PB2001-108554

RESEARCH REPORT 1838-6

MODELING SOAK-TIME DISTRIBUTION OF TRIPS FOR MOBILE SOURCE EMISSIONS FORECASTING: TECHNIQUES AND APPLICATIONS

Harikesh S. Nair and Chandra R. Bhat

CENTER FOR TRANSPORTATION RESEARCH
BUREAU OF ENGINEERING RESEARCH
THE UNIVERSITY OF TEXAS AT AUSTIN

AUGUST 2000

REPRODUCED BY: NTIS
U.S. Department of Commerce
National Technical Information Service
Springfield, Virginia 22161

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/TX-0-1838-6		

4. Title and Subtitle	5. Report Date
MODELING SOAK-TIME DISTRIBUTION OF TRIPS FOR MOBILE SOURCE EMISSIONS FORECASTING: TECHNIQUES AND APPLICATIONS	August 2000
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Harikesh S. Nair and Chandra R. Bhat	0-1838-6
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Center for Transportation Research The University of Texas at Austin 3208 Red River, Suite 200 Austin, TX 78705-2650	11. Contract or Grant No. 0-1838
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Texas Department of Transportation Research and Technology Transfer Section/Construction Division P.O. Box 5080 Austin, TX 78763-5080	Research Report (9/99-5/00)
	14. Sponsoring Agency Code

15. Supplementary Notes

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

16. Abstract

The soak-time of vehicle trip starts is defined as the duration of time in which the vehicle's engine is not operating and that precedes a successful vehicle start. The temporal distribution of the soak-time in an area is an important determinant of area-wide mobile source emissions. In the current paper, we formulate and implement a methodology for modeling soak-time durations. The methodology involves estimation of models using vehicle trip data from household travel surveys and supplementary zonal demographic/land-use data. The effectiveness of the methodology lies in its easy application at the traffic zonal level within a metropolitan region to obtain zone-specific soak-time distributions by time-of-day and origin activity purpose. The methodology is applied to estimate soak-time duration models for the Dallas-Fort Worth area of Texas.

17. Key Words	18. Distribution Statement
soak-time, mobile source emissions, start emissions, operating mode fractions, first starts, non-first starts, MOBILE emissions factor model	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	36	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

MODELING SOAK-TIME DISTRIBUTION OF TRIPS FOR MOBILE SOURCE EMISSIONS FORECASTING: TECHNIQUES AND APPLICATIONS

by

Harikesh S. Nair and Chandra R. Bhat

Research Report Number 0-1838-6

Research Project 0-1838

Transportation Control Measure Effectiveness in Ozone Nonattainment Areas

Conducted for the

TEXAS DEPARTMENT OF TRANSPORTATION

in cooperation with the

U.S. DEPARTMENT OF TRANSPORTATION

Federal Highway Administration

by the

CENTER FOR TRANSPORTATION RESEARCH

Bureau of Engineering Research

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2000

DISCLAIMERS

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 views or policies of the Federal Highway Administration or the Texas Department of Transportation. This report does not constitute a standard, specification, or regulation.

There was no invention or discovery conceived or first actually reduced to practice in the course of or under this contract, including any art, method, process, machine, manufacture, design or composition of matter, or any new and useful improvement thereof, or any variety of plant, which is or may be patentable under the patent laws of the United States of America or any foreign country.

NOT INTENDED FOR CONSTRUCTION,
BIDDING, OR PERMIT PURPOSES

Dr. Chandra R. Bhat
Principal investigator

ACKNOWLEDGMENTS

The authors acknowledge the support of TxDOT Project Director William Knowles of the TxDOT Austin District, George Reeves of the TxDOT Dallas District, Wayne Young of TNRCC, Carol Nixon of the TxDOT Houston District, and Ken Cervenka, Ken Kirkpatrick, Gustavo Baez, Mahmoud Ahmadi, and Chris Klaus of NCTCOG for providing documents and data.

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

TABLE OF CONTENTS

1. BACKGROUND AND SIGNIFICANCE OF WORK 1
2. STATE OF THE ART/PRESENT PRACTICE 2
2.1. MOBILE5 versus MOBILE6 2
2.2. Present practice 3
3. MODEL FRAMEWORK 6
3.1. Model estimation 6
3.2. Model application 8
3.2.1. Soak-time distribution for MOBILE6 8
3.2.2. Start mode fractions for MOBILE5 9
4. DATA PREPARATION 10
4.1. Data sources 10
4.2. Sample formation 11
5. EMPIRICAL ANALYSIS 11
5.1. Sample description 11
5.2. Empirical results 13
5.2.1. Results for first starts vs. non-first starts 13
5.2.2. Results for soak-time duration model for first starts 14
5.2.3. Results for soak-time duration model for non-first starts 15
6. APPLICATION CONSIDERATIONS 17
7. CONCLUSIONS 18
8. REFERENCES 21

LIST OF TABLES

Table 1. Distribution of trip starts by time-of-day 23

Table 2. Distribution of trip starts by activity purpose preceding trip start 23

Table 3. Empirical results for binary logit model for first starts 24

Table 4. Empirical results for soak-time duration model for first starts 25

Table 5. Empirical results for soak-time duration model for non-first starts ... 26

Table 6. Cross-classification of home-based work trips by origin-end activity and time-of-day 27

Table 7. Cross-classification of home-based other trips by origin-end activity and time-of-day 27

Table 8. Cross-classification of non-home-based trips by origin-end activity and time-of-day 27

PROTECTED UNDER INTERNATIONAL COPYRIGHT
ALL RIGHTS RESERVED
NATIONAL TECHNICAL INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE

Reproduced from
best available copy.

1. BACKGROUND AND SIGNIFICANCE OF WORK

The significant contribution of mobile-source emissions to air pollution, combined with the deteriorating air quality in many metropolitan areas, has resulted in a need to determine current/projected emissions inventories, and establish the relative emissions contributions of mobile and other sources. These have to be included in the State air quality Implementation Plan (SIP) for achieving clean air levels consistent with National Ambient Air Quality Standards (NAAQS). In addition, an emissions reduction budget is assigned to each emission source for achieving reasonable further progress (RFP) toward attainment.

The integration of transportation and air quality planning is important for mobile source emissions estimation and for establishing conformity of mobile source emissions to the emissions and emissions-reduction budgets in the SIP. The Environmental Protection Agency (EPA) requires the use of the MOBILE model for such emissions estimations and conformity analyses for all areas except California, which uses the EMFAC7F model (MOBILE5b is the current version of the MOBILE model; MOBILE6 is in development and is likely to become available in Fall 2000).

The emissions factor models (MOBILE or EMFAC7F) require several traffic-related inputs. One of these inputs is the distribution of the soak-time of vehicle trip starts. For its MOBILE6 model, EPA has defined the soak-time as the duration of time in which the vehicle's engine is not operating and which precedes a successful vehicle start (a successful vehicle start is defined as a vehicle start that does not result in a stall). If the soak-time is less than 12 hours, the corresponding engine start is designated a "hot-start". If the soak-time is more than 12 hours, then the engine start is defined as a "cold-start" [1].

MOBILE6 uses the soak-time for a vehicle trip as one of several factors to estimate the emissions associated with the engine start for the trip. The soak-time for a trip is likely to depend on various factors such as the activity purpose preceding the trip start (*i.e.,* the origin activity purpose), the time-of-day of the trip start, and other land-use and socio-demographic characteristics of the zone of trip start. Comprehensive household travel surveys, conducted periodically by Metropolitan Planning Organizations (MPOs),

can be used to extract information on soak-times and can form the basis for the development of models of soak-time duration as a function of the various factors identified above.

In the current report, we formulate and implement a methodology for modeling soak-time durations. The methodology involves estimation of models using vehicle trip data from household travel surveys and supplementary zonal demographic/land-use data. The effectiveness of the methodology lies in its easy application at the traffic zonal level within a metropolitan region to obtain zone-specific soak-time distributions by time-of-day and origin activity purpose. Data from a household travel survey conducted in the Dallas-Fort Worth area of Texas is used in the empirical analysis of the report.

The rest of the report is structured as follows. The next section reviews current practice in modeling trip start emissions and places soak-time duration within this broader context. Section 3 discusses the model estimation and application framework. Section 4 focuses on data sources and data assembly procedures. Section 5 presents the empirical results. Section 6 discusses application considerations related to integrating the soak-time model with travel demand models. The final section concludes the report.

2. STATE OF THE ART/PRESENT PRACTICE

2.1. MOBILE5 versus MOBILE6

MOBILE5 uses the concept of operating mode fractions in determining mobile source emissions. Specifically, the use of MOBILE5 requires the classification of vehicle miles of travel into three operating modes: cold-transient, hot-transient, and hot-stabilized. EPA defines the *transient* mode of operation as all operations before 505 seconds after the start of a trip and the *stabilized* mode as all operations after 505 seconds after the start of a trip. Transient trips are further classified as cold-transient or hot-transient depending on whether the start mode was a cold-start or a hot-start, respectively. For the MOBILE5 model, EPA has defined cold-starts to be starts that occur at least four hours after the end of the preceding trip for non-catalyst-equipped vehicles, and at least one hour after the end of the preceding trip for catalyst-equipped vehicles. Hot-starts are those that occur less than four hours after the end of the preceding trip for non-catalyst

vehicles, and less than one hour after the end of the preceding trip for catalyst equipped vehicles. MOBILE5 recommends using the following percentages as default values for operating mode fractions: cold-transient (20.6%), hot-transient (27.3%), and stabilized (52.1%). The analyst using MOBILE5 also has the option of providing region-specific operating mode fractions. It is important to note here that operating mode fractions represent a combination of start modes (cold versus hot) and running mode (transient versus stabilized).

In MOBILE6, the "start" emissions are estimated separately from the "running" emissions (emissions emitted while the vehicle is being driven). The start emissions are calculated as emission "increments" resulting from vehicle start-ups while the running emissions estimates are based only on hot-stabilized engine operations [1]. The procedure for calculating the start emissions entails partitioning the hot vehicle starts (ranging from 1 minute to 12 hours) into 70 time-bins and assigning an emissions effect to each of the time-bins. From the distribution of soak-times, the proportion of soaks that fall into each time-bin is obtained. Within MOBILE6, the emission value of an average vehicle start is calculated as the sum of the product of the start emission effects associated with each time-bin and the corresponding soak-length activity proportion. The product of this average vehicle start emissions with the number of starts per day represents the start emission level. Estimates of hourly start-emission level values developed using default soak-time distributions representing national average conditions are available within MOBILE6. The user has the ability to accept these default emission level values or develop region-specific estimates by specifying a local distribution of soak-times.

In the next section, we review some of the extant methodologies for modeling start emissions and discuss the distinguishing characteristics of the current study.

2.2. Present practice

The literature on modeling start emissions focuses on estimating start mode fractions (*i.e.,* cold versus hot starts, as defined by MOBILE5) rather than developing the distribution of soak-times. This is because MOBILE5 requires start mode fractions (as

part of developing the operating mode fractions) and not soak-time distributions. As indicated earlier, MOBILE6 advances the practice by emphasizing the underlying disaggregate soak-time distribution for emissions estimation. The authors are not aware of any prior study that models soak-time duration from regional travel data. But, to the extent that the start mode fractions used by MOBILE5 are aggregate representations of soak-time distribution, we will review earlier studies on start mode fractions in this section.

The practice in most MPOs in the country is to accept the default start (and operating) mode fraction values developed by EPA through its Federal Test Procedure (FTP). However, these default values were developed over twenty years ago and recent research [2] suggests that it may no longer adequately represent overall vehicle emission control performance under current driving conditions. In fact, as a general recommendation, EPA suggests the use of locally estimated values of traffic and other inputs whenever possible.

Few studies have attempted to develop locally estimated start mode fractions of trips. Brodtmen and Fuce [3] used field data obtained by direct on-road measurement of engine conditions to develop start mode fractions in New Jersey. Ellis *et al.* [4] analyzed origin-destination data from travel surveys in Alabama to develop aggregate measures of start modes. In another study, Garmen Associates obtained the percentage of cold-starts in North and South New Jersey using a combination of the MOBILE default values and a few local test runs. These methods, however, have the following limitations: a) they make *ad hoc* and strong assumptions regarding start modes by trip purpose (for example, the study by Garmen Associates assumes that all work trips are cold-starts); b) they involve substantial and expensive field data collection; and c) they compute a single set of values to be applied throughout a state or spatially aggregate regions within a state.

More recently, Venigalla *et al.* [5] used data from the National Personal Transportation Survey (NPTS) to model cold and hot-start proportions as a function of trip purpose and time-of-day. The study indicated that trip purpose and time-of-day are the most important variables affecting the variation in start mode fractions. Venigalla *et al.*'s study is an important contribution to estimating start mode proportions, and highlights the value of using travel survey data to determine start mode fractions.

In this study, as in the report by Venigalla *et al.*, we use household travel survey data to determine soak-time durations. However, there are important differences between our approach and that of Venigalla *et al.* First, our focus is on the disaggregate soak-time distributions, while that of Venigalla *et al.* was on the more aggregate start mode fractions. The approach we develop can be used to provide the soak-time distributions needed by MOBILE6 or can be aggregated to provide the start fractions needed by MOBILE5. Second, we examine soak-time durations as a function of zonal land-use and socio-demographic characteristics in addition to time-of-day and purpose. Our analysis, therefore, allows the development of distinct soak-time duration distributions for each zone in a metropolitan area. Venigalla *et al.'s* analysis is at a more geographically aggregate level, allowing variations across metropolitan areas in the country based on urban area size, but not allowing local variations within an urban area. Third, we explicitly distinguish soak-time distributions for first starts of the day and non-first starts of the day, since the distributions for these two types of starts are likely to be very different. Fourth, our purpose taxonomy is based on the activity pursued prior to the trip start rather than the trip purpose taxonomy used by Venigalla *et al.* To highlight this difference, consider a set of trips from home to shop and the corresponding reverse set of trips from shop to home. In the approach used by Venigalla *et al.,* both trips would be classified as home-based other and would be assigned the same start mode fractions. In our approach, we allow the possibility that the soak-time for the trip starts from home to shop are higher than for the trip starts from shop to home, since the soak-duration is likely to be a function of the activity purpose being pursued during that soak-time (in our empirical analysis, we also examined the effect of destination activity purpose of the trip on soak-time duration, but did not find any significant impact of this variable). Finally, our analysis allows us to determine separate soak-time distributions for interzonal trips and intrazonal trips. Assuming that most intrazonal trips are on local streets, the latter distribution provides the zone-specific soak-time distributions for use with local road vehicle miles of travel (VMT) estimates.

3. MODEL FRAMEWORK

The modeling approach in the report uses vehicle trip data from household travel surveys and zonal demographic/land-use data from supplementary data sources. Three steps are involved in model estimation. The first step models whether a vehicle trip start is the first of the day or not, using a discrete binary logit model. The second step analyzes the soak-time distribution for the set of first vehicle trip starts of the day, using a log-linear regression model. The third step models the soak-time distribution for the set of non-first trip starts of the day, also using a log-linear regression model. The use of a log-linear form for soak-time guarantees the non-negativity of soak-time in application of the models.

The application step of the model predicts the soak-time distributions of trip starts for each traffic analysis zone in a metropolitan region, and for each combination of time-of-day and activity purpose. An important characteristic of the proposed method is the ease with which the estimated models from vehicle trip data can be immediately applied to obtain zonal-level soak-time distributions.

In the next section, we present the details of model estimation. In the subsequent section, we discuss model application to obtain zone-specific soak-time distributions for each combination of time-of-day and activity purpose.

3.1. Model estimation

Let q be the index for vehicle trip start, t be the index for time-of-day, and i be the index for activity purpose prior to the trip. Let G be an indicator variable taking the value 1 if a trip start is a first start and 0 otherwise. Define ω_{qti} to be a dummy variable taking the value 1 if trip start q occurs in time-period t and is preceded by activity purpose i, and 0 otherwise; define δ_{qz} as another dummy variable taking the value 1 if trip start q occurs in zone z, and 0 otherwise. Let x_z be a vector of zonal attributes.

We use a binary logit formulation to model whether or not a trip start is the first start of the day. The probability that trip start q is the first of the day is written as:

$$P_q(G=1)=\frac{1}{1+e^{-\left[\left(\sum_{t,i}\gamma_{ti}\omega_{qti}\right)+\beta'\left(\sum_z \delta_{qz}x_z\right)\right]}} \qquad \text{(Eq 1)}$$

where γ_{ti} ($t=1,2,\ldots T$; $i=1,2,\ldots I$) are scalars, and β is a vector of parameters indicating the effect of zonal attributes on the probability of a first start.

Next, we assume the soak-time to be log-normally distributed in the population of trip starts, and develop separate linear regression models for first starts and non-first starts. Let a_q^F be the soak-time for trip start q if it is a first start, and let a_q^{NF} be the soak-time for trip start q if it is a non-first start. Then, we write the log-linear regression equations for first starts and non-first starts as follows:

$$\begin{aligned}
\ln\left(a_q^F\right) &= \eta^F + \sum_{t,i}\alpha_{ti}^F\omega_{qti} + \lambda^F\left(\sum_z \delta_{qz}x_z\right) + \varepsilon_q^F, \varepsilon_q^F \sim N\left[0,\left(\sigma^F\right)^2\right] \\
\ln\left(a_q^{NF}\right) &= \eta^{NF} + \sum_{t,i}\alpha_{ti}^{NF}\omega_{qti} + \lambda^{NF}\left(\sum_z \delta_{qz}x_z\right) + \varepsilon_q^{NF}, \varepsilon_q^{NF} \sim N\left[0,\left(\sigma^{NF}\right)^2\right]
\end{aligned} \qquad \text{(Eq 2)}$$

where α_{ti} ($t=1, 2,\ldots T$; $i=1,2,\ldots I$) are scalars to be estimated, λ^F and λ^{NF} are vectors of parameters also to be estimated, and ε_q^F and ε_q^{NF} are normally distributed random error terms introduced to complete the econometric specification.

In the equation structures of 1 and 2 above, we have not allowed interactions between zonal attributes and time-of-day/activity purpose combinations; however, this is purely for notational convenience and for ease in presentation of the model application step. Such interactions can be included within the model structure without any additional conceptual or estimation complexity. Similarly, the notation structure implies full interactions of time and activity purpose, though more restrictive structures such as single dimensional effects without interaction can be imposed by appropriately constraining the γ_{ti}, α_{ti}^F and α_{ti}^{NF} scalars across the different time/activity purpose combinations. Finally, we also include an intrazonal dummy variable, and interactions of this variable with time-of-day/origin purpose, in our empirical specification. This allows us to accommodate separate soak-time distributions for intrazonal vehicle trips and interzonal

vehicle trips. Again, for simplicity in presentation, we suppress these additional intrazonal variables in the presentation of model structure.

3.2. Model application

This section discusses the application of the estimated models in the previous section to obtain the soak-time distribution for use in MOBILE6 and to obtain the start mode fraction of trips for use in MOBILE5.

3.2.1. Soak-time distribution for MOBILE6

Once the model parameters in Equations 1 and 2 are estimated from disaggregate vehicle trip start data, the soak-time distribution for any zone in the study area by time-period and activity purpose can be determined in a rather straightforward manner. To see this, first consider the binary logit model for first starts. The probability that any trip start q in time-period t preceded by an activity of purpose i, and occurring in zone z, is a first start can be written as:

$$P_{tiz}(G=1)=\frac{1}{1+e^{-\left[\gamma_{ti}+\beta' x_z\right]}} \qquad \text{(Eq 3)}$$

Let M_{tiz} be the total number of trip starts in time-period t preceded by an activity of purpose i, and occurring in zone z. Then, the fraction of first trip starts in time t with purpose i in zone z can be written as:

$$\delta_{tiz}=\frac{M_{tiz} * P_{tiz}(G=1)}{M_{tiz}}=P_{tiz}(G=1) \qquad \text{(Eq 4)}$$

Hence, once the γ_{ti} and β parameters are estimated, the fraction of first starts in time t with purpose i in zone z can be computed using the expression in Equation 3.

Next, the (log) soak-distribution of trip starts in time t with origin activity purpose i in zone z for first starts and non-first starts may be written as:

$$\ln\left(a_{tiz}^{F}\right)\sim N\left[\eta^{F}+\alpha_{ti}^{F}+\lambda^{F}x_z,\left(\sigma^{F}\right)^2\right]=N\left[\Delta_{tiz}^{F},\left(\sigma^{F}\right)^2\right]$$

$$\ln\left(a_{tiz}^{NF}\right)\sim N\left[\eta^{NF}+\alpha_{ti}^{NF}+\lambda^{NF}x_z,\left(\sigma^{NF}\right)^2\right]=N\left[\Delta_{tiz}^{NF},\left(\sigma^{NF}\right)^2\right] \qquad \text{(Eq 5)}$$

The means (Δ_{tiz}^{F} and Δ_{tiz}^{NF}) and variance ($(\sigma^{F})^{2}$ and $(\sigma^{NF})^{2}$) of these distributions can be estimated from the parameter estimates obtained in the estimation stage.

The objective in our effort is to obtain the fraction of soaks in each of 70 time-bins (as needed by MOBILE6) across both first- and non-first starts for each zone, and for each activity purpose and time-of-day combination. Let k be an index for time-bin ($k=1,2,\ldots,70$), and let time-bin k be bounded by the continuous soak-time value of m_{k-1} to the left and by m_k to the right. Then, the fraction of soaks in time-bin k for first starts may be written as:

$$fraction\left(m^{k-1} < a_{tiz}^{F} < m^{k}\right) = \Phi\left[\frac{\log\left(m^{k}\right) - \Delta_{tiz}^{F}}{\sigma^{F}}\right] - \Phi\left[\frac{\log\left(m^{k-1}\right) - \Delta_{tiz}^{F}}{\sigma^{F}}\right] \qquad \text{(Eq 6)}$$

The corresponding expression for non-first starts is:

$$fraction\left(m^{k-1} < a_{tiz}^{NF} < m^{k}\right) = \Phi\left[\frac{\log\left(m^{k}\right) - \Delta_{tiz}^{NF}}{\sigma^{NF}}\right] - \Phi\left[\frac{\log\left(m^{k-1}\right) - \Delta_{tiz}^{NF}}{\sigma^{NF}}\right] \qquad \text{(Eq 7)}$$

Finally, the fraction of soaks in time-bin k across first and non-first starts for zone z, time-of-day t and activity purpose i may be computed as follows:

$$fraction\left(m^{k-1} < a_{tiz} < m^{k}\right) = fraction\left(m^{k-1} < a_{tiz}^{F} < m^{k}\right) * \delta_{tiz} + fraction\left(m^{k-1} < a_{tiz}^{NF} < m^{k}\right) * \left(1 - \delta_{tiz}\right) \qquad \text{(Eq 8)}$$

3.2.2. Start mode fractions for MOBILE5

MOBILE5 requires the mode fractions in cold-starts and hot-starts to compute the operating mode fractions. As indicated in Section 2.1, hot-starts are defined as those with a soak-time of 1 hour or less for vehicles with catalytic converters and with a soak-time of less than 4 hours for vehicles without catalytic converters. If a start is not a hot-start, it is a cold-start. Thus, the fraction of hot-starts among the population of trip starts made by vehicles with catalytic converters for zone z, time-of-day t and original activity i is as follows:

$$fraction(hot\text{-}starts) = \Phi\left[\frac{\log(60) - \Delta_{tiz}^{F}}{\sigma^{F}}\right] * \delta_{tiz} + \Phi\left[\frac{\log(60) - \Delta_{tiz}^{NF}}{\sigma^{NF}}\right] * (1 - \delta_{tiz}) \qquad \text{(Eq 9)}$$

The fraction of cold-starts among the population of trip starts made by vehicles with catalytic converters for zone z, time-of-day t and original activity i is trivially obtained as *[1- fraction(hot-starts)]*.

The expression for the fraction of hot-starts among the population of trip starts made by vehicles not equipped with catalytic converters is the same as Equation 9 with log(240) substituted for log(60).

4. DATA PREPARATION

4.1. Data sources

The data used in the empirical analysis are drawn from two sources: the 1996 Activity Survey conducted in the Dallas-Fort Worth (D-FW) area and the zonal land use and demographics characteristics file for the D-FW area. These data sources were obtained from the North Central Texas Council of Governments (NCTCOG), and are briefly discussed next.

The 1996 activity survey collected information on activities undertaken during a weekday by members of 4,839 households. For non-travel activities, information on the activity type, start and end times of participation, and location was collected (the location of each activity was geo-coded to a traffic analysis process, or TAP, zone; there are 919 TAP zones in the D-FW planning area). For travel activities, information on the mode of travel used, costs incurred, and trip duration was collected. In addition, the survey elicited individual and household socio-demographic information.

The zonal level land use and demographics characteristics file contained land use and demographic data at the level of the traffic survey zone (TSZ) within the D-FW metropolitan planning area. The land use information for each TSZ provides information on total land area and acreage in several individual land use purposes (for example, in manufacturing; in retail, hotel and motel; in institutional buildings such as churches, government, museums, schools, and hospitals; in multifamily households; and in airport

runways/terminals). The demographic information for each TSZ includes population, number of households, population density, median income, average household size, *etc.*

4.2. Sample formation

Several data assembly steps were involved in developing the sample. First, we converted the raw composite (travel and non-travel) activity file into a corresponding person-trip file. Second, we identified person-trips that were pursued using a motorized vehicle owned by the household. Third, we translated the person-trip file into a corresponding vehicle trip file, which provided the sequence of trips made by each vehicle in the household. In this process, we extracted and retained information on the time-of-day of each vehicle trip start, TAP zone of trip start location and trip end location, purpose of activity being pursued during soak-time, and soak-time prior to vehicle trip start. The first trip start in the day for each vehicle was also identified and flagged. To compute the soak-times for these first trip starts, we assume that the soak-time prior to the first trip start is invariant across days for each vehicle (this assumption is necessary because only a single day of diary data is available). The soak-time for the first trip starts can then be computed as the difference in time between the first trip start of the day and the last trip end of the diary day. Fourth, we aggregated the TSZ-level land-use and demographic characteristics to the TAP-level, and appended this information to each vehicle trip start based on the TAP in which the trip start occurs. Finally, we conducted several screening and consistency checks on the resulting data set from the previous steps (a flow chart of this screening process is available from the authors). As part of this screening process, we eliminated observations that had missing data on departure times, activity purposes, and/or on the TAP location of the vehicle trip start.

The final sample used for analysis includes 18,231 vehicle trip start observations. Of these, 4,246 (23.3%) were first starts, and 13,985 (76.7%) were non-first starts.

5. EMPIRICAL ANALYSIS

5.1. Sample description

The dependent variable of interest in model estimation is the soak-time duration distribution of trip starts. The soak-time duration for first starts varies from a minimum of

71 minutes to a maximum of 1,339 minutes (around 22 hours). The mean soak-time for first starts is 835 minutes (about 14 hours) with a standard deviation of about 204 minutes. The soak-time duration for non-first starts varies from a minimum of 1 minute to a maximum of 1,345 minutes (around 22 hours). The mean soak-duration for non-first starts is 161 minutes (about 2.5 hours) with a standard deviation of 196 minutes.

Three types of variables were considered to explain soak-time for first starts and non-first starts. These are: a) time-of-day variables identifying the time of trip start, b) activity purpose variables indicating type of activity pursued prior to the trip start, and c) zonal and trip attributes. Interactions among these three sets of variables were also considered. In the description below, we briefly highlight some of the characteristics of these sets of variables.

The time-of-day of trip start was associated with one of the following six time-periods: morning (midnight-6:30 a.m.), a.m. peak (6:30 a.m.-9:00 a.m.), a.m. off-peak (9:00 a.m.-noon), p.m. off-peak (noon-4:00 p.m.), p.m. peak (4:00 p.m.-6:30 p.m.), and evening (6:30 p.m.-midnight). The time-periods for the a.m. and p.m. peaks were based on the peak period definitions employed by the transportation department of the NCTCOG in the D-FW area. The times for the off-peak periods were determined by splitting the remaining blocks of time at noon and midnight. The distribution of first starts and non-first starts by time-of-day is presented in Table 1. As can be observed, a substantial fraction of first starts occur in the a.m. peak period, reflecting the morning commute trip. Only a small fraction of first starts occur outside the a.m. periods. On the other hand, most of the non-first starts occur in the p.m. periods, reflecting a combination of return-home trips from work and trips to/from other non-work activities.

The distribution of first starts and non-first starts by activity purpose prior to the trip start is provided in Table 2. About 96% of the first starts begin from home (analysis of the destination activity of these "first start" trips indicates that about 60% are destined to work or school, while the remainder are quite evenly distributed across other activity purposes). For non-first starts, the activity purpose prior to the trip start is much more evenly distributed, though close to half of all starts begin at work or home (about 47% of these "non-first start" trips are destined to home, while the remainder are rather evenly distributed across the other purpose categories).

Several zonal (TAP-level) land-use and demographic characteristics were considered in our analysis. Of these, the following zonal attributes were significant determinants of the trip start type (first versus non-first) and/or soak-time duration: acreage in multifamily households, retail employment and service employment, number of households, and population density. The trip-related attribute included in the model was an indicator variable for whether or not the trip corresponding to the vehicle start was an intrazonal trip. Including this intrazonal trip indicator enables the distinction of soak-time duration for intrazonal (local) trip starts and interzonal trip starts. Of the 18,231 trips in the sample, 2,676 (14.7 %) are intrazonal.

5.2. Empirical results

This section presents the empirical results for the estimated models. Section 5.2.1 discusses the estimation results for trip start type (first versus non-first starts). Section 5.2.2 presents the results for soak-time duration of first starts, and Section 5.2.3 presents the corresponding results for non-first starts.

5.2.1. Results for first starts vs. non-first starts

The binary logit model results for first starts vs. non-first starts are provided in Table 3. The base category used is non-first starts. Thus, a positive coefficient on a variable indicates that the variable increases the probability of a first start, while a negative coefficient implies that the variable decreases the probability of a first start. The constant in the model does not have any behavioral interpretation; it adjusts for the range of zonal attribute values in the sample and the sample shares of first starts.

The time-of-day variables are introduced into the model with the morning period as the base time-period. The signs on the estimated coefficients for all other time-periods are negative and increasing in magnitude from the a.m. peak to the evening. This implies that trip starts that occur earlier in the day are more likely to be first starts than those made later in the day.

The activity purpose variables are introduced using the "home" purpose as the base category. The purpose dummy variables for all other activities are negative,

indicating that, everything else being equal, trips from home are most likely to be first starts. A comparison of the magnitudes of coefficients across the activity purpose categories provides additional information regarding the likelihood of first starts among the group of non-home trip starts. Specifically, starts from work are more likely to be first starts than from other non-home purposes; starts after school and social-recreational activities are more likely to be first starts than those after shopping, personal business, and other (for example, drop-off/pick-up) activities.

Among the zonal and trip attributes, the population of the zone of trip start, the number of households, and an indicator of whether the trip was an intrazonal one have significant impacts on the likelihood of the trip start being the first of the day. The effect of zonal population may reflect more opportunities for participation in activities (such as shopping, social-recreational, *etc.*) in highly populated areas, which would result in more trips made per vehicle. As the number of trips per vehicle increases, the fraction of first trip starts has to decline, which would explain the negative effect of zonal population on the likelihood of first starts. On the other hand, for a given zonal population size, a higher number of households would imply more spreading of trips across households. This can lead to a decrease in number of trips per vehicle, resulting in a greater likelihood of first starts. Finally, intrazonal trips are short-distance trips, and stops pursued in such trips may be more likely to be linked with (and pursued after) stops to more distant locations (such as a shopping stop near home on the way back from work). Consequently, starts for intrazonal trips are less likely to be first starts in the day.

The log-likelihood values at convergence and with only the constant are provided toward the bottom of Table 3. The hypothesis that the time-of-day, activity purpose, and zonal/trip variables have no impact on the probability of first starts is strongly rejected by a log-likelihood ratio test. The pseudo-R^2 value for the model is 0.71.

5.2.2. Results for soak-time duration model for first starts

The results for the soak-time duration model for first starts are given in Table 4. The dependent variable in the model is the logarithm of soak-duration.

The time-of-day variables are introduced with the morning period being the base. The p.m. peak and evening periods are combined into a single period because of very few

first starts in these periods (see Table 1). The sign and magnitudes of the time-of-day variables in Table 4 indicate that the soak-time for first starts occurring later in the day is higher than for those occurring earlier in the day.

The activity purpose categories are collapsed into three categories for modeling soak-time duration for first starts: home, work, and non-home/non-work. This aggregation is necessary because very few first starts are preceded by a non-home or non-work activity. Dummy variables for work and non-home/non-work purposes are included with home being the base activity. The sign on the work purpose variable in Table 4 suggests that first starts from work have a smaller soak-duration than those from home. This is quite intuitive, since the first starts from work presumably represent the early morning return-home trips of individuals who came in to work late the previous night. We did not find any generic impact of the "non-home/non-work" purpose variable across time periods; however, the interaction effect of this variable with the morning period is statistically significant. The sign on this interaction suggests that the soak-time duration for first trip starts in the early morning period after non-work and non-home activities (*i.e.,* after social-recreation, shopping, *etc.*) is smaller than for trip starts in the early morning period from home or work.

Among the zonal and trip attributes, acreage in multifamily households is a significant determinant of soak-time duration. The sign on this variable suggests a smaller soak-duration for first starts occurring in zones with large acreage in multifamily households. This may be a consequence of return-home trips in the early morning after a social event in the neighborhood. The intrazonal dummy variable and its interaction with the "home" purpose indicate that the soak-duration of first starts associated with intrazonal trips is smaller than the soak-duration of first starts associated with interzonal trips for trips not originating at home. However, the reverse relationship holds for trips originating at home. The last few rows of the table provide summary fit statistics. The R^2 value for the model is 0.34.

5.2.3. Results for soak-time duration model for non-first starts

The results for the soak-time duration model for non-first starts are presented in Table 5. The dependent variable in the model is the logarithm of soak-duration.

The time-of-day variables are introduced with the morning and a.m. peak period being the base. The morning and a.m. peak periods are combined into a single period because of very few non-first starts in these periods (see Table 1). The results indicate that the soak-time for non-first starts occurring later in the day is higher than for those occurring earlier in the day.

The activity purposes are introduced with home as the base activity. Interaction effects of activity purpose with time-of-day are also introduced. For these interaction effects, we use aggregate classifications for time-of-day and activity purpose, based on extensive empirical testing. Specifically, we classify non-home purposes into two broad categories: work/school and non-work/non-school (the home purpose is the base). We also use only two broad time periods: a.m. off-peak and p.m. (the morning/a.m. peak period is the base).

Referring to Table 5, the main effect for the work/school purposes, considered with the interaction effect for work/school purposes with time-of-day, indicates the following: a) For the morning/a.m. peak periods, non-first starts from work/school have a smaller soak-duration than for non-first starts from home, and b) for other time periods, non-first starts from work/school have a larger duration than non-first starts from home. These results are intuitive, since individuals are likely to spend longer durations at work/school than at home during the day. The main effects for all other non-work and non-school purposes indicate that non-first starts from those purposes are associated with a smaller soak-duration than non-first starts from home, school and work. However, this is less so as the day progresses, presumably because individuals have more time to pursue shopping, personal business, social-recreational and other activities in the latter part of the day.

Among the zonal and trip attributes, the number of people in retail and service employment is a significant determinant of soak-time duration. The sign on this variable suggests a larger soak-duration for non-first starts occurring in zones with large number of people in retail and service employment. The number of people in service and retail employment here may be viewed as indicator variables for the "size" of shopping and service-related opportunities in a zone; bigger the "size", larger is likely to be activity duration of participation and therefore, the soak-duration. The intrazonal dummy variable

and its interaction with the "home" purpose suggest that the soak-duration of non-first starts is larger for intrazonal trips originating at home, but lesser when originating at non-home activities.

The last few rows of the table provide summary fit statistics. The R^2 value for the model is 0.55.

6. APPLICATION CONSIDERATIONS

The models estimated above can be applied as discussed in Section 5.2.3 to obtain zone-specific soak-time distributions for each combination of time-of-day and activity purpose. In this section, we discuss how these soak-time distributions can be used in combination with travel demand models.

Travel demand modeling may be based on an activity-based approach or on a trip-based approach (see Bhat and Koppelman [6] for a detailed discussion). In the activity-based modeling approach, the emphasis is on activities, and trips are considered as the derivative of the need to participate in activities at different locations. The activity-based approach treats time as an all-encompassing continuous entity within which individuals make activity/travel participation decisions (see Kurani and Lee-Gosselin [7]). Thus, if an activity-based approach is used in travel demand modeling, the stops (and, therefore, vehicle trip starts) by purpose type, time-of-day, and zone of origin are modeled explicitly. This information can be immediately used with the zone-specific soak-time distributions by time-of-day and activity purpose developed in this report.

The methodology developed here can also be used with a trip-based approach. In the trip-based approach, the trip-interchanges (zone-to-zone production-attraction matrices) within each trip purpose category are determined first and are subsequently converted into a zone-to-zone origin-destination matrix by time-of-day (fixed factors are typically used in this conversion, though departure time choice models can be used to better serve this purpose; see Steed and Bhat [8]). Depending on the trip purpose classification used, the zone-to-zone origin destination matrix by time-of-day may or may not provide all the information needed for integration with the soak-time model developed in this report. Specifically, if the home-based trips are classified into home-

based work, home-based shopping, home-based school, home-based personal business, home-based social-recreational, and other home-based trips in the trip-based modeling, then information on trip starts by zone of origin, activity purpose prior to the trip start in the classification scheme used in the soak-time model, and time-of-day is available for home-based trips. In addition, if the non-home-based trips are also further sub-classified by origin activity categories in the six non-home activity typology used in the soak-time model, then information on trip starts by zone of origin, activity purpose, and time-of-day is available for all trips, and this can be used with the soak-time distribution model.

Most MPOs do not use the level of disaggregation in trip purposes discussed earlier. In fact, many continue to use only three trip purpose types: home-based work, home-based other, and non-home-based (there is, however, an increasing trend toward using more disaggregate trip purpose categories). Until MPOs use a more disaggregate trip-based classification scheme, or use an activity-based approach, an alternative is to post-classify trips into the origin activity purpose and time-of-day adopted in our soak-time model. This can be done by applying fixed factors obtained from travel surveys. In Tables 6, 7, and 8, we disaggregate home-based work, home-based other, and non-home-based trips by activity at origin end and by time-of-day for the D-FW region. These fractions can be applied to zone-to-zone production-attraction matrices in each of the three broad trip purpose categories to obtain the number of trip starts by zone of origin, activity purpose prior to trip start, and time-of-day. The implementation is particularly straightforward using a Geographic Information System (GIS) platform. This is the method that the research team is using to determine zone-specific soak-time distributions in the D-FW metropolitan planning area as part of an ongoing air quality-related project funded by TxDOT.

7. CONCLUSIONS

The temporal distribution of the engine off (soak) times of trips in a region is an important traffic related input to mobile source emissions models. The soak-time associated with a trip could depend on various factors such as the activity purpose preceding the trip start, the time-of-day of trip start and possibly other land-use and socio-demographic characteristics of the zone of trip origin. More accurate region-

specific estimates of the distribution of soak-time and the operating mode fractions can be obtained by modeling the soak-time as a function of these variables. In the current report, we estimate a model of soak-time durations as a function of these different attributes. The modeling approach in the report uses vehicle trip data from household travel surveys and zonal demographic/land-use data from supplementary data sources. Three steps are involved in model estimation. The first step models whether a vehicle trip start is the first of the day or not using a discrete binary logit model. The second step analyzes the soak-time distribution for the set of first vehicle trip starts of the day with a log-linear regression model. The third step models the soak-time distribution for the set of non-first trip starts of the day, also using a log-linear regression model. Our proposed model framework contributes toward improved mobile source emissions modeling by developing a systematic approach to analyzing soak-time durations. The effectiveness of our methodology lies in its easy application at the traffic zonal level within a metropolitan region to obtain zone-specific soak-time distributions by time-of-day and origin activity purpose. We apply the model framework to obtain zone-specific soak-time distributions for the D-FW region. However, it should be applicable to any metropolitan region after appropriate re-estimations of the trip-level soak-time models.

Notwithstanding the very general nature of our model formulation, it is important to acknowledge two limitations of the application of the formulation in the current report. First, the model does not capture seasonal variations in soak-time distribution or its variation between weekdays and weekends. This is because the travel diary survey in the D-FW area was limited to a single survey day, and hence information on trip-making behavior of individuals across different days/seasons is not available. If and when such data become available, these day-to-day and seasonal variations can be accommodated in a straightforward manner within the framework of this report. Second, since the survey focused on *household* travel, no commercial trips are included in the survey. Extending the empirical application in this report to model the soak-time durations of commercial trips is an important direction for future research. Such an extension will require collection of survey data on commercial vehicle activity.

REFERENCES

1. Brzezinski, D. J., and Newell, T. P., MOBILE6: A Revised Model for Estimation of Highway Vehicle Emissions. U.S. Environmental Protection Agency Office of Mobile Sources, Assessment and Modeling Division, 1998.

2. Federal Test Procedure Review Project: Preliminary Technical Report. Report EPA 420-R-93-007, Certification Division, Office of Mobile Sources, Office of Air and Radiation, U.S. Environmental Protection Agency, May 1993, (http://www.epa.gov/oms/regs/ld-hwy/ftp-rev/ftp-summ.txt).

3. Brodtmen, K. J., and Fuce, T. A., "Determination of Hot- and Cold-Start Percentages in New Jersey." Report FHWA/NJ-84/001, New Jersey Department of Transportation, July 1984.

4. Ellis, G. W., Camps, W. T., and Treadway, A., The Determination of Vehicular Cold and Hot Operating Mode Fractions for Estimating Highway Emissions. State of Alabama Highway Department, 1978.

5. Venigalla, M., Miller, T., and Chatterjee, A., "Start Modes of Trips for Mobile Source Emissions Forecasting." *Transportation Research Record 1472,* pp. 26-34, 1995.

6. Bhat, C. R., and Koppelman, F. S., "A Retrospective and Prospective Survey of Time-Use Research." *Transportation,* Vol. 26, No. 2, pp. 119-139, May 1999.

7. Kurani, K. S., and Lee-Gosselin, M. E. H., "Synthesis of Past Activity Analysis Applications." Activity Based Travel Forecasting Conference, New Orleans, Louisiana, 1996.

8. Steed, J., and Bhat, C. R., "On Modeling the Departure Time Choice for Home-Based Social/Recreational and Shopping Trips." *Transportation Research Record 1706,* pp. 152-159, 2000.

Table 1. Distribution of trip starts by time-of-day

Time-of-day of trip start	Percentage distribution for	
	first starts	non-first starts
morning	14.1%	0.5%
a.m. peak	64.4%	6.9%
a.m. off-peak	14.2%	12.6%
p.m. off-peak	5.7%	30.0%
p.m. peak	1.2%	28.9%
evening	0.4%	21.2%

Table 2. Distribution of trip starts by activity purpose preceding trip start

Activity purpose preceding trip start	Percentage distribution for	
	first starts	non-first starts
Home	96.4%	19.3%
Work	2.2%	29.9%
School	0.0%	3.3%
Social/Recreational	0.8%	15.4%
Shopping	0.1%	10.5%
Personal Business	0.3%	11.2%
Other	0.2%	10.4%

Table 3. Empirical results for binary logit model for first starts

Variable	*Coefficient*	*t-statistic*
Constant	5.186	24.13
Time-of-day variables (Morning period is base)		
a.m. peak	-2.465	-11.736
a.m. off-peak	-4.251	-19.96
p.m. off-peak	-5.932	-27.55
p.m. peak	-7.069	-30.67
evening	-7.780	-30.61
Activity purpose prior to trip start ("Home" purpose is base)		
Work	-3.677	-30.19
School	-4.807	-8.99
Social/Recreational	-4.734	-21.48
Shopping	-5.728	-13.87
Personal Business	-5.402	-20.35
Other	-6.874	-23.20
Zonal and trip attributes		
Population x 10^{-5}	-6.949	-3.68
Number of households x 10^{-4}	1.709	3.58
Intrazonal trip	-0.496	-5.60
Number of Observations	18,231	
Log-Likelihood Function	-3196.08	
Log-Likelihood for Constants only	-10977.08	

Table 4. Empirical results for soak-time duration model for first starts

Variable	*Coefficient*	*t-statistic*
Constant	2.827	689.71
Time-of-day variables (Morning period is base)		
a.m. peak	0.069	15.71
a.m. off-peak	0.163	29.11
p.m. off-peak	0.225	30.63
p.m. peak/evening	0.270	22.06
Activity purpose (Home is base) **and associated interactions**		
Work	-0.091	-8.80
Morning x "non-home/non-work" activity purpose	-0.481	-10.03
Zonal and trip attributes and associated interactions		
Acreage in multi-family households x 10^{-4}	-4.078	-3.64
Intrazonal trip	-0.288	-9.84
Intrazonal x Home origin	0.307	10.38
Number of observations	4,246	
Regression sums of squares	20.24	
Residual sums of squares	38.67	
R^2	0.344	
Adjusted R^2	0.342	

Table 5. Empirical results for soak-time duration model for non-first starts

Variable	*Coefficient*	*t-statistic*
Constant	1.667	45.88
Time-of-day variables (morning-a.m. peak period is base)		
a.m. off-peak	0.086	1.82
p.m. off-peak	0.171	4.51
p.m. peak	0.258	6.79
evening	0.291	7.64
Activity purpose prior to trip start ("Home" purpose is base)		
Work/School	-0.220	-3.42
Social/Recreational	-0.514	-11.96
Shopping	-0.777	-17.81
Personal Business	-0.974	-22.54
Other	-1.214	-29.63
Time-of-day and activity purpose interaction effects		
a.m. off-peak		
Work/School	0.402	5.45
Social-Recreational/Shopping/Personal Business/Other	0.311	5.75
p.m.		
Work/School	0.713	10.95
Social-Recreational/Shopping/Personal Business/Other	0.283	6.59
Zonal and trip attributes		
Number of people in retail and service employment x 10^{-5}	1.004	7.26
Intrazonal trip	-0.121	-8.85
Intrazonal "home" trip	0.082	3.01
Number of observations	13,985	
Regression sums of squares	4279.89	
Residual sums of squares	3452.17	
R^2	0.55	
Adjusted R^2	0.55	

Table 6. Cross-classification of home-based work trips by origin-end activity and time-of-day

Trip purpose	Time-of-day of trip start					
	morning	a.m. peak	a.m. off-peak	p.m. off-peak	p.m. peak	evening
Home	7.82%	33.13%	4.64%	5.09%	1.35%	1.22%
Work	0.94%	0.79%	1.67%	8.79%	26.08%	8.49%

Table 7. Cross-classification of home-based other trips by origin-end activity and time-of-day

Trip purpose	Time-of-day of trip start					
	morning	a.m. peak	a.m. off-peak	p.m. off-peak	p.m. peak	evening
Home	1.36%	13.91%	7.66%	9.77%	8.69%	8.30%
School	0.01%	0.03%	0.38%	2.12%	0.84%	0.81%
Social/Recreational	0.30%	0.29%	1.05%	2.62%	2.09%	8.93%
Shopping	0.06%	0.24%	1.51%	3.29%	3.53%	3.28%
Personal Business	0.13%	0.18%	1.75%	2.72%	2.85%	1.73%
Other	0.10%	2.04%	0.45%	2.51%	2.92%	1.55%

Table 8. Cross-classification of non-home-based trips by origin-end activity and time-of-day

Trip purpose	Time-of-day of trip start					
	morning	a.m. peak	a.m. off-peak	p.m. off-peak	p.m. peak	evening
Work	0.02%	1.39%	8.63%	13.42%	10.43%	2.06%
School	0.00%	0.16%	0.59%	1.50%	0.22%	0.28%
Social/Recreational	0.24%	0.99%	2.91%	9.46%	1.80%	4.22%
Shopping	0.04%	0.46%	1.94%	4.36%	2.16%	1.58%
Personal Business	0.14%	1.68%	4.97%	6.51%	2.69%	1.07%
Other	0.34%	6.67%	1.31%	2.55%	2.16%	1.05%