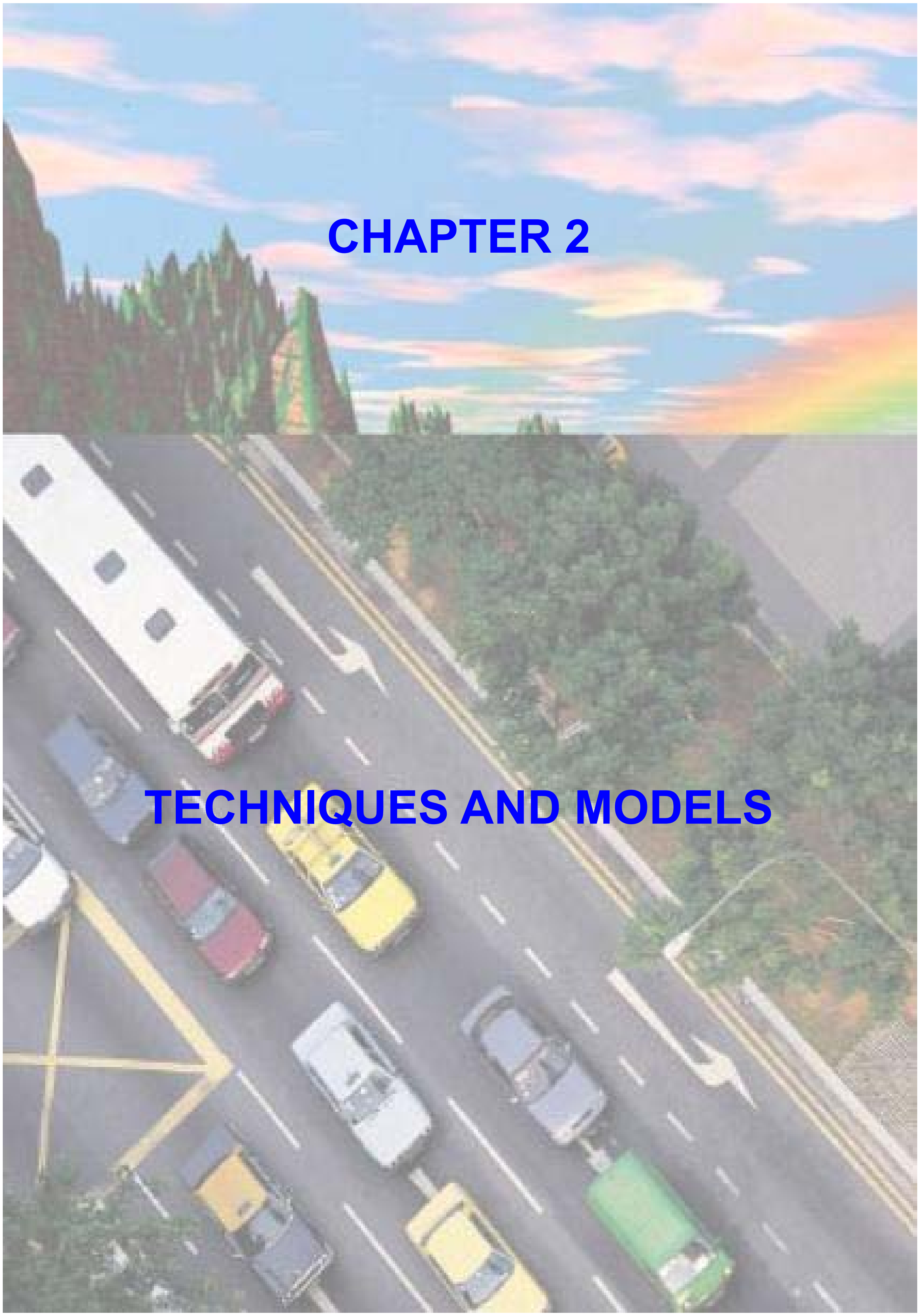
This guide contains three basic chapters that summarize guidance on forecasting traffic characteristics for air quality analysis. The three chapters are as follows:

Chapter 1. Introduction – provides an introduction on forecasting traffic characteristics for air quality analysis.

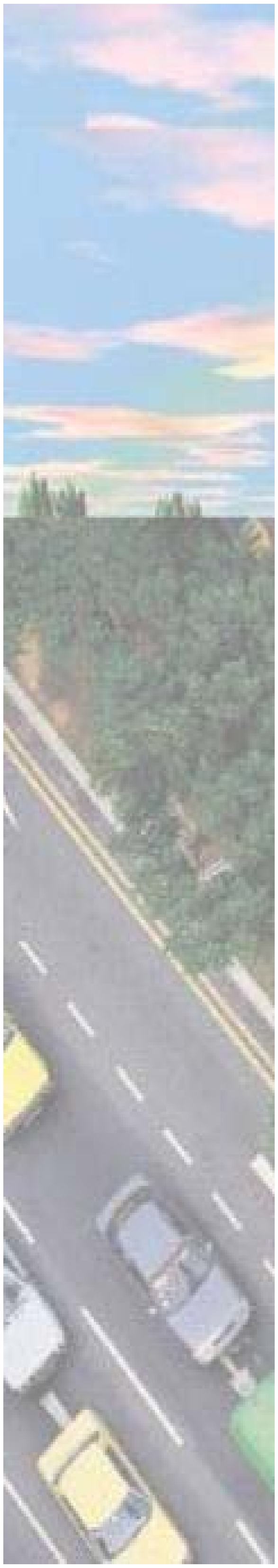
Chapter 2. Techniques and Models – provides descriptions of the techniques and models for estimating and forecasting traffic characteristics.

Chapter 3. Summary and Recommendations – provides summary of techniques and models. This chapter also includes recommendations to implementation plan.

CHAPTER 2



TECHNIQUES AND MODELS



This chapter provides descriptions of techniques and models for forecasting traffic characteristics for air quality analyses. The techniques and models include:

- Forecasting vehicle age distribution by regression modeling;
- Adjusting mileage accumulation rates based on small sample surveys; and
- Improving VMT related variables by correlating link volume with count data and link attributes.

DEVELOPMENT OF MODELS FOR FORECASTING VEHICLE AGE DISTRIBUTION

Methodology Description

Vehicle age distribution modeling system is an object in which variables of different kinds interact and produce observable vehicle age distribution. The modeling system can be illustrated in Figure 1. These input variables could be either the predictable socioeconomic factors, or the complex unpredictable or immeasurable inputs. Disturbances are unexpected inputs to the system, the efforts of which should be eliminated through proper modeling. The predictable socioeconomic indices may include population, average income, household, population density, etc. If the variables are unpredictable or immeasurable, the chronological series can be used as the input of the function.

The 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. The projection of the age distribution for the target year can be obtained when the input variables are supplied.

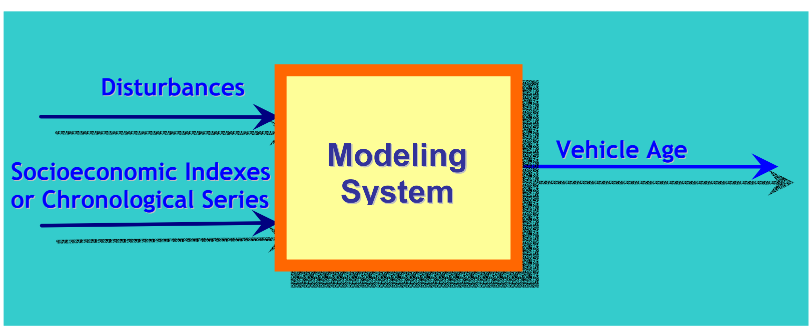
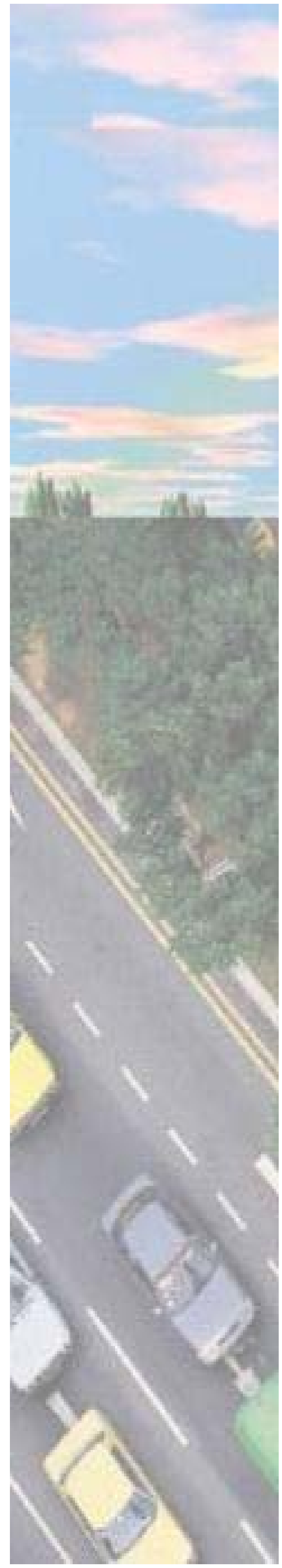


Figure 1 Vehicle age distribution modeling system.



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 models can automatically select the optimal inputs of socioeconomic indices that have higher correlations to the resulted vehicle age distribution. For example in modeling age distribution in 8 Houston- Galveston Area Council (HGAC) counties: Brazoria, Chambers, Fort Bend, Galveston, Harris, Liberty, Montgomery and Waller, a total of 22 socioeconomic indices were provided. The algorithm optimally selected 4 for them for each of the models.

However, it is often impossible to collect enough socioeconomic data that are all predictable. A suboptimal selection of independent variables refers to the situation where less number of socioeconomic data than the ideal situation is selected for inputs to the model. Based on the case studies in 8 HGAC counties and in El Paso, the suboptimal selection of input socioeconomic indices will result in a closer estimation of age distribution to the results by the optimally selected inputs comparing with the national-wide default one. The three general socioeconomic indices that can be easily projected are *total population*, *total employment*, and *personal income*.

Figure 2 presents the produced three emission factors (VOC, CO and NO_x) by the default age distributions and by the forecasted local ones from models for 8 HGAC counties and El Paso in the year 2001. The x-axis in this figure is purely a categorical classification, while y-axis stands for the emission factors. There are differences between two sets of results, especially for CO. In most counties, emission factors (especially for CO) are smaller than the ones that are generated by default age distributions. The only exception is Chambers County, where local emission factors are slightly larger than default ones.

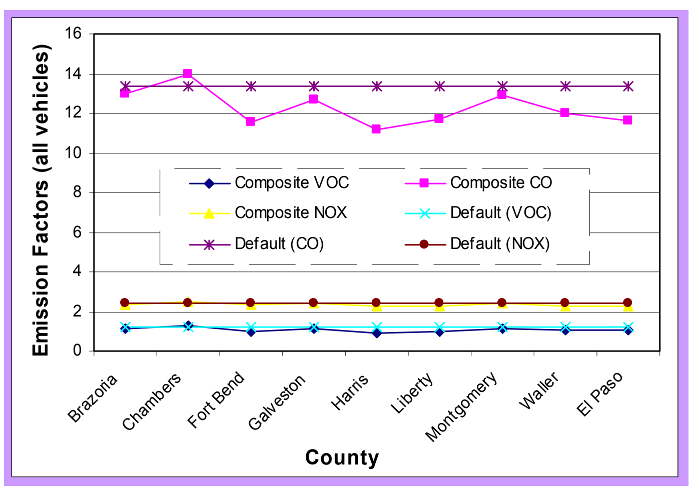


Figure 2 Comparison of emission factors by default and forecasted vehicle age distributions for 8 HGAC counties and El Paso in the year 2001.

Interface of Program MOFAD

The program that can realize the model for forecasting vehicle age distribution is named as MOFAD (MOdeling and Forecasting Age Distribution). It has a Graphic User Interface (GUI) as illustrated in Figure 3.

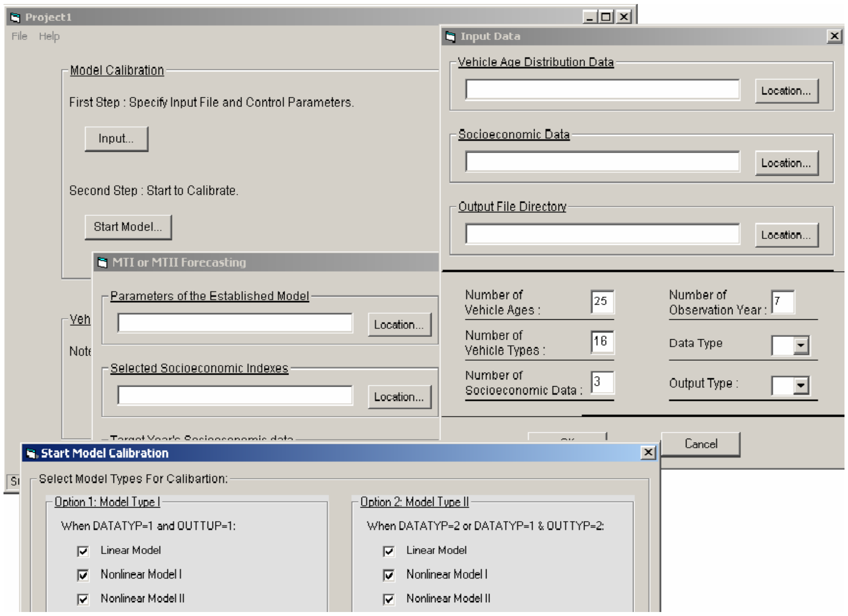


Figure 3 Graphic User Interface (GUI) of program MOFAD

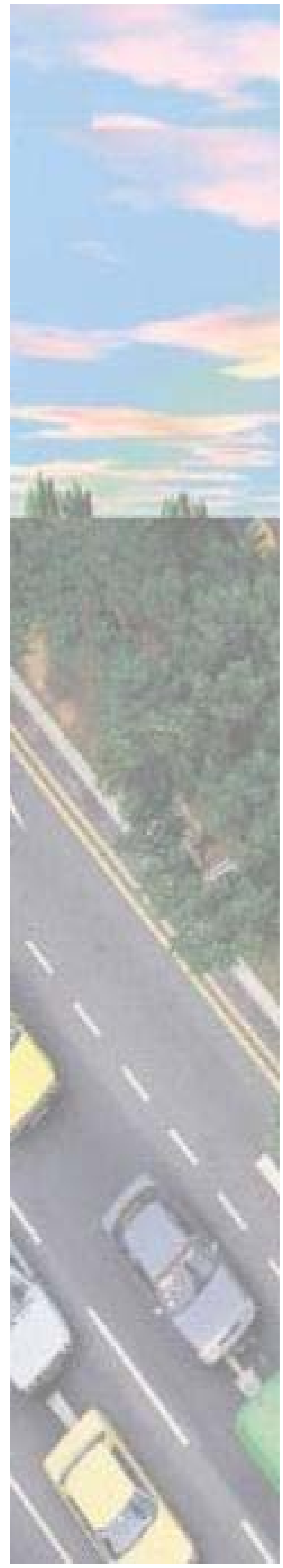
The whole program is divided into two main parts, *model calibration* and *vehicle age distribution forecasting*. Four types of output files can be generated, among which the standard output (outst.txt) can be inputted into MOBILE6 directly.

Model Calibration

Model calibration needs some input files. These include the vehicle age distribution (VAD) file, the socioeconomic index (SI) file, and the parameter file. Table 1 & 3 are descriptions, while Table 2 &4 are samples of the VAD file and SI file.

Table 1 Description of Vehicle Age Distribution (VAD) File

Line	Field	Description
1	1	File tile (up to 40 characters)
2	1	Year for the oldest observation data
3	1	Name of vehicle type
3	2^+	The dependent data for the corresponding vehicle type defined in field 1 at the corresponding year
4^+		Repeat line3, till the dependent data for all the vehicle types are included
...		Repeat line 3-4^+, till the dependent data for all the observation years are included



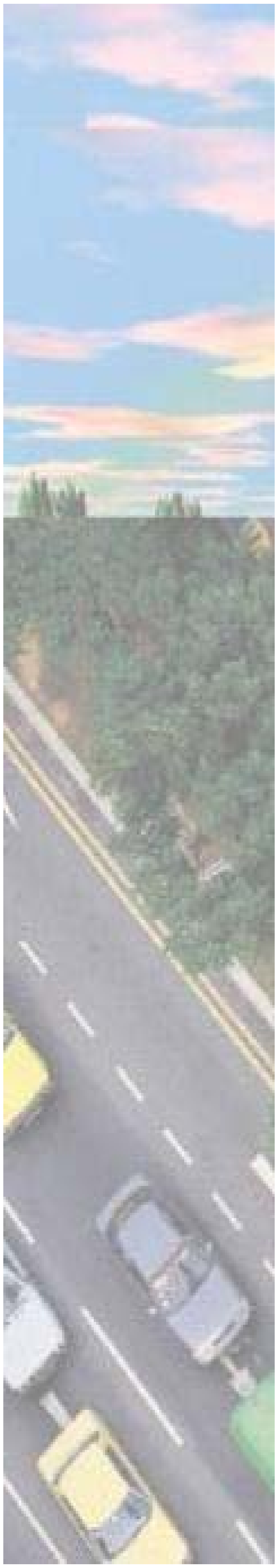


Table 2 Sample Vehicle Age Distribution (VAD) File

This is a dependent file for modeling of sample

1994								
LDV	73684	69949	59259	61020	53868	57266	53992	…
LDT1	7772	5971	5132	5291	4446	4886	4398	…
LDT2	25875	19877	17086	17614	14799	16264	14643	…
LDT3	464	302	178	217	201	227	217	…
LDT4	214	139	82	100	93	105	100	…
HDV2B	319	294	225	254	246	188	214	…
HDV3	28	26	20	22	22	17	19	…
HDV4	21	19	15	17	16	12	14	…
HDV5	19	17	13	15	14	11	12	…
HDV6	58	54	41	46	45	34	39	…
HDV7	61	57	43	49	47	36	41	…
HDV8A	31	29	22	25	24	19	21	…
HDV8B	73	67	51	58	56	43	49	…
HDBS	22	20	16	18	17	13	15	…
HDBT	3	3	2	2	2	2	2	…
MC	1018	866	601	370	409	481	346	…
1995								
LDV	73684	69949	59259	61020	53868	57266	53992	…
…								

Table 3 Description of Socio-Economic Index (SI) File

Line	Field	Description
1	1	File tile (up to 40 characters)
2	1+	Name of the independent variable(s), format **A**(12)
3	1	Year for the oldest observation data
4	1	Socioeconomic data for the corresponding independent variable(s) at the corresponding year
5+		Repeat 3-4, till all the years are included

Table 4 Sample Socio-economic Index (SI) File

Year	Population	Age20	Age20-65	Age65	…
1994	3034628.00	973928.00	1878236.00	227087.00	…
1995	3064758.00	982583.00	1897861.00	232696.00	…
1996	3104190.00	1000059.00	1917360.00	239361.00	…
1997	3202021.00	1026838.00	1986519.00	249669.00	…
…					

Besides the two input data files (VAD file and SI file), there are some control parameters to be inputted by the users. The control parameters are listed in Table 5.

Table 5 Control Parameters

General Parameter

Number of vehicle types [integer]: Number of vehicle types (default: 16).

Number of vehicle ages [integer]: Number of vehicle ages (default value is 25.

Number of socioeconomic indices [integer]: Number of socioeconomic factors.

Number of observation year [integer]: Number of years for which observation data is provided.

Whether optimal selection of socio-economic indices is needed: if x=1 then optimal selection is needed

Input Data Type

If x=1 then the VAD file is in the unit of absolute number of vehicles (TYPE I).

If x=2 then the VAD file is in the unit of percentage. In this case, all data should be within the range of [0.0, 1.0].

With the above data files and control parameters, the program can calibrate corresponding parameters within models. Two files, parameter.txt and socioindex.txt, will be generated and stored in the output directory containing the information of calibrated parameters and name of selected socio-economic indices. When no optimal selection of socio-economic indices is needed, socioindex.txt contains the same information as the user provided before calibration. This happened when users feel confidence based on their own experiences. The can specify socio-economic indices like population, income, etc. However, parameter calibrations are still necessary even in this case.

Vehicle Age Distribution Forecasting

The forecasted vehicle age distribution will be easily obtained if the calibration information as well as the necessary socio-economic indices is provided to the program. The calibrated information contains the calibrated parameters and the name of needed socio-economic indices. The format of the inputted socio-economic indices file for forecasting is the same as that for calibration (Table 4.) However, the number of socio-economic indices will be less if the optimal selection of socio-economic indices were applied.

Outputs from MOFAD

Four types of output files have been used for forecasting.

outst.txt: provides forecasting result with the format specifically designed for MOBILE6 input;

outin.txt: outputs all the input data for users to check;

outsu.txt: provides summary information of the result calculated from the model;

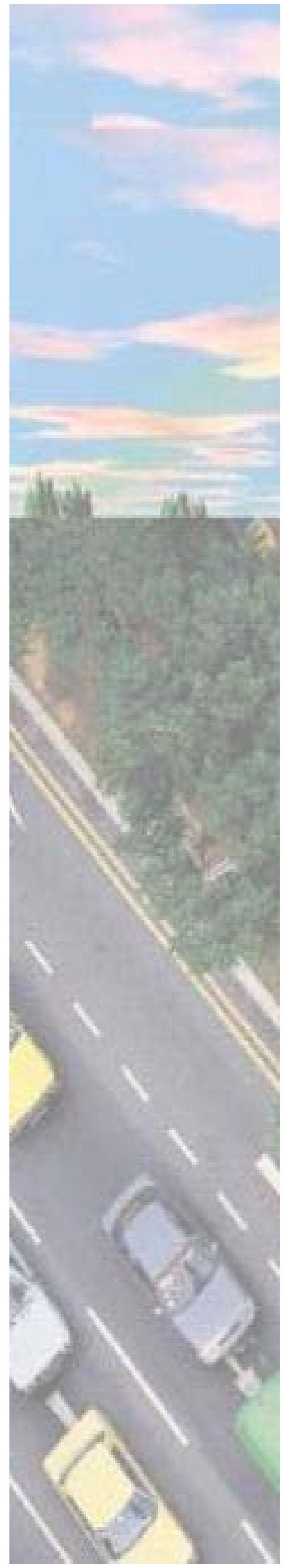
outde.txt: provides detailed statistical result derived from the model calculation.

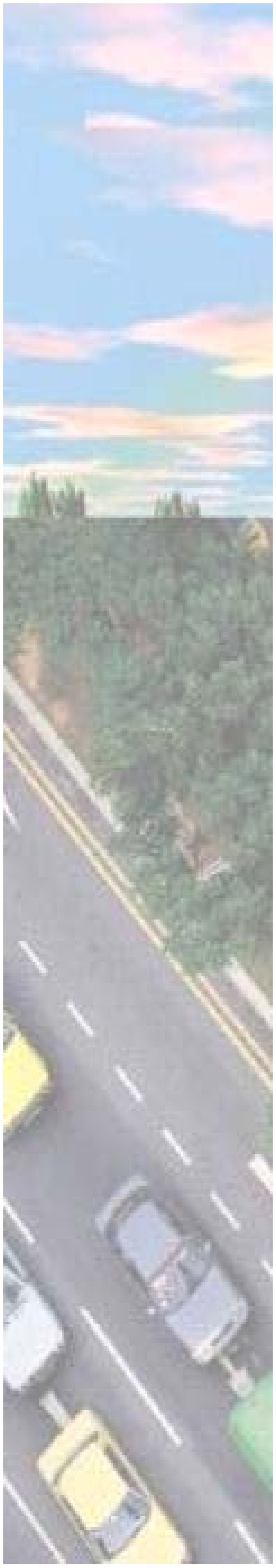
The standard output (outst.txt) can be inputted into MOBILE6 directly, while outin.txt has the same format as that in Table 2 and Table 4.

The samples of the standard output (outst.txt), the summary output (outsu.txt) and the detailed output (outde.txt) are listed in Table 6 through Table 8.

Table 6 Sample Standard Output File (outst.txt)

```
This is a sample standard output file
 1 .08775  .08400  .07996  .07923  .07842  .07624  .06587  .06126  .05567
   .05044  .04529  .04184  .03598  .03103  .02725  .02043  .01339  .00904
   .00755  .00692  .00695  .00635  .00465  .00329  .02120
```





```
 2 .00438  .11418  .09734  .09507  .09121  .09629  .09620  .07742  .00190
   .00173  .08218  .05518  .04774  .04014  .02118  .02188  .00273  .01568
   .00672  .00103  .00062  .00312  .00433  .00391  .01784
 3 .00261  .16353  .13989  .13607  .12992  .00211  .00196  .10888  .00112
   .00103  .00107  .08172  .06974  .00147  .03423  .00169  .00162  .03118
   .01732  .00631  .00151  .00132  .00111  .00985  .05276
 4 .25738  .20378  .11145  .09004  .05278  .04413  .03140  .02211  .01636
   .01497  .01377  .01317  .01067  .01123  .01339  .01150  .00762  .00917
   .01008  .00865  .00801  .00812  .00364  .00419  .02240
   …
16 .12998  .10254  .08339  .07356  .06548  .04994  .03884  .02929  .01472
   .00941  .01245  .01555  .01166  .02098  .03088  .03408  .03454  .03999
   .04092  .03157  .02469  .01922  .01509  .01165  .05955
```

Table 7 Sample Summary Output File (outsu.txt)

```
The criteria (Error Mean [s] Square          ) for the model is:     3580985.000000
 The model for vehicle type 1 and age 1 is:

   Y=     .416E+06+(     .912E+01)X  5+(    -.203E+02)X10+(    -.267E+03)X11+(    -
.109E+01)X16

 The criteria (Error Mean [s] Square          ) for the model is:      116931.500000
 The model for vehicle type 1 and age 2 is:

   Y=   -.304E+06+(    .581E+05)LgX 4+(   -.313E+05)LgX 5+(    .394E+05)LgX13+(
-.175E+05)LgX15

 The criteria (Error Mean [s] Square          ) for the model is:      577033.600000
 The model for vehicle type 1 and age 3 is:

   Y=   -.260E+05+(    .126E+00)X 1+(   -.113E+01)X 5+(   -.471E-02)X13+(    .433E-
02)X14

...
```

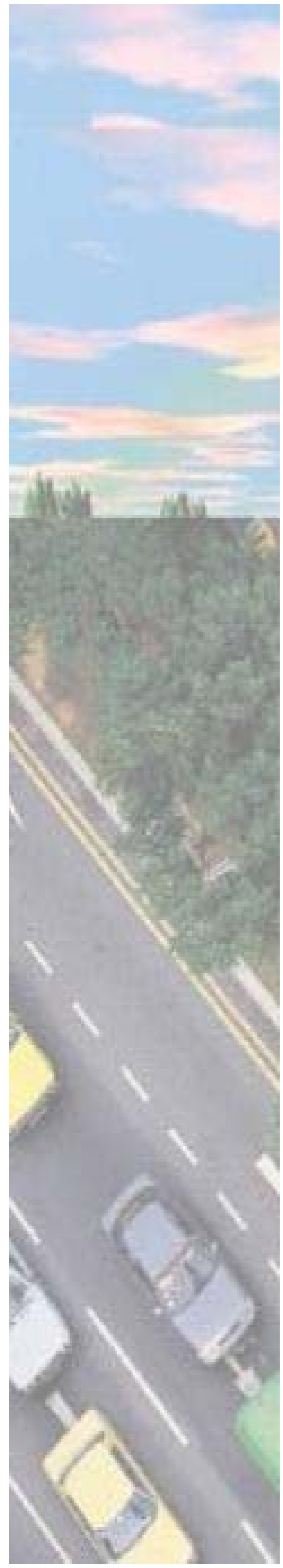
Table 8 Sample Detailed Output File (outde.txt)

```
****************************************************************************
The output of the results for vehicle type 11 and age 11
****************************************************************************
The suitable model is selected from total  7316 candidate models!
The selected input socioeconomic indexes for the final model are:
[11   14   18   19]
R-squared    Adjusted     Est.  Std. Dev.              Coefficient of
(percent)    R-squared     of Model  Error     Mean    Var. (percent)
91.585          74.754            .004332    .03967             10.92

* * * Analysis of Variance * * *
Sum of          Mean                        Prob. of
Source                DF       Squares       Square   Overall F        Larger F
Regression            4    4.085E-04   1.021E-04        5.442             .1612
Residual               2    3.753E-05   1.877E-05
Corrected Total     6    4.460E-04

* * * Sequential Statistics * * *
Indep.     Degrees of       Sum of                         Prob. of
Variable      Freedom       Squares     F-statistic       Larger F
1            1               2.128E-04       11.338          .0780
2            1               1.515E-04         8.073          .1048
3            1               4.319E-05         2.301          .2686
4            1               1.022E-06           .054          .8372
* * * Inference on Coefficients * * *
```

```
                      Standard                  Prob. of   Variance
Coef.   Estimate         Error   t-statistic  Larger |t|    Inflation
  1      .3646           .3511      1.038       .4082     45976.7
  2      .0384           .0436       .880       .4716         2.0
  3     -.0797           .0399     -1.999       .1836         2.0
  4      .0675           .2227       .303       .7906       660.4
  5     -.0404           .1733      -.233       .8372       688.3
   * * * Variance-Covariance Matrix for the Coefficient Estimates * * *
   1          2            3            4            5
   1      .12326   -.010649   -.011228   -.017048     .012338
   2                 .001905    .000375      .0045    -.003604
   3                            .001590   -.002989     .002443
   4                                          .0495    -.038565
   5                                                    .030050
   ************************************************************************
   The output of the results for vehicle type 11 and age 12
   ************************************************************************
   The suitable model is selected from total 7316 candidate models!
   The selected input socioeconomic indexes for the final model are:
   [11   14   18   19]
…
```



AJUSTING VEHICLE MILEAGE ACCUMULATION RATES BASED ON SMALL SAMPLE SURVEY

Methodology Description

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

A realistic approach to solving this problem is to conduct a small sample survey in the concerned area, and then to adjust MOBILE6 default values by incorporating with the local information collected. The small sample survey may not exactly cover all the vehicle types required by MOBILE6.

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 (i.e. real MAR/default) for all sampled vehicle types;

Step 3: Calculate the mean, confidence interval and Maximum Relative Error (MRE) for each surveyed vehicle types;

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.

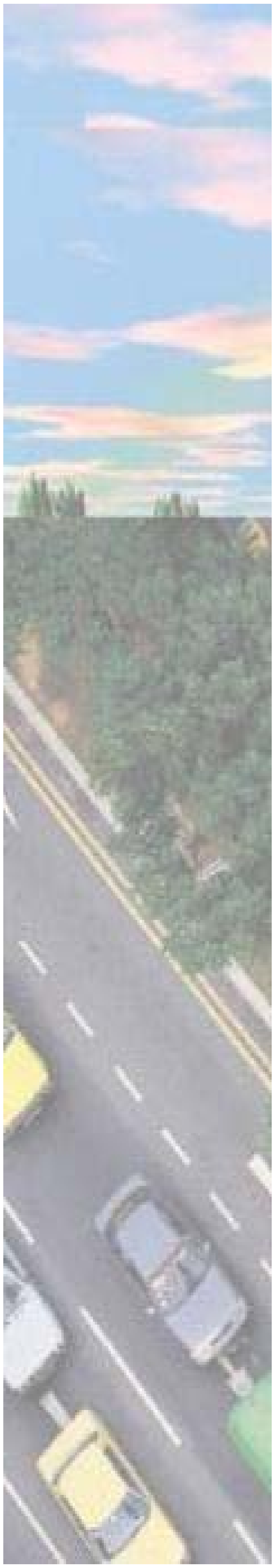


Figure 3 illustrate the whole modeling process.

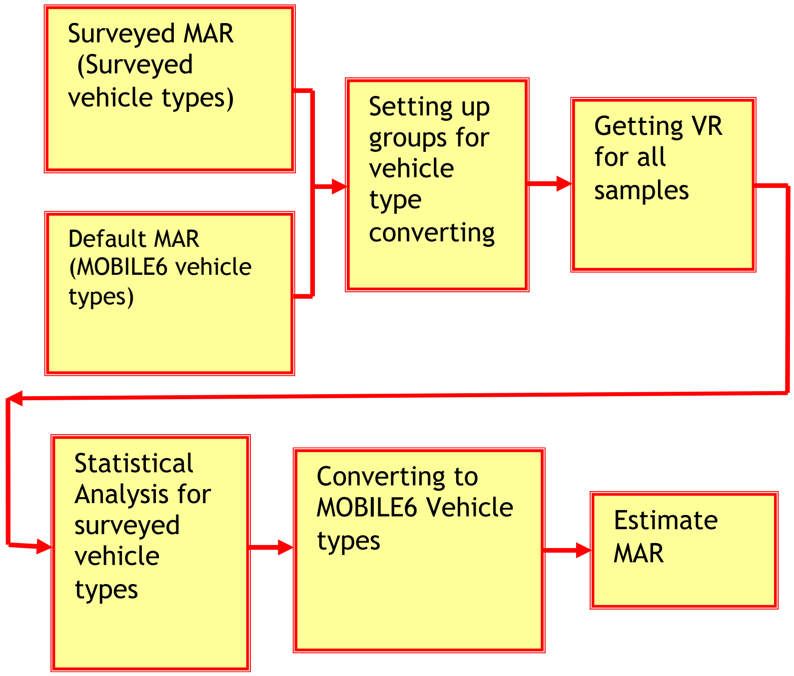


Figure 3 Modeling process for adjusting vehicle mileage accumulation rates

Sample Size Determination

Sample size is a key factor affecting survey errors. The minimum sample size L on a particular surveyed vehicle type is required by the following equation providing both sample mean and sample variance are known with a certain confidence and maximum relative error (MRE):

$$L > \mathrm{int}\left(t_{\alpha/2,(L-1)} \cdot \frac{S}{\mathrm{MRE}} \Big/ \bar{r} \right)^2 + 1 \qquad (1)$$

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

In the above equation, MRE is a pre-required maximum relative error with the confidence $(1-\alpha)$. $\bar{r}$ and S are mean and standard deviation of Variance Ratio (VR). $t_{\alpha/2,L-1}$ can be firstly chosen as $L=\infty$ to get a initial L. So an initial $t_{\alpha/2,\infty}$ can be gotten for updated L.

Here is an example of how to determine the sample size. Suppose we are now interested in the Mileage Accumulation Rate (MAR) for car. The first thing to do is to conduct a sample field survey and get the average mileage accumulation rate in the past year for some (say, 10 or 20) vehicles. Then the Variation Ratio (VR) for each car can be gotten by dividing the surveyed MAR by the nationwide default one, which is the average default values for say, LDGV and LDDV. Then mean $\bar{r}$ and variance S of VR for car can be calculated.

Suppose by calculation from the sample survey, the mean $\bar{r}$ and variance S of VR for car are 1.5 and 1.0, respectively, and the pre-required maximum relative error is 10% with the confidence level needed as 90%=(1-0.1), therefore, $t_{\alpha/2,L-1} = t_{0.05,\infty}$ can be found as 1.645 from the table of *t*-Distribution in the appendix of this Guide book and also in any probability books. So the initially calculated sample size can be obtained as int((1.645*1.0/0.1/1.5)^2)+1=int(120.3)+1 =121. Then by rechecking the table of *t*-Distribution and updating $t_{\alpha/2,L-1} = t_{0.05,120}$ as 1.658, the recalculated sample size can be obtained as int((1.658*1.0/0.1/1.5)^2+1)=int(122.2+1)=123. This is the pre-determined minimum sample size.

Subsequently the real field survey can be conducted with a sample size larger than 123, and the Variation Ratios as well as mean and variance of Variation Ratios can be obtained. For example, the real sample size selected is 150, and the mean and standard variation after survey is 1.6 and 1.1, respectively. So, the corresponding minimum sample size should be int((1.645*1.1/0.1/1.6)^2+1)= int(127.9+1)=128<150. Therefore, the used sample size 150 is good to guarantee tolerable survey errors.

However, additional survey is needed in case the needed minimum sample size is larger than the one used (i.e. 150 in this example.)

Case Study in Houston and El Paso

The following is the case studies in two Texas areas: Houston and El Paso. Houston is a big city area with large population and area; while El Paso is a relatively medium one. 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.

In the case study in Houston and El Paso, four vehicle types were surveyed and the corresponding group setting was listed in Table 9.

Table 9 Group Setting for Adjusting Vehicle Mileage Accumulation Rates for Houston and El Paso Area

Groups	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

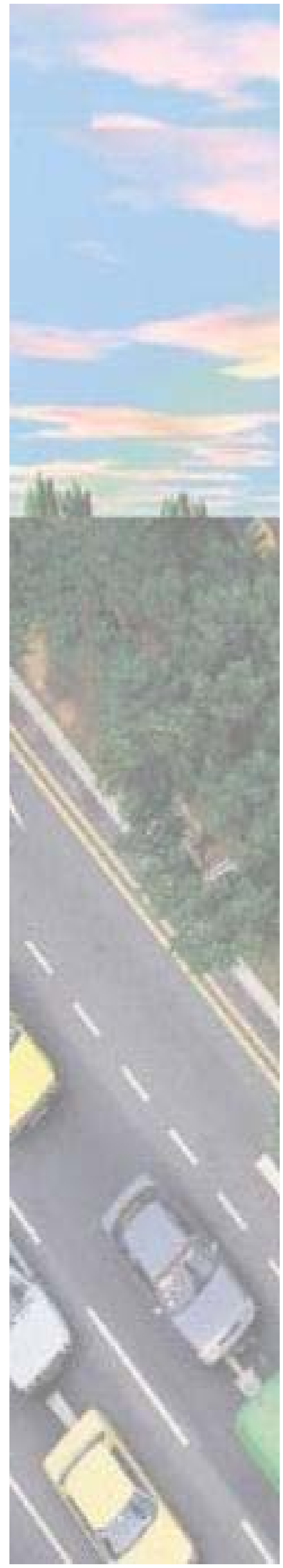


Table 10 & 11 lists the statistical results of Variance Ratio (VR) for the two areas, where MRE (%) is the maximum relative error with 90% confidence level while Std means Standard Deviation.

Table 10 Statistical Result for Houston

	Average	Std	Sample Size	MRE (%)
Car	1.422	1.164	426	6.53
Truck	1.382	1.381	231	9.76
SUV	1.149	0.738	81	11.73
Van	1.214	1.071	60	18.74
SUV, Van	1.177	0.892	141	10.50
All Type	1.367	1.192	798	5.08

Table 11 Statistical Result for El Paso

	Average	Std	Sample Size	MRE (%)
Car	0.690	0.272	93	6.73
Truck	0.476	0.261	15	18.88
SUV	0.615	0.244	20	14.61
Van	0.703	0.267	9	20.86
SUV, Van	0.642	0.250	29	11.90
All Type	0.656	0.273	137	5.84

In the both tables, the Maximum Relative Error (MRE) is calculated by using the following equation:

$$\mathrm{MRE} = t_{\alpha/2,(L-1)} \frac{S}{\sqrt{L}} \Big/ \bar{r} \cdot 100\% \qquad (2)$$

All the variables in the above equation are defined the same as that for determining the sample size. In a whole, the calculated mean and standard deviation of Variation Ratio varies a lot from Table 10 to Table 11. However, slight fluctuations exist among different vehicle types in the same location (i.e. for Houston only or for El Paso only.)

The adjusted vehicle mileage accumulation rates in Houston and El Paso can be obtained based on the surveyed Variance Ratios and the converting matrix for vehicle types in Table 9.

The corrected local mileage accumulation rates in Houston County are much higher than the default ones, while that 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 Houston is a big urbanized area with large population lived in or nearby, while El Paso is a relatively smaller one. People in Houston should travel more than the average U.S. cities, while people in El Paso may travel less than the average.

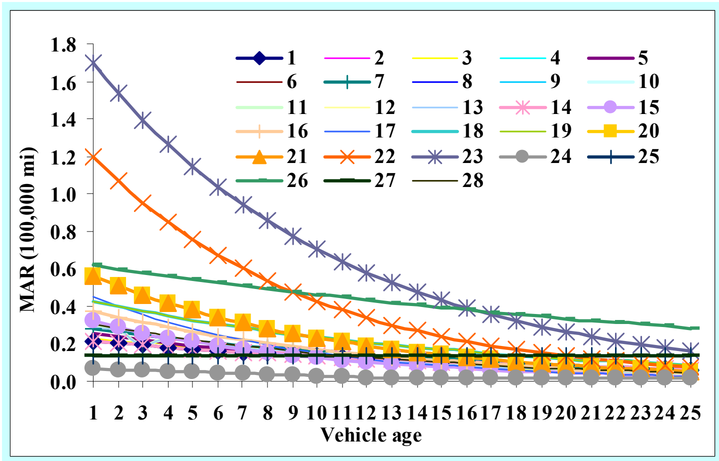


FIGURE 5 Corrected local mileage accumulation rate for Houston.

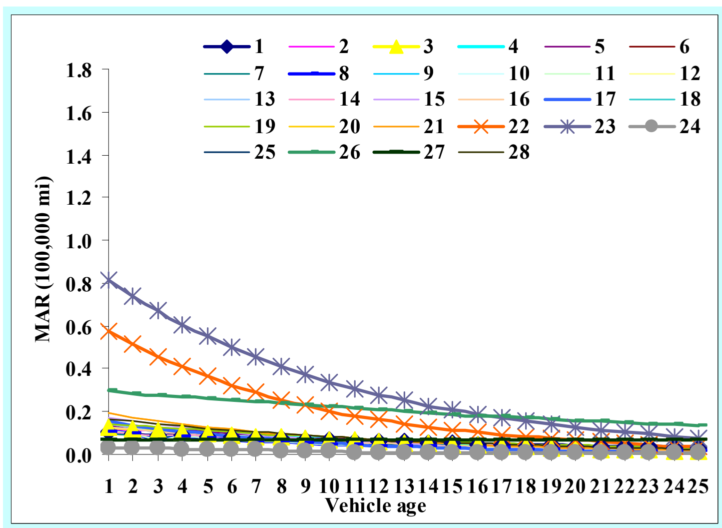
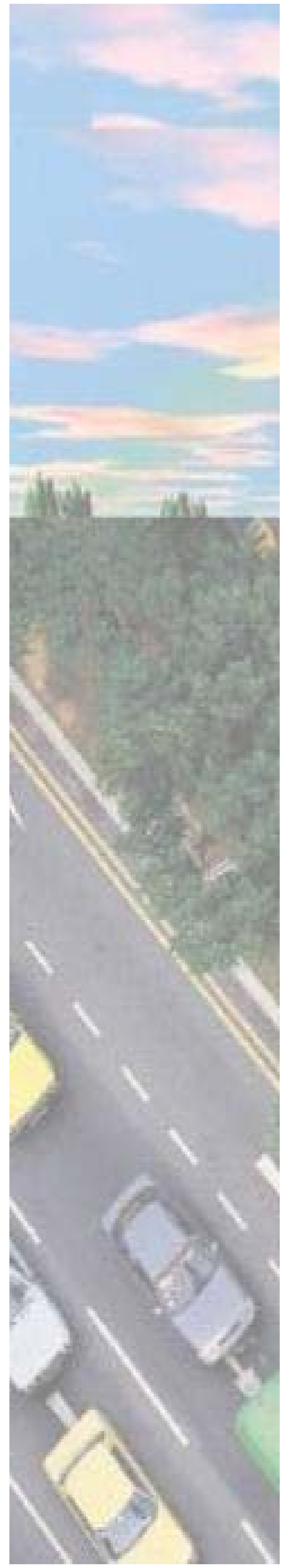


FIGURE 6 Corrected local mileage accumulation rate for El Paso area.

The adjusted mileage accumulation rates in both Houston and El Paso areas were inputted 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 Figure 7. From Figure 7 we can see that the emission factors in Houston are much higher than in El Paso. The trends are similar as that for mileage accumulation rate.



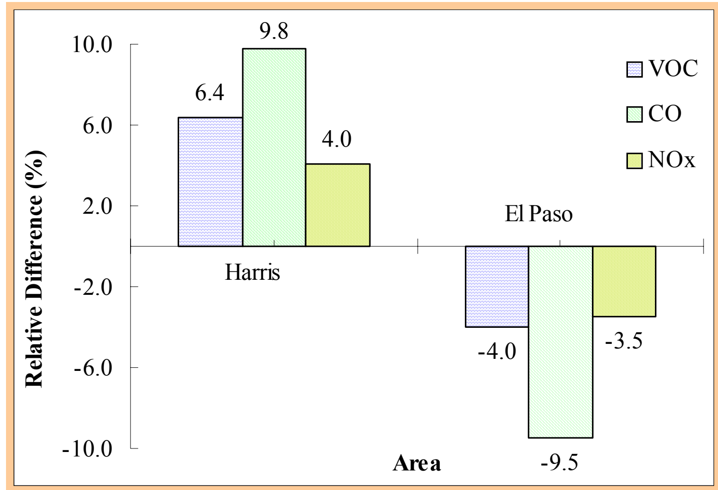


Figure 7 Relative differences of emission factors when compared adjusted mileage accumulation rates with MOBILE6 default values for Houston and El Paso.

IMPROVEMENT TO VMT ESTIMATION FOR AIR QUALITY ANALYSIS

Methodology Description

Currently there are several VMT (& mix) estimation methods. Traffic Count Method (TCM) 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, without consideration of useful traffic counts. The *improved* methodology sets 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's Traffic Count Method since it is a useful and the only method that can disaggregate volume according to the requirements by MOBILE6.

The calibrated coefficients in the improved volume estimation algorithm can be obtained by a multivariate regression analysis using any standard routine.

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

Step 1: Estimate volume at links without traffic counts;
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 v/c ratio using either link-specific capacities or lookup tables;

Step 5: Apply 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 those of Traffic Count Method (TCM) except for the important new feature in Step 1, where the link traffic volume is estimated by both link attributes and count volume instead of the simple extension of the count volume from the station.

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.

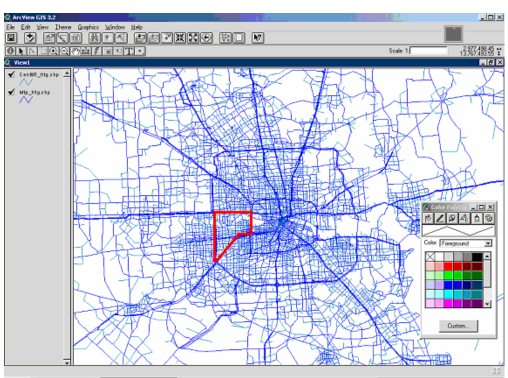
These disaggregating processes of the methodology are conceptually straightforward, although their calculations might be relatively complex. To obtain the hourly VMT, the distribution of link-level volumes by hour of day should be prepared by using the user-provided distribution. If the user does not 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 by the hourly volume, the hourly link VMT can then be obtained directly.

After obtaining the hourly distributions of VMT by speed for all kinds of facilities, it becomes easy to derive 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.

The following sections further illustrate the whole process with the case study in southwest Houston.

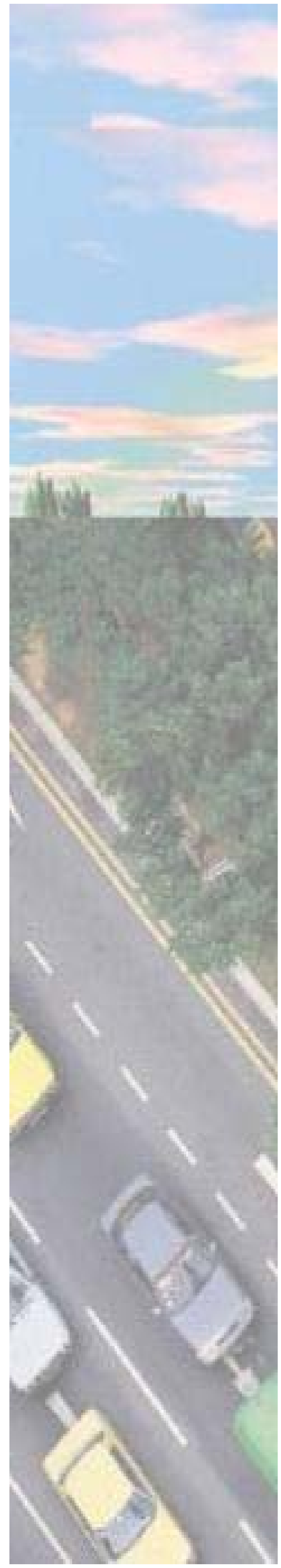
Information Collection for Case Study

The selected sub-network for case study contains 276 links with 34 freeway links, 110 arterial links and 132 local street links. While there is no ramp information available, the final ramp VMT can be a portion of freeway VMT. EPA suggests this portion as 8.7%.



Note: illustrates the area for case study

Figure 8 GIS network for Houston area in the format of ArcView



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.

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.

Link Volume and VMT Estimation

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

The link volume estimation starts from the calibration of parameters $\alpha_{kt}^{j}\ (\forall j = 0, 1,\ 2, \ldots\ n_x)$ in the following equation:

$$v_{kt}^{i_y} = c_{yt}^{i_y} \cdot e^{\alpha_{kt}^{0} + \sum_{j=1}^{n_x} \alpha_{kt}^{j} \cdot x_{kt}^{j}} \qquad (3)$$

In the above equation, $c_{yt}^{i_y}$ is the volume for vehicle type *t* at count station i_y, where $i_y = 1, 2, \ldots, I_y$, I_y is the total number of count stations on road type *y*, *y* is the road type, *t* is the vehicle type.

$v_{kt}^{i_y}$ is the volume for vehicle type *t* on link k with road type *y*. It can be estimated from any of the count station i_y located on the same road type (i.e. type y).

The calibrations of coefficients were conducted among the "known" count data, i.e. the 42 links where count data were available. Following the regular multivariate regression process, the calibrated coefficients are obtained and listed in Table 12. In Table 12, the calibrated coefficients for freeway, arterial, and local streets are different, meaning that the different relationships existed inside.

After the calibration process, the volume on link k for vehicle type t based on the count data from link i_y can be estimated by Equation (3). 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:

$$\hat{v}_{kt} = \frac{1}{I_y} \sum_{i_y=1}^{I_y} \hat{v}_{kt}^{i_y} \tag{4}$$

TABLE 12 Coefficients for Volume Estimation Model Calibrated by Southwest Houston Real Data

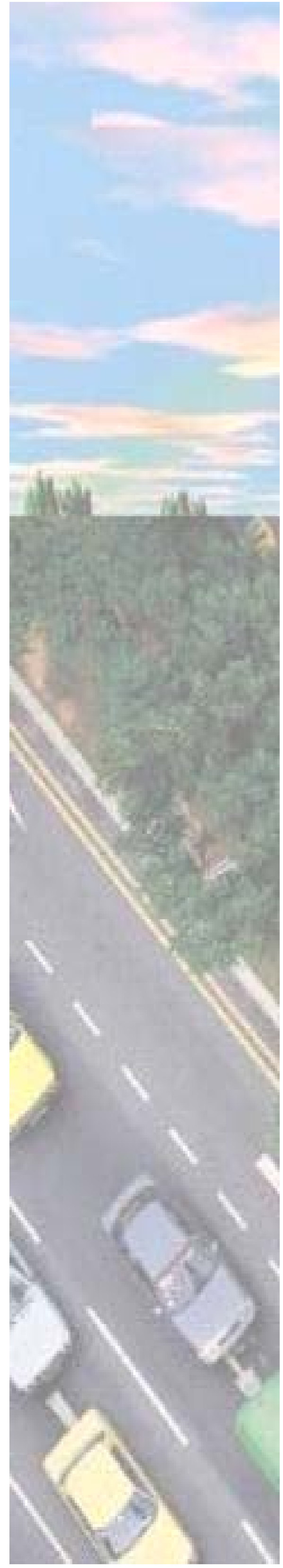
Coefficients	Facility Types		
	Freeway	Arterial	Local Street
α_0	-8.6×10^{0}	-9.2×10^{-1}	2.7×10^{0}
α_1	-3.6×10^{-6}	1.8×10^{-5}	-5.8×10^{-5}
α_2	-3.0×10^{-4}	-4.2×10^{-2}	5.6×10^{-2}
α_3	1.5×10^{-1}	4.8×10^{-2}	-1.4×10^{-1}
α_4	-1.0×10^{-2}	-3.9×10^{-3}	-3.5×10^{-2}

The link VMT could then be obtained by multiplying the estimated link volume with its corresponding link length. The link VMT should be disaggregated according to the hours of day so that hourly VMT can be obtained. 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 will 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.

The speed VMT can be estimated as well. The hourly-congested speeds can be achieved by applying the BPR curve, the standard form of which is:

$$s_k = s_{kf} \Big/ \left(1 + a_k \left(v_k / c_k\right)^{bk}\right) \tag{5}$$

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.



Based on the above estimation, VMT related variables were estimated. Table 13 lists one of the results - the VMT fractions on all facility types. The VMT fractions on facility types by improved method with two scenarios (Improved (I) and Improved (A)) were closer to the real one than the Traffic Count Method (TCM). Improved (I) means that the improved model calibrate each facility type independently; while Improved (A) means that the calibrated model for freeway was used to links of all facility types.

Table 13 Comparison of VMT Split by Facility Types Based on Different Methods in Southwest Houston Area

VMT	Freeway	Arterial	Local	Ramp	Total
Real	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

Impacts to Emission Estimation

In the case study, the final estimated local VMT related variables were inputted into MOBILE6 so that the impacts on emission estimations could be obtained. Figure 9 lists the estimates for three emission factors VOC, CO and NO_x based on different scenarios. Improved (I) and Improved (A) have the same meaning as in Table 13. It is shown in Figure 9 that the proposed improvements have better estimation on the emission factors than both the nationwide default one and TCM estimation. 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. The relative errors of the three emission factors for the improvements are smaller than both TCM and default values.

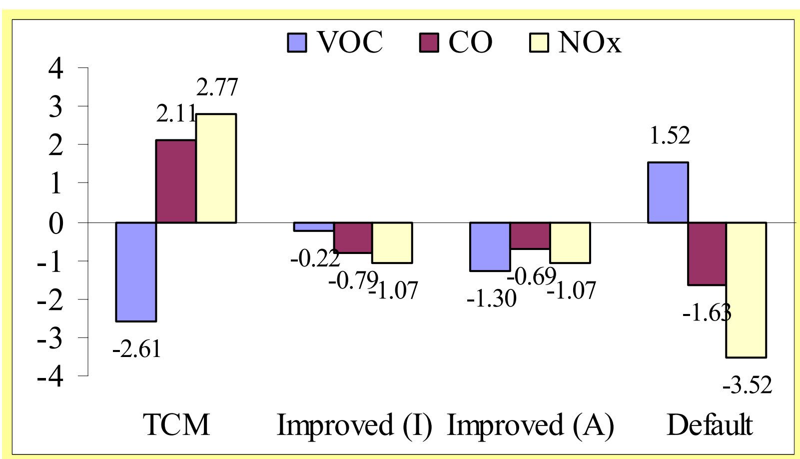
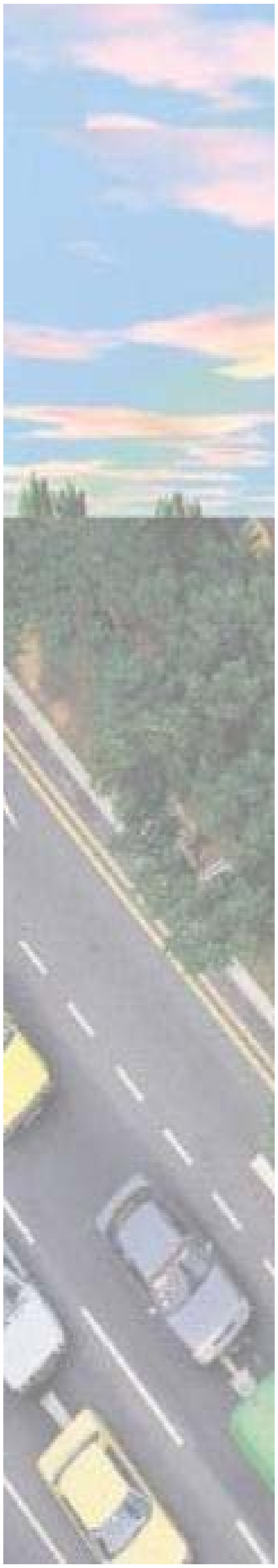


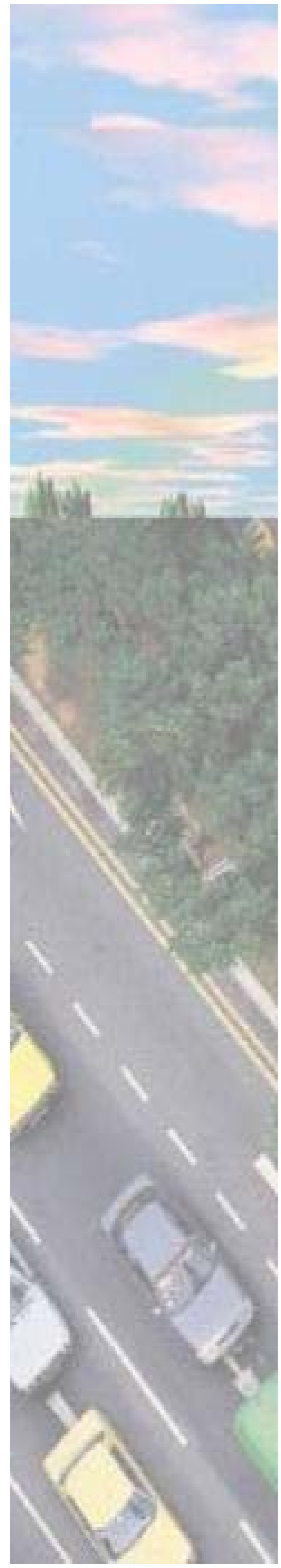
Figure 9 Relative errors (%) for estimates of emission factors based on different scenarios.

CHAPTER 3

SUMMARY AND RECOMMENDATIONS



This chapter provides a summary as well as recommendations for the implementation plan using the techniques and models of forecasting traffic characteristics for air quality analyses.



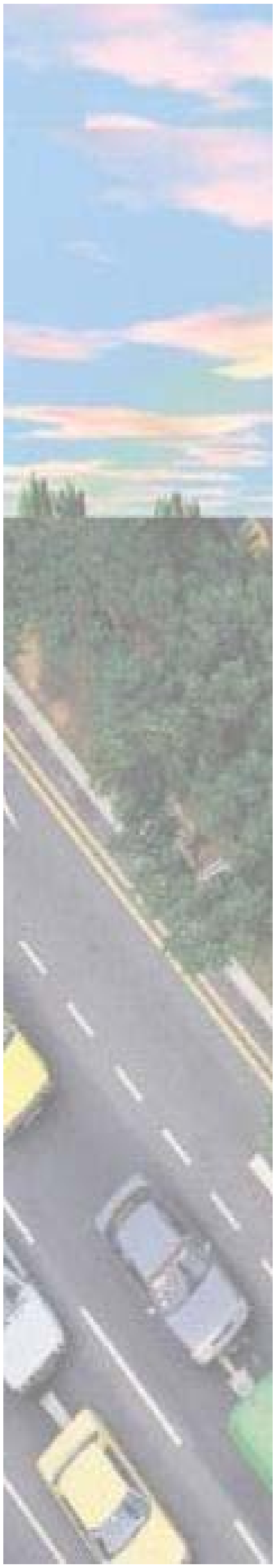
Summary

In this research, techniques for estimating and forecasting the three critical mobile source emission related travel indicators were developed. The three indicators are vehicle age distribution, mileage accumulation rates by vehicle type, and VMT related variables.

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 used for examination.

As for mileage accumulation rates, the modeling of the correcting process for mileage accumulation was developed. 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. From the results, the real mileage accumulation in Houston area is 1.34 times higher than the national-wide default value, and in El Paso it is 0.58 times less than the default value.

As for VMT, there are currently several methodologies on VMT related variables' estimation. However, none of the existing approaches can be directly used for MOBILE6 in Texas. In this research, the improvements were proposed by considering 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 (TCM) and the nationwide MOBILE6 defaults in terms of the estimation of both VMT and emission factors.



RECOMMENDATIONS TO IMPLEMENTATION PLAN

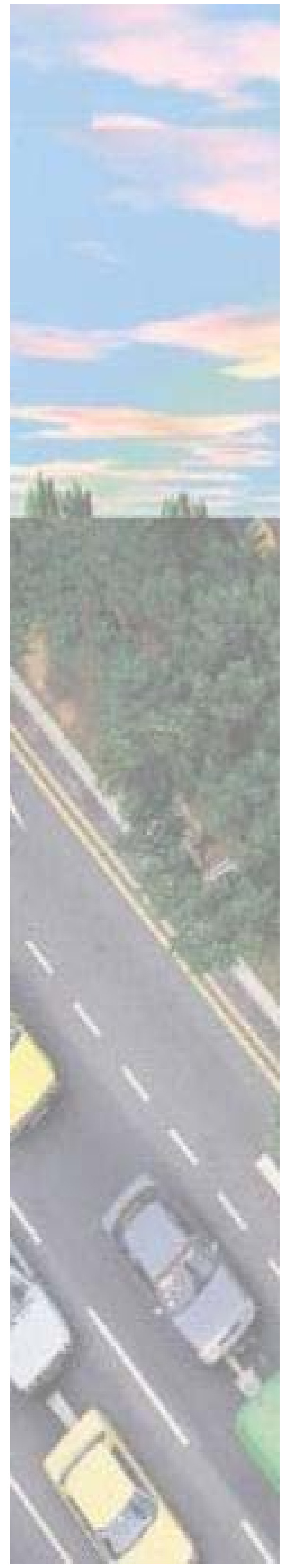
The objectives of implementation are to put developed techniques into practice in Texas; to build up suitable programs for practical use; to modify the proposed model based on more practical information; and to provide suggested default values of three travel indicators for different Texas counties/cities and for entire state of Texas.

For forecasting vehicle age distribution, information for different Texas counties should be collected, including historical vehicle age distribution and local socioeconomic indices. The computer software realizing the techniques will be improved with a user-friendly interface being suitable for different types of users. The software should produce rich output information while state default and county defaults be generated.

For mileage accumulation rate, several prototype cities in Texas will be selected representing different city types. Historical mileage accumulation data mainly on Inspection/Maintenance data for selected cities will be collected. Workable spreadsheet realizing the corresponding algorithm will be developed. Default values of mileage accumulation rates for the analyzed cities will be estimated and extended to all Texas cities including the entire Texas State.

As for VMT related variables, several prototype cities in Texas representing different city categories will be selected first. Then, link attributes and count data for selected cities will be collected. The computer software for VMT estimation is going to be developed with user-friendly interface being suitable for different types of users. The outputs should meet all the requirements by MOBILE6. After obtaining VMT related variables for selected cities, algorithms for extending to all Texas cities will be developed and default values for all the cities and the entire Texas State will be obtained.

REFERENCES



1. 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, DC., January 2000.

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

3. Cambridge Systematics, Inc. Volume IV: Charpter 2. *Use of Locality-Specific Transportation Data for the Development of Mobile Source Emission Inventories*. Final Report, Prepared for Mobile Sources Committee, Emission Inventory Improvement Program. September, 1996.

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

5. 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.

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

7. 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.

8. Environment Protection Agency, 2000. User's Guide to MOBILE6: Mobile Source Emission Factor Model, DRAFT. *Report EPA420-P-00-005*. November 2000.

9. Environment Protection Agency. Draft User's Guide to MOBILE6. *EPA420-D-01-002a*. United States Environment Protection Agency, Feb. 2001. http://www.epa.gov/oms/m6.htm. Accessed July 06, 2002.

10. 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.

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

12. 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.

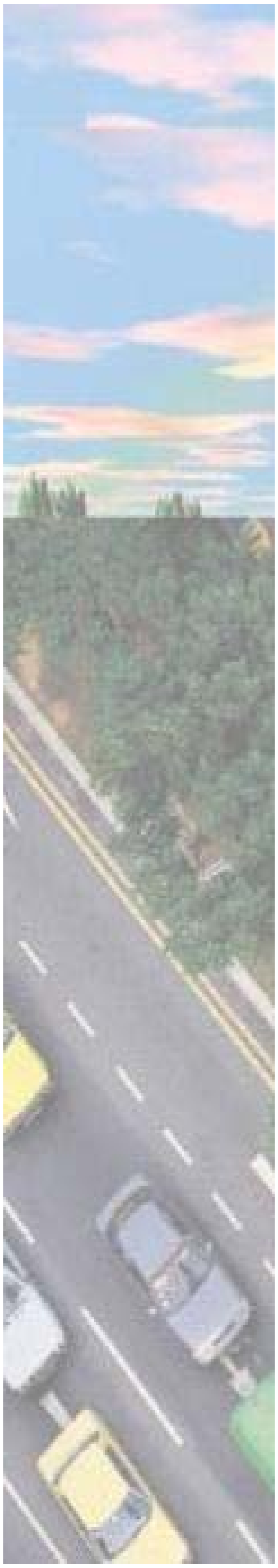
13. Information Services, Oregon State University. Government Information Sharing Project. http://govinfo.kerr.orst.edu/index.html. Accessed Feb. 20, 2001

14. 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, Preprints of the *80th Annual Meeting of Transportation Research Board* (CD-ROM), Washington D.C., January 2001.

15. 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.

16. 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.

17. 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.



18. 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.

19. 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.

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

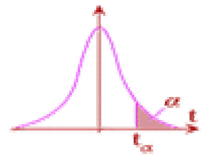
21. U.S. Environmental Protection Agency. *Frequently Asked Questions on MOBILE6*. January 16, 2002

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

23. 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.

24. 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
T-Distribution Table



df	α = 0.1	0.05	0.025	0.01	0.005	0.001	0.0005
∞	1.282	1.645	1.960	2.326	2.576	3.091	3.291
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.920	4.303	6.965	9.925	22.328	31.600
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.689
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.660
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.091	3.291

More information are available in Table 12 of Biometrika Tables for Statisticians, vol. I, edited by E. S. Pearson and H. O. Hartly, Cambridge University Press, Cambridge (1954), and Table III of Statistical Tables for Biological, Agricultural, and Medical Research, R. A. Fisher and F. Yates, Oliver & Boyd, Edinburgh, 1953.