

Mode Choice Methodology in TRANSIMS

Qingying Lu

Thesis submitted to the Faculty of Virginia Polytechnic Institute and State University
in partial fulfilment of the requirements for the degree of

Master of Science
in
Civil and Environmental Engineering

Dr. Antoine Hobeika, Chair
Dr. Antonio Trani
Dr. Hojong Baik

September 2002
Blacksburg, Virginia

Keywords: Transportation planning, TRANSIMS, Mode choice, Disaggregate,
Behavioral

Mode Choice Methodology in TRANSIMS

Qingying Lu

Abstract

TRANSIMS is a disaggregate, behavioral transportation planning package developed under US DOT's and EPA funding at the Los Alamos National Laboratory (LANL). It is an integrated system of travel forecasting models designed to give transportation planners accurate, complete information on traffic impacts, congestion, and pollution by simulating second-by-second movements of every person and every vehicle through the transportation network of a large metropolitan area. There is no built-in module for travellers' mode choices In TRANSIMS. The modes going with the shortest path are always taken. In Portland Study, a mode choice methodology implemented by a series of feedback processes is proposed. However, it uses aggregate, deterministic mode choice model. There is little solid theoretic ground for the format and coefficients of the generalized costs used in the calibration process. The accessibility to a mode, especially to Transit, was also not included in the model. In the thesis, a disaggregate and deterministic mode choice methodology in TRANSIMS is developed. The accessibility to each mode is analyzed and included in the model. The methodology is then implemented on the Blacksburg transportation planning study, namely Blacksburg_Lite. The analysis of the result is based on the indicator of mode choice, mode split between Transit and Auto. The indicator is close to that in survey data and converged fast. Therefore, this mode choice methodology could be used within TRANSIMS framework.

Acknowledgements

Primarily, I want to express my appreciation to my advisor, Dr. Antoine Hobeika for his guidance and encouragement throughout the period of my graduate study at Virginia Tech. Appreciation is also owed to Dr. Antonio Trani and Dr. Hojong Baik, who have served on the review committee of the thesis.

Additionally, I am appreciative of the instructors that I have had here at Virginia Tech, namely Dr. Antoine Hobeika, Dr. Wei Lin, Dr. Antonio Trani, Dr. Sam Tignor, Dr. Gerardo Flintsch, and Dr. Randy Dymond.

The thesis is based on the datasets of Blacksburg_Lite study implemented by Virginia Tech TRANSIMS group. Therefore, I would like to thank other members in TRANSIMS group: Dr. Hobeika, Chaw, Yahong, Mansourah, Ken and Kevin.

I also want to express my gratitude to my seniors, Hojong, Jamal, Yihua, Srinivas and Rajan for their elderly advice. I thank my colleagues, Senanu, Debayan, Nanditha, Sudheer, Charla and Sohn and my juniors, Wang, Bin, Yingjian, Anand, Christnan and Mansourah for their friendships and supports.

Finally and most of all I would like to thank my husband, my parents and my brother for their deep love and continuous encouragement.

Table of Contents

Abstract	ii
Acknowledgement	iii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
1. Introduction	1
1.1 Background	1
1.2 Problem Definition	2
1.3 Thesis Objectives	2
1.4 Organization of Thesis	3
2. Literature Review	4
2.1 Transportation Planning Models	4
2.1.1 Trend Analysis	4
2.1.2 Mathematical Models	4
2.1.3 Disaggregate Models	5
2.2 Mode Choice Models	6
2.2.1 History of Mode Choice Models	6
2.2.2 Comparison between Disaggregate and Aggregate Mode Choice Models	8
3. TRANSIMS Review	12
3.1 Introduction	12
3.1.1 TRANSIMS Framework	12
3.1.2 Population Synthesizer	14
3.1.3 Activity Generator	15
3.1.4 Route Planner	17
3.1.5 Traffic Microsimulator	17
3.1.6 Output Visualizer	18

3.1.7 Emission Estimator_____	19
3.1.8 Feedback Module_____	20
3.2 Mode Choice Methodology in Portland Study_____	23
3.2.1 Feedback Processes in Portland Study_____	23
3.2.2 Mode Choice Methodology in Portland Study_____	25
3.3 The Need for a New Mode Choice Methodology in TRANSIMS_____	27
4. Theoretical Aspects of the Proposed Methodology_____	28
4.1 Introduction_____	28
4.2 Framework of Proposed Mode Choice Model_____	30
4.3 Calibration of Tour Mode Choice on Survey data_____	32
4.3.1 Introduction of Binomial Logit Model_____	32
4.3.2 Introduction of Maximum Likelihood Estimation_____	33
4.3.3 Computational Aspects of Maximum Likelihood Estimation_____	33
4.3.4 Maximum Likelihood Estimation Binomial Logit Model_____	34
4.4 Tour Mode Choice for Synthetic Data in TRANSIMS_____	36
5. Application and Evaluation of the Methodology_____	38
5.1 Application of the Methodology in Blacksburg Study_____	38
5.1.1 Background of the Research_____	38
5.1.2 Specification of Tour Mode Choice Model_____	39
5.1.3 Calibration of Tour Mode Choice Model on Survey Data_____	41
5.1.4 Implementation of Calibrated Model in Blacksburg Study_____	43
5.2 Evaluation of the Methodology in Blacksburg Study_____	48
6. Conclusions_____	51
6.1 Conclusions_____	51
6.2 Recommendations for Further Study_____	51
References_____	53
Vita_____	56

List of Tables

Table 4.1 Trip modes allowed by tour mode _____	28
Table 5.1 Specification of the Utility Function _____	40
Table 5.2 Survey Data _____	41
Table 5.3 Estimated Coefficients _____	42
Table 5.4 Verification of the calibrated model on survey data _____	42
Table 5.5 Activities in a typical day in Blacksburg _____	49
Table 5.6 Breakdown between “ Work” and “Shop & Other” within Alone activities _____	49
Table 5.7 Results of Mode Split between Transit and Auto for Alone “Work” activities _____	49

List of Figures

Figure 3.1 TRANSIMS Framework_____	13
Figure 3.2 STF-3A data for a Block Group_____	14
Figure 3.3 PUMS data for a Census Tract_____	14
Figure 3.4: An example of the classification and regression tree generated by CART algorithm_____	16
Figure 3.5 The Output Visualizer Displaying Vehicles On The Network In A Zoomed-in View_____	19
Figure 3.6 Flowchart of module interfaces for the Selector Module_____	21
Figure 3.7 Relationship between the three model tools_____	22
Figure 3.8 Feedback processes in Portland Study (Portland Study Report)_____	24
Figure 4.1 Tour and trip mode choice model nesting structure_____	29
Figure 4.2 Framework of the proposed mode choice model_____	31
Figure 5.1 Simplified transportation network in Blacksburg-Lite Study_____	38
Figure 5.2 Location of the town of Blacksburg_____	39
Figure 5.3 Group activities by shared rides_____	44
Figure 5.4 Generate tour database with tour type_____	44
Figure 5.5 Convert all activities to general transit mode and get travel plans_____	45
Figure 5.6 Divide them by completeness and collect travel time for complete tours_____	46
Figure 5.7 Collect in-vehicle time and out-of-vehicle time for Auto tours_____	46
Figure 5.8 Obtain household income per person_____	47
Figure 5.9 Generate feedback file for complete tours_____	47
Figure 5.10 Generate feedback file for incomplete tours_____	47
Figure 5.11 Generate new activity file with chosen tour modes_____	47
Figure 5.12 Positions to withdraw results for iterations_____	50

Chapter 1 Introduction

1.1 Background

Mobility has become an integral part of agriculture, manufacturing, education, the life of a city, access to sports and recreation, and the conduct of international trade, travel and investment. The more complex life becomes, the more indispensable are the elements of the transportation systems.

The accomplishments of transportation are notable, but there is growing concern over congestion, pollution, physical deterioration, deteriorating services and resource constraints. How can the transport-dependent society cope with the growing demand within the resource constraints? These are momentous issues to be addressed by the planning of transportation facilities.

Transportation planning has gain people's attention for a long time and it has become a hot research topic as early as 1950s. In the 1960s and 1970s, the standard four-step procedure was the accepted method for carrying out planning studies. It is one of aggregate transportation planning models, which have been used widely and gained a great amount of accomplishments. These models at base attempt to represent the behavior of a group of travelers instead of a single individual.

The aggregate transportation planning models have been severely criticized for their inflexibility and inaccuracy. Disaggregate models appeared in 1980s offer substantial advantages over the aggregate counterparts. Disaggregate models at base represent the behavior of individuals. The disaggregate models are clearly superior to the grouping of behavior zonally and by predefined segments in aggregate models.

However, disaggregate transportation planning model being used for a large metropolitan area has just been on the research screen, not in the real world because it would involve a huge amount of data, which were neither available, nor manageable by computers at that time period. Benefiting from the boom of computing capability of modern computers and more and more detailed census data, disaggregate transportation planning models have become possible to be used in the real world.

TRANSIMS is a disaggregate, behavioral transportation planning package developed under US DOT's and EPA funding at the Los Alamos National Laboratory (LANL) over the last eight years. It is an integrated system of travel forecasting models designed to give transportation planners accurate, complete information on traffic impacts, congestion, and pollution. It brings the recent advances in the computer hardware and software to the field of transportation modeling. At the core of TRANSIMS is an agent-based simulation system capable of simulating second-by-second movements of every person and every vehicle through the transportation network of a large metropolitan area.

TRANSIMS is rather a modeling system than a model. TRANSIMS is composed by a series of Modules, which could be executed in any desired order by a set of scripts. Models are refined and calibrated through a series of feedback loops between TRANSIMS modules that changes the behavior of selected individuals in the synthetic population.

There is no built-in mode choice module in TRANSIMS. The transportation mode for chained trips, conferred here as a tour, is decided in the shortest path algorithm in Router Planner Module. The allowed mode on the paths with the smallest generalized cost (usually time), will be selected and fed into the simulation along with the paths.

Los Alamos National Lab (LANL) has developed a mode choice methodology in TRANSIMS in Portland Study. A calibration process on mode choice was incorporated. Feedback processes were used heavily in the process. There were two parts. First, a mode choice model was calibrated on sample synthetic data in base year. Then, the calibrated model was applied to the forecast year. The calibration was made towards the mode split in real life. Generalized cost composed of travel time, travel cost and income were used in the mode choice calibration process.

1.2 Problem Definition

In TRANSIMS, there is no built-in module for travelers' mode choices. The transportation modes are taken from path with the smallest generalized cost in Route Planner Module.

In Portland Study, a mode choice methodology implemented by a series of feedback processes is proposed. However, this methodology still uses an aggregate, deterministic model. Besides, There is little solid theoretic ground for the format and coefficients of the generalized costs used in the calibration process. The accessibility to a mode, especially to Transit, was not included in the model.

1.3 Thesis Objectives

The objective of this thesis is to provide a disaggregate, stochastic methodology on mode choice with the accessibility to a mode modeled in TRANSIMS. The model is then applied to transportation planning study in Blacksburg, VA. The results of the study are explained and evaluated.

1.4 Organization of Thesis

The thesis first presents a literature review, followed by a review on the upcoming transportation-planning package, TRANSIMS, and the mode choice methodology in the Portland Study using TRANSIMS. Subsequently, the theoretic aspect of the proposed

mode choice methodology in TRANSIMS is presented. The application of the methodology to Blacksburg and the evaluation of the results are stated afterward. Finally, conclusions and recommendations for further research are summarized.

Chapter 2 contains insight gleaned from the current literature on transportation planning models, and aggregate and disaggregate mode choice models. In addition, some of the lessons learned and recommendations of others for further research are presented.

Chapter 3 begins with an introduction on TRANSIMS. The feedback module, which is critical in the mode choice process, is then presented in detail. The mode choice methodology used in Portland Study by TRANSIMS is explained and the need for a new mode choice methodology is stated.

Chapter 4 focuses on the explanation of the theoretic foundation of the proposed mode choice methodology in TRANSIMS, which includes two parts: calibration on survey data and implementation on synthetic data in TRANSIMS.

Chapter 5 presents the application of the proposed mode choice methodologies in Blacksburg Study. The evaluation of the application's result is then discussed.

Finally, Chapter 6 presents a brief discussion and summary of the findings.

Chapter 2 Literature Review

2.1 Transportation Planning Models

Transportation planning has gain people's attention for a long time and it has become a hot research topic as early as the 1950s.

2.1.1 Trend Analysis

At the early stage of transportation planning, trend analysis was used a lot. It was implemented by extrapolating past trends to estimate future travel. Such an approach is still used occasionally for estimating future traffic on a single facility, in a relatively isolated area, where only moderate and uniform growth or change in development pattern is anticipated. A more sophisticated extension of the approach is to incorporate the relationship between the traffic trend and the development trend. An understanding of such a relationship can be obtained by comparing the past traffic trend to the trend of development during the same period. This relationship is than extrapolated in the future to obtain the forecast. (Daganzo, 1979)

2.1.2 Mathematical Models

The use of mathematical models in transportation planning is usually more difficult, more time-consuming and more costly, but provide more accuracy. Such models are usually developed with data from travel surveys and are verified using actual traffic counts on numerous facilities during the same period as the travel survey. There are two principal kinds of transportation planning models: direct demand models and sequential models. (Daganzo, 1979)

1. Direct-demand models

Direct demand models are complicated mathematical formulations that estimate roadway traffic and mass transit usage directly from the urban development forecast. The philosophy of these models is that a traveler makes decision on a trip concerning all aspects, such as destination and mode, simultaneously. The implication in this approach is that all characteristics of a trip are interactive and are viewed together as available options by the traveler.

Direct-demand models have not received wide acceptance among transportation planning agencies primarily because of their complexity. The mathematical formulations of these models are complicated, and they are difficult to be successfully understood, developed, and applied. The mathematical structure usually becomes very intricate in order to incorporate the variables that represent factors considered to affect travel behavior. This intricacy can cause considerable problems with model preparation because it is difficult to determine which variables or parameters are effective for providing accurate estimates.

As a result, direct-demand models have not been accepted to the degree that their philosophical attractiveness would seem to indicate.

2. Sequential models

From 1960s, the sequential models started to be widely accepted to carry out transportation planning. The four-step procedure is a kind of typical sequential models. There are major modules in four-step model, each representing a different aspect of the traveler's decisions. The *trip generation* module estimates how many trips will be made to and from individual analysis zones. The *trip distribution* module estimates where trips originating in various zones will be destined, linking the trip origins and destinations from the generation model. The *mode choice* module identifies whether each trip will be made by motor vehicle or mass transit, and it may estimate the size of carpools. The *traffic assignment* module allocates the trips for each mode to the most likely route, whether a roadway or a transit line.

One of the principal reasons for the much wider acceptance of the sequential modeling approach is that it seems to be easier to understand. Breaking the decision process into steps facilitates analysis of the several factors that influence travelers' decisions although it is somewhat contrived. This breaking down also eases the development and verification of the models because there are fewer factors of influence in each module. The relationship between the cause and effect of those factors are more likely to be apparent. The federal and many state governments provide technical support, including computer programs, documentation, and training, only for sequential models.

2.1.3 Disaggregate Models

All of the models listed above can be categorized as aggregate transportation planning models, which have been used widely and gained a great amount of accomplishments.

However, the aggregate transportation planning models have been severely criticized for their inflexibility and inaccuracy. These models at base attempt to represent the average behavior of a group of travelers instead of a single individual.

Disaggregate models appeared in 1980s offer substantial advantages over the aggregate counterparts. Disaggregate models at base represent the behavior of individuals. The disaggregate models are clearly superior to the grouping of behavior zonally and by predefined segments in aggregate models.

However, disaggregate transportation planning model being used for a large metropolitan area has just been on the research screen, not in the real world because it would involve a huge amount of data, which were neither available, nor manageable by computers at that time period. Benefiting from the boom of computing capability of modern computers and more and more detailed census data, TRANSIMS as a disaggregate transportation planning model has become possible to be used in the real world.

2.2 Mode Choice Models

2.2.1 History of Mode Choice Models

Many different approaches have been used in the transportation field to forecast the usage of a transportation facility or service. In the early days, the techniques in transportation planning were very rudimentary and there was no mode choice concept proposed. Practitioners usually developed planning models with a very specific application or problem in mind.

Mode choice models emerged as a module in the sequential model structure and it was usually the most complicated one. Typically these models estimate how many persons will ride public transit and how many will use private vehicles. Further sophistication of these models may include identifying submode choice among different transit services and estimating the number of car pools or van pools of various sizes for high occupancy vehicle facilities.

Mode choice is only implemented on users with access to more than one travel mode. These users are called choice users. The others are called captive users. Households without automobiles, for example, are captive public transportation users and have little or no mode choice, as are households in areas that are not easily accessible to points of public transportation service: they are captive automobile users.

Although choice can be reflected by the number of people choosing each alternative, a more appropriate measure is the proportion of a population that makes each choice. Using proportions as measures has the advantage that it permits the analysis of choice independent of the total number of people in a population.

Proportions are meaningful in the case of deterministic choice at a market level only. For deterministic choice at an individual level it suffices to identify the alternative chosen using some meaningful rule. In the case of stochastic choice, the proportions are thought of as probabilities measured over a population and a choice set. Stochastic choice probabilities can be measured at an aggregate market level, indicating for each alternative the probability that it is the choice of an individual chosen at random from the population, when the sampling is defined over a given period of time during which it is postulated that a choice will be made. At the individual level, the choice probability of an alternative defines the number of times the alternative is chosen by the individual relative to the total number of times a choice is made if the individual faces exactly the same choice environment.

Experience with travel behavior analysis suggests that the deterministic model of choice may be limited in its replication of real life situations. Three primary reasons suggest that a stochastic model of choice may be preferable. One is that the behavior of individuals may not always follow the rational rules of choice exactly and that the traveler behavior

cannot be anticipated in a deterministic model. The second is that it is usually not possible to include in the choice function that all the variables possibly influencing choice. If such a function were possible, it would no doubt be so complicated as to render it impractical. The third reason is that the typical potential traveler is not likely to have perfect information about the transportation system and the alternatives it offers. Thus the set of alternatives I , identified by the analyst, may be larger than that encountered in fact by the traveler, or the function may contain variables about which the information as perceived by the traveler may be absent or incomplete.

These reasons suggest that a good model of choice might be one in which the choice function is considered a random function that takes on different values with certain probabilities. This random function reflects that alternatives are perceived differently by different individuals or by the same individual on different occasions even given values of the choice function. The perceived utility, call it $U(\cdot)$, is then a random function. So the probability that an alternative is chosen is defined as the probability that it has the greatest utility among the available alternatives.

Transportation engineers and planners began to borrow techniques from econometrics and applied them methodically in mode choice problems. Application of discriminant analysis to modal split is such an example (Beesley, 1965; Quarmby, 1967).

As researchers looked for more efficient and reliable ways of predicting how transportation facilities and services are used, it became clear to them that some of the "off-the-shelf" techniques could be improved and new econometric techniques began to emerge. The Multinomial Logit (MNL) model is, perhaps, the most noticeable such technique. It has been popularized by econometricians and researchers with interest in transportation problems and has been used in many transportation studies (Theil, 1969; Rassam *et al.*, 1971; McFadden, 1973, 1974).

The appeal of MNL model basically results from the fact that it's easy to calculate. In spite of its computational advantage, its Independence of Irrelevant Alternatives (IIA) property is also known to be a weakness. This axiom on selection probabilities means that the relative probability of choice between two alternatives depends only on their measured attractiveness and is independent of all other available alternatives. It's certainly not the case in the real life. For example, if one is dealing with urban mode choices, then this property might imply that the relative odds of taking an automobile over taking a bus is independent of whether there is train service to the same destination. This is not likely since the presence of a train as a third alternative is likely to affect the probability of choosing the bus more than that of choosing the automobile and hence likely to change their relative odds. However, there is little evidence that models free from IIA property will significantly improve forecasting results.

The Multinomial Probit Model (MPL) is another technique, which has a more decent theoretic background. However, only a few, very limited applications of the model have appeared in the travel demand literature, and there is still no evidence to suggest in which

situations the greater generality of MPL is worth the additional computational problems it brings. (Ben-Akiva and Lerman, 1985)

Because the Multinomial Logit Model is not always suitable when some of the modes are inherently more similar (e.g., bus and rail may be considered more similar to each other than either of them to car), the so-called nested Logit may also be used, precisely to reflect the similarities. Within the nested structure, the first choice may be between the car and public transport, while, conditional on choosing public transport, there may be a further choice between the bus and rail. (John Bates, 2000)

Before 1980s, aggregate mode choice models dominated in the field of transportation. Basically, these models attempt to represent the average behavior of a group of travelers instead of a single individual. It impairs models' flexibility and accuracy.

From 1980s, the disaggregate modeling technique began to be used on mode choice as well as in the other travel demand models. A significant effect of the additional precision gained from this approach has been the ability to obtain reliable estimates for models with a much broader range of explanatory variables since the models capture the ongoing human decision-making mechanisms more accurately than the aggregate ones. The Multinomial Logit model can also be used on the disaggregate level. The model is estimated directly from information of individuals without first aggregating it to a zonal level. Its output is the probability to choose an alternative for a specific individual.

Not matter in the aggregate mode choice model or in the disaggregate one, the variables used to explain mode choice always belong to two categories: socioeconomic demand variables and level of service or supply variables. The former includes Income, Age and role in household, car ownership, and household size. The latter includes the travel time and travel cost. An important concept related to the latter is Accessibility. In broad terms, accessibility has been defined as "the ease and convenience of access by means of a choice of travel" by the Department of Environmental, Policy and Procedure Guidelines, PPG6, 1996(3). The accessibility for private cars can be measured by access time to and from parking lot, whereas that for public transport can be measured by walking and waiting time as well as the frequency of transport service. Accessibility plays a very important role in travelers' mode choice (Schoon, McDonald, and Lee, 1999).

In the next section, we will compare the disaggregate mode choice models with their aggregate counterparts.

2.2.2 Comparison between disaggregate and aggregate mode choice models

In the area of modeling travelers' behavior in a multimode situation, the debate between aggregate and disaggregate models has captured most of the researchers' attention from 1970s'. The result is the widespread recognition of the desirability and efficiency of the disaggregate multinomial Logit model. Detailed description of the model and its

statistical properties are given in McFadden (1973) and in Richards and Ben-Akiva (1975), respectively. Some of the main factors which contributed to the above result include the fact that the disaggregate approach rectifies several deficiencies associated with the aggregate approach, and that it is more behaviorally sound. These factors are explained in more detail below. (Ghoneim and Sargious, 1986)

- Prediction ability

How does a model simulate the phenomenon under study and to produce reliable forecasts for future planning purposes is the key to measure the ability of a modeling technique. Empirical evidence indicates that disaggregate mode choice models are capable of performing these tasks successfully.

Watson (1973, 1974) demonstrated that the disaggregate mode choice model provides a better statistical explanation of mode choice behavior than its aggregate counterpart. He found that prediction errors associated with the aggregate approach are over 12 times larger than those associated with the disaggregate approach. His conclusions were further supported by Parody (1977) whose results showed that disaggregate models can be used successfully to predict future travel behavior.

- The behavior aspect

As indicated by Richards (1982), the behavioral aspect of a model is a required property which reflects the analysts' understanding of the mechanisms influencing travel behavior and which depends on the model's ability to capture causal relationships. Daly (1982) demonstrated that the disaggregate statistical analysis is the most useful technique in the quantitative analysis of travel behavior. In this technique, theory of behavior is assumed priori and then empirically tested and refined. Disaggregate models are based on the observable behavior of individual decision-makers. Therefore, statistically, they are efficient in extracting the maximum possible information through the averaging procedure. Disaggregate models are more apt to replicate passengers' travel behavior through causal factors and interrelationships of passengers and system attributes.

- Model calibration

Empirical tests have shown that aggregate parameter estimates in model calibration process are less reliable than the disaggregate estimates derived from the same data (Berkowitz et al. 1985). Parameter estimation bias can lead to serious planning deficiencies and limitations when misleading information or erroneous parameter signs are obtained. A subsequent result of the latter is possible omission of important variables from the analysis.

Aggregation also hinders the specification of the models and the measurement of the independent variables. At an aggregate level, it is extremely difficult to identify causal relationships between variables and observed passenger behavior. Generating measurements of level of service variables (e.g., the travel time) or a meaningful concept

to incorporate socioeconomic variables (e.g., income) in aggregate models is also a very difficult task to undertake.

- The transferability

Based on the assumption that people with similar characteristics and under similar circumstances behave similarly, disaggregate models have potential for transferability. In order to allow transferability, disaggregate models can now be modified in such a way as to overcome previous problems of large variances in prediction errors. Koppelman and Rose (1983) have presented some adjustment methods which, if applied to these models, will enhance their transferability. When a disaggregate model is used in another location, it can be updated with disaggregate or even aggregate data. In a study by Koppelman *et al.* (1983), four models of different specifications were transferred between geographic sectors. Updating the disaggregate models with aggregate data consistently gave transferability improvements close to those achieved when these models were updated with disaggregate data. On the other hand, aggregate models are highly dependent on the zoning structure. Therefore, they are not easily transferable, geographically, and may not even apply to the same area if changes in the zoning system occur. It must be borne in mind, however, that this transferability property of disaggregate models can only be attained if they are properly selected and calibrated, and that, unfortunately, proper validation of the choice probability function is a very difficult task requiring enormous data-gathering efforts. (Daganzo, 1979)

- Data requirements

Disaggregate data contain more information than aggregate data and are therefore more valuable for calibration purposes. This can be seen by means of the following illustrative argument: If we want to study the behavior of wealthy people, it is not sufficient to study the average reaction of a town with high average income because even in a very rich town there will be significant fraction of poor and middle-income people, which will filter the results. Some transportation studies have indicated that much of the dispersion in some attributes, such as income and age, is lost when one used aggregate data [see, for example, Fleet and Robertson(1968) and McCarthy(1969)]. Since the choice probability function of a disaggregate demand model is perfectly suited for calibration with high-quality disaggregate data, it is possible to use just a few hundred data points in order to obtain accurate estimates of parameters. The efficiency with which the information in the data is handled with disaggregate demand models is perhaps their most desirable feature, since it greatly simplifies the data-collection process. Berkowitz *et al.* (1985) demonstrated that a disaggregate model performs as well as an aggregate model (comparable goodness of fit results) when employing as little as 13.3% of the data used to develop the latter.

Contrary to this quantitative efficiency, only high quality data are adequate for disaggregate analysis. The lack of such data to date has been the major obstacle to employ disaggregate models. However, the improvements in disaggregate model

transferability as noted in Koppelman and Rose (1983), and other ongoing research on improving the method of modeling human choice behavior, should help to reduce the problem of insufficient data.

Miller (1985) outlines the important features of an ideal data base for disaggregate transportation analysis. Three key attributes are presented: first, comprehensiveness in covering all travel modes, purposes, duration, and characteristics of the trip and trip-maker; second, consistency in the data collection methodology and timing and in the definition of terms and variables throughout all phases of the travel demand analysis; and third, spatial disaggregation by designing survey forms which provide information about trips and trip-makers at the individual level and by constantly updating these surveys. Data availability does not make aggregate models any better than the disaggregate ones. It only makes it easier to continue employing them in practice. Therefore, there is an urgent need to rectify this data problem and further improve model transferability in order to benefit from the disaggregate approach and achieve the best possible results by integrating it in the complete demand forecasting process.

Chapter 3 TRANSIMS Review

3.1 Introduction

The TRansportation ANalysis SIMulation System (TRANSIMS) is a set of new transportation and air quality analysis and forecasting procedures developed to meet the Clean Air Act, the Intermodal Surface Transportation Efficiency Act, Transportation Equity Act for the 21st Century, and other regulations. It consists of mutually supporting simulations, models, and databases that employ advanced computational and analytical techniques to create an integrated regional transportation system analysis environment. By applying advanced technologies and methods, it simulates the dynamic details that contribute to the complexity inherent in today's and tomorrow's transportation issues. The integrated results from the detailed simulations will support transportation planners, engineers, decision makers, and others who must address environmental pollution, energy consumption, traffic congestion, land use planning, traffic safety, intelligent vehicle efficiencies, and the transportation infrastructure effect on the quality of life, productivity, and economy. (<http://transims.tsasa.lanl.gov/>)

3.1.1 TRANSIMS Framework

TRANSIMS is composed by a series of Modules, Population Synthesizer, Activity Generator, Route Planner, Traffic Microsimulator, and Feedback Module, which could be executed in any desired order by a set of scripts. Besides, Emission Estimator and Output Visualizer are two other modules for analyzing and displaying TRANSIMS output data. The framework is shown in Figure 3.1.

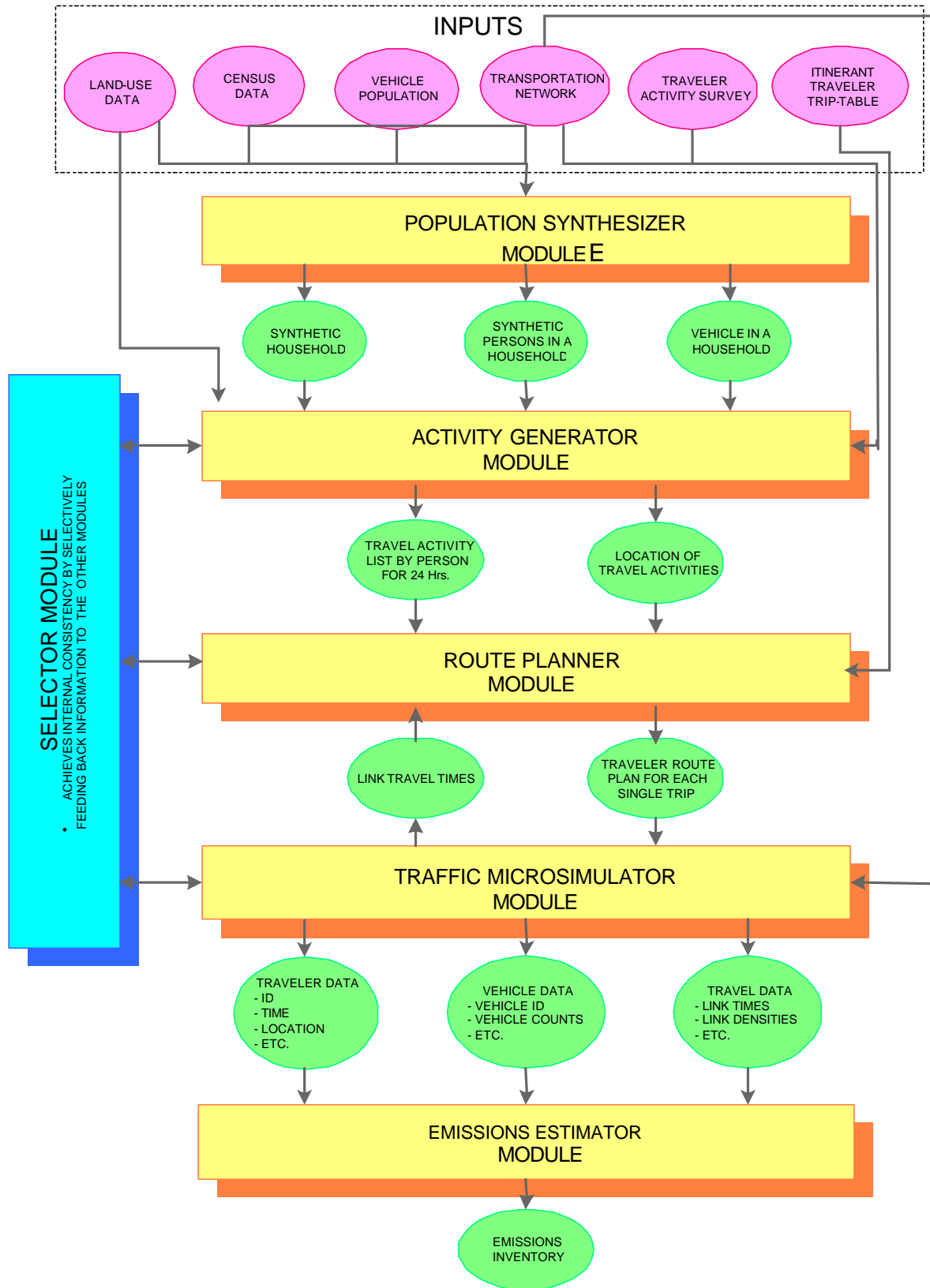


Figure 3.1 TRANSIMS Framework

3.1.2 Population Synthesizer

The Population Synthesizer module creates synthetic households that represent every real individual household in the metropolitan region under study. These synthetic households have synthetic persons that match the Census Block Group population demographics obtained from the census bureau.

A metropolitan area is comprised of many PUMA (Public Use Microdata Area). Each PUMA is comprised of many block groups. The complete census records (long forms) are sampled for the households within the PUMA and are called PUMS. The PUMS usually consists of a 5% representative sample of all the households in a PUMA. Each block group has its own summary of marginal demographic data for all the households referred to as STF-3A. The STF-3A does not provide the cross-tabulation of the demographic data. The Iterative Proportional Fitting (IPF) procedure is implemented to create a detailed cross-tabulated demographic data, whose marginal values correspond with STF-3A summary statistics for each block group from the PUMS data.

To generate synthetic persons for future scenarios, STF-3A forecasting marginal statistics need to be used.

STF-3A and PUMS data are shown in Figure 3.2 and Figure 3.3, respectively; whereas the question marks are detailed cross-tabulated demographic data need to be obtained from the process.

Householder's Age								
Workers	15-24	25-34	35-44	45-54	55-64	65-74	>74	Total
0	?	?	?	?	?	?	?	30
1	?	?	?	?	?	?	?	57
2	?	?	?	?	?	?	?	224
>2	?	?	?	?	?	?	?	14
Total	32	82	72	54	33	24	28	325

Marginal values

Figure 3.2 STF-3A data for a Block Group

Householder's Age								
Workers	15-24	25-34	35-44	45-54	55-64	65-74	>74	Total
0	3	12	14	11	21	51	54	166
1	9	47	63	44	59	45	14	281
2	18	56	113	83	45	17	6	338
>2	2	17	70	88	55	17	3	252
Total	32	132	260	226	180	130	77	1037

Figure 3.3 PUMS data for a Census Tract

Next, we need to match each household in the STF-3A cell to an actual-sampled household in the corresponding PUMS cell with similar demographics. Once the match is finalized, the full properties, and characteristics of the sampled household in PUMS are assigned to the synthetic household in STF-3A. The matching could be conducted by randomly selecting or based on some kinds of weights.

The synthetic households are located on the network using land use data. There might be more than one activity location in a Block Group. The probability to choose an activity location will depend on its land use data. For example, the activity locations that have greater residential area should have greater possibilities to be chosen to locate household locations.

Each synthetic household is created with a number of vehicles assigned to it. The number of vehicles owned by each household is given in the PUMS and is therefore transferred to the synthetic households through the matching process. The vehicle type assigned to each vehicle is determined at random according to a national distribution of vehicle emission types.

The last step to generate a synthetic population is to assign a unique number to each household and person in the population.

3.1.3 Activity Generator

The Activity Generator module creates activity lists for every synthetic person by matching synthetic households with survey households, whose travel and activity participation information for each household member is known, based on similar demographic characteristics.

Firstly, the survey households are classified according to their demographic data using CART algorithm. The CART algorithm is a technique for modeling a regression relationship between one or more dependent variables Y and independent variables X_1, X_2, \dots, X_K . Survey households will be put in the terminal nodes of the classification and regression tree in a way total variance in the value of dependent variables over the entire survey population is minimized at the end of CART algorithm.

An example of the classification and regression tree obtained by CART algorithm is shown in Figure 3.4.

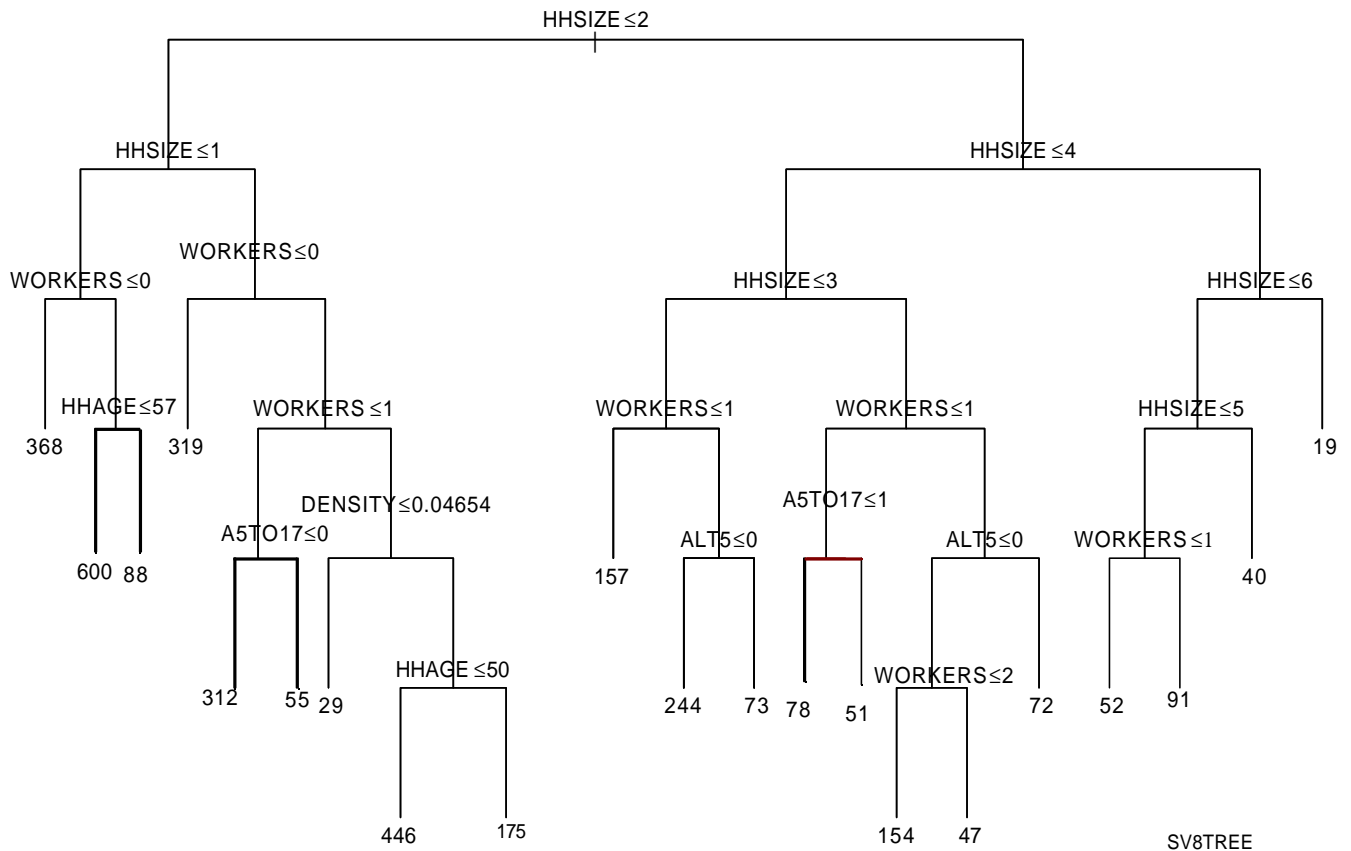


Figure 3.4: An example of the classification and regression tree generated by CART algorithm

Next, each synthetic household (obtained from the Population Synthesizer Module) will be assigned to a unique terminal node of the tree built from the survey households based on the same demographic data.

A survey household within the terminal node will be selected to match with the given synthetic household that is assigned to this node. The selection of the survey household could be done randomly, or weights could be assigned to survey households for flexibility.

The following step in Activity Generator is to match the assigned synthetic household members with the selected survey household members based on the following four demographic variables: *relate*, *work*, *gender*, and *age*. The synthetic persons can then take over the activity patterns from the matched survey ones. Each activity of the synthetic person is given a preferred start-time, end-time, and duration based on skeletal patterns of the survey person.

Each activity of the synthetic person must have a specific location within the region under consideration. First, a discrete choice model is used to select an appropriate zone

for each activity. This is carried out based on the zone data that specifies the attractiveness of each activity type within each zone. The selection of zones for work activities is conducted before that of the zones for all other types of activities. Second, a location for each particular activity is randomly specified within the selected zone using the weights given in the activity location table.

3.1.4 Route Planner

The Route Planner Module develops the route plans for each traveler based on the demand represented in the Activities data file. A time-dependent label-constrained shortest path for each trip executed by a traveler in the system is developed.

The technique adopted by TRANSIMS to identify a suitable travel route for any user is a variant of Dijkstra's procedure for finding shortest paths, which is suitably modified to accommodate time-dependent travel times and label sequence constraints. The underlying problem is referred to as Time-Dependent Label-Constrained Shortest Path Problem (TDLSP) and it is unique to TRANSIMS applications.

Once the plans are generated for all the travelers, they are simultaneously fed into the Traffic Microsimulator Module.

If no path is found that obeys the mode constraints of the traveler (*Path Anomaly*), or that satisfies the time bound constraint (*Time Anomaly*) or the *Invalid Shared Ride Time* anomaly, then the traveler and the trip request are marked for a *Path/Time/Shared Ride Time Anomaly* feedback to the Feedback Module.

3.1.5 Traffic Microsimulator

The TRANSIMS Microsimulator module executes individuals' travel plans as provided by the Route Planner module at the start-time specified by the plan. Plans that overlap in time are executed simultaneously by the Microsimulator, resulting in overall transportation dynamics.

The TRANSIMS Microsimulator module simulates the movement and the interactions of travelers in the transportation system of the study area. In this module every traveler tries to execute his/her travel movements, link by link along the intermodal networks, according to the plans generated from the Route Planner Module.

The TRANSIMS Microsimulator is capable of handling intermodal travel plans, multiple travelers per vehicle, multiple trips per traveler and vehicles with different operating characteristics. Realistic traffic behavior is simulated in Microsimulator by making decisions about lane changes, passing slow vehicles and evaluating interactions with other vehicles.

The Microsimulator first reads in the network data, which includes details such as links, nodes, lane connectivity, parking locations, etc. In the next step, the vehicle type, location

and traveler plans are read. Then, travelers and vehicles are placed on the network. Step three includes the movement of the travelers and vehicles on the network in tandem with their travel plans using the Cellular Automata approach.

The amount of computation necessary for simulating a large transportation network at a level of detail down to an individual traveler and an individual vehicle is huge. Hence a coarse simulation approach entitled “Cellular Automata” (CA) is used to keep up with the fast computational speed necessary to simulate a whole region.

The Cellular Automata approach essentially divides every link on the network into a finite number of cells. At every time step each of these “cells” is scanned for a vehicle presence. If a vehicle is present, the vehicle position is advanced to another cell using a simple rule set. These rules are concerned with vehicle movements in the same lane, performing lane changes, turn pocket lanes and vehicle movements at intersection. The rule set is made simple to increase the computational speed necessary for a large simulation.

Reducing the size of the “cell”, expanding the rule set and adding vehicle attributes increases the fidelity of the Microsimulator but would greatly affect the computational speed. The size of 7.5 meters length and a traffic lane’s width is chosen as a default value for the “cell”.

TRANSIMS Microsimulator also provides the capability of using multiple computer processors (CPU’S) to support a large number of travelers and a considerable size transportation network.

3.1.6 Output Visualizer

The output from TRANSIMS can be pictorially represented and viewed by the Output Visualizer. The Output Visualizer enables the user to display various input and output sets. It also provides tools to facilitate the analysis of the data sets. The tools encompass plotting, GIS, statistics, and animation. The Output Visualizer displays include those of travel plans, vehicles, signals, and intersection queues, among other data types.

The graphical user interface (GUI) of the Output Visualizer allows the user to manipulate the three-dimensional objects and presents the output in a graphical way. A user could be interested in the travelers’ plans (single, aggregated or filtered) that could be overlaid on a network as shown in Figure 3.5. As can be seen from the figure, such a representation would clearly help in the understanding of the plan data characteristics as overlaid over the network.

Apart from the above-mentioned features, the Output Visualizer is capable of animating the simulation as it progresses. Essentially, the snapshot data (that is output from the Microsimulator) feeds the frames for animation. It should be noted that an animated simulation for the whole network would include a lot of data in the form of Snapshot and would result in very huge files which require a lot of memory space.



Figure 3.5 The Output Visualizer Displaying Vehicles On The Network In A Zoomed-in View

3.1.7 Emission Estimator

The Emissions Estimator module translates traveler behavior into consequent energy consumption and pollutant emissions of nitrogen oxides, hydrocarbons, carbon monoxide, carbon dioxide, and particulate matter.

The Emissions Estimator, in the present version of TRANSIMS, produces estimates for:

- 1) tailpipe emissions from light-duty vehicles (LDVs),
- 2) tailpipe emissions from heavy-duty vehicles (HDVs), and
- 3) evaporative emissions

as a function of vehicle fleet composition, fleet status, and fleet dynamics. This section explains tailpipe emissions from LDV, and HDV. The evaporative emissions will be explained in a separate section.

The Emissions Estimator module thus requires information regarding:

- the fleet composition developed from the Population Synthesizer,
- inspection and maintenance test results obtained from local and national databases, and

- traffic patterns produced by the Traffic Microsimulator module.

The emission inventory obtained from TRANSIMS form the basis for the computation of pollutant concentrations in a metropolitan area based on atmospheric conditions, local transport and dispersion, and chemical reactions.

3.1.8 Feedback Module

TRANSIMS is a framework of transportation modeling. It offers the planner the capability to establish the specific and efficient model he needs. First, the TRANSIMS modules (Population Synthesizer, Activity Generator, Route Planner and Emission Estimator) can be arranged in different procedures. There is no fixed procedure with which must be complied. Second, feedback iterations provide a natural way to tailor models (of activity locations, mode selections, route planning, etc.) to specific, possibly overlapping, subpopulations. Feedback enables the overall computational system to reflect “learned” behavior within the simulated population represented. It drives the system toward Nash equilibrium: Information about each individual’s experiences is used to select those with sub optimal choices and aggregated to estimate the costs of other choices. The Feedback Module, also called Selector Module, collects, groups and sends data among different modules in the feedback process in TRANSIMS.

An outline for the module interfaces involving the Feedback Module is displayed in Figure 3.6.

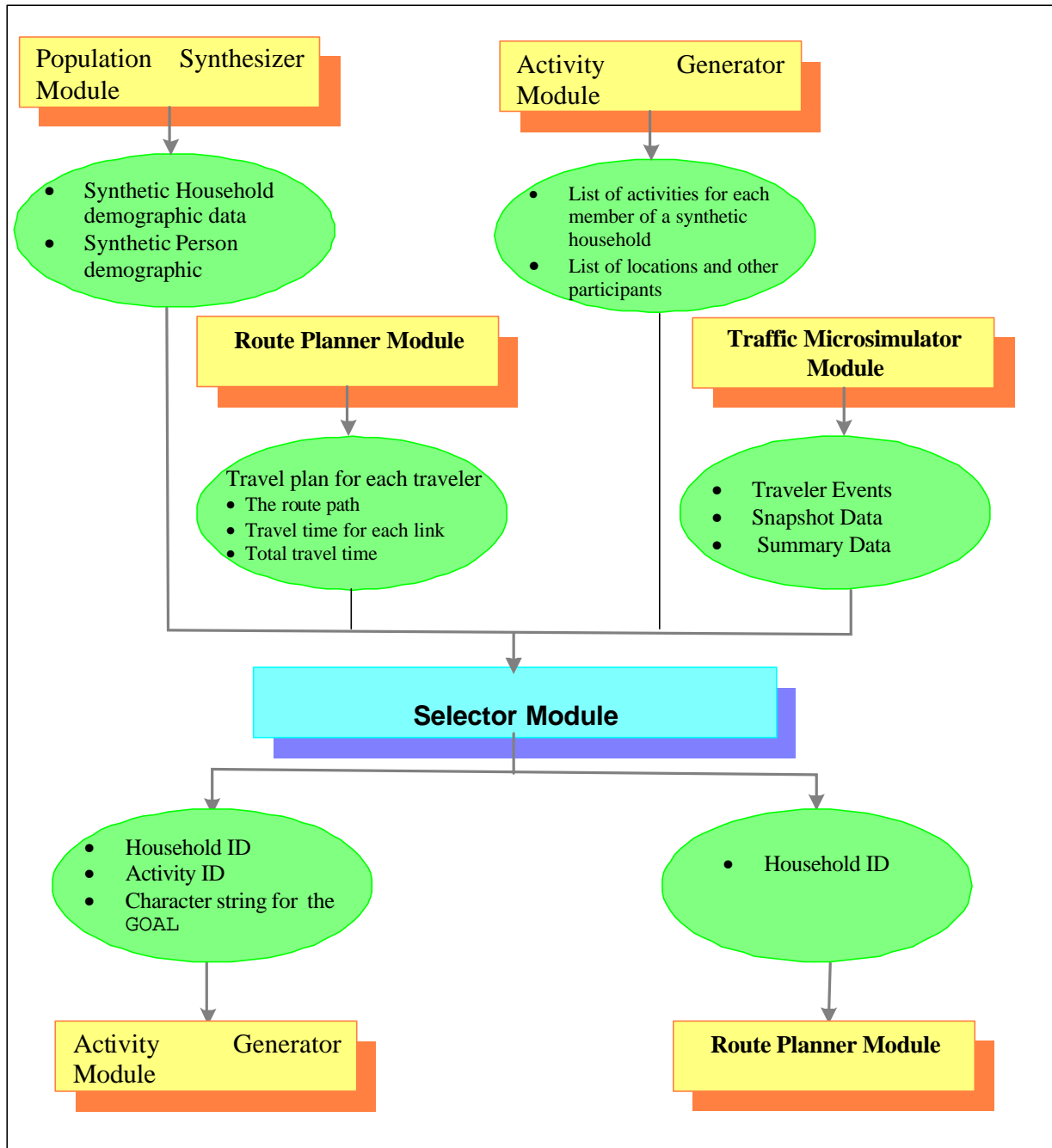


Figure 3.6 Flowchart of module interfaces for the Selector Module

The Feedback Module, as shown in Figure 3.7, has three modeling tools, Collator, Stratifier and Selector.

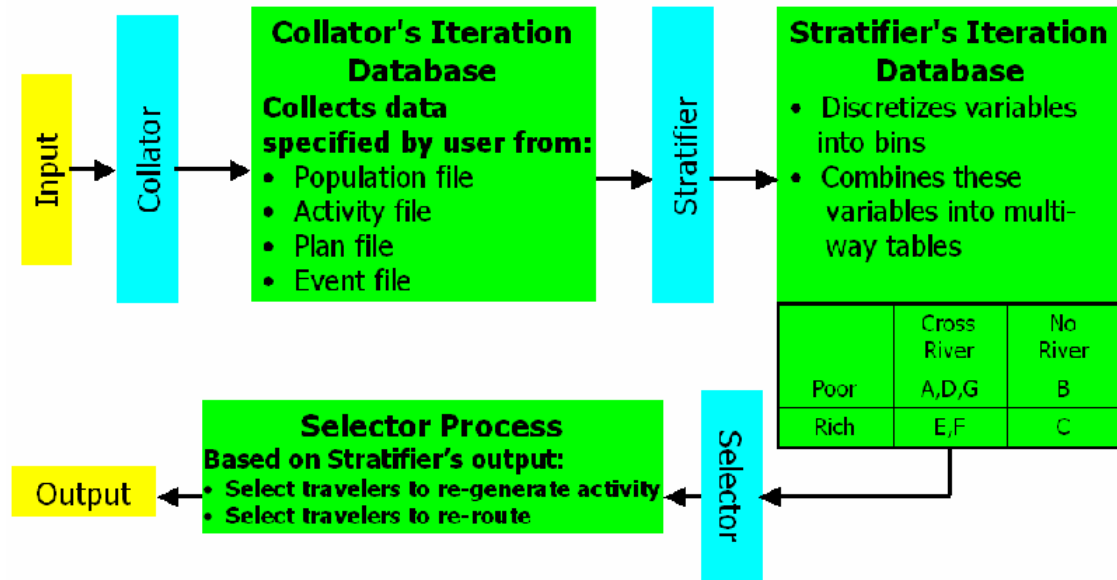


Figure 3.7 Relationship between the three model tools

The Collator gathers data from some input files of TRANSIMS, i.e., Network file and Transit Route file. It also extracts data from output files of Population Synthesizer, Activity Generation, Route Planner, and Traffic Microsimulator. All of the data collected are then put into Collator Iteration Database. Iteration Database is the archive of information about travelers across iterations.

The Stratifier divides trips into groups based on the criteria the user sets. The numeric identifier of each group is added to the iteration database and the database is developed into Stratifier Iteration Database.

It is the Selector's responsibility to pick up a subset of travelers from the Stratifier Iteration Database. It is done in two steps. After the cells are picked up from the group, a subset of travelers or trips is selected within the cell. The output of Selection on each subset will consist of one Activity Feedback File and one Route Planner Feedback File. The Activity Feedback File is then sent to the Activity Regenerator, and the Route Planner feedback File is sent back to the Route Planner. The user is free to use either or both of them.

The rules guarding how these three model tools manipulate data can be set by configuration keys in the configuration file. An iteration script is used to control the overall TRANSIMS Framework. A typical TRANSIMS study involves repeated iteration between modules. There is no single, "standard" iteration script because different study designs involve different iteration schemes.

The information about travelers available to TRANSIMS consists of the traveler-specific data contained in population, activity, plan, vehicle, and simulation output files. For feedback purpose, Feedback Module can extract data from the output files of all other TRANSIMS modules, except Emission Estimator. After some kinds of analysis, selected data will be sent back to Activity Generator and Route Planner module.

3.2 Mode Choice Methodology in Portland Study

3.2.1 Feedback processes in Portland Study

Los Alamos has done major studies on Albuquerque and Dallas transportation and completed a microsimulation of auto traffic patterns in 25 square miles of Dallas that represented about 200,000 vehicles over a five-hour period. A TRANSIMS model of Portland transportation is also done in 2000. The Portland study focuses on forecasting activity demand and predicting trips that use multiple modes of transportation. The goal of TRANSIMS is to develop technologies that can be used by transportation planners in any urban environment.

Usually, we have two ways on modeling: establishing a precise model, and establishing a somewhat coarse model and refining it by feedback processes. The first kind of models always requires strict prerequisites. If there are some changes, it is very likely that they won't work any more. The second kind of models are much more robust on the changes. TRANSIMS belongs to the second kind.

In TRANSIMS, models are developed from a series of feedback loops between TRANSIMS modules that changes the behaviors of selected individuals in the synthetic population. Feedback and controlling feedback are important functions of the Framework. Basically, feedback processes can be used in three aspects:

- Feedback can be used to fix unrealistic transportation movements. For example, after running Activity Generator, many travelers were assigned the walk mode from the survey whereas the activity locations were chosen randomly. It resulted long walk trips, which is greater than 2500 meters. To eliminate long walk trips, the activities with long walk trips will be feedback to Activity Generator and activity locations for these trips will be changed.
- Feedback can also be used in model calibration. An example of it is to calibrate bridge crossing screen-line counts. A model is developed for activity location choice. Given a set of values for parameters in the model, the crossing screen-line counts can be obtained. Adjust the values of parameters through feedback loops until the realistic counts are reached. The model can then be applied for forecasting.

- Thirdly, feedback processes can be used to obtain traffic stabilization. The Router Planner determines the shortest path for a trip by travel time without considering other trips' influences explicitly. After running Traffic Microsimulator, the updated travel time for each link is then fed back to Router Planner and new sets of shortest paths for every traveler can be found. In this way, the combined effects of all travelers on transportation network can be included in the travel times. After several rounds, traffic will be stabilized and generate realistic result.

TRANSIMS is a very flexible model framework. It allows modules to be executed in any desired order by a set of scripts. It is users' responsibilities to design the order in which the modules are called and how the feedback processes are executed to fulfill their objectives.

The model for the Portland Study is outlined in Figure 3.8.

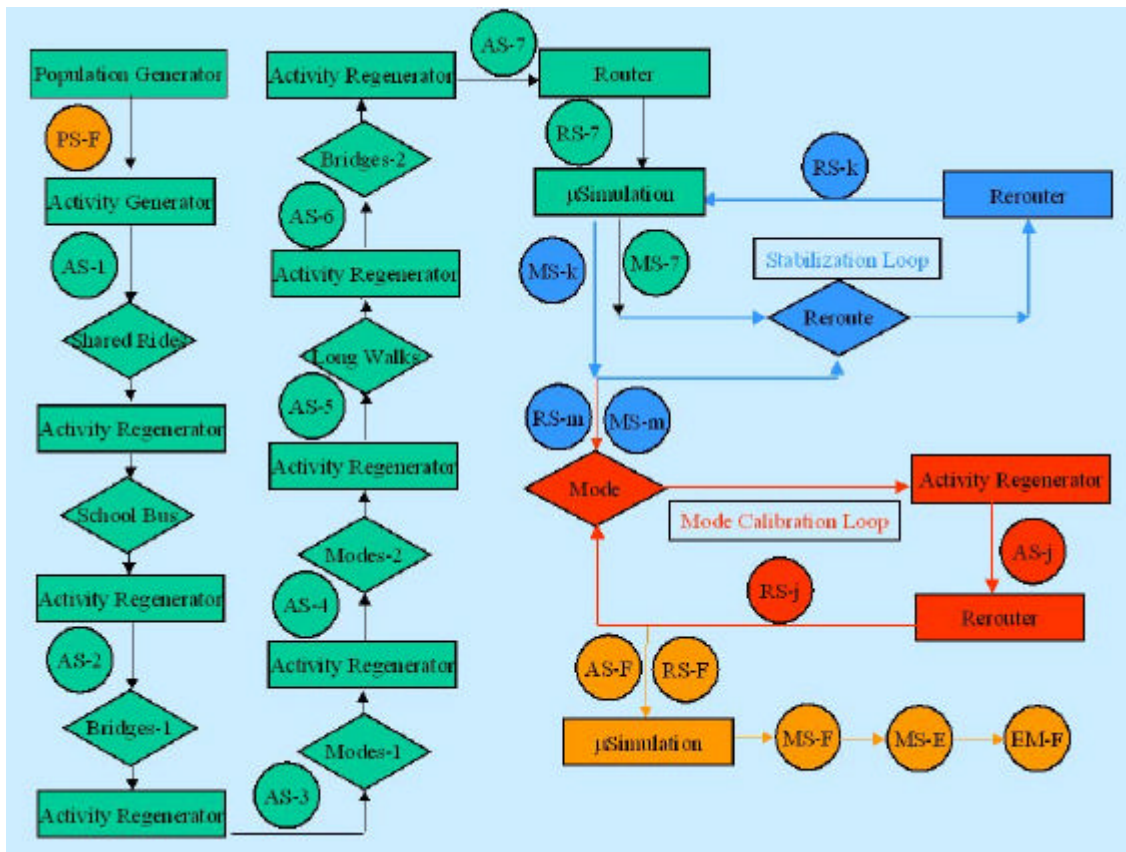


Figure 3.8 Feedback processes in Portland Study (Portland Study Report)

We can divide these feedback processes in Portland Study into three groups based on their positions:

- Inner loops of Activity Generator. They are feedback processes from Activity Generator back to Activity Generator itself. In these loops, unrealistic trips on

shared rides, long walks/bikes to school, impossible transit, OFDN (Origin far from a transit stop, destination near a transit stop) transit tours, and long walks were fixed. Besides, calibration on bridge cross-line counts was also made. These processes are shown in green in Figure 3.8.

- Traffic Stabilization. Selected data are fed from Traffic Microsimulator to Route Planner. It is shown in blue in Figure 3.8.
- Mode Choice Calibration. In this process, Activity Generator and Route Planner are used. It will be explained in detail in the next section. It is in red in Figure 3.8.

3.2.2 Mode Choice Methodology in Portland Study

There were two stages in mode choice procedures in Portland Study. First, a mode choice model was calibrated on sample synthetic data in base year. Then, the calibrated model was applied to the forecast year. The mode choice calibration was made towards the mode split in real life. Since the mode split is an aggregate attribute, the mode choice calibration methodology was still an aggregate procedure.

There are six steps in the process of base year calibration:

1. The activity lists for every traveler were obtained after the execution of Activity Generator. Activity file were converted into a tour-based database instead of a trip-based one. All of the analysis for mode feedback was done with tour-based databases. The reason is that tours are the only way that a mode choice makes sense without adding all of the complexity necessary to change the activity pattern. For example, a traveler on a bus tour could not have an auto trip placed in the middle of the tour without walking all the way home to get the car.
2. Tours were then stratified by the type of tour and other characteristics that may influence mode choice. Three grouping criteria were used in Portland Study.
 - Work/non-work
 - Transit distance – ONDN (Origin Near transit, Destination Near transit), OFDN (Origin Far away from transit, Destination Near transit) and DF (Destination Far away from transit). Distance to transit within 1000 meters is looked as near transit.
 - Home urbanization – very-urban, urban and suburban
 - Work urbanization – very-urban, urban and suburban
3. Since it was not necessary to test all tours to calibrate the model, only a fraction of the tours was selected for analysis. Pick up a subset of tours at uniformly random from each group.
4. Put all of the tours selected onto Transit network. It means convert all Auto trips to General Transit trips. An activity feedback file was sent to Activity Generator with the updated mode, general transit. Execute Activity Generator and Route Planner and collect travel time and DollarCost, a name for transit fare, for each tour.

5. Put the same tours as step 4 onto Auto network. Run Activity Generator and Route Planner with mode Auto and collect travel time and DollarCost for each tour. DollarCost is parking price for Auto tour.
6. Decide the mode for each tour according to the difference of generalized cost between General Transit and Auto. The difference of generalize cost was defined as follows in Portland Study

$$d = C_{Transit} - C_{Auto} = \alpha \cdot \log_{10}(1.01 + \max(0, Income)) \cdot (Time_{Transit} - Time_{Auto}) + (DollarCost_{Transit} - DollarCost_{Auto}) \quad (3.2)$$

There is only one parameter, α , in the Eq. (3.2). Given a value of α for each group specified in step 2, we can get the difference of the generalized cost, d . If d is greater or equal to zero, mode Auto should be chosen for the tour. Otherwise, mode Transit should be chosen. A fraction of auto tours to total number of tours N_{Auto}/T_{Total} (mode split) could be obtained in each group. In Portland Study, a manual fit was used, where the analyst chose the value of α for each group until the target mode split for each group was achieved.

Having calibrated the mode choice model, we would apply it in forecasting. There were still six steps in doing so in Portland study.

1. Convert activity data into a database by tours.
2. Stratify tours based on the same criteria as in base year.
3. Put all of the tours onto Transit network and collect travel time and DollarCost for each tour.
4. Put all of the tours onto Auto network and collect travel time and DollarCost for each tour.
5. Calculate the generalized cost difference between Transit and Auto for each tour according to the Eq. (3.2) and α value obtained from base year data. Decide the mode of the tour based on whether the generalized cost difference is greater than 0.
6. Fixing tours with problems. In the Study, an infeasible transit is any transit tour lasting longer than 24 hours or having more than three transit transfers on any trip within the tour. An infeasible auto trip is any auto tour lasting more than 24 hours. Each infeasible tour is re-moded. If it was an auto tour, its mode was changed to General Transit, and vice versa. All changes were counted, and the same number of tours was changed to the other mode, so that the net count of tours for each mode remains unchanged.

After the mode choice is done, all of the trips are then sent to Route Planner, Traffic Microsimulator and Emission Estimator to finish the whole planning process.

3.3 The Need for a New Mode Choice Methodology in TRANSIMS

From the mode choice process explained above, we can find the mode split between Auto and Transit was used as the only input from outside of TRANSIMS. Mode split was an aggregate indicator, so the mode choice calibration model in Portland Study was still an aggregate one. However, TRANSIMS is a disaggregate, behavioral transportation planning package. The process doesn't make full use of the great amount of data on every trip for every traveler in TRANSIMS.

Besides, mode choice calibration in Portland Study is a deterministic process, where the mode with the shortest generalized cost always was chosen for each tour. However, since similar individuals may make different choices and there must be some unexplained variations in the model, we should use stochastic procedure to model mode choice.

In Portland Study, the concept of generalized cost was used in the mode choice calibration. Although the generalized cost is composed of not only travel time and travel cost but also some socioeconomic data, there is no solid theoretic ground for its format and coefficients.

As we know, accessibility to Transit is significant in mode choice problem. In Portland study, the accessibility to Transit was used in such a way that the distance to a Transit stop was only used as a criterion to stratify tours. The accessibility to a mode was not included in the model.

Furthermore, the mode choice process was implemented after routing and simulating trips. It did not comply with the traditional Four-Step Model, where the mode is specified before the route could be chosen. This incompliance might bring obstacle for acceptance of TRANSIMS by transportation planners.

Therefore, there is a need to develop a new methodology for mode choice in TRANSIMS to overcome these limitations.

Chapter 4 Theoretic Aspects of the Proposed Methodology

4.1 Introduction

Travelers can be divided into two groups: captive users and choice users. Captive users are the ones who can only access one travel mode. For example, households without automobiles are captive public transportation users and have little or no mode choice. Households in areas that are not easily accessible to public transportation service are captive automobile users. Choice users, oppositely, have access to more than one travel mode, and can make mode choice decisions based on the relative advantage of one mode over another. In our research, captive users will use the only mode they can access. Thus, only choice users are within the horizon of the research.

As we discussed before, disaggregate models have higher capability to capture casual relationships and simulate the phenomenon in the real life than aggregate models. At the same time, it requires more detailed data for each traveler. TRANSIMS is a disaggregate, behavioral transportation planning package. A great amount of data on second-by-second movements of every person and every vehicle through the transportation network are available in it. Therefore, implementing a disaggregate model on mode choice in TRANSIMS is efficient and possible.

In many applications, it is necessary to consider choices made according to some sequence. It also happens when a choice is made between groups of alternatives and then followed by a choice within the chosen group.

The mode choice process in the research is implemented in two stages. First, a mode is chosen for a tour between Auto and General Transit. A tour is defined as a series of trips of a traveler within which the first trip begins from home and the last trip ends at home. The direct movement between two activity locations is a trip. Tours are the only way that a mode choice makes sense without adding all of the complexity necessary to change the activity pattern. For example, a traveler on a bus tour could not have an auto trip placed in the middle of the tour without walking all the way home to get the car. The Binomial Models will be used for tour mode choices between Auto and General Transit.

Then, the mode is specified for every trip. If the tour mode is Auto, the trip modes could be Auto and walk; whereas the trip modes could be Walk, Bus, Rail, and Trolley, and etc., if the General Transit is taken as the mode of the tour. We will call them tour mode choice stage and trip mode choice stage in the following sections, respectively.

Table 4.1 Trip modes allowed by tour mode

Tour Mode	Trip Mode				
	Walk	Car	Bus	Rail	Trolley
General Transit	X		X	X	X
Auto	X	X			

By implementing the mode choice in two stages, we can decrease the complexity of the proposed model and increase the computing efficiency. If we make choice among all modes in one step, we will have to use Multinomial Models. Multinomial Models are much more complicated and time-consuming in computation than Binary Models. Things will be even worse if we have to consider the combined modes as options. In our case, tours made by more than one Transit modes, i.e., Bus and Rail, Bus and Trolley, Rail and Trolley all present in the real life. So we will have to deal with at least 7 alternatives. That will take too many efforts on modeling and too much time on computation.

Generally speaking, mode choice models for different types of tours will be different. Therefore, we stratify tours by the type, i.e., work, shop and other, in the tour mode choice stage. If there are any work, school, or college type activities in the tour, then the tour type is classified as “work.” If there are no work, school, and college activities, but there is at least one shopping activity, the tour is a “shop” type tour. Otherwise, the tour type is “other.” The calibration of mode choice model will be implemented upon each type of tours respectively and these models will be used on each type of synthetic tours in TRANSIMS.

Since it is indicated that people’s perceptions of travel time and costs are similar for different transit modes, we will choose trip modes, in the trip mode choice stage, among bus, rail, trolley by the Route Planer’s capability of “rational” choice. In other words, the trip mode will be taken along the path with the smallest generalized cost, composed of time, cost, etc., among allowed trip modes by tour mode.

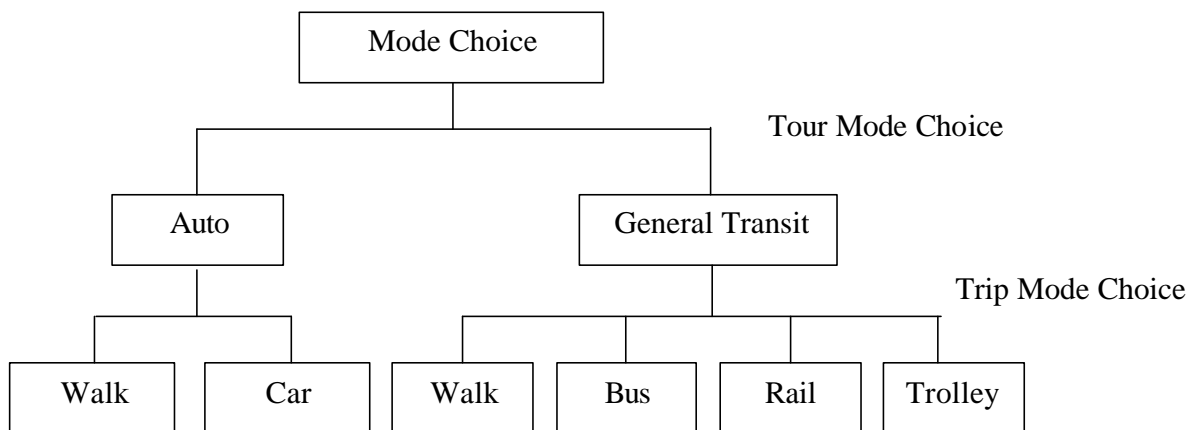


Figure 4.1 Tour and trip mode choice model nesting structure

4.2 Framework of Proposed Mode Choice Methodology

The framework of the proposed mode choice methodology in TRANSIMS is outlined in Figure 4.2. Basically, there are five steps:

1. Generate synthetic persons and vehicles, as well as activity list for every synthetic person from survey data by running Population Synthesizer Module and Activity Generator Module. Travel modes of synthetic activities obtained from survey activities are also shown in activity file. However, those modes will not be suitable any more since the activity locations are chosen randomly. Therefore, we will just discard the initial mode preference in the activity file.
2. Calibrate Tour Mode Choice Model on the survey data of the area under study. Different Models will be obtained from data of different types of tours. Binomial Logit will be used to model tour mode choice.

Decide the mode for every tour based on calibrated Binary Logit models, travel time obtained from Route Planner Module, travel cost from Cost Table and socioeconomic data from Population Synthesizer. In TRANSIMS version after 2.0, variable transit fare for different boarding and exiting places is supported.

We will explain tour mode choice process in detail in Sections 4.3 and 4.4.

3. Decide the mode for every trip among ones allowed by the tour mode. It is depends on the capability of Router Planner Module on choosing the path with the smallest generalized cost. The generalized cost is not supported by the latest TRANSIMS version. Only travel time is used to choose the path. However, the travel time, travel cost and other data can be incorporated in the future version.
4. Simulate all trips with the modes chosen in Traffic Microsimulator. The new travel time on each link is updated after the process. We will run multiple iterations between step 3 (Route Planner) and step 4 (Traffic Microsimulator) to stabilize the traffic, i.e., the new travel times are fed back to the Router Planner Module, which uses them to generate a new set of shortest path for travelers. In this way, the effects of congestion caused by other travelers could be taken into account. It is noted the tour mode choice between Auto and Transit for a traveler is not changed during this process.
5. Go back to step 2 to feed updated transportation network information back to mode choice process and collect new information for every tour. Then go all the way down to step 4. It is noted that we don't need to put all of the tours on Transit and collect information again since fixed schedules are used for Transit in the present TRANSIMS version. If there is no big difference on some indicators, such as mode split, network congestion, between consecutive iterations, terminate; otherwise, go back to step 2 again.

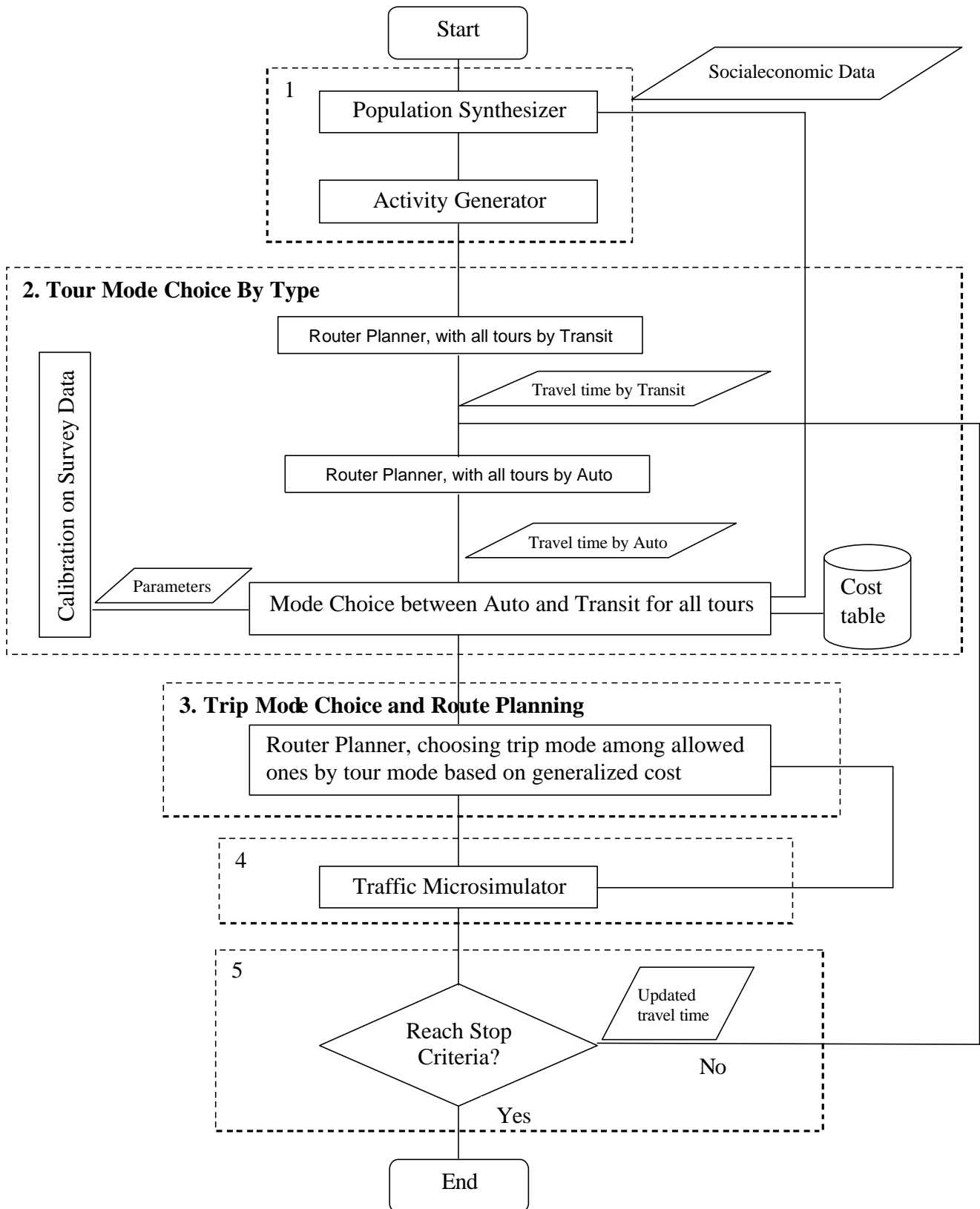


Figure 4.2 Framework of the proposed mode choice model

4.3 Calibration of Tour Mode Choice Model on Survey Data

4.3.1 Introduction of Binomial Logit Model

In the tour mode choice stage, a mode choice calibration process is implemented on the survey data of the area under study. If we could find out the decision pattern on mode choice from survey people and use the same one on the synthetic persons in TRANSIMS, the mode choice process can be thought as reasonable.

Since a Binary Logit Model will be used, we need to define Utility Functions for both modes, Auto and General Transit, in the following format:

$$U_n = \mathbf{b}_0 + \mathbf{b}_1 \times x_{1n} + \mathbf{b}_2 \times x_{2n} + \dots + \mathbf{b}_K \times x_{Kn} = \mathbf{B}'X_n, \quad (4.1)$$

Where $x_{1n}, x_{2n}, \dots, x_{Kn}$ represent the variables, which are significant in the model, for the n^{th} traveler. $\beta_0, \beta_1, \dots, \beta_K$ are coefficients. n represents the n^{th} traveler, and K is the number of variables.

The choice probability for traveler n choosing alternative i is given by

$$P_n(i) = \frac{e^{\mu U_{in}}}{e^{\mu U_{in}} + e^{\mu U_{jn}}} \quad (4.2)$$

In the case of linear-in-parameters utilities, parameter μ could not be distinguished from the overall scale of the β 's. For convenience we generally make an arbitrary assumption that $\mu = 1$.

Eq. 4.2 can also be written as:

$$P_n(i) = \frac{1}{1 + e^{-(U_{in} - U_{jn})}} = \frac{1}{1 + e^{-\mathbf{B}'(X_{in} - X_{jn})}} \quad (4.3)$$

From Eq. 4.3, we can conclude that **only** the difference between utilities of two alternatives matters to the choice decision in the Binomial Logit Model.

To find out the individual traveler's decision pattern on mode choice, we need to get estimators of these parameters. Then we can use these patterns on synthetic travelers in TRANSIMS to make a reasonable mode choice.

4.3.2 Introduction of Maximum Likelihood Estimation

Maximum likelihood estimation is probably the most general and straightforward procedure for finding estimators. Stated simply, a maximum likelihood estimator is the value of the parameters for which the observed sample is most likely to have occurred.

If, as is commonly the case, one can assume that the different individuals of the population act independently, the likelihood function is

$$L^*(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_K) = \prod_{n=1}^N \prod_{i=1}^I P_n(i) \quad (4.4)$$

Where β_1, \dots, β_K are constant parameters, N is the number of observations, and I is the number of alternatives for one observation. $P_n(i)$ is the probability of the n th observation being the i^{th} alternatives, and it's a function of parameters β_1, \dots, β_K . The maximum likelihood estimation method is to find the values of parameters, $\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \dots, \hat{\mathbf{b}}_K$, when the likelihood function reaches its maximum.

It is usually more convenient to find $\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \dots, \hat{\mathbf{b}}_K$ by maximizing the logarithm of the likelihood function, the log-likelihood function

$$L(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_K) = \sum_{n=1}^N \sum_{i=1}^I \log(P_n(i)) \quad (4.5)$$

because it is a sum (rather than a product) of small numbers and its maximum coincides with the maximum of Eq. (4.4).

We will then solve for the maximum of L by differentiating it with respect to each of the β 's and setting the partial derivatives equal to zero. Thus, we seek estimator $\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \dots, \hat{\mathbf{b}}_K$ that solve

$$\max L(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_K), \quad (4.7)$$

which, if they exist, must satisfy the necessary conditions that

$$\frac{\partial L}{\partial \mathbf{b}_k} = 0, \text{ for } k = 1, \dots, K. \quad (4.8)$$

4.3.3 Computational Aspects of Maximum Likelihood Estimation

Numerical techniques must generally be used to find the maximum solution of Eq. (4.5).

The expression of $P_n(i)$ in Eq. (4.5) is typically nonlinear, so the solution of the first-order conditions can prove to be a significant computational problem. But if the second derivative can be computed without great difficulty and the likelihood function is

globally concave, the Newton-Raphson method is simple to implement and computationally efficient.

The Newton-Raphson method is operated in the following steps:

STEP 0: Choose $\hat{\mathbf{b}}_0 = [\mathbf{b}_{01}, \mathbf{b}_{02}, \dots, \mathbf{b}_{0K}]'$ as an initial guess for the parameters. ($\hat{\mathbf{b}}_0 = 0$ is often convenient unless some better guess is available.)

STEP 1: Linearize the function $\nabla L(\mathbf{b})$ around $\hat{\mathbf{b}}_w$. The approximate first-order conditions are given by

$$\nabla L(\hat{\mathbf{b}}_w) + \nabla^2 L(\hat{\mathbf{b}}_w)(\hat{\mathbf{b}} - \hat{\mathbf{b}}_w) = 0 \quad (4.9)$$

Solve the linearized form for

$$\hat{\mathbf{b}}_{w+1} = \hat{\mathbf{b}}_w - [\nabla^2 L(\hat{\mathbf{b}}_w)]^{-1} \cdot \nabla L(\hat{\mathbf{b}}_w) \quad (4.10)$$

STEP 3: Check if $\hat{\mathbf{b}}_{w+1} - \hat{\mathbf{b}}_w$ is “small”. Typical criteria are

$$\left[\frac{1}{K} \sum_{k=1}^K (\hat{\mathbf{b}}_{w+1,k} - \hat{\mathbf{b}}_{wk})^2 \right]^{1/2} < e_1 \quad (4.11)$$

If these conditions are satisfied, terminate with $\hat{\mathbf{b}}_{w+1}$ as the solution. Otherwise, set $w = w + 1$, and go to STEP 1.

4.3.4 Maximum Likelihood Estimation of Binomial Logit Model

Consider the likelihood of any sample of N travelers. As they are drawn randomly from the whole population, the likelihood of the whole sample is the productivity of the likelihood of every person. We define L^* as the likelihood function:

$$L^*(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_K) = \prod_{n=1}^N P_n(i)^{y_{in}} P_n(j)^{y_{jn}} \quad (4.4)$$

Where y_{in} and y_{jn} are indicator variables. If person n chooses alternative i , y_{in} equals to 1 and y_{jn} equals to 0, and visa versa. In other words, the sum of y_{in} and y_{jn} always equals to 1.

The logarithm of L^* , denoted as L can be written as follows:

$$L(\mathbf{b}_1, \dots, \mathbf{b}_K) = \sum_{n=1}^N [y_{in} \log P_n(i) + y_{jn} \log P_n(j)] \quad (4.5)$$

Noting that $y_{jn} = 1 - y_{in}$ and $P_n(j) = 1 - P_n(i)$, function 4.5 can also be written in:

$$\begin{aligned} L(\mathbf{b}_1, \dots, \mathbf{b}_K) &= \sum_{n=1}^N \{y_{in} \log P_n(i) + [(1 - y_{in}) \log(1 - P_n(i))]\} \\ &= \sum_{n=1}^N \left\{ y_{in} \log \left(\frac{e^{\mathbf{b}'X_{in}}}{e^{\mathbf{b}'X_{in}} + e^{\mathbf{b}'X_{jn}}} \right) + y_{jn} \log \left(\frac{e^{\mathbf{b}'X_{jn}}}{e^{\mathbf{b}'X_{jn}} + e^{\mathbf{b}'X_{in}}} \right) \right\} \end{aligned} \quad (4.6)$$

We will then solve for the maximum of L by differentiating it with respect to each of the β 's and setting the partial derivatives equal to zero. Thus, we seek estimator $\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \dots, \hat{\mathbf{b}}_K$ that solve

$$\max L(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_K), \quad (4.7)$$

which, if they exist, must satisfy the necessary conditions that

$$\frac{\partial L}{\partial \mathbf{b}_k} = \sum_{n=1}^N \left\{ y_{in} \frac{\partial P_n(i) / \partial \hat{\mathbf{b}}_k}{P_n(i)} + y_{jn} \frac{\partial P_n(j) / \partial \hat{\mathbf{b}}_k}{P_n(j)} \right\} = 0, \text{ for } k = 1, \dots, K. \quad (4.8)$$

In many cases of practical interest it is shown that likelihood function is globally concave, so that if a solution to the first-order conditions exists, it is unique.

For the sake of notational simplicity we define

$$X_n = X_{in} - X_{jn} \quad (4.13)$$

so that each element of X_n is defined as

$$x_n = x_{in} - x_{jn}, \text{ for } k = 1, \dots, K \quad (4.14)$$

In this notation,

$$P_n(i) = \frac{1}{1 + e^{-\mathbf{b}'X_n}} \quad (4.15)$$

and

$$P_n(j) = \frac{e^{-\mathbf{b}'X_n}}{1 + e^{-\mathbf{b}'X_n}} \quad (4.16)$$

We also note that

$$\frac{\partial \log P_n(i)}{\partial (\mathbf{b}' X_n)} = \frac{e^{-\mathbf{b}' X_n}}{(1 + e^{-\mathbf{b}' X_n})^2} \cdot (1 + e^{-\mathbf{b}' X_n}) = P_n(j) \quad (4.17)$$

and

$$\frac{\partial \log P_n(j)}{\partial (\mathbf{b}' X_n)} = -\frac{1 + e^{-\mathbf{b}' X_n}}{e^{-\mathbf{b}' X_n}} \cdot \frac{e^{-\mathbf{b}' X_n}}{(1 + e^{-\mathbf{b}' X_n})^2} = -P_n(i) \quad (4.18)$$

So

$$\begin{aligned} \frac{\partial L(\mathbf{b})}{\partial \mathbf{b}_k} &= \sum [y_{in} P_n(j) x_{nk} - y_{jn} P_n(i) x_{nk}] \\ &= \sum_{n=1}^N [y_{in} (1 - P_n(i)) - (1 - y_{in}) P_n(i)] x_{nk} \\ &= \sum_{n=1}^N [y_{in} - P_n(i)] x_{nk} \end{aligned} \quad (4.19)$$

Thus, to find the maximum likelihood estimate, we need to solve K nonlinear equations of the form

$$\sum_{n=1}^N [y_{in} - P_n(i)] x_{nk} = 0, \quad k = 1, \dots, K \quad (4.20)$$

The second derivatives can also be solved as follows:

$$\frac{\partial^2 L}{\partial \mathbf{b}_k \partial \mathbf{b}_l} = -\sum P_n(i) (1 - P_n(i)) x_{nk} x_{nl} < 0 \quad (4.21)$$

Since the second derivatives are negative for all value of β , the log likelihood is globally concave.

Apply Eq. (4.19) and Eq. (4.21) in Eq. (4.10), which is the second step in Newton-Raphson method, we can get the estimates of parameters in the Binary Logit Model.

4.4 Tour Mode Choice on Synthetic Data in TRANSIMS

Force all tours take General Transit and choose the shortest paths for them in Route Planner Module. Collect the information that is required in the disaggregate mode choice model calibrated from survey data, such as travel time and travel cost etc., for every tour.

Execute the same process for Auto. Put all of the tours on Auto network, i.e., force all tours take Auto and choose the shortest path for then in Router Planner Module. Collect the same kind of information as for General Transit.

There are always some General Transit tours that cannot be routed since General Transit is not accessible everywhere. Mode Auto will be chosen for these tours. For others, calculate the probability to choose Auto mode for a tour P_A using information obtained above, based on the calibrated mode choice model as Eq. 4.1 and 4.2. A uniformly distributed random number a is generated. If the probability P_A is greater than the random number a , the Auto mode is chosen for the tour. Otherwise, the General Transit Mode will be used.

The calibrated tour mode choice model can be applied on synthetic data in base year for model validation as well as in future year for forecasting.

When applying this methodology for mode choice forecasting, we need to assume that the travelers' behaviors in forecast year will remain the same as in the base year. In such a case, the calibrated mode choice model obtained from survey data will still hold.

Chapter 5 Model Application and Evaluation

5.1 Application of the Methodology in Blacksburg Study

5.1.1 Background of the research

Los Alamos has done major studies of Albuquerque and Dallas transportation and completed a micro simulation of auto traffic patterns in 25 square miles of Dallas that represented about 200,000 vehicles over a five-hour period. A TRANSIMS model of Portland transportation is also done in 2000. The Portland study focuses on forecasting activity demand and predicting trips that use multiple modes of transportation. The goals of these studies are to develop technologies that can be used by transportation planners in any urban environment.

To support future TRANSIMS marketing and training requirements, Virginia Tech implemented the transportation planning on Blacksburg in TRANSIMS under the contract with Federal Highway Administration. The synthetic population was generated in TRANSIM at the same scale as the real one. However, a simplified transportation network in Blacksburg was used in the study (Figure 5.1). It is named Blacksburg-Lite. The research on mode choice methodology in the thesis was based on the study.

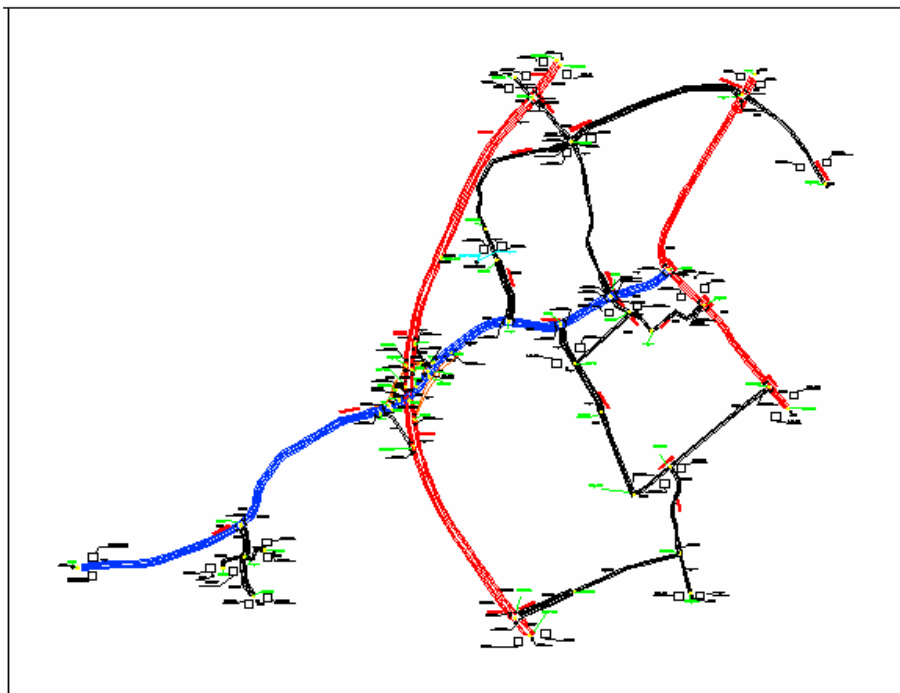


Figure 5.1 Simplified transportation network in Blacksburg-Lite Study

Blacksburg is the largest town in Virginia with a population of approximately 39700 citizens living on just over 12,000 acres of land at the foot of the Washington-Jefferson

National Forest. It is located in Southwestern Virginia, 42 miles southwest of Roanoke served by U.S. 460, Route 11, and Interstate 81 (Figure 5.2). Virginia Tech, the Town's major employer and university, has exceeded its goal of 25,000 students. Some of the students live on-campus, and the others live off-campus in apartments, townhouses and houses. Most of the off-campus students live in apartments within 15 minutes to Campus by car.

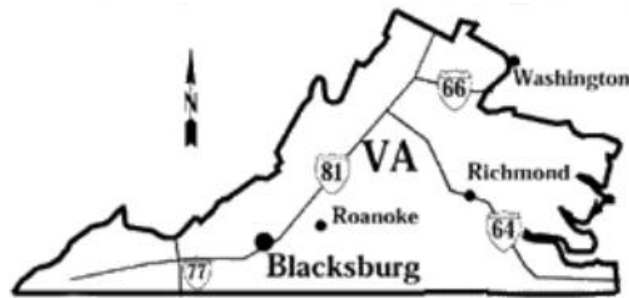


Figure 5.2 Location of the town of Blacksburg

An online one-day activity survey in Blacksburg was undertaken. Data from 124 persons in 62 households were used in the research. Since the survey was implemented before the start of the research, it was lack of some detailed information required by the study. For example, we need in-vehicle travel time and out-of-vehicle travel time while we only have travel time as a whole in the survey. Therefore, we added in some detailed information based on our estimates. Although those data might be examined as inadequate and inaccurate, they were qualified for the purpose of research on methodology.

The available transportation modes in Blacksburg are Car and Blacksburg Transit. We will use bus to represent Blacksburg Transit for simplicity. Bus can be accessed in most of the apartments, campus area and commercial areas in Blacksburg. There were 4 bus routes coded in the research.

As we discussed in Chapter Four, there are two stages in the proposed mode choice methodology, tour mode choice in Stage One and trip mode choice in Stage Two. Tour mode choice is made between Auto and General Transit; whereas trip mode choice needs to be done between walk and car if Auto is the tour mode, and between walk and bus if General Transit is the tour mode.

5.1.2 Specification of Tour Mode Choice Model

Before we calibrate the Binomial Logit Model for tour mode choice based on the survey data, we need to specify the dependent variables in the Utility Function by logic analysis.

Travel time of a tour is one of the most significant factors influencing traveler's mode choice on this tour. In our research, travel time has been divided into two parts, in-vehicle time and out-of-vehicle time. This division reflects the observation that travelers prefer to spend one minute on traveling in a vehicle than a minute waiting for or walking to one. For the transit tour, out-of-vehicle time includes walking to a bus stop, waiting for the vehicle, transferring between bus lines and walking from a bus stop to the destination. Auto out-of-vehicle travel time only includes time walking to and from a vehicle. The out-of-vehicle time in a tour represents the accessibility of the person to a mode.

Apparently, travel cost is another significant factor in the individual's mode choice process. The transit fare is the travel cost for a transit tour. The travel cost for an auto tour mainly includes the parking cost and the gas cost. Note that the coefficients of time and cost are the same in the two utilities. This represents an assumption that a unit of time or cost has the same marginal utility whether it is incurred on the auto or transit mode.

Socioeconomic demand variables are also used to explain mode choice in the study. Since disaggregate data are available at the individual level in TRANSIMS, it is preferable to consider the mode choice behavior of each individual separately, and hence the variables describing the household as a whole might be less relevant. Instead, income per capita is used as a variable in the utility function in the model. It is well known people with lower income are more likely to take transit. Because the difference between alternatives really matters, we'd like this variable appearing in the function of difference of the utilities. Therefore, we defined this variable in only one alternative.

We also included an alternative-specific constant in our model. The alternative-specific constant reflects the difference of the utilities between the two alternatives when all else variables are equal. It should be noted it makes no sense to have one alternative-specific constant for each, since all the matter is their difference. Therefore, 1 was used in the auto alternative and 0 in the transit alternative for constant.

The specification of the Utility functions is shown in table 5.1.

Table 5.1 Specification of the Utility Function

	β_1	β_2	β_3	β_4	β_5
Auto Utility, $U_{A,n}$	1	Auto in-vehicle time (min)	Auto out-of-vehicle time (min)	Auto out-of-pocket cost (\$)	Income per capita (10k\$)
Transit Utility, $U_{GT,n}$	0	Transit in-vehicle time (min)	Transit out-of-vehicle time (min)	Transit fare (\$)	0
Estimated coefficients	?	?	?	?	?

The Utility Functions for Auto and General Transit in the research can be written in the following format:

$$U_{A,n} = \mathbf{b}_1 + \mathbf{b}_2 \cdot inVehicleTime + \mathbf{b}_3 \cdot outVehicleTime + \mathbf{b}_4 \cdot travelCost + \mathbf{b}_5 incomePerPerson \quad (5.1)$$

$$U_{GT,n} = \mathbf{b}_2 \cdot inVehicleTime + \mathbf{b}_3 \cdot outVehicleTime + \mathbf{b}_4 \cdot travelCost \quad (5.2)$$

where $U_{A,n}$, $U_{GT,n}$ are utilities for Auto and General Transit, respectively. $\beta_1, \beta_2, \beta_3, \beta_4$ and β_5 are coefficients.

As we explained in Chapter 4, only the difference between utilities of two alternatives is significant on the choice decision in the Binomial Logit Model. The difference of utilities is obtained from Eq. 5.1 and 5.2:

$$\Delta U_n = \mathbf{b}_1 + \mathbf{b}_2 \cdot \Delta inVehicleTime + \mathbf{b}_3 \cdot \Delta outVehicleTime + \mathbf{b}_4 \cdot \Delta travelCost + \mathbf{b}_5 incomePerPerson \quad (5.3)$$

Then the probability to choose Auto can be expressed as follows:

$$P_{A,n} = \frac{1}{1 + e^{-\Delta U_n}} \quad (5.4)$$

From Eq. 5.4, we can find that the bigger the difference between utilities, the higher the probability to choose Mode Auto.

5.1.3 Calibration of Tour Mode Choice Model on Survey Data

After we got the specification of the Binomial Logit Model by logic analysis, we used the Maximum Likelihood Estimation to get the estimators of parameters. The calibration process is explained in this session.

Part of the Blacksburg survey data used in the research is shown in Table 5.2.

Table 5.2 Survey Data

Tour		Constant	In-vehicle time (min)	Out-of-vehicle time (min)	Cost (\$)	Income/person (10k\$)	Mode Chosen
1	Auto	1	14	18	1.9	1.5	X
	Transit	0	24	14	1	0	
2	Auto	1	5	20	1	1.5	
	Transit	0	6	5	1	0	X
3	Auto	1	10	6	1.5	1.5	X
	Transit	0	24	8	1	0	
4	Auto	1	13	4	1.8	1	X
	Transit	0	24	15	1	0	

5	Auto	1	14	10	1.9	1	X
	Transit	0	24	12	1	0	
6	Auto	1	20	20	2.5	1	
	Transit	0	26	8	1	0	X
7	Auto	1	15	8	2	2	X
	Transit	0	24	12	1	0	
8	Auto	1	12	25	1.7	2	
	Transit	0	16	6	1	0	X
...	
...	

Based on these survey data, we obtained the estimators of parameters using Maximum Likelihood Estimation technology and Newton-Raphson method as explained in Chapter 4. A Matlab file was used to implement this process. It took just several seconds to get the results. The estimators of parameters are listed in Table 5.3.

Table 5.3 Estimated Coefficients

	β_1	β_2	β_3	β_4	β_5
Auto Utility, $U_{A,n}$	1	Auto in-vehicle time (min)	Auto out-of-vehicle time (min)	Auto out-of-pocket cost (\$)	Income per capita (10k\$)
Transit Utility, $U_{GT,n}$	0	Transit in-vehicle time (min)	Transit out-of-vehicle time (min)	Transit fare (\$)	0
Estimated coefficients	0.5127	-0.0260	-0.1346	-0.7374	0.3268

We used the calibrated model to verify on the survey data themselves. The probability of a tour to choose Auto was calculated as $P_{A,n}$ in table 5.4. If the probability greater than 0.5, Auto should be chosen; otherwise, transit should be chosen. The resulting modes obtained from the model were coincident with the modes used in the survey.

Table 5.4 Verification of the calibrated model on survey data

Tour	$P_{A,n}$	Auto chosen?	Tour	$P_{A,n}$	Auto chosen?
1	0.5152	Y	5	0.6693	Y
2	0.2709	N	6	0.1511	N
3	0.7803	Y	7	0.7688	Y
4	0.8825	Y	8	0.1415	N

Therefore, the Utility Functions for both tour modes can be written as Eq. 5.5 and 5.6.

$$U_{A,n} = 0.5127 - 0.0260 \cdot inVehicleTime - 0.1346 \cdot outVehicleTime - 0.7374 \cdot travelCost + 0.3268 \cdot incomePerPerson \quad (5.5)$$

$$U_{GT,n} = -0.0260 \cdot inVehicleTime - 0.1346 \cdot outVehicleTime - 0.7374 \cdot travelCost \quad (5.6)$$

The difference of the Utilities is in the following format:

$$\Delta U_n = 0.5127 - 0.0260 \cdot \Delta inVehicleTime - 0.1346 \cdot \Delta outVehicleTime - 0.7374 \cdot \Delta travelCost + 0.3268 \cdot incomePerPerson \quad (5.7)$$

As expected, the coefficients of both time variables are negative. It means people are likely to choose the mode with shorter travel time, either in vehicle or out of vehicle. Besides, the coefficient of out-of-vehicle time is much bigger than that of in-vehicle time. It complies with the notion that out-of-vehicle time is more significant than in-vehicle time on travelers' mode choices.

The coefficient of travel cost is also negative. It means people are likely to choose the mode with cheaper cost.

The coefficient of income per person is positive. It also explains that richer people are more likely to travel in their own car.

From the above analysis, we would conclude that the model is reasonable on individual tour mode choice.

5.1.4 Implementation of Calibrated Model in Blacksburg Study

We implemented the proposed methodology in TRANSIMS using base year data for model validation after we got it from calibration process.

STEP 0: Preparation on synthetic persons and activities

Run Population Synthesizer and Activity Generator Module to get the synthetic travelers, vehicles and their activity lists. Blacksburg STF-3a, PUMS data and activity survey data were used in the process. The initial activities were stored in the file <activities.0>. We are of special interest on activity files because a mode, which is our research focus, is defined for each activity in them.

Any tour involving shared rides is completely untouched in mode choice process because of the complexity in arranging the shared rides. Therefore, we divided initial activities in <activities.0> into two groups: all of the activities in tours involving shared rides and others. They were stored in <activities.share> and <activities.alone>. <activities.alone> would be used for mode choice analysis in the following steps.



Figure 5.3 Group activities by shared rides

Generate a tour database <tour> from <activity.alone>, from which the type of each tour among “Work”, “Shop” and “Other” is decided and put into <tour>. The activity locations of the tour are also listed in <tour> for the convenience of following procedure.

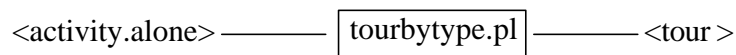


Figure 5.4 Generate tour database with tour type

STEP 1: Collection of the values of required variables

Step 1.1: In- vehicle time and out-of-vehicle time

We will try to get travel time for all tours with General Transit Mode first.

Change the modes of all activities in <activities.alone> to General Transit. Put changing mode command (M) in the activity feedback file named <activities.feedback>. Then send the feedback file to Activity Regenerator, which is a part of Activity Generator. The activity lists in <activities.feedback> were obtained from <activities.alone>.

<activities.feedback> was in the following format:

```

<activities.feedback>
1 1 M 3
1 2 M 3
1 3 M 3
...
  
```

Where the first column is household ID, and the second column is activity ID. M in the third column means changing mode and 3 in the fourth column represents the mode walk-transit-walk as defined in the input file “modes”.

Run Route Planner and get the trip plan file named <plans.transit>. It was in the following format:

```

<plans.transit>
101 0 2 2
19776 3101 3 3102 3
1234 19776 1 0 0
0 [1]
  
```

1
1
...

where the framed 1 represents the mode transit. Change the plan file into a more machine-readable format and give it a new name `<planitdb.transit>`. A perl script “changeplan.pl” was implemented in the process. `<planitdb.transit>` was in the following format:

TRAVID	TRIPID	LEGID	DURATION	COST	DRIVER_FLAG	MODE	FROMLOCID	TOLOCID
101	2	1	7527	1	0	2	1071	3101
101	2	2	1234	1	0	1	3101	3102
101	2	3	2644	1	0	2	3102	1008
102	2	1	6024	1	0	2	1067	3101
102	2	2	1041	1	0	1	3101	3102
102	2	3	6403	1	0	2	3102	1018

...

The above process is shown in Figure 5.5.

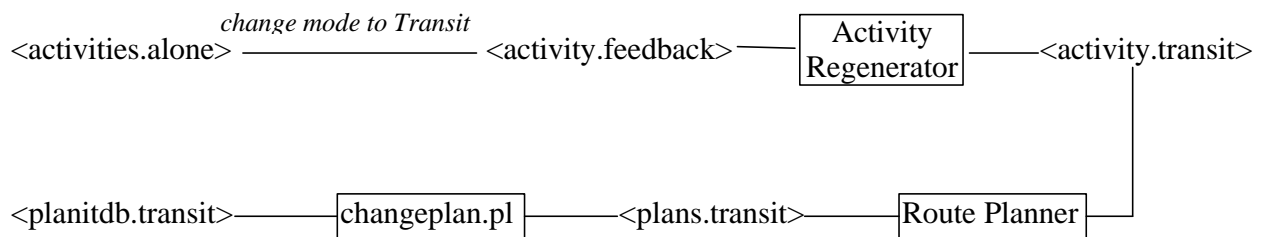


Figure 5.5 Convert all activities to general transit mode and get travel plans

Since transit is not always accessible in everywhere, it is very possible that Route Planner cannot find paths for some tours to go back home within the time windows specified in activity file. These tours are called incomplete tours. The tour database `<tour>` is thus divided into two files: one with incomplete tours and the other with complete tours. Their names are `<incomplete.tour.transit>` and `<complete.tour.transit>`.

For complete tours, collect in-vehicle travel time and out-of-vehicle time from the format-modified plan file `<planitdb.transit>`. The durations of all trips with mode transit in a tour are summed up for the in-vehicle travel time for this tour. The durations of all trips with mode walk in a tour are also summed up. The sum of walk time and the waiting time plus transferring time comprise of the out-of-vehicle time for transit tour. Since the waiting time and transferring time are very difficult to collect from TRANSIMS and they are closely related to frequencies of transit lines, we just use the one-third of the headway of Blacksburg Transit as the expected waiting and transferring time in the study.

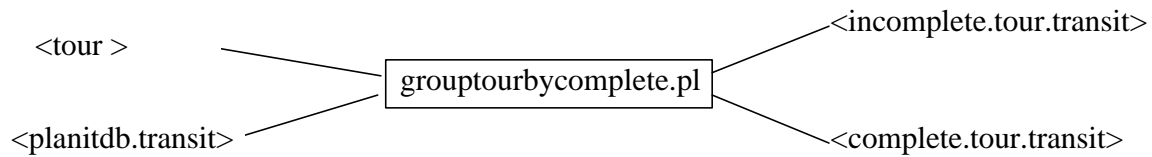


Figure 5.6 Divide them by completeness and collect travel time for complete tours

The same method is used to collect in-vehicle time and out-of-vehicle time for Auto tours as in Figure 5.7.

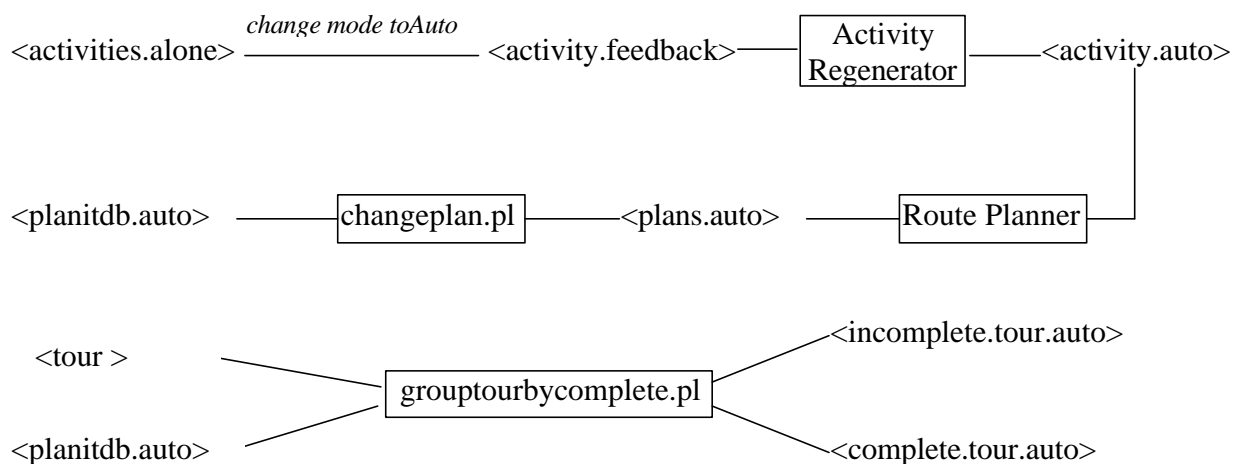


Figure 5.7 Collect in-vehicle time and out-of-vehicle time for Auto tours

STEP 1.2: Travel cost

Travel fare for transit is \$0.50 per trip in Blacksburg. No transferring fee occurs between different bus routes. Travel cost for Auto is obtained by multiplying in-vehicle time by \$0.10. Parking cost for downtown area is \$0.5 in Blacksburg. Parking costs for other areas are just free. If a tour has at least an activity happening in downtown area, \$0.5 parking fee is added to the cost.

Step 1.3 Household Income Per Person

Both household income and the number of persons in a household can be obtained in <pop.located.NISS>, which is the output file of Population Synthesizer Module.

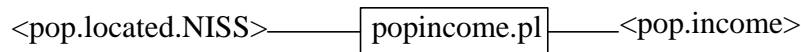


Figure 5.8 Obtain household income per person

STEP 2: Making Tour Mode Choice

Calculate the probability to choose mode Auto by using in-vehicle time, out-of-vehicle time, travel cost and income, based on the calibrated mode choice model as in Eq. 5.1 and Eq. 4.3 for each complete work tour. Decide the modes by comparing the probability with a random number. The mode chosen was added in activity feedback file <activity.feedback.complete>. This process is shown in Figure 5.9.

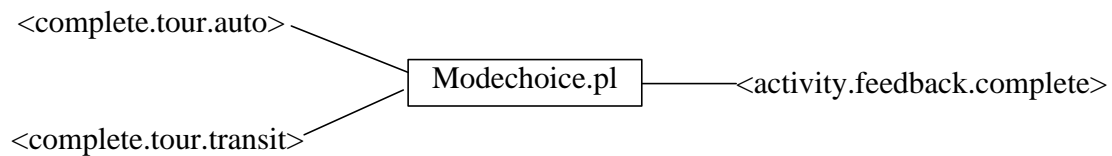


Figure 5.9 Generate feedback file for complete tours

Generate activity feedback <activity.feedback.incomplete> for all of the activities in incomplete tours as shown in Figure 5.10.

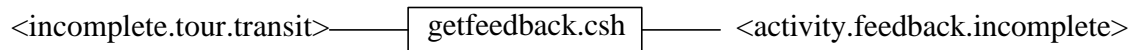


Figure 5.10 Generate feedback file for incomplete tours

Concatenate <activity.feedback.complete> and <activity.feedback.incomplete> into one file and sort activities by household number and activity number. The new file is named <activity.feedback> and sent back to Activity Regenerator. A new activity file with the chosen modes will be obtained at the end of Step 2.

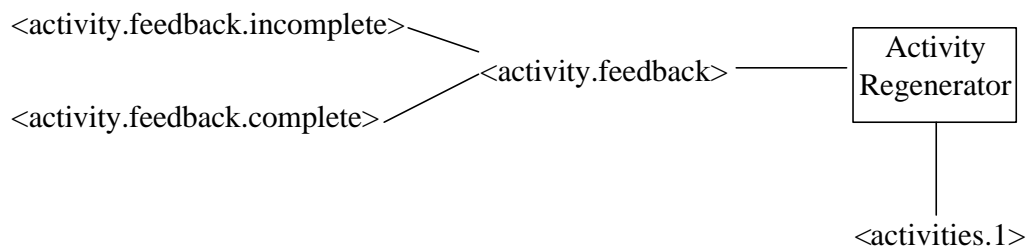


Figure 5.11 Generate new activity file with chosen tour modes

STEP 3: Making Trip Mode Choice and Routing

Run Route Planner with the most updated activity file as input. The trip modes were decided within the limitation of the tour modes along with the shortest paths between activities. In our study, bus and walk are allowed trip modes for tour mode General Transit, and car and walk are allowed trip modes for tour mode Auto. A plan file with specific starting time, ending time and duration is the output of Route Planner.

STEP 4: Simulating the Traffic

Simulate the movements of every traveler and every vehicle seconds by seconds in Traffic Microsimulator according to their travel plans from Route Planner. Since Route Planner does not take the interaction between travelers on congestion into account, we'd better to run several loops between Route Planner and Traffic Microsimulator to get stabilized traffic. Route Planner Module will use the updated travel time obtained from simulation by Traffic Microsimulator Module automatically since the configuration key ROUTER_LINK_DELAY_FILE is defined.

STEP 5: Iterations

After running Traffic Microsimulator Module, the travel time on every link, also probably the travel cost, in the network will change compared to the empty network. Go back to STEP 1 to collect required information and made tour mode choice and trip mode choice. Simulate the traffic with the new mode and the new travel plan. Repeat the process until there is no big difference between results of consecutive iterations. By now, the whole transportation planning process with mode choice methodology is over.

5.2 Evaluation of the Methodology in Blacksburg Study

The results of our research are shown in Table 5.5, Table 5.6 and Table 5.7. Totally there are 79374 activities in a typical day in Blacksburg, in which 63225 activities happened in the alone tours and 16149 activities happened in shared tours. Within the alone activities, 12208 activities are belonging to type "Work", whereas others are type "Shop" and "Other". Any activities in shared tours stay untouched in the mode choice process because of their complexity. For alone rides, we only made tour mode choice on "Work" tours between Auto and Transit to test the methodology. We presumed all of the alone "Shop" and "Other" tours would take auto for simplicity. It's thought to be close to the real situation in Blacksburg.

The result for Iteration Zero was taken after running Population Synthesizer Module and Activity Generator Module. By that time, the mode choice process was not implemented yet. Activity Generator transferred the mode preferences in the activity file from survey data.

The result for Iteration One was taken after execution of mode choice process, but still before Route Planner and Traffic Microsimulator. The mode for each activity is decided based on the calibrated Binomial Logit Model in the mode choice process.

After that, the inner iterations of Route Planner and Traffic Microsimulator were implemented to route, simulate trips and stabilize traffic. Then, we fed the updated travel time and travel cost back and made tour mode choice again. The result was recorded for Iteration Two. The result for Iteration Three was obtained after another outer iteration. The positions to draw results in the study are shown in Figure 5.12.

Table 5.5 Activities in a typical day in Blacksburg

Alone activities	63225
Shared activities	16149
Total	79374

Table 5.6 Breakdown between “ Work” and “Shop & Other” within Alone activities

Alone “ Work” activities	12208
Alone “Shop& Other” activities	51017
Total Alone activities	63225

Table 5.7 Results of Mode Split between Transit and Auto for Alone “Work” activities

Iteration	0	1	2	3
Alone “Work” activities by Transit	6694	6726	6706	6706
Alone “Work” activities by Auto	3798	3764	3789	3789
Mode Split (T/A)	1.763	1.787	1.77	1.77

From Table 5.7, we can find the Transit Work activity to Auto Work activity ratio remained the same between Iteration two and Iteration three. The ratio converged very fast. It only took about 20 minutes to reach the convergence for Blacksburg_Lite study even though the micro simulation was supposed to be time-consuming. The final mode split between Transit and Auto is 1.770, which is recognized as reasonable.

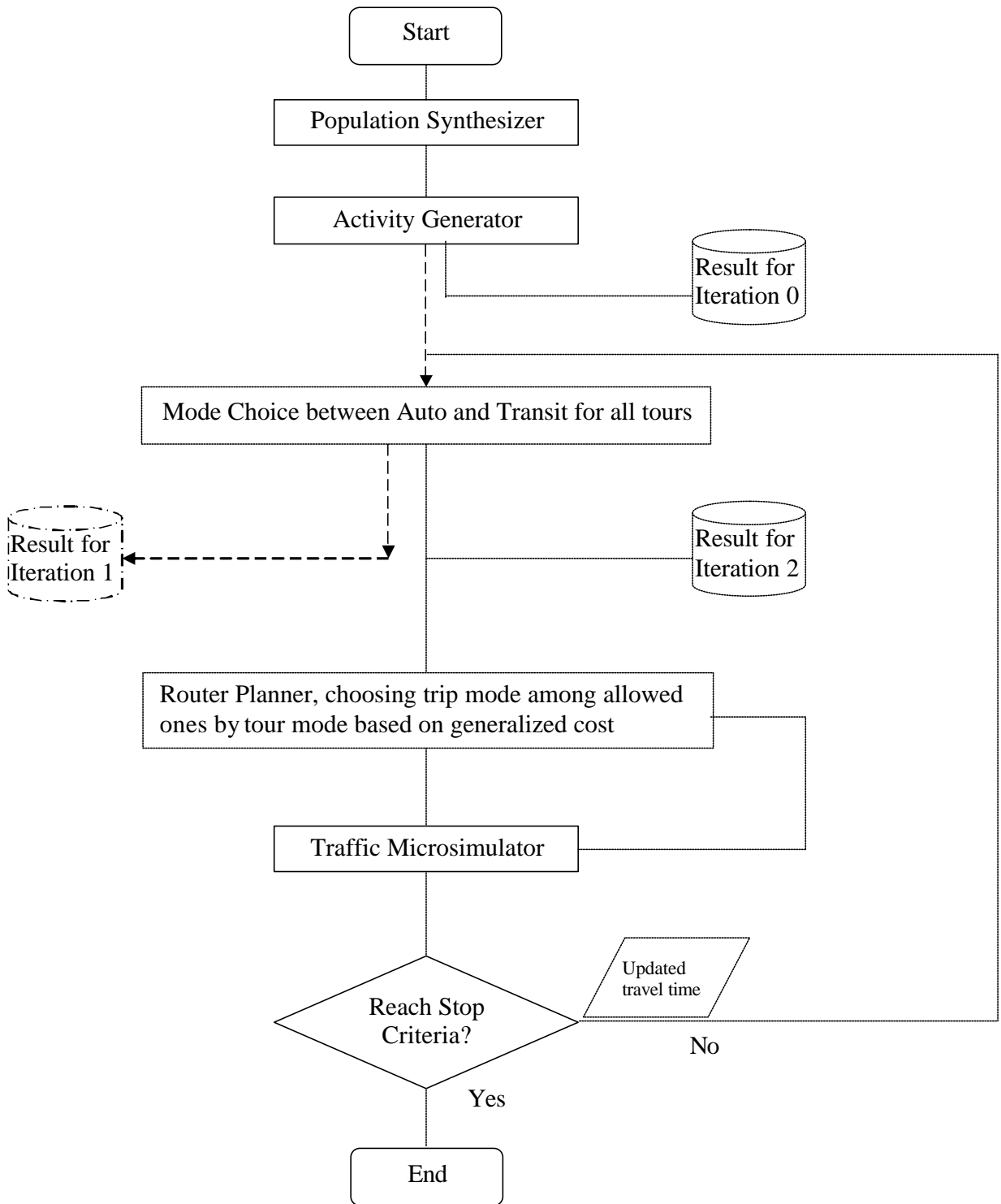


Figure 5.12 Positions to withdraw results for iterations

Chapter 6 Conclusions

6.1 Conclusions

This research provides a disaggregate, stochastic methodology on mode choice with the accessibility to a mode modeled in TRANSIMS. The previous chapters have outlined the theoretic foundation of the proposed mode choice methodology in TRANSIMS and the application of the methodology in Blacksburg_Lite Study. The application's result is then evaluated in Chapter 5. In general, the following conclusions could be drawn from the research:

- 1) The proposed methodology in the research is significant. The mode choice behavior is based on each traveler's own characteristics. It's a disaggregate model so that it provides higher accuracy, data efficiency and geographic transferability. Binary Logit Model used in this methodology is suitable for the stochastic process on mode choice. The accessibility to a mode is also involved in modeling by the variable out-of-vehicle time in the methodology.
- 2) In the proposed methodology, mode choice process is executed after generating activities and before routing and simulating trips. It complies with the procedure of the traditional Four-Step Model, which transportation planners are very comfortable with.
- 3) The result is reasonable. The final mode split between Transit and Auto is close to that in the real life.
- 4) The running time is endurable. The iterations converged fast. It only took about 20 minutes for the complete process for Blacksburg_Lite Study.

6.2 Recommendations for Further Study

The research is mainly concentrated on the development of the methodology. We used some approximation for some variables in it, such as waiting time and transferring time as parts of out-of-vehicle time for each Transit traveler, and travel cost for both Transit and Auto modes. Further research could be conducted on getting more detailed and precise data for those variables from the outputs of TRANSIMS modules.

In order to have a better understanding of the significance of the proposed methodology, it is recommended that more data be collected for Blacksburg network and tour mode choice be implemented on not only "Work" tours but also "Shop" and "Other" tours. In that situation, the accuracy of the result and efficiency on computing time could provide a better test on the methodology.

Another point of interest is the geographic transferability of the methodology. Model transferability is always a concern for transportation planners. Some research indicated that disaggregate models should have good transferability. But no test has ever been implemented on that. If it could be proven that the methodology has good geographic transferability, a great amount of time and capital efforts on transportation planning could be saved. It can also help the acceptance of TRANSIMS by transportation planners.

References

1. Bates, J. (2000). History of Demand Modelling, in “Handbook of Transport Modelling (Hensher, A. and Button, K. ed.), pp. 11-34. 2000 Elsevier Science Ltd., New York.
2. Beesley, M.E. (1965). The value of time spent in travelling: Some new evidence, *Economica* 32, pp. 174-185
3. Beilner, H. and Jacobs, F. (1972). Probabilistic aspects of traffic assignment, Proc. 5th Intern. Symp. On the theory of Traffic Flow and Transportation, Berkeley, 1971 pp. 183-194
4. Ben-Akiva, M., Lerman, S. (1975). “Discrete Choice Analysis”, The MIT Press, London
5. Berkowitz, M. K., Gallini, N. T., Miller, E. J., and Wolfe, R. A. (1985). Disaggregate modelling of Canadian household vehicle ownership, trip-making and fuel usage. Final report, prepared for Transportation Energy Division, Energy, Mines and Resources Canada. Joint Program in Transportation, University of Toronto/York University, Toronto, Ont.
6. Coyle, J., Bardi, E., Cavinato, J. (1990). “Transportation”, West Publishing Company, MN
7. Daganzo, C. (1979). Multinomial Probit, Academic Express, New York.
8. Daly, A. 1982. Applicability of disaggregate models of behaviour: a question of methodology. *Transportation Research—A*, 16A(5-6): 363-370.
9. Ghoneim, N. S., and Sargious, M. (1987). Review of advantages and applications of disaggregate mode choice models in intercity passenger travel. *Canadian Journal of Civil Engineering*. V.14 n6 Dec 1987 pp. 763-770.
10. Hensher, D., Stopher, P. (1977). “Behavioral Travel Modelling”, Croom Helm London
11. Institute of Transportation Engineers, (1992). “Transportation Planning Handbook”, P T R Prentice-Hall, Inc., NJ
12. Kanafani, A. (1983). “Transportation Demand Analysis”, McGraw-Hill, Inc.
13. Koppelman, F. S., and Rose, G. (1983). Geographic transfer of travel choice models: evaluation and procedures. The transportation Center, Northwestern University, Evanston, IL.

14. McCarthy, G. M. (1969). Multiple regression analysis of household trip generation—A critique, *Highw. Res. Rec.* 297, pp. 31-43.
15. McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior, in “Frontiers of Econometrics” (P. Zarembka, ed.), pp. 105-142. Academic Press, New York.
16. McFadden, D. (1974). The measurement of urban travel demand, *J. Public Econ.* 3, pp. 303-328.
17. Miller, E. J. (1985). The data base for intercity passenger travel analysis: current state and future prospects. *Canadian Transportation Research Forum. Proceedings of the 20th Annual Meeting.* Toronto, Ont., pp. 138-154.
18. Ortuzar, J., Willumsen L. (1990). “Modeling Transport”, John Wisley & Sons.
19. Parody, T. E. (1977). Analysis of Predictive qualities of disaggregate modal-choice models. *Transportation Research Record* 637, Transportation Research Board, Washington, DC, pp. 51-57.
20. Quarmby, D. A. (1967). Choice of travel mode for the journey to work: Some findings, *J. Transp. Econ. Policy* 1, pp. 273-314.
21. Rassam, P., Ellis, R., and Bennett, J. (1971). The N-dimensional logit model: Development and application, *Highw. Res. Rec.* 369, pp. 135-147.
22. Richards, M.G., and Ben-Akiva, M.E. (1975). A disaggregate travel demand model. *Saxon House Studies*, Lexington Books, Lexington, MA.
23. Richards, M. (1982). Disaggregate demand models—promises and prospects. *Transportation Research—A*, 16A(5-6): pp. 339-344.
24. Schoon, J. G., McDonald, M., and Lee, A. (1999). Accessibility indices: pilot study and potential use in strategic planning. *Transportation Research Record.* n 1685 Nov 1999. pp. 29-38.
25. Theil, H. (1969). A multinomial extension of the linear logit model, *Int. Econ. Rev.* 10, pp. 251-259.
26. Watson, P. L. (1973). Predictions of intercity modal choice from disaggregate, behavioural stochastic models. *Highway Research Record*, No. 446, pp. 28-35.
27. Watson, P.L. (1974). Comparison of the model structure and predictive power of aggregate and disaggregate models of intercity mode choice. *Transportation Research Record*, No. 528, Transportation Research Board, Washington, DC, pp. 59-65.
28. TRANSIMS_LANL website, <http://transims.tsasa.lanl.gov/>

29. TRANSIMS_DOT website, <http://www.transims.net/>

Vita

Qingying Lu was born on May 6, 1974 in Tianjin, China. She graduated with a Bachelor of Science in Civil Engineering and a Master of Science in Port Planning from Dalian University of Technology, Dalian, China in 1996 and 1999 respectively. She started her work towards her Master's Degree in Transportation at Virginia Polytechnic Institute and State University in the fall of 2000. She is currently working with IBM Business Consulting Services at Fairfax, Virginia since August 2002.