An Activity-Based Maryland Statewide Transportation Model – MSTM Version 2

DRAFT FINAL REPORT

March 31, 2018

Prime Contractor: University of Maryland (UMD)
Subcontractors: Cambridge Systematics (CS), University of Memphis (UM)

UMD PRINCIPAL INVESTIGATOR

Name: Lei Zhang
Title: Herbert Rabin Distinguished Professor and Director of the National Transportation Center
Organization: Dept. of Civil and Environmental Engineering, University of Maryland
Phone: 301-405-2881
Email: lei@umd.edu

UMD ADMINISTRATIVE POINT OF CONTACT

Name: Takeia Bradley
Title: Contract Manager
Organization: Office of Research Administration, University of Maryland
Phone: 301-405-8061
Email: tbradle3@umd.edu

SHA PROJECT CHAMPION

Name: Subrat Mahapatra
Title: Chief, Innovation Performance Planning Division
Organization: Maryland State Highway Administration
Phone: 410-545-5649
Email: smahapatra@sha.state.md.us

PROJECT TEAM

Maryland State Highway Administration: Subrat Mahapatra, Mark Raovic

University of Maryland: Lei Zhang, Sepehr Ghader, Di Yang, Arash Asadabadi, Aref Darzi

Cambridge Systematics: Thomas Rossi, Feng Liu

University of Memphis: Sabya Mishra

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. EXCECUTIVE SUMMARY

Statewide models are used by planners to provide travel demand estimates for various statewide, rural, and intercity planning studies, similar to how regional travel demand models provide such estimates for urban areas. Many agencies have been developing activity-based models to address planning issues for which conventional aggregate models can offer only limited analytical capabilities. Many of these issues are critical for the state of Maryland, which has maintained a statewide model to perform a variety of planning analyses. This report describes development of the Maryland Statewide Transportation Model version 2 (MSTM2) for the Maryland State Highway Administration (MD SHA). MSTM2 is composed of a statewide activity-based travel demand model for short-distance travel at its core and two supporting models: an agent-based micro-simulation travel demand model for long-distance travel; and a tour-based travel demand model for freight travel. MSTM2 is a microsimulation model system within the activity-based framework for short-distance passenger travel, and tour-based framework for long-distance passenger travel and freight travel. Each model component is described in detail in this report. An integrated model and its calibration/validation are also described in separate chapters.

2. INTRODUCTION

2.1. Overview of the Maryland Statewide Transportation Model Version 2 (MSTM2)

Statewide travel demand models offer planners a tool to estimate the number of passenger cars and freight flows on major transport facilities within the state and satisfy significant planning needs. Transportation policies that affect more than one metropolitan area cannot be sufficiently studied with metropolitan models. Statewide long-range planning, intercity travels, and planning for rural areas require a statewide model. Statewide models consider various modes such as passenger car, transit, airports, seaports, and railroads.

Early statewide models had limited capability for realistic modeling of land development and travel patterns because of computational difficulties in covering large geographic areas in detail. The development of databases and database tools, in addition to significant growth in computational power over the last two decades, had a significant effect on improving statewide travel demand models (NCHRP, 1998). Some good examples of statewide travel demand models were developed for Ohio (ODOT, 2010), Oregon (ODOT, 2010), Michigan (MDOT, 2006), California ((Caltrans), 2011), Kentucky (KTC, 1997), Arizona (Erhardt, 2012), Connecticut (ConnDOT, 1997), Florida (FDOT, 2008), and Vermont (VAT, 2010).

The Maryland Department of Transportation State Highway Administration (SHA) has developed and maintained a statewide model to perform various planning analyses. Policies concerning road pricing, environmental issues, and managed lanes were among the issues analyzed by the statewide model. The importance of rail and significance of freight needed to be considered in the model. The Maryland Statewide Transportation Model (MSTM) was the first statewide travel demand model developed for the state of Maryland. It was a trip-based, four-step model primarily developed between 2009 and 2012. A summary of the MSTM can be found in (mishra, et al., 2013).

Conventional, trip-based, four-step models have limiting analytical capabilities as they only produce aggregate results. Many agencies have started developing tour-based or activity-based micro-simulation models that simulate all individuals. Disaggregate simulation of travel demand not only increases the accuracy of prediction by reducing aggregation errors, it also enables the planner to consider more variables that affect travel demand. The time-of-day choice component of such micro-simulation models is capable of modeling activities with a long duration rather than assigning the trips into one single time point. This capability is advantageous for statewide applications that require modeling long trips.

SHA noted the need to update the four-step statewide model, and determined that to address the required planning analyses, it was preferable to use an activity-based approach. Baltimore Metropolitan Council (BMC) had just completed the development of an activity-based model ("InSITE") for the Baltimore region (including much of metropolitan Washington) (Cambridge_Systematics, 2013); SHA took advantage of the new activity-based model to leverage the development of a new activity-based statewide model, dubbed MSTM2. Since developing an activity-based model from scratch can be a time-consuming and resource-intensive process, adapting an existing activity-based model that covers much of the state is an attractive option for

states to develop their own activity-based models, at least in states where a regional activity-based model exists.

MSTM2 leveraged the following existing works for the model development: BMC InSITE activity-based travel demand model (Cambridge_Systematics, 2013); SHRP2 C10 Maryland Integrated Travel Analysis Modeling System (MITAMS); SHRP2 C20 Freight Demand Modeling and Data Improvement Strategic Plan (Chase, et al., 2013); and NTC2014-SU-R-21 U.S. National and Inter-Regional Travel Demand Analysis: Person-level Microsimulation Model and Application to High-speed Rail Demand Forecasting (Zhang, et al., 2014). Furthermore, MSTM2 leveraged the efforts at SHA to improve important input data items of models (e.g. network, land use, and others).

MSTM2 considers the unique geographic and demographic characteristics of the state of Maryland. The majority of Maryland's population lives in the Baltimore-Washington area and a significant proportion of the state's planning activities occur within it. The differences between this area and other areas were appropriately taken into account in the model. Although Washington, D.C. is outside the jurisdiction of the state of Maryland, it is an important source and destination of trips. In addition, there is a fair amount of trip making between Maryland and Northern Virginia jurisdictions. Areas outside the Baltimore-Washington area, such as Eastern Shore and Western Maryland needed special consideration. The Eastern Shore is connected geographically and economically with Delaware and parts of Virginia. Similarly, Western Maryland is significantly linked to other states while maintaining a major corridor within Maryland. It was important to decide which areas outside of Maryland to fully include within MSTM2. A major part of this effort was obtaining the relevant data for the included area and formatting it consistently. Figure 2-1 shows the MSTM2 study area:
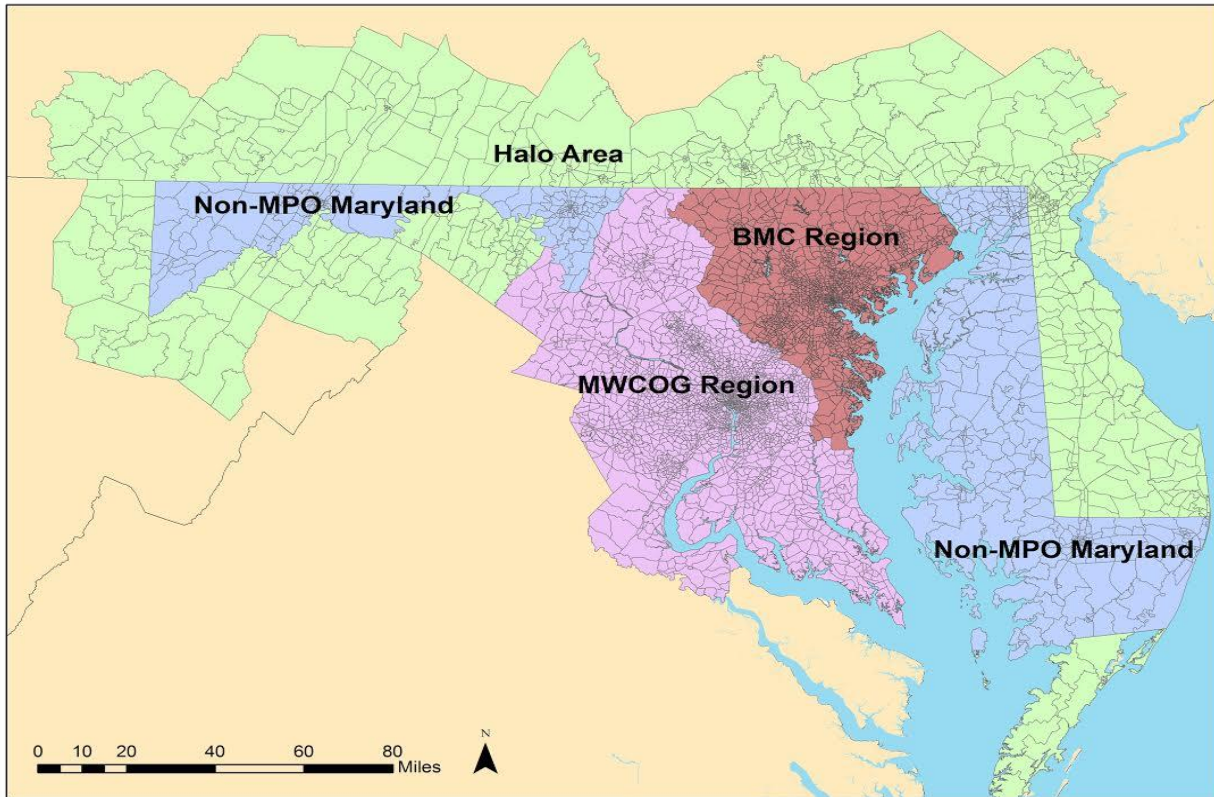
Figure 2-1. MSTM2 study area.

2.2. Value and Applications

MSTM2 took advantage of the modeling efforts, outlined previously, to answer key policy issues including but not limited to:

- Long-distance trips to and from Maryland
- Corridor management policies, including HOV-HOT lanes and active traffic management (ATM) strategies for corridors such as I-270, I-95, and I-295
- Multimodal congestion mitigation strategies, including travel time reliability improvements, integrated corridor management (ICM), and active travel demand management (ATDM)
- The importance of freight to Maryland, with particular focus on the Port of Baltimore, structures of supply chain, and truck tours
- The importance of rail and other transit modes within and through Maryland.
- Economic development and cumulative impact assessment
- Road pricing/managed lanes
- Behavioral implications of non-recurrent congestion and unreliability on travel demand
- Environmental issues, including energy consumption and air pollution emissions, and accessibility changes due to deployment of policies such as transit-oriented development

16

- Non-motorized travel analysis capabilities

MSTM2 is a fully disaggregated (i.e. each traveler is modeled individually) travel demand model that captures the complex travel behavior disregarded by traditional trip-based models. MSTM2 acknowledged that demand for travel is derived from the demand for activities, and that travelers plan their trips considering the possible dependencies (e.g., limited time, budget) between them.

In summary, MSTM2 has the following capabilities:

- Demand disaggregation at all levels
- Modeling participation decisions for activities
- Modeling trip chaining according to temporal and spatial constraints
- Modeling decisions based on lifestyle choices (e.g., full-time workers vs. retired workers)
- Fine time resolution for the demand within the 24-hour day
- Modeling value of travel time savings for each traveler and activity type (e.g., work)
- Modeling activity-based accessibility measures (i.e., accessibility to all activities in which a traveler participates, considering constraints such as scheduling, and travel characteristics such as trip chaining)

These capabilities of MSTM2 enable policy makers to observe effects of their policies on different segments of the population (e.g., income groups), which leads to more equitable policies. In addition, they help monitor how travelers' behavior and their activities change after a certain transportation or land use policy is implemented.

2.3. Model Structure

Figure 2-2 summarizes the structure of MSTM2 and its components. There are two sets of components: population generation modules and travel demand modules. The base year is 2015 and the horizon year is 2040.
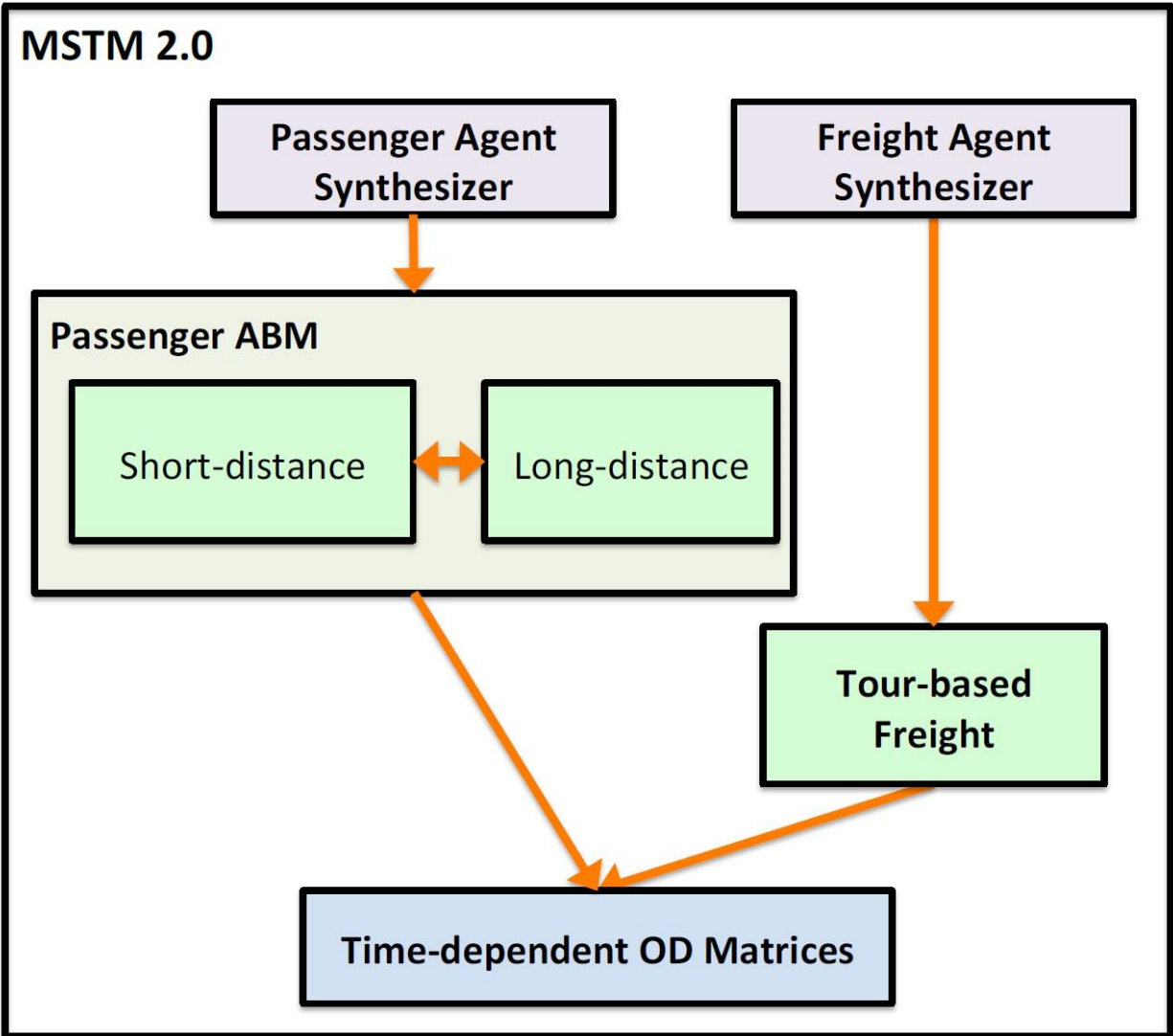
Figure 2-2. Overview of the MSTM2 structure.

The population modules are as follows:

- *Passenger Agent Synthesizer:* This module generates a synthetic population for passenger travel for the state of Maryland and all the other states. Population synthesis is a procedure that expands the sample drawn from a population to the full population such that the synthesized population can be representative of the actual population at various aggregate levels (Ryan, et al., 2009). The main concept of population synthesis is to combine the census sample data (both household and person) with available up-to-date aggregate distribution or margins data (Beckman, et al., 1996). There are many population synthesizers, either standalone software packages or components of microsimulation activity-based travel demand models, most of which function based on iterative proportional fitting procedure (IPF). IPF estimates a distribution of control

variables so that the number of individuals in given categories matches the corresponding margins while maintaining the correlation structure of the seed. This process produces a population with smooth and continuous distribution for all socio-economic variables. Among the various kinds of population synthesizers, PopGen (Konduri, et al., 2015) was selected. PopGen is a standalone, open source software package developed by Arizona State University to generate the population of the whole U.S. by using distributions of household and person variables of interest and a sample of person and household data.

- *Freight Agent Synthesizer:* This module generates a synthetic population of agents for freight travel. The definition of this agent (e.g., firm) is based on work done in the SHRP2 C20 Freight Demand Modeling project (Chase, et al., 2013).

The Travel demand modules are as follows:

- *Passenger Model:* The passenger model is divided into two models: short-distance and long-distance.
  1. The short-distance model is based on the BMC InSITE activity-based model (Cambridge_Systematics, 2013). BMC's InSITE model was extended to cover the entire MSTM model area. (note that original InSITE only covered Baltimore City, Washington, D.C., and Anne Arundel, Baltimore, Carroll, Harford, Howard, Frederick, Montgomery, and Prince George's Counties). The model structure of InSITE was retained. The original InSITE parameters were also initially retained, though it was necessary to examine model results in the areas of the MSTM not originally covered by InSITE and calibrate the parameters.
  2. The long-distance model is based on the project completed by the National Transportation Center at University of Maryland, entitled "U.S. National and Inter-Regional Travel Demand Analysis: Person-level Microsimulation Model and Application to High-Speed Rail Demand Forecasting" (Zhang, et al., 2014). A tour-based model for long-distance travel within the United States was developed in that project.

  The same agent must be present in both short-distance and long-distance models. Both models use the same synthesized agents for consistency. In addition, a consistent definition in terms of the activities for long-distance vs. short-distance travel was required. The long-distance model is responsible for modeling all non-recurrent trips longer than 50 miles. The short-distance model is responsible for modeling all recurrent trips (such as commute trips) and all non-recurrent trips shorter than 50 miles.

- *Tour-based Freight:* This module generates the freight travel demand based on work completed in the SHRP2 C20 Freight Demand Modeling.

3. DATA

3.1. Zone System

At the statewide level, there are 5,314 Statewide Model Level Zones (SMZs) that cover Maryland, Delaware, Washington, D.C., and parts of Pennsylvania, Virginia, and West Virginia, as shown in Figure 2-1. There are also 151 Regional Model Zones (RMZs) covering the full U.S., Canada, and Mexico. RMZs are used for the multi-state commodity flow model and the long-distance passenger model only. They are used as external zones producing flows that are assigned to the network at the Maryland-focused (SMZ) level.

The traffic analysis zones in the national travel demand model system are the Metropolitan Statistical Area (MSA) and non-MSA; that is, the remaining area of a state not belonging to an MSA. Therefore, the number of non-MSA is equal to the number of states in the U.S. A non-MSA is usually larger than a MSA, and it would be desirable to divide the non-MSA into smaller zones (e.g., county or city) in the national travel demand model. The finest geographic resolution in the ATS data that is primarily relied on for model estimation is MSA and non-MSA. The United States include 3202 counties and the MSA/Non-MSA consists of multiple counties or cities. Aggregating the counties or cities can give us the MSA and Non-MSA. Figure 3-1 shows the traffic analysis zone (TAZ) system in the national travel demand model. There are a total of 380 zones (excluding Puerto Rico) covering the mainland of the United States, Alaska, and Hawaii. In the 1995 ATS sample data, there are a total of 208 zones (161 MSAs and 47 non-MSAs) used for model estimation.
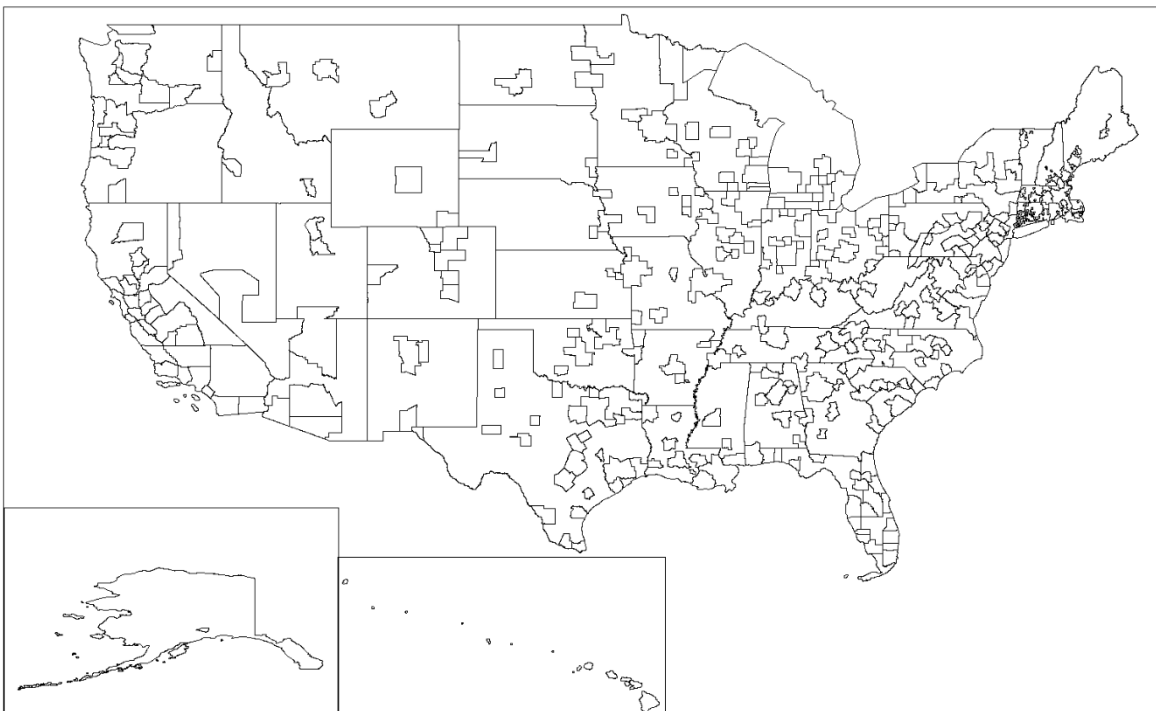
Figure 3-1. National travel demand model traffic analysis zone system.

3.2. Highway Network

The highway network was obtained by joining multiple planning networks from the following sources: Baltimore Metropolitan Council (BMC), Metropolitan Washington Council of Governments (MWCOG), Delaware Department of Transportation (DelDOT) Peninsula Model, and the National Highway Planning Networks (NHPN). This network is in GIS format and includes all necessary attributes for demand modeling, including but not limited to: number of lanes, functional classification, posted speed, and link prohibitions. The spatial coverage of the network is at the national level, with increased resolution within Maryland. The network attributes cover the following time periods: 2012, 2015, 2020, 2025, 2030, 2035, and 2040. In addition, this network is linked to the HERE network, and thus it also has access to HERE attributes.

The main attributes of this network include but are not limited to:

- Distance, in miles
- Free flow speed, in miles per hour
- Functional class
- Initial congested speed, in miles per hour
- Maximum daily lane capacity divided by 50 (service level "E")
- Off-peak toll, in cents
- Peak toll, in cents
- Posted speed limit, in miles per hour
- Route name

3.3. Transit Network

Transit network is coded in Cube and includes all the transit facilities within the study area such as all the MTA and WMATA buses and commuter rails. The main components of the transit network are as follows:

- Transit lines including Bus, Express Bus, Commuter Rail and Metro for peak and off-peak periods;
- Non-Transit legs for peak and off-peak period for both drive and walk access;
- Access and egress links;
- Transfer penalties;
- Fare system for different facilities

3.4. Land Use

3.4.1. Economic Data

This section documents the development process of socio-economic data used in the MSTM statewide model components. Travel demand is derived from economic and demographic activities—primarily households by type and employment by industry. Socioeconomic data by SMZ were developed for the entire statewide model area with consistent categories and definitions, given the availability of source data. The socio-economic data for the SMZs were developed for 2015 and then extrapolated to develop 2040 (future year) model inputs.

3.4.1.1.   Data Sources

For employment data, Longitudinal Employment and Household Dynamics (LEHD) data is used. LEHD data provide categorized employment for at census block. *The American Community Survey (ACS) 2012 5-yr Estimate for Household* is used for household data. The resolution of data is each census block group. Socio-economic data for BMC and part of MWCOG was received from SHA, which covered SMZs 2001-4934 and 10 counties out of a total 65. Socioeconomic data for MWCOG were obtained from Round 9.0 Cooperative Forecasting: employment, population, and households, which used COG TAZ3722 zone system.

LEHD Data

LEHD provides 20 categories of employment, of which seven were created using the following break-downs. Total employment is the summation of all seven categories.

- *Retail: NAICS sector 42, 44*
- *Office: NAICS sector 51, 52, 53, 54, 55, 56, 92*
- *Industrial: NAICS sector 31, 32, 33, 48, 49, 22*
- *Education: NAICS sector 61*
- *Health: NAICS sector 62*
- *Food: NAICS sector 72*
- *Others: NAICS sector 11, 21, 23, 71, 81*

ACS 5-yr Summary File

ACS 5-yr summary file provides a five-year estimate between 2011-2015. This estimate is used because 1-yr or 3-yr estimates are not provided at block group level (their largest unit is county). The "Table B25001: Housing units" is used to obtain household and population data. The table has two components:

- Estimates for each block group. Unique ID is LOGRECNO (Logical Record No.)
- Geographic file, which provide a 12-digit Geographic Identifier (GEOID) for each block group and also serves as the crosswalk between LOGRECNO and GEOID.

<u>(Dis)Aggregation of Data</u>

Aggregation and disaggregation of data were necessary to develop the SE data at a consistent resolution. Since different data sources had different units (e.g., census blocks, census block groups, etc.), appropriate resolution level was chosen to accomplish the task.

<u>Census Block/Block Groups to SMZ Crosswalk</u>

- Use the shapefile for each state containing census blocks/ block groups
- Use the shapefile provided by SHA
- Perform "Spatial join" in ArcGIS to identify which SMZ each census block/ block group belongs to
- Perform the procedure for every state and combine into a single spreadsheet illustrating the crosswalk for all the SMZs.

3.4.1.2. <u>Methodology</u>

<u>Base Year -2015</u>

- **Employment**

  o <u>BMC</u>: At the SMZ level, total employment and distribution of employment category is based on BMC Activity Based Model (ABM) 2012 (Round 8b).

  o <u>MWCOG-within Maryland</u>: At the SMZ level, the distribution of employment category is based on Round 9.0 Cooperative Forecasting: employment, population and households, which used COG TAZ3722 Zone system.

  o <u>MWCOG-outside Maryland</u>: At the SMZ level, the distribution of employment category is based on Round 9.0 Cooperative Forecasting: employment, population and households, which used COG TAZ3722 Zone system.

  o <u>Non-MPO Region Maryland</u>: At the SMZ level, the total employment is the aggregated value of the seven categories of employment. The LEHD data is used for the employment data and the distribution of employment category is based on NAICS codes, as mentioned in Section 3.4.1.1.

  o <u>Halo regions</u>: At the SMZ level, the total employment is the aggregated value of the seven categories of employment. The LEHD data is used for the employment data and the distribution of employment category is based on NAICS codes, as mentioned in Section 3.4.1.1.

<u>Base Year -2015</u>

- **Household**

- o <u>BMC</u>: At the SMZ level, total household is based on BMC Activity Based Model (ABM) 2012 (Round 8b).

- o <u>MWCOG-within Maryland</u>: At the SMZ level, the household data is based on Round 9.0 Cooperative Forecasting: employment, population, and households, which used COG TAZ3722 Zone system.

- o <u>All other areas in MSTM </u>are based on census and ACS 2012

3.4.2. Population Data

A critical input to an activity-based travel demand model is a synthetic population with socio-economic and demographic characteristics for each individual person and each household in the modeling region. MSTM2 employs the synthetic population generator PopGen to synthesize the population based on detailed samples of individual person and household records, and marginal control distributions at the aggregate level. This subsection focuses on introducing major data sources for population synthesis and the methodology for preparing sample data and control totals.

3.4.2.1.   <u>Data Sources</u>

<u>Census 2010</u>

The decennial U.S. Census provides information on all persons and households in the country and can be viewed at census block level geographic resolution (similar to the transportation analysis zone (TAZ) level). The 2010 Census focused on population and housing questions. The 2010 SF1 data is used to examine univariate distributions of households and persons across particular variables (e.g., households by household size and persons by sex or age).

<u>American Community Survey (ACS)</u>

The ACS is conducted continuously by the Census Bureau and provides abundant information that can be used for validation. Because the ACS is conducted continuously, the Census Bureau can make data available every year rather than every 10 years like the decennial census (though for a smaller sample of the population). The ACS samples about one in every 40 addresses every year, or 250,000 addresses every month. For areas with large populations (65,000 or more), survey estimates are based on 12 months of ACS data. For all areas with populations of 20,000 or more, the survey estimates are based on three years of ACS data. The Census Bureau produces estimates for all areas, down to the census tract and block group levels, based on five years of ACS data. One, three, and five-year estimates based on survey data, including the data from 2011, are currently available.

The ACS provides data on housing and population not available from the 2010 Census. Information includes:

- Population characteristics
- Age
- Sex
- Relationship to head of household
- Income
- Employment information including labor force status, industry, and occupation
- Journey to work information
- Household characteristics
- Vehicles available
- Income
- Tenure
- Housing value
- Rent

ACS Public Use Microdata Sample (PUMS)

The Public Use Microdata Sample (PUMS) consists of individual records of the characteristics for a sample of person and housing units in the ACS. The PUMS dataset includes variables for almost every question in the ACS. The Census Bureau produces the PUMS files so that data users can create custom tables that are not available through pre-tabulated (or summary) ACS data products. ACS PUMS files are available for one individual year, three years, and five years, and contain approximately 1, 3, and 5% of the United States population, respectively.

3.4.2.2.   Methodology for Population Synthesis

Sample Data

We use the 2006-2010 five-year American Community Survey (ACS) Public Use Microdata Sample (PUMS) as the sample data.

Marginal Data

For the base year 2015, Census 2010 and 2011-2015 five-year ACS data are used to generate control totals, since ACS data are available at the Census Block Group (BG) level. The following steps are taken to ensure ACS information can apply to the MSTM2 TAZ level:

- Rule 1: If BG and TAZs boundaries match exactly, then use census seeds and MSTM 2.0 Households
- Rule 2: If multiple TAZs are within a BG, then apply seed of one BG to all TAZs
- Rule 3: If multiple BGs are within a TAZ, then take average of the seeds and apply to one TAZ

For the future year 2040, control totals are derived from the base year control totals and the future year land use projections. The base year proportions are multiplied with the 2040 land use projections to generate future year margins.

Control Total Variables

At the county level:

- Household variables: household number of workers by household income (20 categories)
- Person variables: person gender by age (36 categories)
- Group quarter variables: total number of group quarters

At the TAZ level:

- Household variables: household size, household income (the order of the two variables does matter)
- Person variables: personal employment status
- Group quarter variables: total number of institutional group quarters, total number of non-institutional group quarters

Agent types: household, group quarter, and person.

Control level: region level control; county level and geo level control; TAZ level

4. ACTIVITY-BASED TRAVEL DEMAND MODEL FOR SHORT-DISTANCE PASSENGER TRAVELS

4.1. Overview of the InSITE Activity-based Model

The short-distance model of MSTM2 adopts the basic structure, functionality and platform of the activity-based model developed for the Baltimore Metropolitan Council (BMC) called InSITE. Figure 4-1 shows the model structure, components, and their interrelationship.
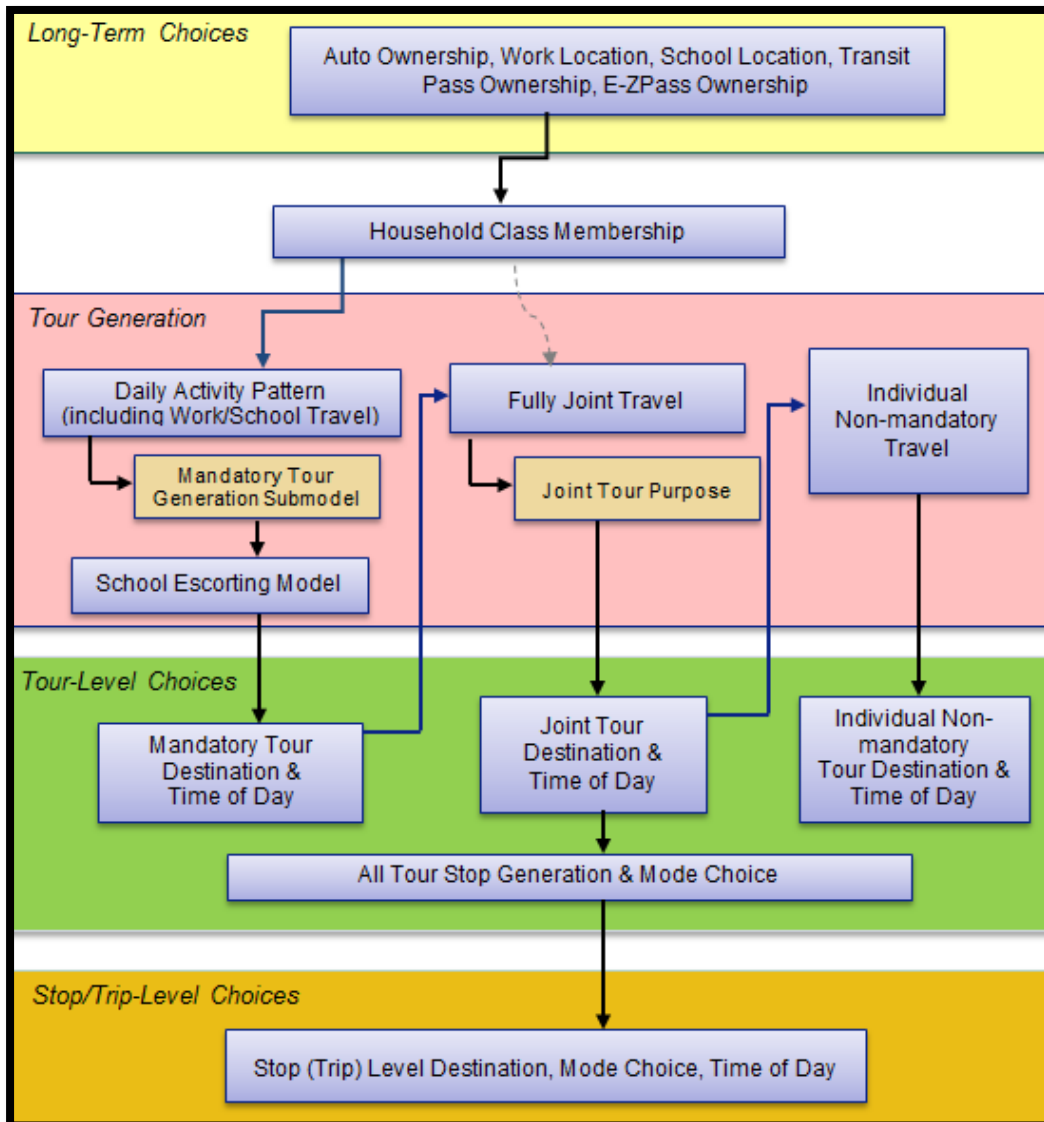
Figure 4-1. InSITE model structure.

Table 4-1 summarizes the components of InSITE activity-based model.

Table 4-1. InSITE Model Components

Model Name	Level	What is Predicted
Synthetic Population Generator	Households	Household size and composition, household income, person age, gender, employment status, student status
Regular Workplace Location	Workers	Workplace location zone
Regular School Location	Students	School location zone
Auto Ownership	Households	Number of autos owned
Transit Pass Ownership	Households	Whether the household owns a transit pass
E-ZPass Ownership	Households	Whether the household owns an E-ZPass transponder
Daily Activity Pattern	Person Day	Zero, one, or two tours for each activity purpose; zero, one, or two stops for each activity purpose
Joint Travel	Households	Number of fully joint tours with two or more household members where household members participate in each joint tour
School Escorting	Person (Household) Day	Where half of tours a student is escorted to/from school, which household member escorts the student, and whether escorting is done on a mandatory tour
Work Tour Destination Choice	Work Tours	For work tours – regular workplace or other work location (and its zone)
Work-Based Sub-tour Generation	Work Tours	Number and purpose of any sub-tours made during a work tour
Work Mode Choice	Work Tours	Main tour mode
School Mode and Time-of-Day Choice	School Tours	Main tour mode, the time period arriving at school and the time period leaving school (all school tours are to regular school location)
Work Time-of-Day Choice	Work Tours	The time period arriving at work and the time period leaving work
Other Tour Time-of-Day Choice	Other Tours	Time period arriving at the primary destination and the time period leaving the primary destination
Other Tour Mode and Destination Choice	Other Tours	Primary destination zone and main tour mode

Intermediate Stop Generation	Half-tour	Number and activity purpose of any intermediate stops made on the half-tour, conditional on day pattern
Intermediate Stop Location	Trip	Destination zone of each intermediate stop, conditional on tour origin and destination, and location of any previous stops
Trip Mode Choice	Trip	Trip mode, conditional on main tour mode
Trip Departure Time	Trip	Departure time, conditional on time windows remaining from previous choices
Special Generators	Zone	Number of trips, trip end location, mode choice
Commercial Vehicle	Zone	Number of trips, trip end location
External Travel	Zone	Number of trips, trip end location
Highway Assignment	Vehicle Trip Table	Link volumes and travel times/speeds
Transit Assignment	Person Trip Table	Transit trips/boardings by route/stop

4.2. Model Implementation

The activity-based short-distance model is implemented and delivered in TourCast, Cambridge Systematic's activity-based modeling platform. TourCast combines an extremely powerful computational engine with an established Python-based specification for configuring individual model components and model sequences.

TourCast was developed over a period of several years in response to ABM model development efforts for the Denver Regional Council of Governments, SHRP2 Project C10B, and ABM deployments for the Houston-Galveston Area Council and Met Council. It provides an ideal balance between configurability and model execution speed. TourCast is designed so that it can operate independently or be integrated into commercial travel demand modeling platforms such as Cube, etc.

TourCast models critical travel behaviors such as time shifting, telecommuting, transit use, and the interactions of household members. TourCast was designed for practitioners and provides extraordinary control, flexibility, and ease of use. In summary, TourCast offers:

- Extremely fast model execution from a finely-tuned model engine
- Integration with commercial modeling software (including Cube)
- Outputs (intermediate and final) that can be sent to databases, GIS, or report generators
- A modular, scalable software design with a service-oriented architecture
- Desktop, server, and cloud configurations

As indicated in Table 4-2, there are several model components implemented in TourCast. These components are all interlinked, where the outputs of a model component are the direct inputs into

a subsequent component. The sequencing of these components is different in application and in estimation. The estimation process begins with tour level mode choice models, which are used to compute logsums by trip purpose for use as key explanatory variables (accessibility or impedances) in several other model components. These are followed by stop- or trip-level models, as these are conditional upon some of the choices made by the tour-level mode choice models. Next, the tour-level time-of-day and destination choice models are estimated, followed by person level models such as daily activity pattern, joint travel, and school escorting models. The mandatory tour purposes are estimated before the non-mandatory purposes, while all the long-term choice models are estimated towards the end of the estimation process.

As shown in Table 4-2, after population synthesis the application of these components starts with the long-term choice models, followed by daily activity pattern and tour-level models. After the tour-level models, the trip-level models are applied, which determines the location, time-of-day, and mode choice for every half-tour (or trip) in each individual tour. This final step produces a roster of trips for every individual in the population.

Table 4-2. InSITE Implementation in Tourcast

Application Sequence	Model Component	Model Applied
0	Population Synthesis	Once for entire region
1	Usual Work Location Choice Model (Long-term)	For every worker
2	School Location Choice Model (Long-term)	For every child & (university student)?
3	Vehicle Availability Model (Long-term)	For every household
4	Daily Activity Pattern Model	For every individual
5	Tour Destination Choice Model – Mandatory	For every work tour
6	Transit Pass Ownership	For every household
7	E-ZPass Ownership	For every household
8	Tour Time-of-day Choice Model – Mandatory	For every mandatory tour
9	School Escorting Model	For every child making school tour
10	Joint Tour Model	For every household (with at least two traveling members)

11	Tour Destination Choice Model - Joint Tours	For every joint tour
12	Tour Time-of-day Choice Model - Joint Tours	For every joint tour
13	Joint Tour Participation	For every traveler in household
14	Tour Generation - Individual Non-Mandatory Travel	For every individual (not stay-at-home)
15	Tour Destination Choice Model - Non-Mandatory Tours	For every non-mandatory tour
16	Tour Time-of-day Choice Model - Non-Mandatory Tours	For every non-mandatory tour
17	Stop Generation Model	For every individual half-tour (where stop was indicated in DAP phase) and all joint half-tours
18	Tour Mode Choice Model – Mandatory	For every mandatory tour
19	Tour Mode Choice Model - Joint Tour	For every joint tour
20	Tour Mode Choice Model - Non-Mandatory	For every non-mandatory tour
21	Stop Destination	For every stop
22	Trip Mode Choice	For every trip
23	Stop Time-of-day	For every stop

4.3. Model Components

4.3.1. *Workplace Location Choice*

The purpose of this model is to simulate the usual workplace location for each full-time or part-time worker in the synthetic population. The choice of the usual workplace location is a two-step modeling process:

1. **Workplace Type Model**, which simulates whether the worker has a usual workplace and whether that workplace location is inside the model region, or whether the worker usually works at home;
2. **"Usual" Workplace Location Choice Model**, if the result of the binary choice model is the "usual" workplace.

The workplace type model is essential to determine who opts for the "no usual" workplace location choice. These individuals are then subject to the tour level (home-based) work destination choice model.

The "usual" workplace location choice model captures the usual location a person commutes to for work and work-related purposes. The final specification for both these models was reached by testing a wide range of model variables.

- For the workplace type model, most of the variables include person and household attributes, land use variables, and proximity to the external boundary of the region. The person and household attributes are the most significant indicators as to whether an individual desires to work at a specific location.
- For the location choice model, the variables entered were similar to those used in most destination choice models, including level of service (LOS), socioeconomic and demographic (SED) and land use (LU) variables. The models also include size functions, which are additive logarithmic functions of various types of employment, households and enrollment. This model has a similar structure to that of the work destination choice model.

4.3.2. School Location Choice

The purpose of this model is to simulate the school location for each pre-school or school-age child in the synthetic population. The model captures the location where a child attends preschool, daycare, or K-12 school. This model handles the entire school location choice for a child; there is no tour destination choice model for school tours. This is different from work tours, which have a usual work location, determined by the work location choice model, and a tour destination choice model, which captures the destination choice of workplace for each person who does not have a usual workplace, and the destination choices of workers who do not travel to their usual workplaces on the travel day.

Children of a similar age in the same household are likely to attend the same school. "Lining up" the usual school location decisions among children in the same household makes the school escort activity by an adult in the household more likely to be coordinated. The school locations in a household are simulated sequentially from the youngest to the oldest child. Therefore, the usual school locations of the younger children in the household are known when simulating the school locations of older children. This information is used in the model to group the school location choice for children in the same household.

The final specification was reached by testing a wide range of model variables—level of service, socioeconomic and demographic and land use. The models also include size functions, which are additive logarithmic functions of elementary, middle school, and high school enrollment, and households (for pre-school age children only).

4.3.3. Vehicle Availability

This model estimates the number of vehicles owned by each household in the synthetic population. It is a multinomial logit model with alternatives representing ownership of zero, one, two, three, four, or five or more vehicles.

4.3.4. Transit Pass Ownership

For each household in the synthetic population, this model simulates whether the household has a transit pass. It is a binary model with alternatives representing possession of a transit pass or not.

4.3.5. E-ZPass Transponder Ownership

This model predicts whether a household in the synthetic population has an E-ZPass transponder to use when driving on toll roads. It is a binary model with alternatives representing possession of a transponder or not.

Unlike most of the other models in InSITE, this model was not estimated merely from the household survey data set. Information on whether the household had a transponder was not recorded in the household survey. The alternate data source was a list of transponder owners by zip code provided by E-ZPass administration. This indicates that household ownership of a transponder was not known. What was known was an estimate of the probability of having a transponder given the zip code of the household location, estimated as the number of households in the zip code in the transponder owner database divided by the total households in the zip code.

Because there is no "choice variable" in the data set, the estimated probabilities were used in defining the likelihood functions for the model estimation. The likelihood function is computed as follows:

1. Compute probabilities of toll transponder ownership for each household in the survey as in any logit model.
2. Take the weighted average probability across each zip code (weighted based on the expansion factors for households).
3. Compute log likelihood = nhhs1 * log(P1) + nhhs2 * log(1-P1)
   Where:
   nhhs1 is the number of households with toll transponders in the zip code
   nhhs2 is the number without (equal to total minus those with) in the zip code
   P1 is the average probability of toll transponder ownership for the zip code from the survey respondents.

The final log likelihood is then scaled by (nzipcodes/total households). This ensures valid statistics on the coefficient estimates.

4.3.6. Daily Activity Pattern Models and Fully Joint Tour Generation

This section discusses the daily activity pattern (DAP) and fully joint tour generation models. This includes several distinct model components related to DAP and joint travel for the household, including the latent class model structure. The models described in this section are identified in Figure 4-2, which shows a portion of the overall model structure presented in Figure 4-1.
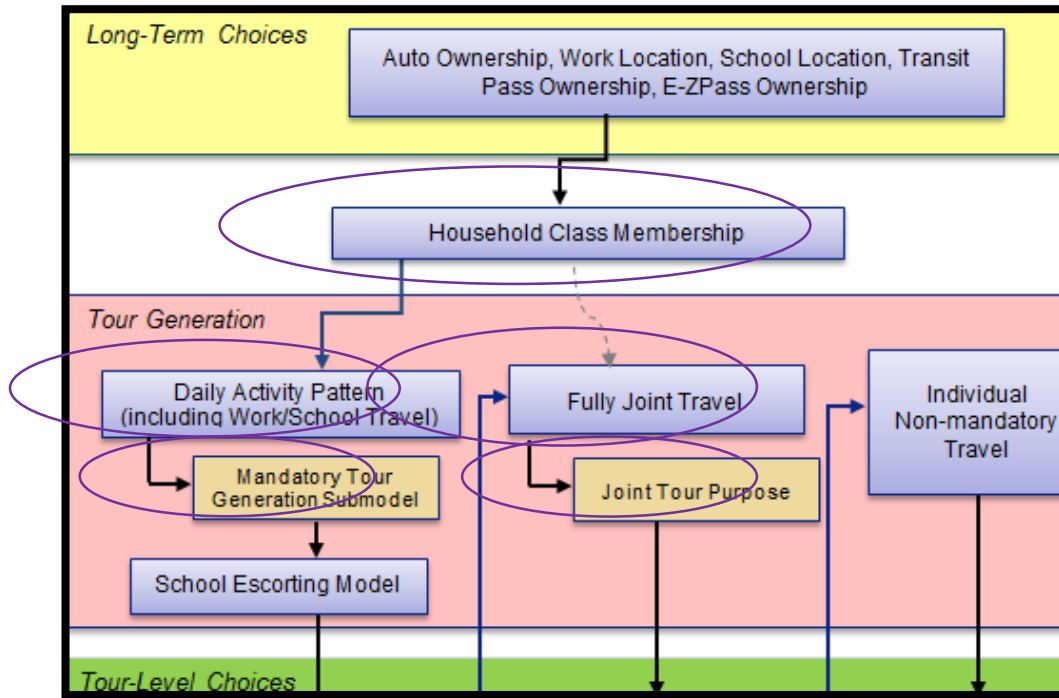
Figure 4-2. Model components related to daily activity pattern.

The daily activity pattern (DAP) determines whether individuals travel on the travel day, and the number and purpose of mandatory tours generated for the individual. These two components are split across two separate models: a high-level DAP model and a mandatory tour enumeration model. The fully joint tour generation models consist of three distinct models: the first model determines the number of fully joint tours for the household; the second determines the purpose of each tour generated for the household; and the third determines the members of the household that will participate in the joint tour.

Note that in this case, joint tours are restricted to non-mandatory tours. Examination of the survey data show that few joint mandatory tours are made among household members. The school location choice model considers the probability that children of similar ages attend the same school, and the school escorting model does consider the "bundling" of children in escorting decisions.

In InSITE, each person in the synthetic population is classified as one of eight "person types":

- CH1 – Child 1 (age 0-4)
- CH2 – Child 2 (age 5-15)
- CH3 – Child 3 (age 16-17)
- FTW – Full time worker
- PTW – Part time worker
- NWA – Non-working adult
- STD – Adult student (e.g., university)
- SEN – Senior (age 65+)

Note that in InSITE, the use of age as a variable is not restricted to the person type definitions. Many model components use the person's specific age, sometimes in connection with a specific age range (e.g., age 5-10 or age greater than 40) to explain travel behavior.

The models detailed in this section include five separate estimated models. They are: the latent class DAP and joint tour generation model, the adult mandatory tour enumeration model, the child mandatory tour enumeration model, the joint tour purpose model, and the joint tour participation model.

4.3.6.1. Latent Class DAP and Joint Tour Model

The first components of both the DAP and fully joint travel modules were estimated simultaneously in a latent class model structure. The high-level DAP model is a choice model for each individual in the household with three alternatives:

- Mandatory pattern
- Non-mandatory pattern
- Stay home pattern

A mandatory pattern consists of at least one mandatory (work or school) tour, and possibly other non-mandatory tours. A non-mandatory pattern consists of at least one non-mandatory tour and zero mandatory tours. A stay home pattern indicates that the individual did not generate any travel within the model region on the travel day. Note that a person with a "stay home" pattern may not actually be at home all day; he or she might be traveling outside the region with no travel within the region (for example, out of town on a business trip). Basically, the intention is to identify those persons who do not travel within the model region on the travel day.

The fully joint tour model is a choice model that also offers three alternatives: zero, one, or two joint tours. These two models are estimated simultaneously in a latent class model structure. As part of this model structure, a secondary household class model is also estimated, which is another discrete choice model. Each of these three discrete choice models estimated simultaneously here (i.e., the high-level DAP model, the joint tour generation model, and the household class model) are multinomial logit models. The interaction between these three models is shown in Figure 4-3.
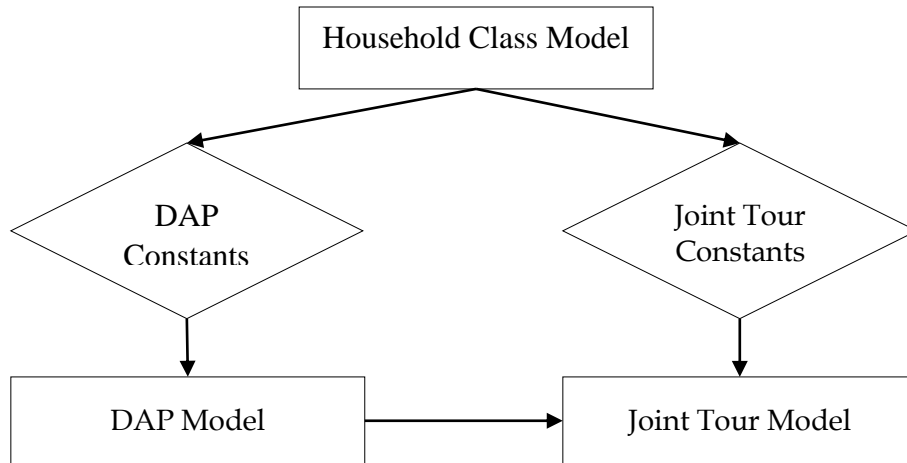
Figure 4-3. Latent class DAP and joint model interaction.

As shown in the Figure 4-3, the household class model determines only the constants that are used for the DAP and joint tour models. The constants are based on the class selected for the household. If the household selects class 1, the constants used for household members' DAP models are class 1 DAP constants and the constants used for the joint tour model are class 1 joint tour constants. Besides the constants in the DAP and joint tour models, all other variables are generic to the household class. That is, the coefficients for other variables do not depend on the household class. Note that the joint tour model also depends on the outcomes of the DAP model. This is a consequence of the sequencing of these two models, not of the simultaneous model estimation.

There are several reasons for choosing this model structure. First, we can accommodate interactions between individual DAP choices and household level joint tour generation choices via the household class model. In some activity-based models, DAP is assumed to be independent from joint travel, which is not ideal. Second, interactions between DAP choices of household members are accommodated seamlessly and without the need to test large numbers of two-way interaction terms in the DAP utility functions. Instead, the DAP constants themselves determine the interactions across household members. This is done by having different constants for each person type. For one class, the constants may show it to be very likely that a young child and full-time worker both stay home, while in another class, the constants may show it to be more likely that the young child stays home, and the full-time worker goes to work. These sorts of interactions are accommodated through DAP constants for eight mutually exclusive person types. Third, application of this group of models is extremely simple. It is a sequence of three simple multinomial logit (MNL) models, two of which (the DAP and joint tour model) have only three alternatives, with the other has a small number of alternatives as well (in this case, six).

4.3.6.2.   Mandatory Tour Generation Sub-Model

The high-level DAP model only determines whether an individual will have a mandatory day, non-mandatory day, or stay home. Within the choice of mandatory DAP, each individual must also select the number and purpose of mandatory tours to engage in. The mandatory tour generation sub-models do just this. Two separate models were estimated, one for adults and the other for children. The main reason for segmenting adults and children is that they have different education purpose options; children can have school activities while adults can have university activities.

Note the non-mandatory tours for persons with non-mandatory patterns are generated in the individual non-mandatory tour generation model. Non-mandatory tours, if any, made by persons with mandatory DAPs are also generated by this model.

While nested logit model structures were tested, the MNL model structure worked the best. The alternatives are defined by the number of mandatory tours (one or two), the number of stops (zero, or one or more), and the types of tours (work, school, or university). There are 13 alternatives in the model, but only a subset is considered available for any particular person type. Table 4-3 shows the model alternative availability matrix for each person type.

Table 4-3. Mandatory Tour Options by Person Type[3]

Alternative	CH1	CH2	CH3	STD	FTW	PTW	NWA	SEN
1 Work0[4]			X	X	X	X	X	X
1 Work1			X	X	X	X	X	X
2 Work0					X	X		
1 Work0 + 1 Work1					X	X		
2 Work1					X	X		
1 Work0 + 1 Univ.				X	X	X		
1 Work1 + 1 Univ.				X	X	X		
1 Work0 + 1 School			X					
1 Work1 + 1 School			X					
1 Univ.				X	X	X	X	X
2 Univ.				X				
1 School	X	X	X					
2 School	X	X	X					

3. Person types are as follows: CH1 = child less than 5 years, CH2 = child 5-15 years, CH3 = child 16+ years, STD = college student, FTW = full-time worker, PTW = part-time worker, NWA = non-working adult < 65 years, SEN = non-working adult 65+ years.

4. Work0 is a work tour with no stops, Work1 is a work tour with 1+ stops.

Joint Tour Purpose Model

The joint tour purpose model determines the tour purpose for each joint tour generated by the household in the joint tour generation model. The alternatives of the model include the following tour purposes:

- Meal
- Shopping
- Personal business
- Social/recreation

The tours are modeled in sequence for each household, rather than simultaneously. This reduces the number of alternatives that must be considered in the model from 16 (for two tours with up to four purpose options for each) to four. It also means that the joint distribution of purposes for households with two joint tours may not exactly match the observed distribution (e.g., the models may suggest it is more likely than observed to have households with two joint meal tours). However, we did check the distributions of one- and two-joint tour households and found that the distribution of purposes for two-joint tour households was close to what one would expect if the purposes of each joint tour were independent of one another. Moreover, we tested variables in the model that accounted for the household having two joint tours (rather than just one).

4.3.6.3.   Joint Tour Participation Model

The joint tour participation model determines the household members that will engage in a joint tour that was generated by the joint tour generation model. It does this by sequencing through all household members based on the person type in the following order:

- Child 1 (age 0-4)
- Non-working adult (less than 65 years old)
- Senior
- Child 2 (age 5-15)
- Child 3 (age 16+)
- Part-time worker
- Full-time worker
- Adult student

This order is chosen because it represents an ordering of person types based on their typical availability to take part in joint activities (e.g., workers and college students typically have more limited availability, while non-workers and preschool-age children have more flexibility).

The model has only two alternatives—either the person can be part of the joint tour or not. All household members with an active DAP (mandatory or non-mandatory) are eligible to participate in the joint tour. Scheduling of the joint tour is simulated after participation (rather than before) and is based on the time availability overlaps of participating household members. Note that the

time-of-day has already been simulated for mandatory tours and for school escorting activities for every person in the household prior to the application of the joint tour participation model.

No strict requirements are enforced on the estimation to ensure that model application will result in a valid fully joint tour. In other words, application of the model could result in zero or one household members participating. In such cases, the model will be rerun until a valid tour is constructed.

4.3.7. School Escorting

The school escorting model captures the choice of whether a child is escorted to/from school, and if so, by whom. The school escorting model is sequenced after mandatory tour destination choice, but before mandatory tour time-of-day choice. By sequencing the model in this way, the model is independent of timing constraints (instead, timing considerations are conditional on whether there is school escorting).

For each child with a school activity, the escorting model simulates escorting for both the outbound half-tour (travel to school) and the return half-tour (travel home from school). The alternatives in the model include the no escort alternative and up to three adults (if they exist in the household). For each adult, two alternatives are specified. The first is a mandatory tour option, where the adult would escort the child as a stop on the adult's mandatory tour (typically work). If the adult does not have a mandatory tour, then the mandatory option is not available as an alternative in the model. The second is a stand-alone school escorting tour option, where the adult generates a new tour for the specific purpose of escorting the child to/from school.

Since each school tour has two travel components (outbound and return), there are two choice dimensions, handled simultaneously in the model. In other words, each alternative represents the joint choice for outbound half-tour escort choice and return half-tour escort choice.

One issue in modeling school escorting behavior is how to handle multiple children in the same household traveling together to or from school. We call this bundling of the escort decision, since it suggests the escort choices of each child are not independent. To accommodate this behavior, child bundles were considered in the escorting model. Child bundle formation consists of two parts. First, the school tours for each child must share a common school location (school locations are predicted by the school location choice model). If this condition is not met, the children will not be bundled. Second, a probability distribution based on observed household survey data is used to simulate whether two children form a bundle or not (if the first condition is also met). The probability distribution depends on the ages of the children being considered. The distributions for outbound and return half tours are shown in Table 4-4.

Table 4-4. Percentages of Children from Same Household Having Same Escort Decision
Outbound Half Tours

Age Range	0 to 4	5 to 10	11 to 13	14 to 18
0 to 4	98%			
5 to 10	76%	98%		
11 to 13	71%	84%	85%	
14 to 18	0%[1]	19%	62%	70%

Return Half Tours

Age Range	0 to 4	5 to 10	11 to 13	14 to 18
0 to 4	98%			
5 to 10	68%	93%		
11 to 13	43%	75%	75%	
14 to 18	0%	19%	56%	56%

Source: Baltimore/Washington region household survey data set.

[1] In the household survey, there were only two child pairs in this category (who happened to have the same escort decision). It is asserted that this value should be 0% for application purposes.

Pairs of children are simulated from youngest to oldest. For instance, in the case of a 3-year old, 8-year old, and 15-year old, the 3-year old and 8-year old would be compared first. From the table, we see the probability of bundling among these two children on the outbound half tour is 76%. Second, the 8-year old and 15-year old would be compared, having a probability of bundling of 19%. This probability holds regardless of whether the 3-year old and 8-year old were already bundled. In total for this case, there are four possible outcomes, each with the following probabilities:

- No bundling: P = (1-0.76) * (1-0.19) = 0.194
- 3 years, 8 years bundled: P = (0.76) * (1-0.19) = 0.616
- 8 years, 15 years bundled: P = (1-0.76) * (0.19) = 0.046
- All bundled: P = (0.76) * (0.19) = 0.144

Note that it is not possible for the 3- and 15-year old to be bundled without the 8-year old; this is by design. Also note that bundling on the outbound half tour and on the return half tour are handled independently from one another. Children need not be bundled on both half tours. Since children are bundled for the purposes of the school escorting choice, the decision-maker is the child bundle.

4.3.8. Tour Destination Choice

The purpose of this model is to simulate the location of each primary activity on a tour. A destination choice model was estimated for each tour purpose[1]. A total of five models were estimated, defined as follows:

- Individual work
- Individual university
- Individual non-mandatory (including escort purpose)
- Individual work-based sub-tours
- Fully joint non-mandatory tours

Note that by definition, all tours begin and end at home, except work-based sub-tours.

4.3.9. Individual Non-Mandatory Tour Generation

The individual non-mandatory tour generation model predicts the number and type of individual non-mandatory tours (as opposed to joint tours among household members, which are modeled separately) for each individual in the simulated population. This model is a function of the number of available time windows, activity patterns of other household members, transportation accessibility, and other household and person attributes.

4.3.10. Tour Time-of-day Choice

The tour level time-of-day choice models are estimated for the following tour purposes:

- Work
- University/school
- Individual non-mandatory
- Joint non-mandatory
- Work-based sub-tour

Note that by definition, all tours begin and end at home, except work-based sub-tours. Time-of-day choice is predicted in 30-minute intervals (48 periods across the day) beginning and ending at 3:00 a.m. The models predict—jointly—the time interval an individual arrives at and the time interval they leave from the tour primary activity location. Thus, for each arrival period alternative, there are multiple return period options. For instance, for an arrival in the 8-8:30 a.m. period, one

---

[1] A school tour purpose model was not estimated. A regular school place location model will be among the long-term choice models, and each child person type will have a school place assigned. When a child generates a school tour, it will automatically be assigned to the regular school place location; thus, no tour-level model is needed.

could potentially return at any time interval thereafter (up until the last period of the day). Each arrival/return pair represents an alternative, for a total of 1,176 alternatives.

The main approach is to use alternative specific constants for various groupings of arrival periods, departure periods, and durations, plus shift effects that push arrivals earlier or later and durations of stay longer or shorter (as a result, shifting the departure period as well). The models also make extensive use of the concept of available time windows for scheduling tours. As each tour is simulated, the periods that are used by the tour are made either fully or partially unavailable for any other tours, and the length of remaining time windows available during the day is calculated.

Finally, a logsum variable from the mode choice model is used to estimate the effects of transportation level of service, including road congestion and tolling, on time-of-day choice.

4.3.11. Tour Destination Choice

The purpose of this model is to simulate the location of each primary activity on a tour. A destination choice model was developed for each tour purpose[2]. This is analogous to having gravity models for each trip purpose in a four-step model. A total of five models are estimated, based on travel purpose, and are defined as follows:

- Individual work
- Individual university
- Individual non-mandatory (including escort purpose)
- Individual work-based sub-tours
- Fully joint non-mandatory tours

Note that by definition, all tours begin and end at home, except work-based sub-tours.

4.3.12. Stop Generation

These models estimate the number of intermediate stops (i.e., excluding the primary activity for the tour) for each tour in the simulated daily activity patterns, and the purpose for each stop. They estimate the number of stops on each half tour, to a maximum of three.

Separate models are estimated for different tour purposes. Note that school escorting stops are not modeled here since these are modeled earlier in the model chain. Nor are "change mode" stops, such as transit transfers or park-and-ride and kiss-and-ride stops (although the stop for the person

---

[2] A school tour purpose model was not estimated. A regular school place location model will be among the long-term choice models and each child person type will have a school place assigned. When a child generates a school tour, it will automatically be assigned to the regular school place location; thus, no tour-level model is needed.

driving a transit rider to a stop is modeled as an escort stop if it is not the primary activity of the (escort) tour). Stop generation is modeled for the tour purposes listed below:

- Work
- School/university
- Individual non-mandatory (including escort and fully joint tours)
- Work-based sub-tours

Therefore, there are a total of eight models: a number-of-stops model and a stop-purpose model for each of the four purposes listed above.

4.3.13. Work-Based Sub-tour Generation

The work-based sub-tour generation model predicts the number and type of work-based sub-tours for each home-based work tour (referred to as the "parent tour") that is simulated in the daily activity pattern model. This model is a function of characteristics of the parent work tour, other activities the traveler has during the day, the distance from the home to the workplace, highway and transit accessibility, and other household and person attributes.

4.3.14. Tour Mode Choice

Four tour mode choice models were estimated, based on tour purposes or groups of purposes (the first three models represent home-based tours):

- Work
- School/university
- Non-mandatory, individual and joint (including shopping, meal, personal business, social/recreation, and escort tours)
- Work-based sub-tours

The final model specifications were reached by testing a range of model variables, nesting structures, and model constraints, though we were guided by our experience with similar models in other activity-based model systems. Types of variables included attributes of the traveler and his or her household, tour characteristics, level-of-service attributes (e.g., travel cost and travel time), and land-use variables. In each model, the cost variable was segmented by household income level.

4.3.15. Stop Location Choice

The purpose of this model is to simulate the location of each activity that is not the primary activity on a tour (and is not a school escorting activity). Prior to the application of the stop location choice model, the following information is known about the tour:

- The home location of the traveler
- All information about the primary activity (purpose, location, start and end time periods)
- The chosen tour mode
- The number and purpose of stops generated on each half tour (outbound and return)

The objective of the stop location choice model is to simulate the location of each of these stops. If the stop is the only stop on a half tour, the stop location is simulated based on the locations of the home and the primary activity. In cases where multiple stops exist on a single half tour, the stops are modeled in sequence starting with the one closest in temporal proximity to the primary activity. For outbound stops (first half tour), this means stops are modeled in reverse chronological order, and for return stops (second half tour), this means stops are modeled in chronological order. In such cases, a second stop's location is simulated based on the location of the first stop's simulated location (rather than that of the primary activity) and the home location, and a third stop's location is simulated based on the location of the second stop's simulated location and the home location.

The main approach of the model is similar to that used for tour destination choice. Size variables, defined similarly to those used in tour destination choice, reflect the amount of activity, and therefore, attractiveness of destinations. The main differences are that 1) the "detour generalized accessibility" and "detour distance," rather than the direct home-to-activity location time and distance, are used as the impedance measures, and; 2) a single model is used for stop location choice, rather than separate models by activity purpose (although some model parameters vary by stop and/or tour activity purpose).

4.3.16. Stop Time-of-day Choice

Time-of-day choice is predicted in 30-minute intervals (48 periods across the day) beginning and ending at 3:00 a.m. Unlike the tour time-of-day choice models, the stop time-of-day choice model predicts only a single time period for each stop activity. This period corresponds to the activity start time if the stop occurs on an outbound half-tour, or the activity end time if the stop occurs on a return half-tour.

At the point in which stop time-of-day choice is applied in the model chain, the start and end times of a tour's primary activity have already been simulated. Thus, the difference between stop-start time and primary activity start time (outbound half-tour) or stop-end time and primary activity end time (return half-tour) represents the duration of that stop (inclusive of the travel time between activity locations). Of course, an outbound half-tour stop must occur earlier than the primary activity start time and a return half-tour stop must occur later than the primary activity end time.

In cases where multiple stops exist on a single half-tour, the stops are modeled in sequence, starting with the one closest in temporal proximity to the primary activity (the stop location choice model sequences stops the same way). For outbound half-tour stops, this means stops are modeled in

reverse chronological order, and for return half-tour stops, this means stops are modeled in chronological order. In such cases, the second stop's timing would be bounded by the start/end time of the first stop's simulated time period (rather than that of the primary activity), and a third stop's timing would be bounded by the start/end time of the second stop's simulated time period. As an example, the modeling sequence and timing bounds of three outbound half-tour stops to a primary activity that begins in the 9:00-9:30 a.m. period are shown in Figure 4-4.
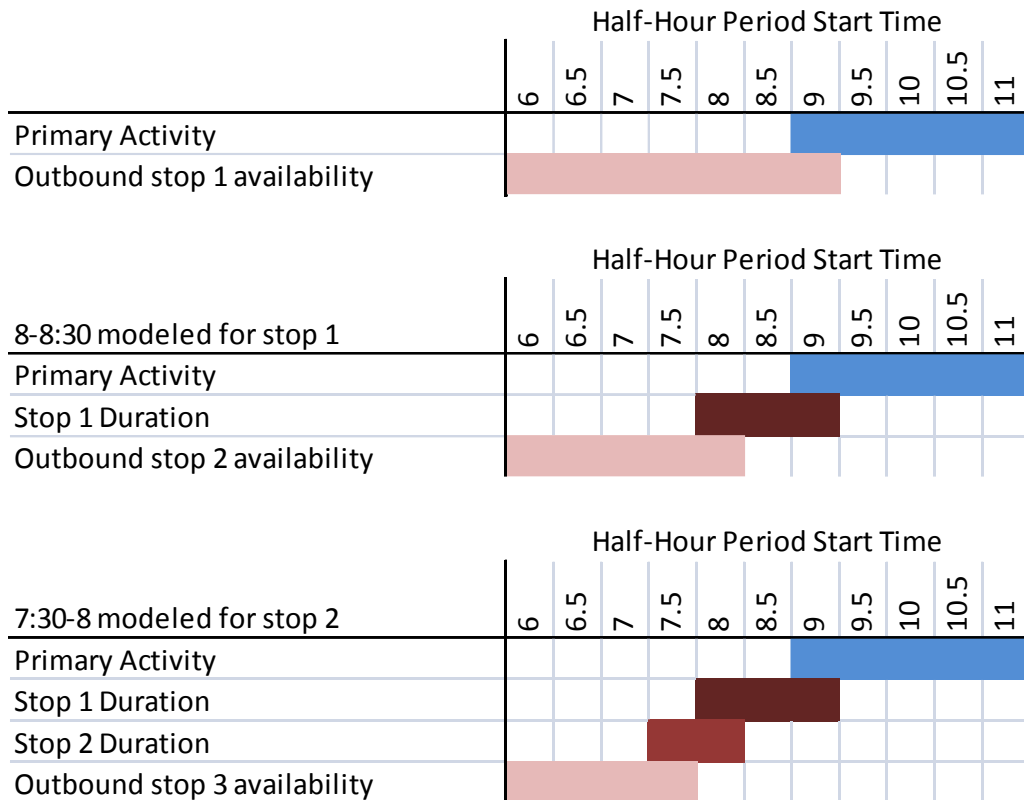
Figure 4-4. Example of stop modeling sequence.

In addition to knowing the primary activity's start and end time, the start and end times of other tours for an individual are also known when stop time-of-day choice is applied. The timing of these other tours binds the available timing of stops from the opposite direction. For instance, if an individual has a work tour that ends in the 4:30-5:00 p.m. time period and a meal tour that begins in the 7:00-7:30 p.m. time period, and that meal tour has an outbound half-tour stop, the timing of that stop is bound between 4:30 and 7:30 p.m., as shown in Figure 4-5.

	Half-Hour Period Start Time
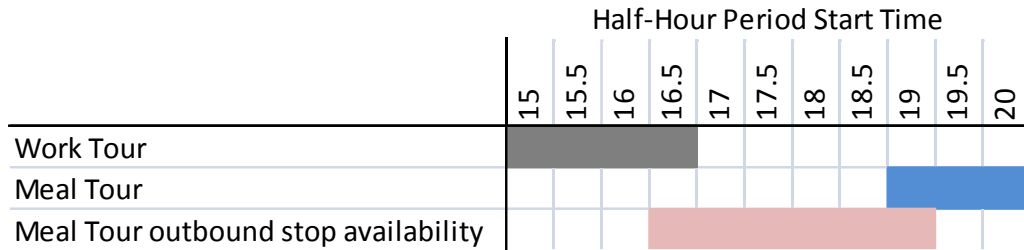
Figure 4-5. Example of time-of-day constraints with multiple tours.

As indicated above, the difference in stop time-of-day and the time-of-day of the subsequent or previous activity represents the duration of that stop. This duration includes both the actual stop activity's duration and the travel time between the stop activity and subsequent/previous activity. It is worth noting that no specific requirements are placed on the stop time-of-day choice to ensure consistency between activity timing choices and travel times. In other words, it would be possible (though unlikely) for the stop time choice to be such that there is insufficient time to travel between activity locations and engage in the stop activity (e.g., tour ends in 4:00-4:30 period, stop choice is the 4:30-5:00 period, and travel time between locations is 60 min). There are several reasons for this. First, the 30-minute time periods are a bit fuzzy, in that choice of a specific time period could indicate an exact time at the beginning of that 30-minute period or the end. In the case of activities that lie one period apart (e.g., 4:00-4:30 and 4:30-5:00), the exact times chosen could be as little as 1 minute apart or as large as 59 minutes apart. This means any restrictions would have to be very broad regardless. Second, while tour mode would be known when the model is applied, trip modes may be different than tour modes. The choice of a specific mode's skims to use in such cases is unclear. Third, given these first two considerations, implementation of such requirements could be rather complicated. Last, we felt these inconsistencies would generally be rather rare, and were able to include some variables in the model specification that work to discourage these potential inconsistencies.

The main approach of the model is similar to that used for tour time-of-day choice. Alternative specific constants are used for various durations (segmented by half-tour) and shift effects can encourage longer or shorter stop durations. The models also use the concept of available time windows.

4.3.17. Trip Mode Choice

A single trip mode choice model was estimated for all tour/trip purposes. There are indicator variables for tour purpose that in effect create different constants by mode for each tour purpose. Note that by definition, all tours begin and end at home, except work-based sub-tours, which begin and end at work. There may be trips of different purpose than the tour purpose.

The final model specifications were reached by testing a range of model variables, nesting structures, and model constraints, though we were guided by our experience with similar models in other activity-based model systems. Types of variables included attributes of the traveler and his or her household, tour characteristics, transportation level of service (represented by the

generalized accessibility from the tour mode choice model), land use variables, tour mode-related variables, and previous trip purpose variables related to escort activities.

4.4. Model Enhancement for MSTM2 Implementation

This chapter describes the changes made to enhance the BMC InSITE TourCast activity-based model implementation to support region-specific segmentation of parameters for the Maryland Statewide Transportation Model implementation. The series of high-level checks performed to verify correct model operation are described in this section.

4.4.1. Introducing Segmentation for Non-BMC Area

As part of the model enhancement, a regional segmentation has been implemented into the model according to the following sub-regions:

- BMC Modeling Domain
- Panhandle (Washington/Allegany/Garrett Counties)
- Chesapeake West (Charles/St. Mary's/Calvert Counties)
- Eastern Shore
- Virginia and West Virginia
- Delaware and Pennsylvania

The following components and parameters have been enhanced with region-specific segmentation:

- Workplace type choice model: no usual workplace constant, external workplace constant, work from home constant
- Workplace location choice model: intra-zonal indicator coefficient, highway distance coefficient, size function coefficients
- School location choice model: intra-zonal indicator coefficient, highway distance coefficient, size function coefficients, TAZ county different from home county coefficient
- Vehicle availability model: all alternative-specific constants
- Mandatory tour enumeration: all alternative-specific constants
- Individual non-mandatory tour enumeration: all alternative-specific constants
- Fully joint non-mandatory tour enumeration: all alternative-specific constants
- Work-based tour enumeration: all alternative-specific constants
- Tour destination choice model: intra-zonal indicator coefficient, highway distance coefficient, size function coefficients
- Tour time-of-day model: Shift variables

- Tour mode choice model: All alternative-specific constants

The segmented terms are implemented in the TourCast configuration files. The segmented terms are implemented independently as marginal parameters in the alternative utility with an initial value of zero. Therefore, the current implementation effectively does not vary the model parameters by region, but the user may adjust all terms individually by region if necessary during model calibration.

The base model parameter values are consistent with the BMC InSITE Beta Version 005 model.

4.4.2. Model Verification Tests

The MSTM2 model was initially run with a 20% sample followed by a 100% sample to verify proper operation of all components. The following summaries were conducted using the 20% sample outputs, and the results of each were compared to the BMC InSITE model results and by a super-set of the regions: BMC; Chesapeake West / Panhandle / Eastern Shore; and the other states (VA/WV/DE/PA).

Population: household makeup (workers, children, size, income); person distributions (age, gender, income).

Vehicle Ownership: distribution of the number of vehicles and auto sufficiency; average vehicles by income, area type and region.

Daily Activity Pattern: distribution of general DAP (mandatory, non-mandatory, no travel); distribution of general DAP by person type.

Tours: distribution of tour type and activity purpose; tour purposes by adult person type.

Tour Distribution: tour length distributions segmented by mandatory and non-mandatory; tour duration distributions segmented by mandatory and non-mandatory; trip time-of-day distributions.

Mode: mode share; drive-alone and walk-transit mode share by income.

4.4.3. Special Demand Generators

The following are the special generators in the MSTM2 halo area. Each corresponding total employment is adjusted based on the data reported on the respective website. For MSTM2, there will not be any individual component of treating special generators. However, the employment can be adjusted based on the validation results. The validation task consists of simulated volume to count volume comparison on the links conecting the special generators. Table 1 shows the list of special generators with SMZ number and employment. Figures 1 through 10 show the spatial location of special generators.

Table 4-5. List of Special Generators in MSTM Halo Area

Facility	SMZ	FIPS	County	Employment
Wilmington Hospital	6089	10003	Kent, DE	48,879
Baltimore-Washington Int'l Airport	354	24003	Anne Arundel, MD	11,054
John Hopkins university	119	24510	Baltimore City, MD	4,879
Fort George G. Meade (Army Station)	469	24003	Anne Arundel, MD	54,419
University of Maryland	2316	24033	Prince Georges, MD	10,000
George Washington University	1456	11001	District of Columbia	748
Regan International airport	5070	51013	Arlington, VA	7,633
Dulles International Airport	5604	51107	Loudoun, VA	18,146
Andrews AFB (Air Force Station)	2746	24033	Prince Georges, MD	2,709

Legend

Facilities level-II

- H Hospital
- Military Station
- University
- Airport
- SMZ_Level_II_Revised
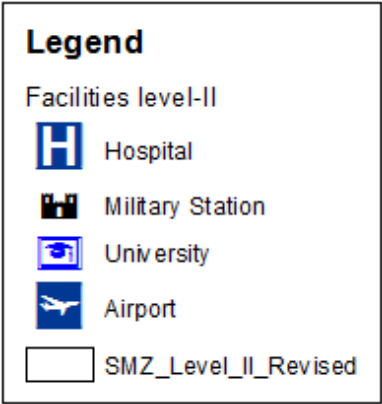
Figure 4-6. Wilmington Hospital, DE.

Figure 4-7. Baltimore-Washington Int'l Airport, MD.
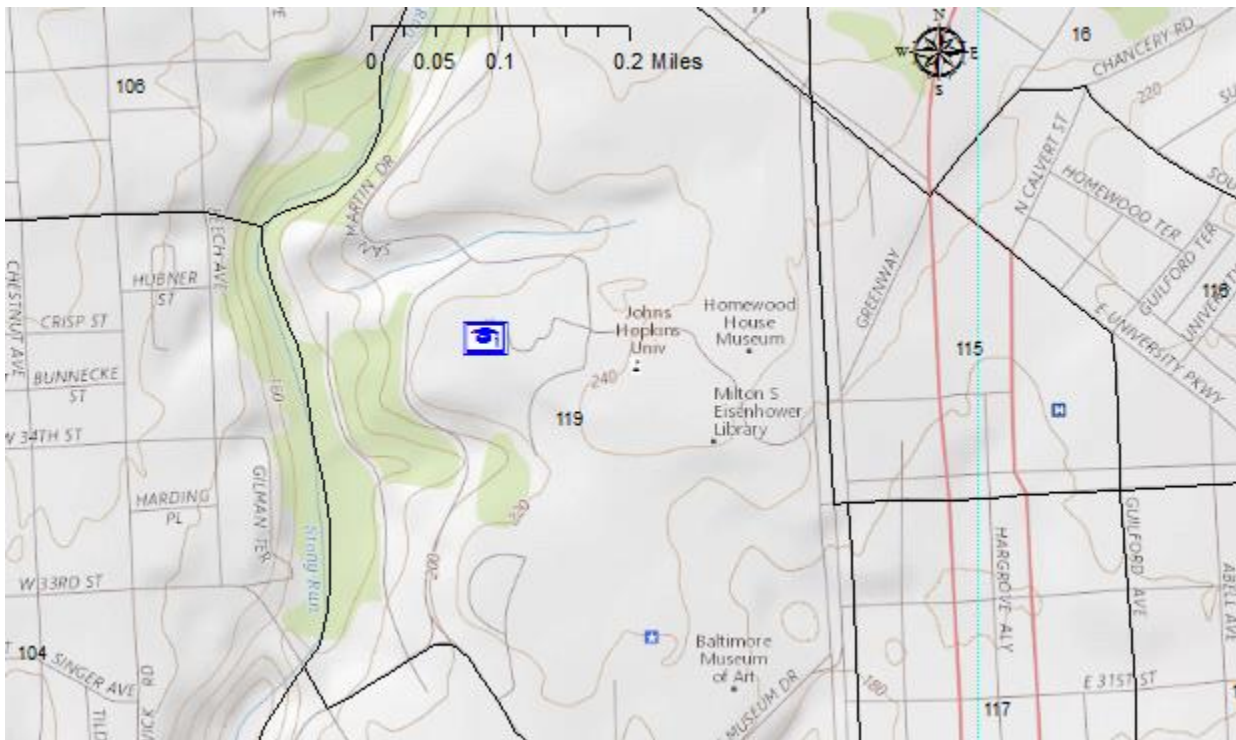
Figure 4-8. John Hopkins University, MD.
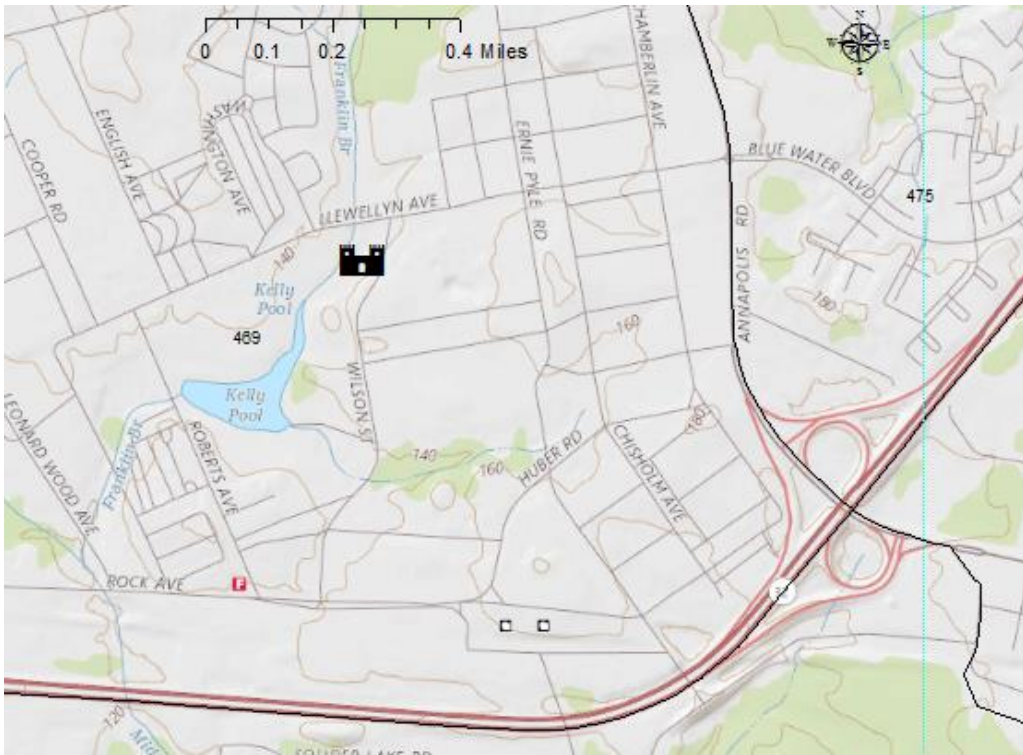
Figure 4-9. Fort George G. Meade, MD.
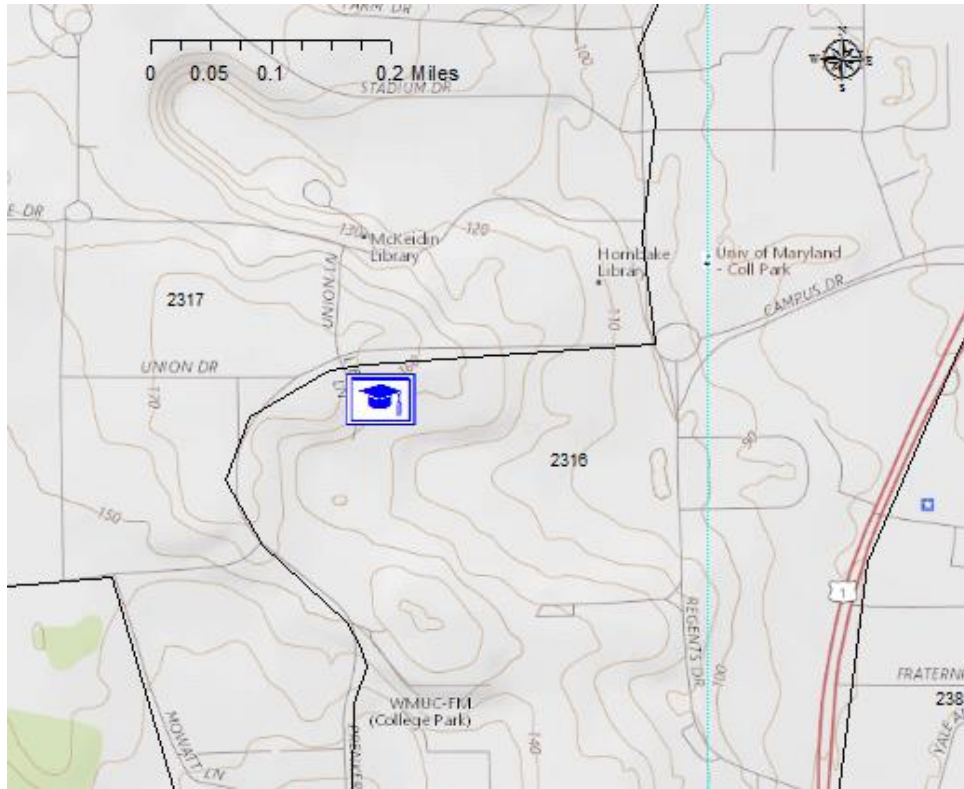
Figure 4-10. University of Maryland, College Park, MD.
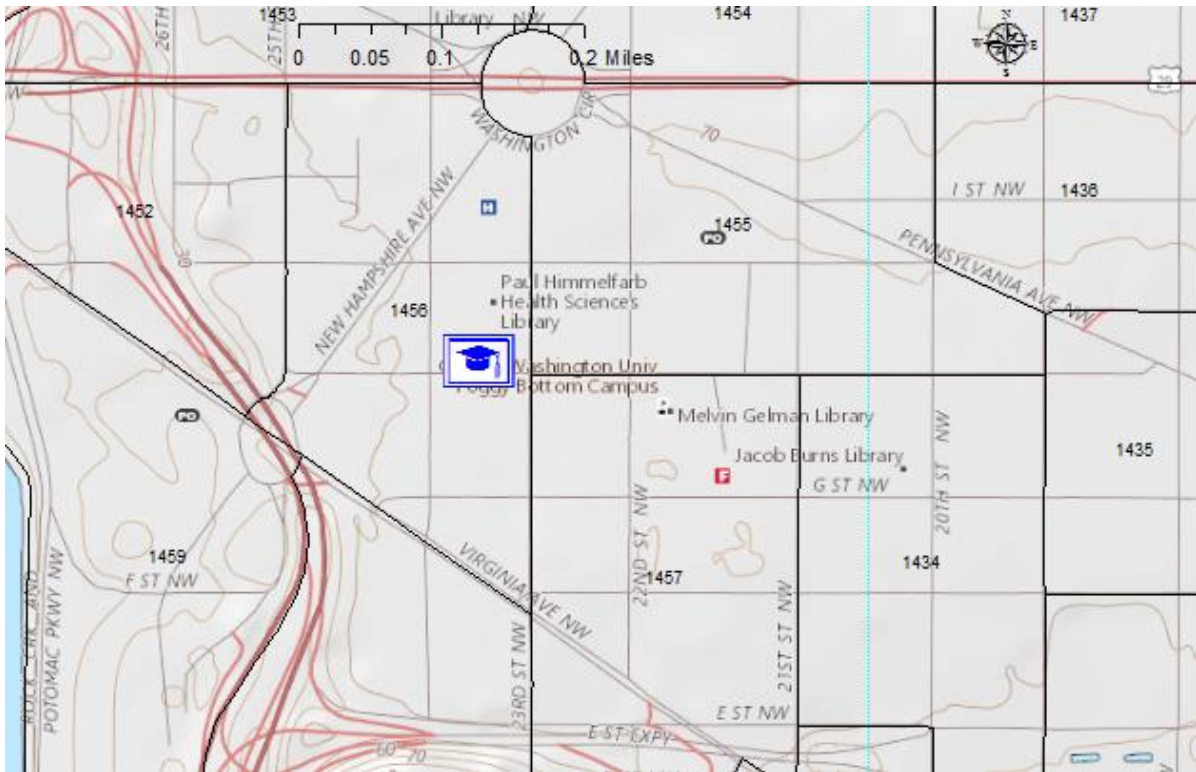
Figure 4-11. George Washington University, DC.
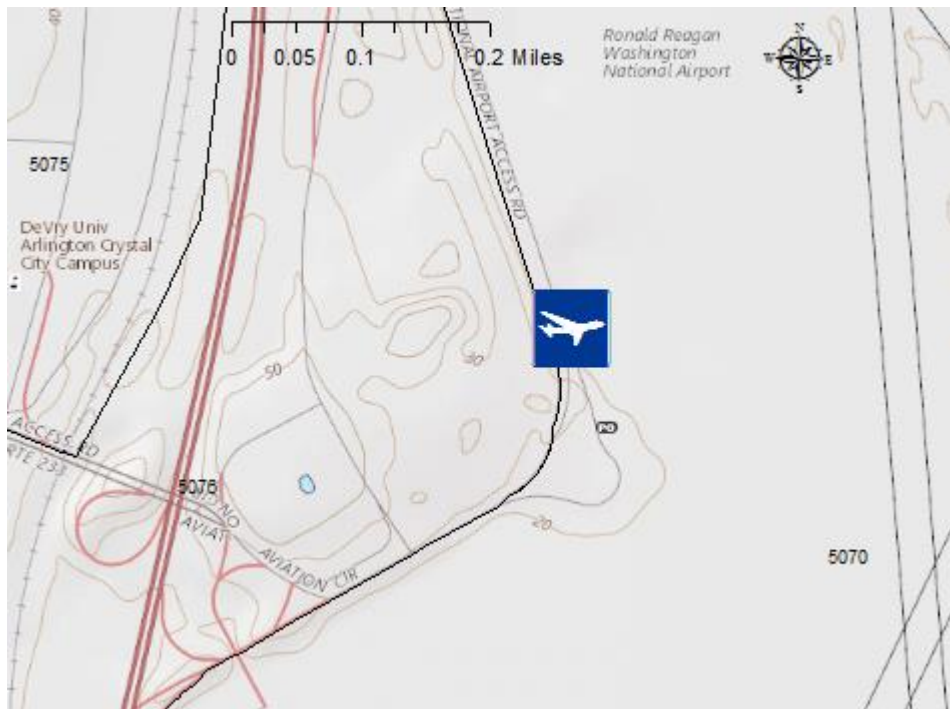
Figure 4-12. Regan International airport, VA.
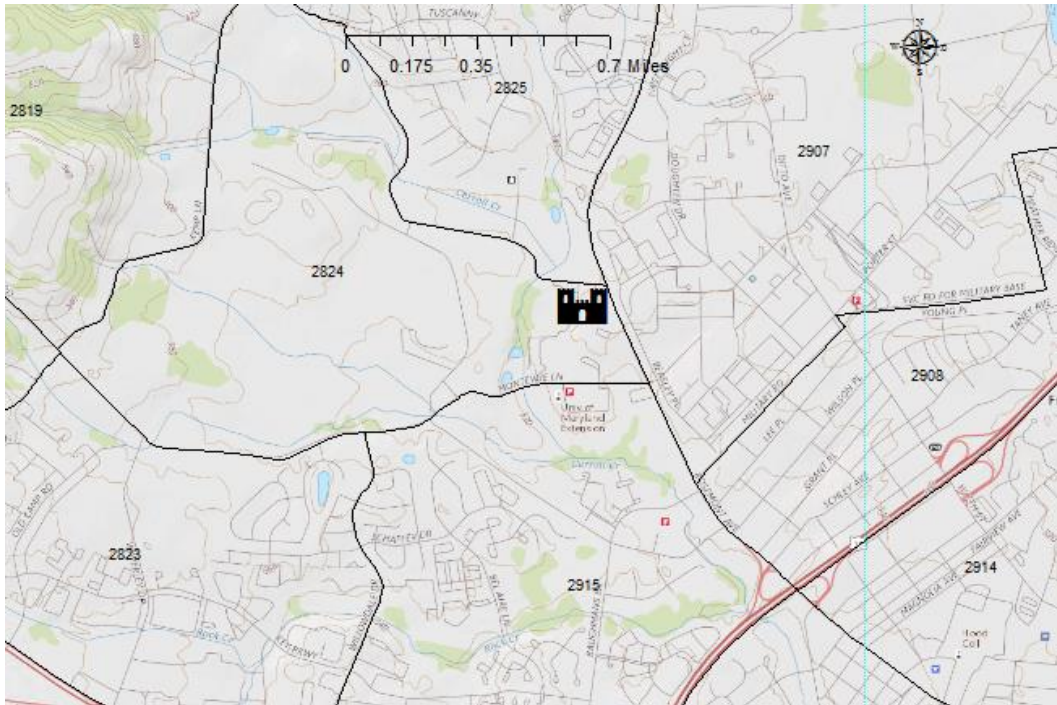
Figure 4-13. Fort Detrick, MD.
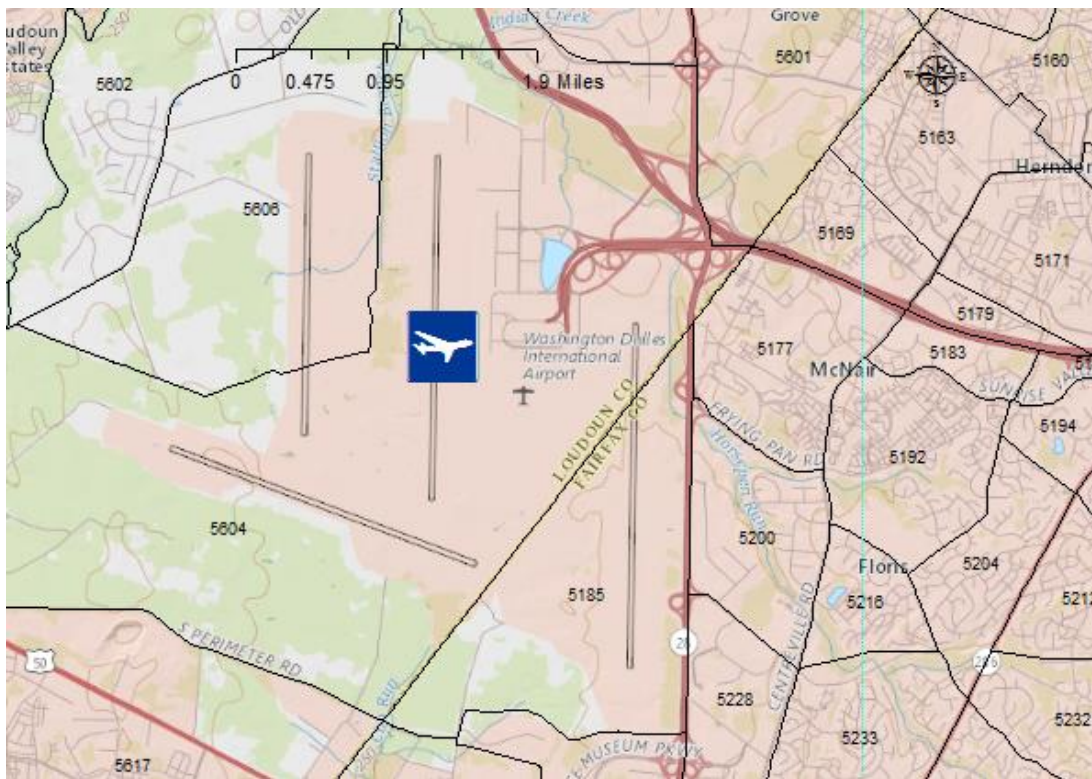
Figure 4-14. Dulles International Airport, VA.
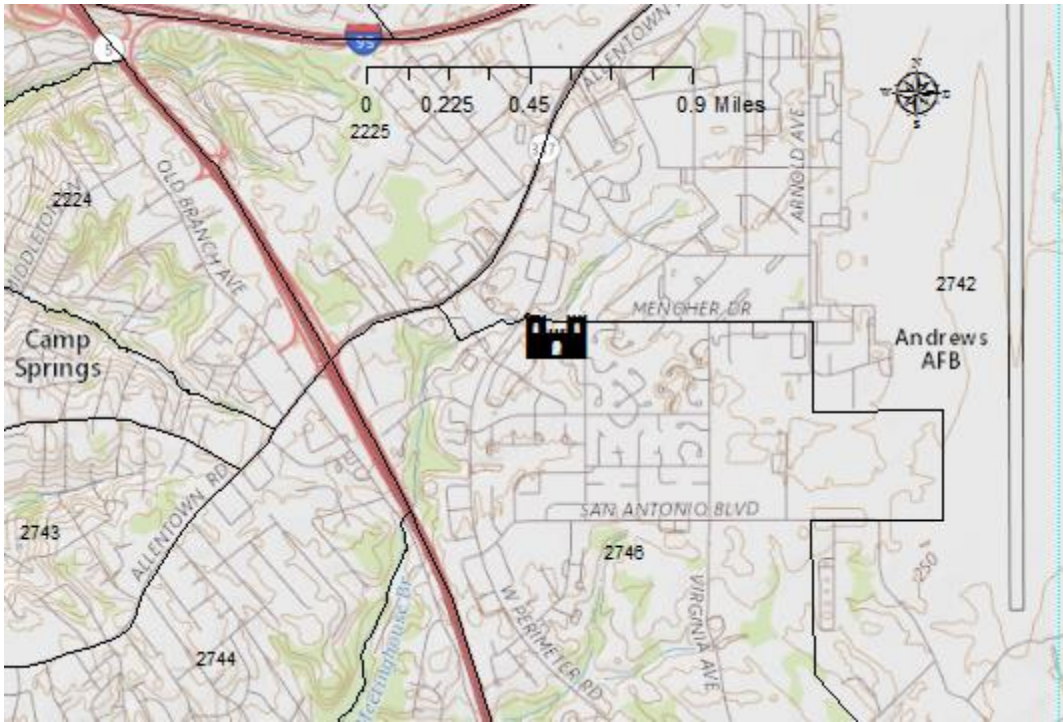
Figure 4-15. Andrews AFB, MD.

5. AGENT-BASED MICROSIMULATION TRAVEL DEMAND MODEL FOR PASSENGER LONG-DISTANCE TRAVELS

5.1. Overview

The increasing interest in national transportation policies—from strategic infrastructure investment to infrastructure operation and management with regard to efficiency, sustainability, and safety—has incited researchers and decision makers to call for advanced and policy-sensitive analysis tools. Highway infrastructure investment, high-speed rail, and airport development all depend on national travel markets. Any infrastructure investments or operational and management improvements should be evaluated through a capable national travel analysis tool instead of region-level, corridor-level, or state-level models, which are mainly used in these types of analyses.

After the Intermodal Surface Transportation Efficiency Act (ISTEA) was enacted in 1991, many state departments of transportation began developing statewide travel demand models as critical analysis tools in addressing legislative requirements in statewide planning. However, the statewide models are weak in external trips, which are usually generated with information from federal and neighboring states instead of available socioeconomic data. A national long-distance travel demand model can provide external trips for statewide models in base-year and future-year. According to Giaimo and Schiffer's review (Giaimo, et al., 2004) of statewide travel demand modeling developments, most statewide travel demand models in the U.S. do not consider long-distance travel.

However, this situation has started to change. Today more than 35 states have active modelling efforts to meet statewide policy and legislative development needs. The California Statewide Travel Demand Model (CSTDM) is a tour-based travel demand model. It can forecast all types of travel, as well as long-distance trips made by California residents. Ohio's statewide travel demand model uses a state-of-the-art tour-based modeling approach. It includes person travel for both short-distance and long-distance (larger than 50 miles). The long-distance travel models are developed based on the Ohio DOT's sponsored long-distance travel survey data and can estimate the frequency and characteristics of long-distance travel for assignment on the transportation network. The Florida statewide travel demand model (FLSWM) employs the traditional four-step model methods for passenger and freight travel in the state of Florida. The Maryland Statewide Transportation Model 1.0 (MSTM 1.0) is a four-step travel demand model designed to generate link-level assignment for personal and freight travel. In terms of passenger travel, it models short-distance or urban personal trips for residents in the study area. The regional level model includes a long-distance travel model for both residents and visitors with trips longer than 50 miles one-way. The Kentucky Statewide Travel Demand Model (KYSTM) models the long-distance trips (over 100 miles) in Kentucky and parts of the neighboring states. The model adopts a modified four-step travel demand model, removing the mode choice module.

National long-distance passenger travel demand analysis has been an understudied area in transportation planning. The lack of multimodal, long-distance, origin-destination data has seriously limited planners' ability to conduct quantitative analysis of operational effectiveness and infrastructure investment. As the nation and various states engage in funding transportation

infrastructure improvements to meet future long-distance passenger travel demand (interstate highway tolling/expansion, high-speed rail, and next-generation passenger air transportation system that relies more on smaller airports and aircrafts), developing a national Multimodal Travel Analysis System (MTAS) and an American Long-distance Personal Travel Program (LDPT) have become the priorities and fundamental work for planners when conducting national travel analysis.

The developed national long-distance model used in this project is a person-based, activity-based national travel demand model for national travel analysis. All major behavioral dimensions of long-distance travel are considered. Compared to the traditional four-step approach, activity-based techniques offer several advantages: (1) it is easier to consider tours, multi-day and multi-stop trips, and intermodal access/egress transfers that are important for long-distance travel modeling; (2) households and persons are the basic units of analysis, which enable detailed behavioral representations and interactions; and (3) it provides a rich framework for analysis of travel as a multi-day, monthly, quarterly, or yearly pattern of behavior, derived from activity participation. There are also significant differences between long-distance trips considered in the current developed activity-based model and daily/weekly trips in metropolitan/state-level tour/activity-based models developed in previous research. For instance, long-distance trips usually take days or weeks and may involve car, airplane, train, bus, or a combination of modes. It is often the case that households first choose travel time for long-distance vacation trips based on time and budget before selecting destination and mode. Categorization of trip purposes is also different for long-distance trips. Cost of travel for long-distance trips is not just travel disutility, but also includes lodging, food, etc. The same applies to the total travel time for long-distance, which usually covers not only in-vehicle travel time but also the ingress/egress time, transfer time, and lodge time. The lower frequency of long-distance travel may also imply a different decision-making process.

The developed long-distance model used in this project is the first attempt to develop an integrated activity-based travel demand model system for an individual's quarterly/yearly long-distance or national activities and travel in the U.S at the Metropolitan Statistical Area (MSA)/Non-MSA level, which is the highest geographic resolution in the long-distance travel survey data. The model system is developed considering the specific attributes of the long-distance travel, such as low frequency, long activity duration, long activity duration at intermediate stops on the tour legs, different sets of mode alternatives, etc. Therefore, the model system not only takes into account an individual's long-distance travel at the tour level, but also at the stop level. National-level travel survey data are used to estimate the model components and provide the parameters for simulation. The model is implemented in our developed micro-simulation platform, which simulates each individual's yearly long-distance activities and travel in the U.S. over the course of one year.

The developed activity-based national travel demand model can forecast the long-distance passenger trips made by auto, air, and train in the U.S. over a one-year period in a micro-simulation framework. The long-distance trip in the model system is defined as trips greater than or equal to 50 miles, one-way. For each long-distance activity, there is only one tour destination or primary destination, and during each leg of the tour, there could be multiple intermediate stops. Econometric model developments are conducted for the multiple model components, which guarantee the maximum behavior realism.

The yearly long-distance activity schedule can be presented as a set of different long-distance activities per year. The activity pattern model is responsible for generating the long-distance trip rates by activity type (business, personal business, pleasure) for each person using the multiple

classification analysis (MCA) method. The estimation is done based on 1995 American Travel Survey data (ATS). The ATS data provides detailed long-distance travel information, including the origin and destination of the trip, stops along the way to and from the destination, side trips originating at the destination, the means of transportation, the reason of the trip, the lodging type, the number of nights spent away from home, and travel party size.

After deriving how many long-distance trips each person makes, the next step is to predict the tour characteristics, namely mode, destination, time of year, and duration. Each of these attributes are predicted using choice models. Figure 5-1 shows the model hierarchy for different trip purposes. The solid arrow indicates that the output of the upper level can be used as an explanatory variable at the lower level, while the dash arrow means that the expected utility of the lower-level models can affect the choices at the upper level. The hierarchy structure is different for business and pleasure trips; the long-distance pleasure activity requires people to consider their time availability prior to other decisions. When they have a period of time for pleasure, they will decide when to spend it, where to go, and how to go sequentially. In contrast, people engaging in long-distance business and personal business activities usually give priority to the decision of the activity location and time, including the time of year and duration, followed by the tour mode choice. All of these model components use either discrete choice models or duration model.

After individuals have made decisions about their travel to the main destination, they will make plans for their trips on the way to and from the destination. The stop level structure predicts the information of the intermediate stops people would make during their inbound/outbound legs of a long-distance tour. The stop frequency model at the higher level determines the number of intermediate stops people will make on the way to/from the tour destination. Once the number of stops on each tour leg is obtained, the purpose for each stop must be determined; the stop purpose category follows the same tour-level activity type, which are business, personal business and pleasure. At the lower tier of the stop-level structure, the location for each stop will be predicted with the similar method employed in the primary destination choice at the tour level.
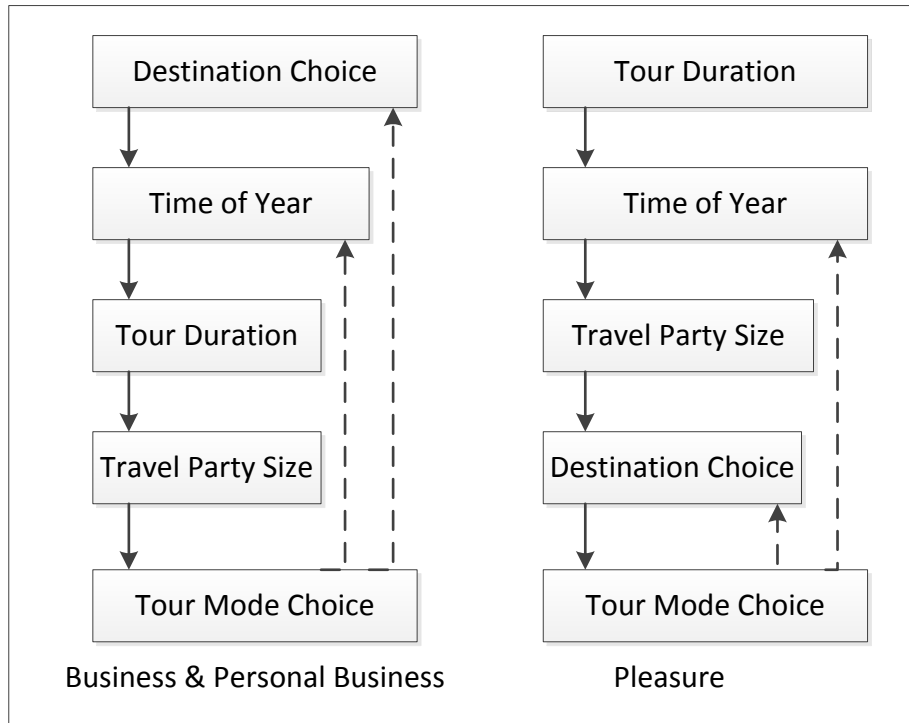
Figure 5-1. Tour level procedure and model components.

The model system is implemented in the developed micro-simulation platform, which simulates the individual's yearly long-distance activities and travel in the U.S. with the input of the population data, the associated transportation OD skim data, and economic/demographic data.

5.2. Model Components

The model system consists of three tiers (see Figure 5-2): 1) the yearly long-distance activity pattern level, which estimates the number of different types of activities a person will choose during one year; 2) the tour level model system, which contains choices of tour destination, time of year, tour duration, and tour mode; 3) the stop level model system, which estimates the intermediate stop frequency, and the purpose and location of each stop made during the inbound and outbound legs of the tour.
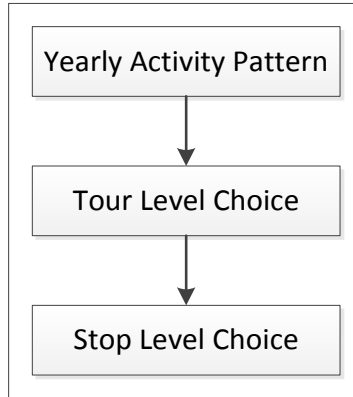
Figure 5-2. Activity-based long-distance travel demand model system.

5.2.1. *Activity Pattern Level Model*

The demand for long-distance activities and travel can be considered a choice among all the possible bundles of activities and travel annually. The model system adopts a timeframe of one year because of less frequent long-distance travel, and days, weeks, and even months of activity duration. Unlike a regular urban-level activity schedule, people choose their long-distance activities with few interactions within one year due to much less frequent long-distance travel. As shown in Figure 5-3, the yearly long-distance activity schedule can be presented as a set of different long-distance activities per year, and all the long-distance activities at this level are the primary activities. Therefore, the yearly long-distance activity pattern can be presented as {B-x, PB-y, P-z}, where B, PB, and P stand for the activity type of business, personal business, and pleasure respectively, and x, y, z are integers (x, y, z ≥ 0) referring to the number of the corresponding activities during one year. In the model, the multiple classification analysis (MCA) method, which is a frequently-used trip generation method in traditional four-step travel demand model, is employed to estimate the long-distance trip rates by activity type.
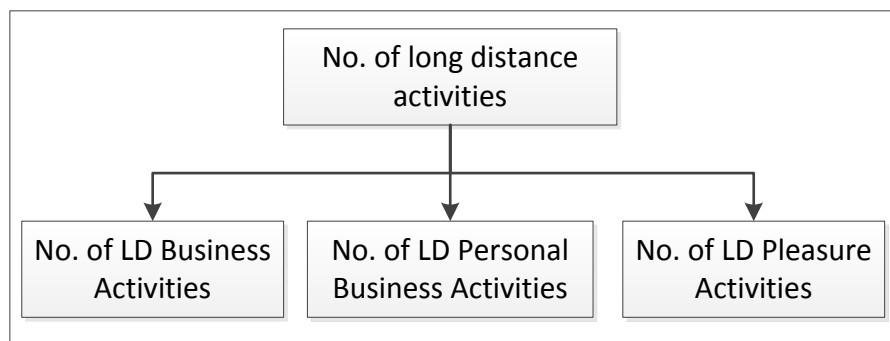
Figure 5-3. Yearly long-distance activity pattern level.

5.2.2. *Tour Level Structure*

Each long-distance activity schedule has a primary tour and may have zero or more intermediate stops or side stops during the legs of the tour and at the destination. In our model system, the secondary tours or the side stops that occur based on the long-distance primary destination are ignored due to the data limitation and its coverage of urban- or metropolitan-level travel. The tour-level model system defines the characteristics of the primary tour of each long-distance activity, such as the tour destination, time of year, tour duration, travel party size, and tour travel mode. When we develop and estimate each model component at the tour level, it is assumed that the outcomes of the upper-level model, the household, personal characteristics, and mobility attributes are already known. Therefore, the solid arrow in Figure 5-1 indicates that the output of the upper level can be used as an explanatory variable at the lower level, while the dash arrow means that the expected utility of the lower-level models can affect the choices at the upper level. Meanwhile, the upper level decisions of tour duration and destination will constrain the travel mode choice at the lower level; for instance, if individuals from Washington, D.C., only have one day to travel to California and get back, it is unlikely they will drive. In reality, the proportion of the long-distance tour travel time in total duration varies by person. As the 1995 ATS data have limitations on the activity duration information at the primary destination, we simplified the temporal constraint and made an assumption that for each person the total tour travel time should not exceed half of the total tour duration. When people decide to pursue a long-distance activity, they usually have different priority considerations and decision procedures for different activity types. In the research, we made a set of assumptions about people's decision-making process of long-distance travel at the tour level. For example, the long-distance pleasure activity (a discretionary activity) requires people to consider their time availability prior to other decisions. When they have a period of time (days, weeks or months) for pleasure, they will decide when to spend it, where to go, whom to go with, and how to go sequentially. In contrast, people taking long-distance business and personal business activities usually give priority to the decisions of the activity location and the time (including time of year and duration), followed by travel party size and tour mode choice. Therefore, two different tour level structures are proposed for business/personal business and pleasure (Figure 5-1). According to the direction of the dash lines in Figure 5-1, both time of year models and destination models should include the expected utility variable from the mode choice model (mode choice logsum).

All the model components are estimated mainly based on the 1995 ATS data, and most of the models are estimated using discrete choice model (multinomial logit model) except the tour duration model. In our research, we use 80% of the data sample to conduct the model estimation, while the remaining 20% is used to validate the estimated model.

5.2.2.1. Travel Mode Choice Model

Three travel modes are modeled at the tour level, i.e. {(car, car), (air, air), (train, train)}, and no combination of different travel modes is considered due to the small sample size in the ATS data. Multinomial logit model is employed to develop the travel mode choice model, and a piecewise linear utility function, shown in **Error! Reference source not found.**, is adopted.

$$U_{ij} = \alpha_1 \cdot tc_{ij.r1} + \alpha_2 \cdot tc_{ij.r2} + \ldots + \alpha_n \cdot tc_{ij.rn} + \beta \cdot tt_{ij} + \varepsilon_{ij} \qquad (5\text{-}1)$$

U_{ij}: the utility value of the person choosing travel mode i for long-distance activity j between a specific OD;

i: refers to the three tour level travel modes, {(car, car), (air, air), (train, train)};

j: one of the three long-distance activities (business, personal business, and pleasure);

$tc_{ij.r1}, tc_{ij.r2}, \ldots, tc_{ij.rn}$: total travel cost of mode i for jth long-distance activity when travel cost falls into the range of r1, r2, ..., rn;

tt_{ij}: total travel time using mode i for jth long-distance activity;

$\alpha_1, \alpha_2, \ldots, \alpha_n$: the coefficients of total travel cost for different travel cost ranges;

β: the total travel time coefficient;

ε_{ij}: error term capturing the factors that affect utility but are not observable by the researcher.

5.2.2.2.   Time of Year Choice

Because the finest temporal resolution in ATS is a quarter, our proposed model system functions at a time resolution of three-month or one-quarter. The three-month increments begin in January and end in December, thus four quarters in total. In the ATS sample data, few records are observed that depart from and arrive at home across quarters. Consequently, we adopt a choice set of only four alternatives {(Q1, Q1), (Q2, Q2), (Q3, Q3), (Q4, Q4)} for each person when he/she decides what time of the year to travel and what time of the year to get back, where Q1, …., Q4 refer to Quarter 1 to Quarter 4. Multinomial logit model is adopted for time of year choice model, and the model employs the personal and the zonal characteristics, most of which are generic across the four time alternatives. Since transportation network LOS attributes vary by time period, especially airfare with large price fluctuations in different seasons, these variables are specified as alternative-specific based on the four time alternatives. The general form of the TOY model utility can be represented in **Error! Reference source not found.**.

$$U_i = \alpha \cdot logsum_i + B_i \cdot X + \varepsilon_i \qquad (5\text{-}2)$$

$$logsum_i = ln \sum_k e^{V_{ik}}$$

Where,

U_i: the utility value of person choosing to travel during the time period i, where i=1, 2, 3, 4, refering to (Q1, Q1), (Q2, Q2), (Q3, Q3), (Q4, Q4);

$logsum_i$: mode choice logsum during time period i, and represents the total ease of travel between two TAZs across all available travel modes during time period i;

α: mode choice logsum coefficient;

X: vector of person's characteristics;

B_i: vector of person's characteristics coefficients for time alternative i;

V_{ik}: representative mode utility for the tour by mode k during time i.

At the tour level for pleasure long-distance activity, a simplified time of year choice model that only takes into account the person's characteristics is first developed and applied when the pleasure destination is not known. The time period assigned from this simplified model will serve as an input or known attributes for the destination choice model. Once the destination is chosen, a full time-of-year model that considers the mode choice logsum is re-run to choose the final time period (Figure 5-4).
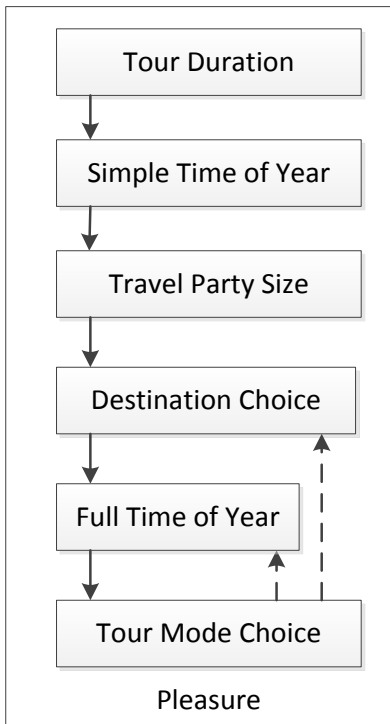
Figure 5-4. Re-simulating time-of-year choice model

5.2.2.3.   Tour Duration Choice Model

Different from the urban- or metropolitan-level travel demand model systems, the duration of long-distance trip is measured in days away from the origin and covers the entire time period starting from the origin and ending at the origin. The tour duration is modeled, as it could affect the travel distance and travel mode that people will choose when they plan long-distance travel. Hazard

duration model (survival analysis) analyzing the time to the occurrence of event is employed for the tour duration model. According to the feature, where the long-distance travel duration is recorded as days, the discrete time survival analysis method is utilized for tour duration choice (Gokovali, et al., 2007). In the discrete time survival analysis for tour duration, we consider each long-distance tour as a subject, and all the subjects are uncensored in a one-year calendar period. The longest duration for a long-distance tour is set at 31 days. The time unit is measured as day. The survival time T is a discrete random variable with probabilities:

$$f(t) = \Pr(T = t) \tag{5-3}$$

Where t represents the time interval. The discrete time survival function, which describes the chance that a person will survive beyond the time period t in question without experiencing the event, is given by **Error! Reference source not found.**, while the failure function giving the probability that the event has occurred by duration t is given by **Error! Reference source not found.**.

$$S(t) = \Pr(T \geq t) = \sum_{n=t}^{\infty} f_n \tag{5-4}$$

$$F(t) = 1 - S(t) \tag{5-5}$$

The hazard rate, which represents the probability that an event occurs given that one has survived to that time t, is:

$$h(t) = \Pr(T = t | T \geq t) = \frac{f(t)}{S(t-1)} \tag{5-6}$$

Given the hazard rate, the discrete time survival function can also be written in **Error! Reference source not found.**:

$$S(t) = (1 - h_1)(1 - h_2) \dots \dots (1 - h_{t-1})(1 - h_t) \tag{5-7}$$

The probability of the event occurring during the time interval t is:

$$\Pr(t - 1 < T \le t) = F(t) - F(t - 1) = S(t - 1) - S(t) \tag{5-8}$$

Two functions, including logistic regression function and complementary log-log function, can be used to fit the discrete-time hazard models (Allison, 1982); we adopted the logistic regression function for the hazard rate in our analysis.

$$\log\left(\frac{h_{it}}{1 - h_{it}}\right) = \alpha_t + \beta \cdot X_{it} \tag{5-9}$$

where h_{it} is the hazard rate and is the probability of an event occurring given that one (i) has survived to the time (t); i (1, 2,…n) refers to individual; t, taking on positive integer value, refers to the discrete time; α_t is the baseline hazard function; β is the coefficient vector of the covariates; X_{it} is the covariates or explanatory variables of individual i at time t. In the duration model, the explanatory variables or covariates are known features of the long-distance tour, person, and household characteristics. It is less likely that these attributes will change over time during the period of the long-distance tour. Therefore, the covariates are assumed time-independent. Meanwhile, multiple baseline hazard functions including log(time) and polynomial in time are tested. Little difference is shown in the model estimated coefficients (for the same covariates) and the model without sample validation results. In our model, the polynomial function of time was employed as the baseline hazard function. Therefore, the hazard rate function can be represented as:

$$\log\left(\frac{h_{it}}{1 - h_{it}}\right) = \gamma t + \theta t^2 + \beta \cdot X_{it} \tag{5-10}$$

$$h_{it} = \frac{1}{1 + \exp(-\gamma t - \theta t^2 - \beta \cdot X_{it})} \tag{5-11}$$

where $\gamma, \theta,$ and β are coefficients which need to be estimated.

5.2.2.4.   Travel Party Size Choice Model

The travel party size choice is modeled for each long-distance tour, and it determines how many persons participate in the tour. It also is assumed that no one will get on or off the tour during the long-distance travel. The model is a multinomial logit model, and each person will have a choice set of four alternatives (travelling alone, travelling with 1 person, with 2 persons, or with 3 or more persons) for all three long-distance activities (business, personal business, and pleasure). An individual travelling alone is set as the base alternative for all the three models. The explanatory variables mainly include the person and the household characteristics. According to the tour-level

66

model structure, the tour destination is determined and known before people decide the travel party size for business and personal business travel. Therefore, the zonal attributes can be utilized in the travel party size choice model for long-distance business and personal business tours.

5.2.2.5. Tour Destination Choice Model

The destination choice determines the location of the long-distance tour's primary destination. It works at the zonal level and each person is assigned a TAZ as his/her primary destination according to the multinomial logit destination choice model. In the 1995 ATS sample data, there are a total of 208 TAZs, which means that each person faces a choice set of 208 alternatives. In order to reduce the estimation time and complexity, the method of simple random sampling (SRS) is implemented due to the independency of irrelevant alternatives (IIA) property of a multinomial logit model (Nerella, et al., 2004). Consequently, each person has a destination choice sub-set of 10 alternatives, with one representing the person's chosen zone and nine randomly selected from the rest of 207 zones. Because people tend to take their long-distance pleasure and business trips in large cities, tourist attractions, or vacation locations, we added several dummy variables in the long-distance business and pleasure destination choice models to control people's preference for certain places. The zonal attractiveness variables in the destination choice models are primarily the zonal total employment, the number of households, and the dummy variable, indicating a metropolitan statistical area. As seen in **Error! Reference source not found.**, destination is first determined for long-distance business and personal business travel, and the travel time period is unknown. Therefore, in the destination choice model for business and personal business, the mode choice logsum is calculated with the average travel time and travel cost of each travel mode across four seasons.

5.2.3. Stop Level Structure

After individuals have made decisions about their travel to the main destination, they will make plans for their trips on the way to and from the destination based on the remained time. It is assumed that people have the same logic to determine their stops or trips during the tour legs regardless of the main activity types. Consequently, the same model structure at the stop level is applied to all three tour-level activity types (business, pleasure, personal business) (Figure 5-5). The stop level structure generates the information of the intermediate stops people would make during their inbound/outbound legs of the long-distance tour. A stop during the tour is defined as one people make for a certain purpose like business, personal business, or pleasure. Stops for rest or transfer in the same travel mode or across multiple travel modes are not the ones we analyze, and are excluded from the data set. At the stop level, the information about each long-distance tour such as the tour duration, travel mode, travel party size, and tour origin and destination are already known.
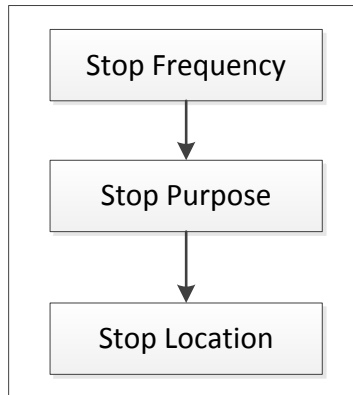
Figure 5-5. Stop level procedure and model components.

5.2.3.1.   Stop Frequency Choice Model

The stop frequency model at the higher level determines the number of intermediate stops people will have on the way to/from the tour destination. In each direction, a maximum number of four stops can be made, which results in a maximum of five trips on each tour leg. The stop frequency choice model for each half tour leg is developed using multinomial logit model, and each person faces a choice set of five alternatives (0, 1, 2, 3, 4) for each tour leg. Zero stops is set as the base alternative for both models. The models utilize the long-distance tour characteristics as the explanatory variables such as tour duration, tour mode, the activity type of the long-distance tour, time of year, and distance between tour origin and destination.

5.2.3.2.   Stop Purpose Choice Model

Once the number of stops a person will make during the long-distance tour is obtained, the purpose of each stop is determined (the middle-level model component in Figure 5-5) through the stop purpose choice model. The stop purpose category follows the same tour-level activity types that include business, personal business, and pleasure. The model is developed for each half tour using multinomial logit model, and the pleasure purpose is set as the base alternative. Both obtained stop-level and tour-level characteristics can be used as the explanatory variables in the stop purpose choice model for each tour leg, such as the sequence of the stop, the long-distance primary activity type, travel party size, and the tour travel mode.

5.2.3.3.   Stop Location Choice Model

At the low tier of the stop-level structure, the location for each stop is estimated with the similar method employed in the primary destination choice at the tour level. Before the stop location, we know the number of stops and the sequence of the stops during each half tour, the stop purpose, and tour origin and destination. As we assume that people only take one of the three modes (air, car, train) and do not transfer among different modes for the entire tour, the travel mode for each trip on each half leg will remain the same as the one estimated at the tour level. In the stop location choice, the distance between stop origin and stop location should be larger than 50 miles. The

68

short-distance travel based on the stop origin will not be incorporated and modeled. The same methodology as the one used in tour destination choice is employed. However, different from the tour-level primary destination choice, the impedance of travel to an intermediated stop in the stop location choice model should measure the additional impedance between the tour origin or stop origin and the tour primary destination, if it is an outbound trip (Bradley, et al., 2006). The main variables in stop location choice model are the out-of-direction or detour generalized travel cost and detour travel distance. For instance, the level of service (LOS) variables for the first stop on the way to the tour primary destination are based on the additional impedance between the tour origin and the tour destination (Figure 5-6), and the LOS for the following stop is based on the additional impedance between the previous stop and the tour destination (Figure 5-7). In the meantime, the tour origin becomes the stop origin for the stop, in the situation of Figure 5-6, and stop i will be called the stop origin of stop j in the Figure 5-7 situation. The same method works with the stops of the inbound direction but in an opposite way, as the anchor point is the tour origin instead of tour primary destination.

Detour travel cost = Cost (O, S) + Cost (S, D) − Cost (O, D)
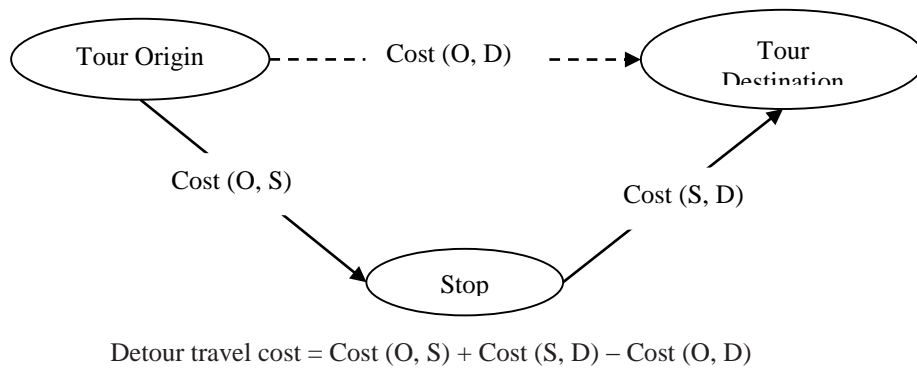
Figure 5-6. LOS estimation for the first stop during outbound tour leg.

Detour Travel Cost for Stop j = Cost (S_i, S_j) + Cost (S_j, D) − Cost (S_i, D)
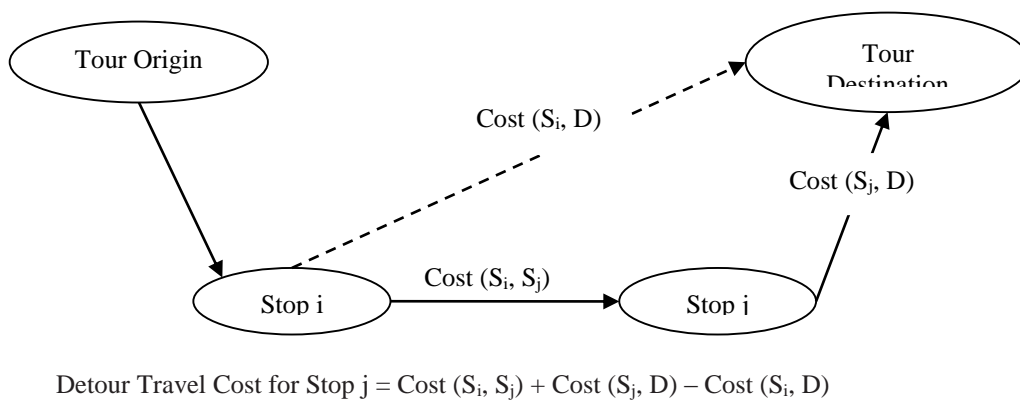
Figure 5-7. LOS estimation for the _j_th stop during outbound tour leg.

The detour generalized travel cost combines the detour travel cost and travel time components according to the time of value obtained from the tour level mode choice model.

6. FREIGHT MODEL

6.1. Overview

The freight model is developed out of a recently completed SHRP2-C20 project and integrated into MSTM2. The model is a designed to be multi-layer, where freight flows and the resulting vehicles movements at the local level are informed by and are sensitive to changes at the upper level. For example, changes over time in the freight flows to and from a region influence the demand for long-distance truck travel, but also the need for additional local truck movements to facilitate local deliveries and pickups of shipments.

The model represents the complete supply chain of good movements flowing in and out of the model region. This includes global suppliers and buyers who produce imports and consume exports, as well as business establishments across the country. That structure supports scenario testing to understand the impacts of changes in the economy over time, including different patterns of long haul domestic flows, and imports and exports as domestic and international trading partners change.

The model also moves from a traditional trip-based approach to a more realistic tour-based approach, where the travel patterns of trucks are modeled to mimic the routings that dispatchers and drivers develop. These more realistic travel patterns are inherently more responsive to the effects of congestion in the transportation system.

6.2. Model Components

Figure 6-1 shows the architectural design of the tour-based statewide multimodal freight model. The proposed approach consists of three components. The first component is a national supply chain model that constitutes of synthesis of firms, buyer-supplier selection, distribution channel selection, mode choices, and shipment size. The second component is a regional truck tour-based model that consists of local delivery and pick-up of shipments. Stop generation, choice, sequence, duration, and other tour characteristics are determined in this stage. Similarly, commercial vehicle tours are captured in the regional model. The third component is the statewide level, which integrates the output from the national and regional models to obtain an origin-destination matrix with higher resolution in the core focus area and lower resolution outside of the core area of MSTM.
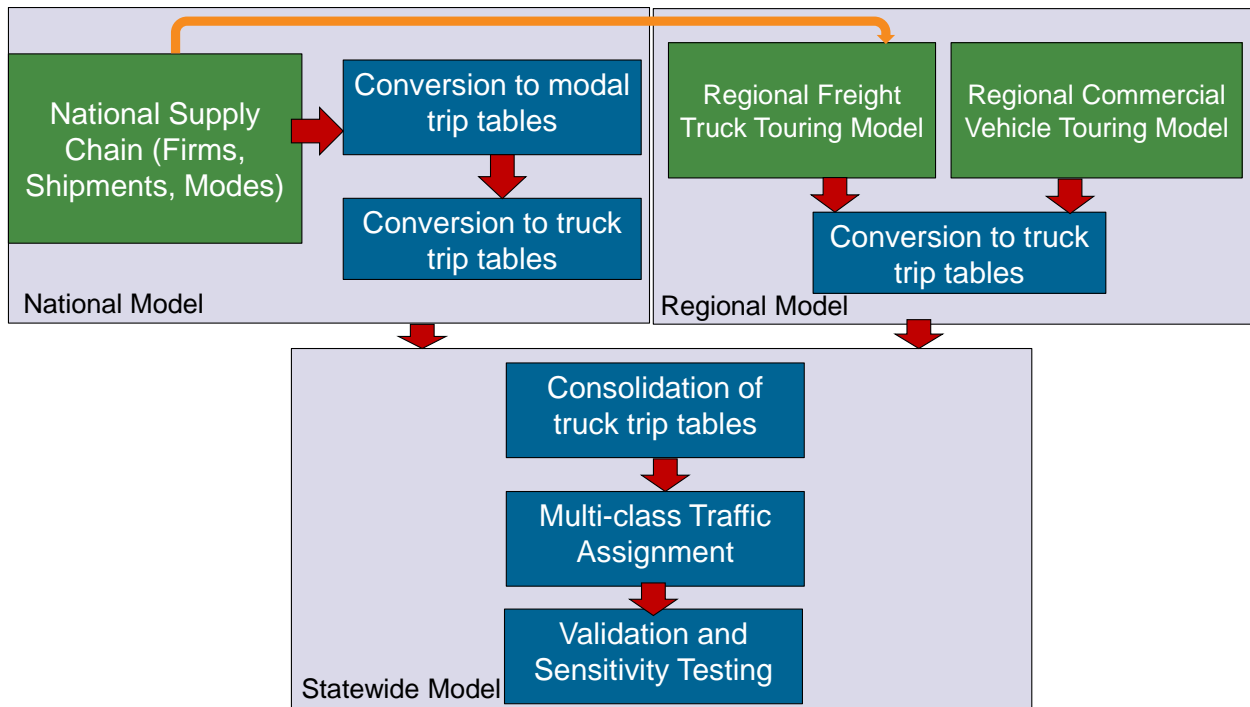
Figure 6-1. Freight model components.

6.3. Input Data

Table 6-1 shows five broad categories of input data needed for the freight model. Each of the data types are explained in the following sub-sections.

6.3.1. Employment Data

The employment data used in the MSTM models is broken down to a seven-category system. The seven categories cover the industrial, retail, office, education, health, and food employment sectors. The seventh category includes other employment classified as administrative, support, waste management, arts, entertainment, recreation, hotels, and other services. Each of the seven employment categories are broken down into different classification types.

Table 6-2 shows each employment category and how they are classified.

Table 6-1. Data Inventory for MSTM 2.0

Input Data Type	Input Data Need and Purpose	Source	Description	Availability
Employment Data	County Business Pattern (CBP) Data	U.S. Census	Establishments and employment by industry	Available
	Maryland Department of Planning (within MD)		Parcel level employment	Needed
	NETS (outside MD and within MSTM)	NETS	National Employment Time Series (NETS) point level employment	Needed
	Local Employment and Household Dynamics (LEHD)	U.S. Census	Employment by quarter with 15 NAICS code	Available
Economic Data	Input-Output Make-and-Use Tables	U.S. Bureau of Economic Analysis	Values of commodities exchanged between industries	Available
	NAICS6 to I/O Industry Correspondence	U.S. Bureau of Economic Analysis	Correspondences between NAICS6 industries and NAICS input-output industries	Available
Freight Flows	FAF3 Commodity Flows (or FAF4 if available)	FHWA	Commodity flows between FAF3 zones	Available
	Truck Trip Origins and Destinations	American Transportation Research Institute	Sample of truck trips by county and TAZ origin-destination	Needed
	Commodity-Flow Survey (CFS)	FHWA	Shipment size distributions by commodity	Available
Network Elements	Network Links	ORNL and US Army Corps of Engineers'	Rail (ORNL), and waterway network (U.S. Army Corps of Engineers') links	Needed
	Transport and Logistics Nodes (TLN)	ORNL, BTS	Intermodal nodes provided connections between networks	Needed
	Distribution Center Locations	InfoGroup	Warehousing and distribution locations for routing shipments	Needed
Validation and Other Data	Truck Counts	State DOT	Medium and heavy truck counts	Available
	Weigh-in-Motion Data	State DOT	Truck weight distribution	Needed
	Carload Waybill Data	USDOT (via State DOT)	Complete restricted dataset of carload waybills	Needed

	T-100 Data	BTS	Air freight segment and market data	Needed
	PIERS Data	PIERS	Import/export shipment data by Port	Needed

Table 6-2. MSTM Employment Categories and Classifications

Variable	NAICS Code	Description	Model Description
Industrial Employment	11	Agriculture, Forestry, Fishing, and Hunting	Industrial
	21	Mining	Industrial
	22	Utilities	Industrial
	23	Construction	Industrial
	31	Manufacturing	Industrial
	32	Manufacturing	Industrial
	33	Manufacturing	Industrial
	42	Wholesale Trade	Industrial
	48	Transportation	Industrial
	49	Warehousing	Industrial
Retail Employment	44	Retail Trade	Retail
	45	Retail Trade	Retail
Office Employment	51	Information	Office
	52	Finance and Insurance	Office
	53	Real Estate	Office
	54	Professional, Scientific, and Technical	Office
	55	Management of Companies and Enterprises	Office
	92	Government	Office
Education Employment	61	Educational Services	Education
Health Employment	62	Health Care	Medical Services
Food Employment	722	Food Service	Food Service
Other Employment	56	Administrative, Support, Waste Management	Other Services
	71	Arts, Entertainment, and Recreation	Other Services
	721	Hotels	Other Services
	81	Other Services	Other Services

The MSTM model has a geographical extent covering North America; the geographical extent can be seen in Figure 6-2. The MSTM model has three resolutions, where the low resolution includes country size, Canada, and Mexico—for example, a medium resolution of zone sizes at the state level, and a high resolution where zone sizes range from about 30 zones per county to over 300 zones per county within Maryland. Figure 6-2 highlights the MSTM buffer zone total employment for 2012.
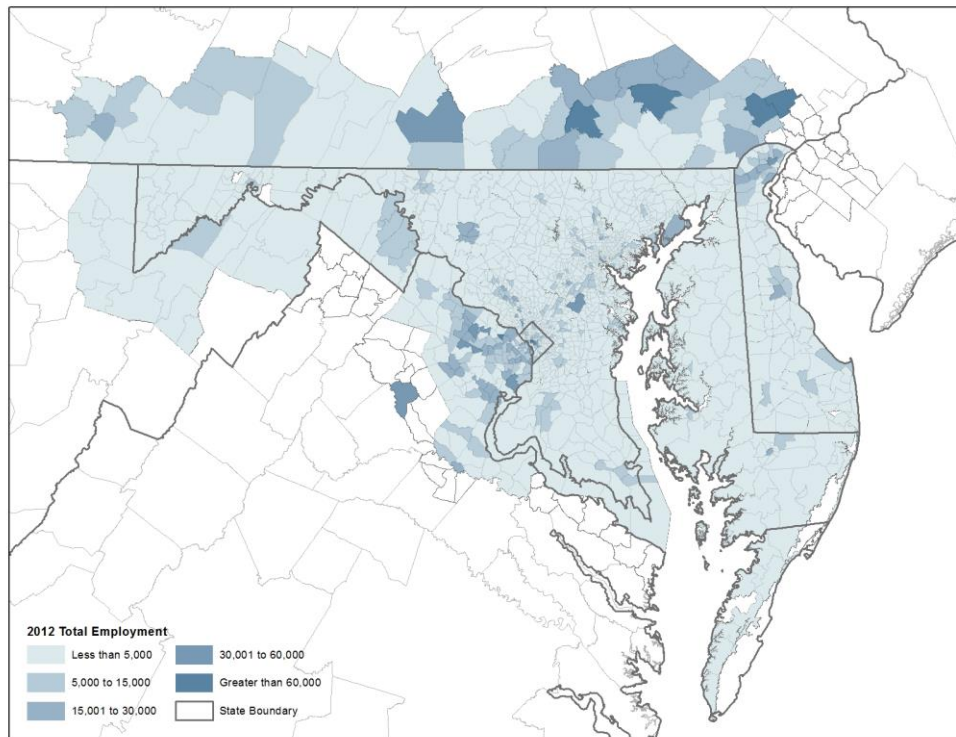
Figure 6-2. MSTM Zone Total Employment 2012

6.3.1.1. National Employment Data

County-level employment data for the United States outside of Maryland, in the form of CBP data, are used to synthesize firms for all of the model area except for Maryland. For each county, this dataset contains the number of firms in each category, defined by industry and number of employees. Industry is defined based on the North American Industrial Classification System (NAICS) using six-digit classifications. The dataset is an annual series that provides sub-national economic data by industry. This series includes the number of establishments, employment, first quarter payroll, and annual payroll.

In addition to the CBP data, the model used Infogroup business data covering Maryland. The Infogroup database has more detailed employment and business location data for Maryland than the CBP data. The Infogroup data also overcome the limitations of the CBP data for agricultural and construction employment. This dataset contains information on the geographic location of the

business, primary business conducted at the location, number of employees, and other useful information.

The two sets of employment data are used in conjunction with information from the Make and Use table to apportion FAF4 flows. The Make and Use table provides the value of goods that are traded between different industries. The resulting values are weighted by employment for apportioning the flows. Table 6-3 shows the number of firms by employment category and size.

Table 6-3. Number of Firms by Employment

NAICS2	INDUSTRY DESCRIPTION	1 TO 19	20 TO 99	100 TO 249	250 TO 499	500 TO 999	1K TO 5K	5K TO 10K	>10K	TOTAL
11	Agriculture, etc.	3468	271	56	24	9	2	0	0	3830
21	Mining, Quarrying, etc.	503	60	13	5	0	2	0	0	583
22	Utilities	869	152	36	9	5	2	0	0	1073
23	Construction	74609	5000	686	116	28	6	1	0	80446
31	Manufacturing	25658	3761	655	162	84	27	0	0	30347
42	Wholesale Trade	35036	2683	427	112	40	17	1	0	38316
44	Retail Trade	107606	7616	1705	294	98	22	0	0	117341
48	Transportation & Warehousing	17934	1479	268	66	37	14	0	0	19798
51	Information	13338	1056	297	87	27	19	1	0	14825
52	Finance & Insurance	40016	2077	222	66	35	28	0	0	42444
53	Real Estate, etc.	49239	2367	249	48	12	8	0	0	51923
54	Professional Services, etc.	76629	3084	320	68	21	11	0	0	80133
55	Management of Companies, etc.	595	63	5	1	0	2	0	0	666
56	Administrative Services, etc.	43140	1972	341	93	26	16	0	2	45590
61	Educational Services	9995	3967	1520	114	29	16	1	2	15644
62	Health Care & Social Assistance	67314	4545	924	247	108	90	8	2	73238
71	Arts, etc.	12603	1161	207	50	17	10	2	0	14050
72	Accommodation Services, etc.	35079	10384	952	136	67	26	0	0	46644
81	Other Services	87136	2433	235	41	16	6	1	0	89868
92	Public Administration	14206	3514	644	240	101	47	2	1	18755
	Total	714973	57645	9762	1979	760	371	17	7	785514

6.3.1.2. Foreign Employment Data

The CBP data does not contain foreign employment data. The primary objective of including foreign firms in the model is to ensure that international flows between Maryland and foreign countries can be allocated to either buyers or supplier firms at the foreign country end. The firm synthesis model accomplished this by generating a representative firm in each of the three largest

employment size groups for each type of industry by NAICS6 code in each of the foreign FAF4 zones. A total of 9,096 foreign firms are generated for use in the model.

6.3.1.3.   Supplemental Employment Data

Longitudinal Employer-Household Dynamics (LEHD) is a national longitudinal job frame that combines data from state and federal sources to create a linked employer-employee dataset. These data are collated by the Census Bureau and cover approximately 90% of employed persons.

Agriculture data on farms by size and sales were derived from U.S. Department of Agriculture data to provide supplemental information in understanding agricultural production locations. While these data were not used in the development of the model, they are available to support refinements to the employment data in future work.

6.3.2. Economic Data

The Bureau of Economic Analysis' Input-Output Make and Use tables (2007 benchmark tables) is used to create supply chains in the supplier firm selection component. For each production industry, the table reports the value of goods consumed by each buyer industry. The model uses this information to identify the most important commodities that are consumed for each buyer industry and their associated supplier industries.

The reported values of goods exchanged between producers and consumers are also used to apportion FAF4 flows by commodity type spatially (among the constituent counties within a FAF4 zone in Maryland) and by industry. These apportionments use information from the Make and Use table to determine the total volume of a commodity that is produced or consumed by a particular industry, compared to other industries that produce or use the commodity. Table 6-4 shows an example of the detailed use table. This table shows the commodities used for "oilseed farming" and "coal mining" industries by the producer's and the purchaser's value.

Table 6-4. Detailed Use Table Sample Data View

Commodity	Commodity Description	Industry	Industry Description	Producer Value	Purchaser Value
1111A0	Oilseed farming	1111A0	Oilseed farming	1025.2	1137.6
325320	Agricultural chemical manuf.	1111A0	Oilseed farming	508.4	702.9
324110	Petroleum refineries	1111A0	Oilseed farming	413.4	462.4
1111B0	Grain farming	1111A0	Oilseed farming	320.4	320.4
325310	Fertilizer manufacturing	1111A0	Oilseed farming	269.8	316.6
212100	Coal mining	212100	Coal mining	1199.4	1970.7

333120	Construction machinery manuf.	212100	Coal mining	628.7	760.7
212310	Stone mining and quarrying	212100	Coal mining	396.8	691.8
331110	Iron, steel mills & ferroalloy manuf.	212100	Coal mining	144.9	177.7
336300	Motor vehicle parts manuf.	212100	Coal mining	143.4	214.3

6.3.3. Commodity Flow Data

6.3.3.1. Freight Analysis Framework (FAF4)

The model relies on the FAF4 database, which is a FHWA freight data product that has United States domestic and international freight flows for calendar year 2012. FAF4 reports the annual tons moved and the monetary value of the shipments. Maryland FAF4 zones are shown in Figure 6-3. FAF4 categorizes the freight into 43 Standard Classification of Transported Goods (SCTG) commodity classes and reports movements by seven modes (truck, rail, waterway, air [includes truck-air], multiple modes and mail, pipeline, and other/unknown) for each origin and destination (FAF4 zones). The 43 SCTG commodity groups (Table 6-5) were used in different groupings throughout the model and are described in the respective model steps. About 237,000,000 tons originate from Maryland, of which about 38.5% stays within Maryland, according to FAF4 data for 2012. About 98% of Maryland-to-Maryland movements by weight and 95% by value are by truck. These figures are 88% and 85% for Maryland to other states movement, respectively. Freight destined for Maryland totals 160,600,000 tons, of which 80% is by truck. Figure 6-4 shows freight flows from Maryland and Figure 6-5 shows freight flows to Maryland.
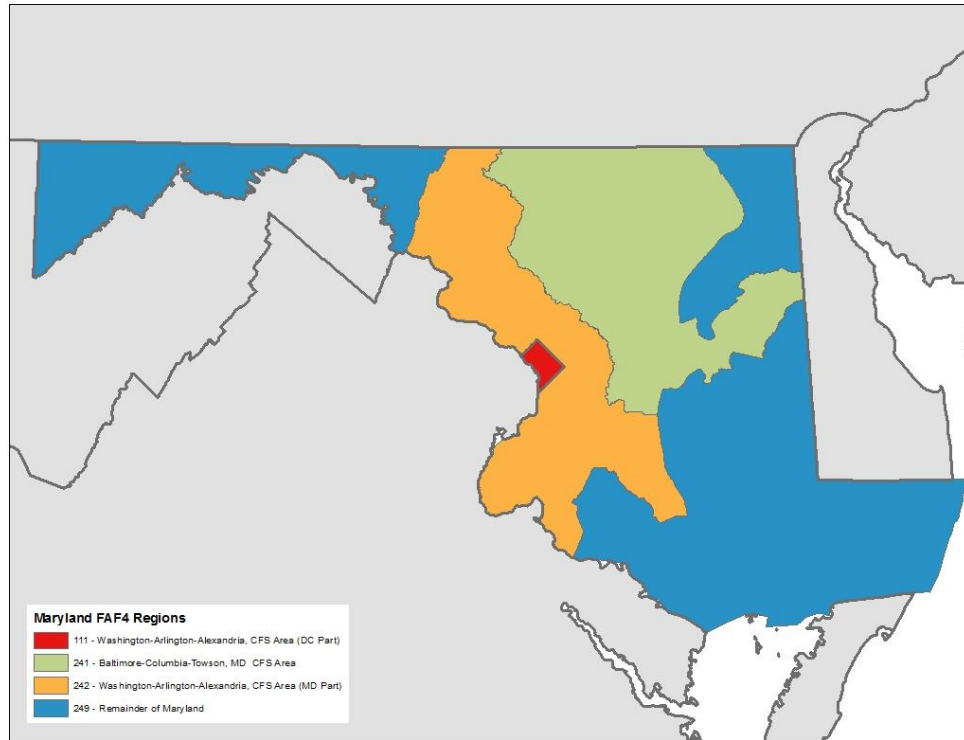
Figure 6-3. Maryland FAF4 regions.

Table 6-5. Commodity Groups and Categories

1	Live animals/fish	Live animals and live fish	Animals	Functional/Innovative
2	Cereal grains	Cereal grains	Bulk natural resource (BNR)	Functional
3	Other ag. prods.	Other agricultural products	Bulk natural resource (BNR)	Functional/Innovative
4	Animal feed	Animal feed and products of animal origin, n.e.c.	Intermediate processed goods (IPG)	Functional
5	Meat/seafood	Meat, fish, seafood, and their preparations	Finished goods (FG)	Functional/Innovative
6	Milled grain prods.	Milled grain products and preparations, bakery products	Finished goods (FG)	Functional
7	Other foodstuffs	Other prepared foodstuffs and fats and oils	Finished goods (FG)	Functional
8	Alcoholic beverages	Alcoholic beverages	Finished goods (FG)	Functional
9	Tobacco prods.	Tobacco products	Finished goods (FG)	Functional
10	Building stone	Monumental or building stone	Bulk natural resource (BNR)	Functional

11	Natural sands	Natural sands	Bulk natural resource (BNR)	Functional
12	Gravel	Gravel and crushed stone	Bulk natural resource (BNR)	Functional
13	Nonmetallic minerals	Nonmetallic minerals n.e.c.	Bulk natural resource (BNR)	Functional
14	Metallic ores	Metallic ores and concentrates	Bulk natural resource (BNR)	Functional
15	Coal	Coal	Bulk natural resource (BNR)	Functional
16	Crude petroleum	Crude petroleum	Bulk natural resource (BNR)	Functional
17	Gasoline	Gasoline and aviation turbine fuel	Intermediate processed goods (IPG)	Functional
18	Fuel oils	Fuel oils	Intermediate processed goods (IPG)	Functional
19	Coal-n.e.c.	Coal and petroleum products, n.e.c.	Bulk natural resource (BNR)	Functional
20	Basic chemicals	Basic chemicals	Intermediate processed goods (IPG)	Functional
21	Pharmaceuticals	Pharmaceutical products	Finished goods (FG)	Functional/Innovative
22	Fertilizers	Fertilizers	Intermediate processed goods (IPG)	Functional
23	Chemical prods.	Chemical products and preparations, n.e.c.	Intermediate processed goods (IPG)	Innovative
24	Plastics/rubber	Plastics and rubber	Intermediate processed goods (IPG)	Functional
25	Logs	Logs and other wood in the rough	Bulk natural resource (BNR)	Functional
26	Wood prods.	Wood products	Intermediate processed goods (IPG)	Functional
27	Newsprint/paper	Pulp, newsprint, paper, and paperboard	Intermediate processed goods (IPG)	Functional
28	Paper articles	Paper or paperboard articles	Intermediate processed goods (IPG)	Functional/Innovative
29	Printed prods.	Printed products	Finished goods (FG)	Functional
30	Textiles/leather	Textiles, leather, and articles of textiles or leather	Finished goods (FG)	Functional/Innovative
31	Nonmetal min. prods.	Nonmetallic mineral products	Finished goods (FG)	Functional/Innovative

32	Base metals	Base metal in primary or semi-finished forms and in finished basic shapes	Intermediate processed goods (IPG)	Functional
33	Articles-base metal	Articles of base metal	Intermediate processed goods (IPG)	Functional
34	Machinery	Machinery	Finished goods (FG)	Functional/Innovative
35	Electronics	Electronic and other electrical equipment and components and office equipment	Finished goods (FG)	Innovative
36	Motorized vehicles	Motorized and other vehicles (including parts)	Finished goods (FG)	Innovative
37	Transport equip.	Transportation equipment, n.e.c.	Finished goods (FG)	Functional
38	Precision instruments	Precision instruments and apparatus	Finished goods (FG)	Innovative
39	Furniture	Furniture, mattresses and mattress supports, lamps, lighting fittings, and illuminated signs	Finished goods (FG)	Functional/Innovative
40	Misc. mfg. prods.	Miscellaneous manufactured products	Finished goods (FG)	Innovative
41	Waste/scrap	Waste and scrap	Other	Functional
43	Mixed freight	Mixed freight	Finished goods (FG)	Functional/Innovative
99	Unknown	Commodity unknown	Other	Functional

The freight flow data are used in two ways. First, the OD pairs reported in the FAF4 data are used to identify the location of candidate suppliers for every buyer during supplier firm selection. Second, the flow data are apportioned to individual supplier-buyer pairs in goods demand.
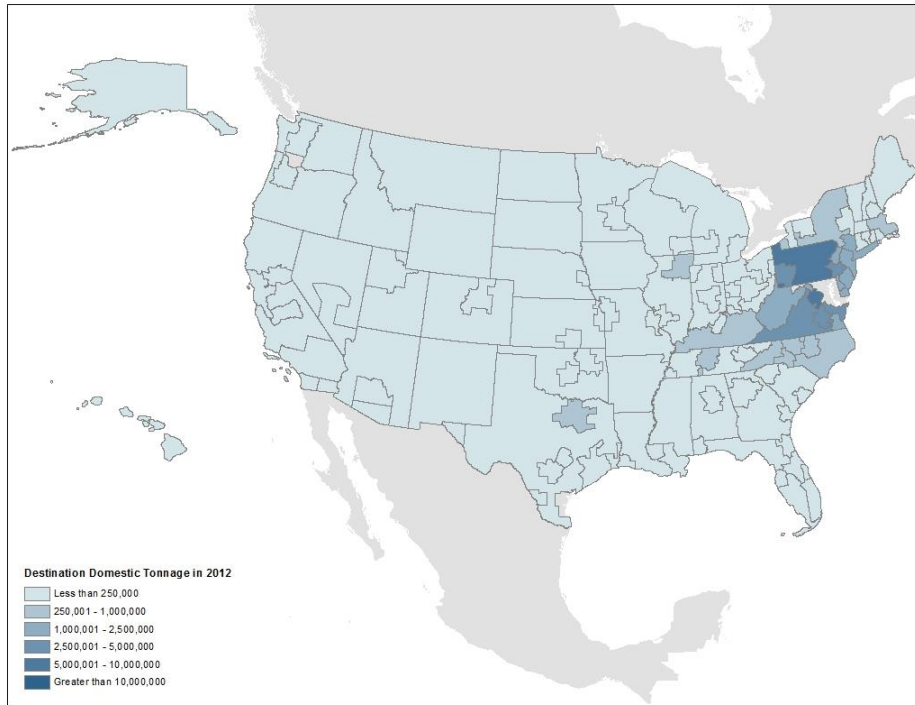
Figure 6-4. 2012 freight flows from Maryland.
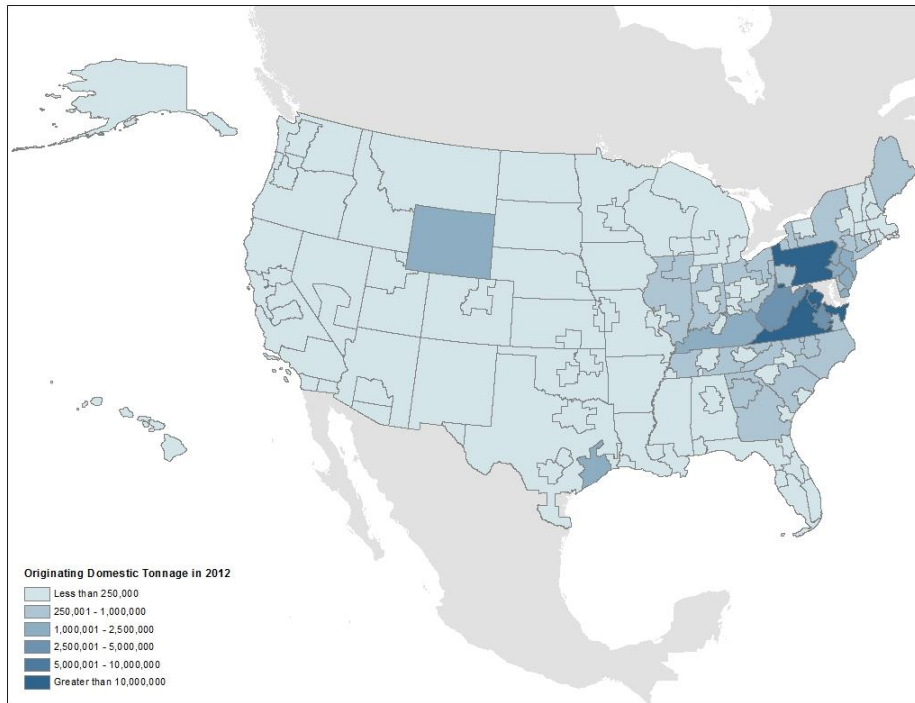
Figure 6-5. 2012 freight flows to Maryland.

Table 6-6 presents the most important commodities that are transported within Maryland by weight. This reflects that gravel, non-metal mineral products, and gasoline products are the top commodities by weight transported within Maryland.

Table 6-6. Top II Commodity Movement by Weight

Commodity	Tons	Percent of Total
Gravel & Crushed Stone	26,281,417	29%
Non-Metallic Mineral Products	9,766,901	11%
Gasoline Products	8,170,177	9%
Natural Sands	6,502,947	7%
Waste & Scrap	6,479,888	7%
Other Prepared Foodstuffs	4,531,981	5%
Other Coal and Petroleum Products	3,385,777	4%
Wood Products	3,358,363	4%
Other Non-Metallic Minerals	2,779,453	3%
Animal Feel, Eggs, Honey & Other Animal Products	2,465,287	3%

6.3.4. Network Elements

This section of the document discusses each dataset used for the development of the freight network. The freight network for MSTM2 was built from five components. These components include:

- Railroad network
- Waterway network
- Railroad and waterway network nodes
- Network connectors
- Intermodal Facility locations

  o Airports

  o Freight terminals

  o Sea Port Facilities

  o U.S. Major Sea Port Facilities

Figure 6-6 shows the multimodal network.
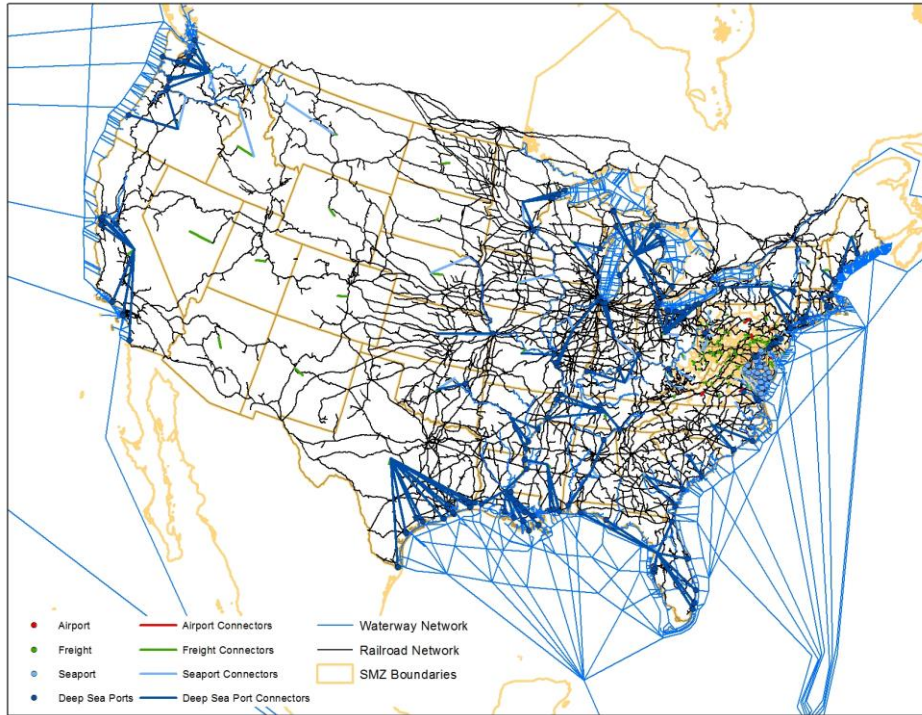
Figure 6-6. Multimodal freight network.

6.4. Model Results

The model produced 32 trip tables encompassing four vehicle classes (commercial vehicles, short distance single-unit trucks, short distance multi-unit trucks, and long-distance trucks) and eight times of the day. The resulting truck trip tables were assigned to MSTM network and Figure 6-7 shows the assigned volumes. We observe that the higher assigned volumes are shown on major corridors as expected.
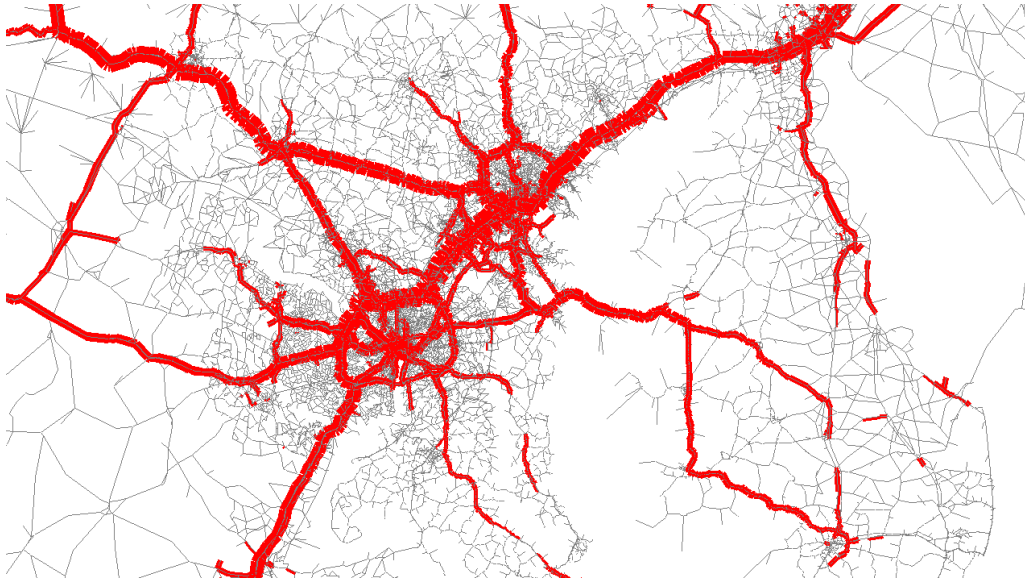
Figure 6-7. Total daily assigned trucks.

7. INTEGRATED MODEL

7.1. Overview

A statewide transportation model involves both passenger and freight travels covering short-distance (shorter than 50 miles) and long-distance (longer than 50 miles) trips. Previous sections have introduced each individual model component of MSTM2. This section discusses how MSTM2 integrates the aforementioned model components and models the impacts of various types of travels on the statewide network. Generally, MSTM2 contains three main model components: an activity-based short-distance person travel model, an agent-based microsimulation long-distance person travel model, and a tour-based freight model. The outputs of these three model components, typically in the form of origin-destination (OD) matrices, are integrated and assigned to the statewide network so that the influences of different travel forms on the Maryland region can be captured jointly.

Unlike MSTM version 1, which was aggregated in nature, all model components in MSTM2 are at the individual level. In other words, all three model components in MSTM2 simulate the travel behavior of each individual travel agent. MSTM2 is also the first statewide model to feature individual-level travel analyses when modeling both passenger and freight trips. This feature enhances our understanding of individual travel behavior in the region and allows studies of person-based policy scenarios, such as variable-rate vehicle mileage pricing.

MSTM2 provides two model running options. In the first option, only the short-distance person travel component is executed (long-distance and freight components are not executed for the interest of runtime). Demand inputs (i.e., OD matrices) from both long-distance and freight models are joined with the short-distance demand matrices before the multiclass assignment. This option applies to policy scenarios that impose insignificant impacts to long-distance and freight travels. In the second option, all three model components are executed. This option requires longer runtime but can reflect the effects of proposed scenarios on not only short-distance travels but also long-distance and freight travels.

7.2. Implementation and Applications

MSTM2 is implemented in CUBE Catalog (Figure 7-1). Since all three model components are developed in different coding environments (e.g., C# and JAVA), CUBE Catalog provides a powerful platform to call various components, and takes care of network skimming and other intermediate steps. Additionally, CUBE Catalog comes with a very user-friendly graphical user interface (GUI), which makes it easier to manage various model components. It is very convenient to customize a model run regardless if one wants to run MSTM2 as a whole or only run a specific model component under the CUBE Catalog environment.
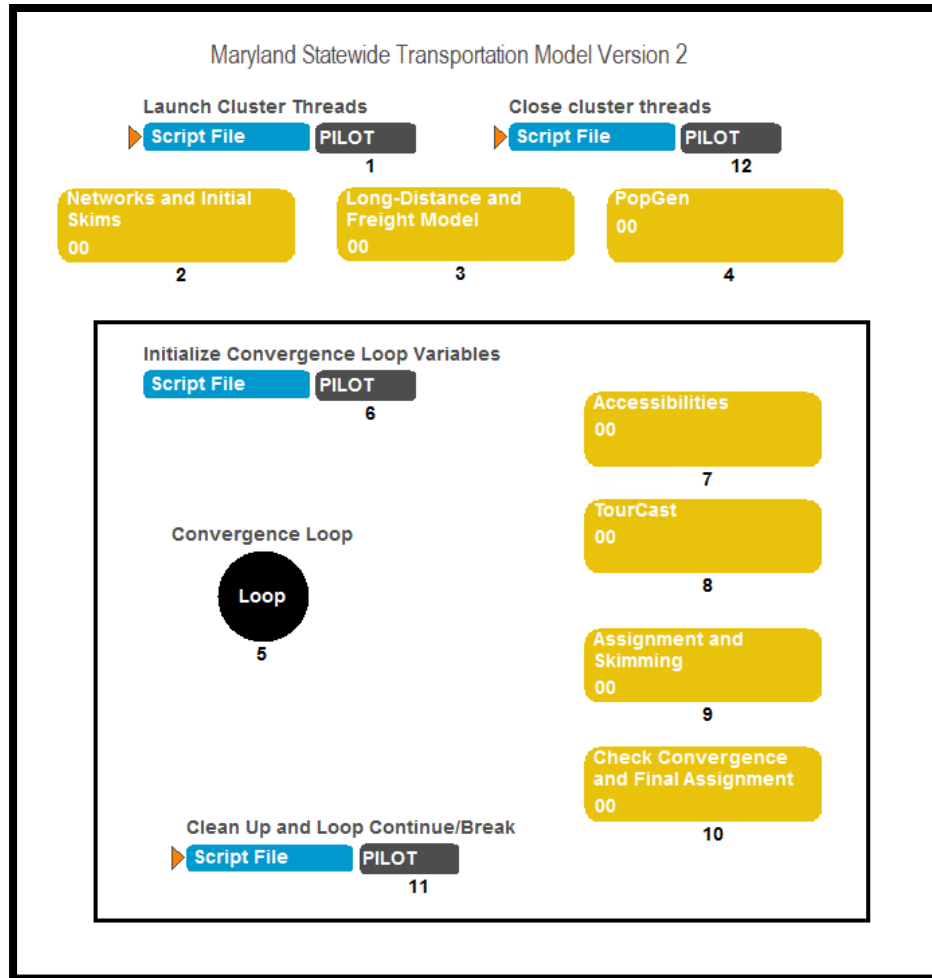
Figure 7-1. MSTM2 graphical user interface.

MSTM2 is executed in numeric order as shown in Figure 7-1. Model components outside the black box are for preparation purposes and only performed for the first iteration. Specifically, Module 2 skims the initial transportation network and generates skim matrices for both highway and transit options based on the free-flow condition. Module 3 either calls both long-distance and freight models or processes OD matrices from these two model components, depending on which run option is specified by the user. Module 4 calls the population synthesizer PopGen to generate the synthetic population for the modeling area for the short-distance model. Modules inside the black box are executed iteratively until a pre-defined iteration number or specific model convergence criteria are met.

Within the black box, Module 7 calculates accessibility measures for the short-distance person travel model based on different network conditions. Module 8 calls TourCast, which is the model engine for the short-distance model. Module 9 combines demand inputs from all three model components and performs a multi-class assignment on the statewide network. Module 10 checks two criteria: model convergence requirement (typically gap-based convergence measures) and pre-

specified iteration number. If either criterion is met, MSTM2 model run is terminated. Otherwise, modules within the black box will be executed for another iteration.

8. CALIBRATION, VALIDATION, AND SCENARIO TESTING

8.1. Short-Distance Model Validation

This chapter summarizes the short-distance activity-based model component validation. The tests consisted of comparisons of model results in the BMC area for various market segments to the expanded BMC household survey data. These comparisons were done in Excel spreadsheet files. The model application software, TourCast, outputs .dbf files that were imported into a relational database and processed with stored procedures using MySQL. The processed summaries were exported to comma delimited files that can be read directly into the Excel spreadsheets, which were populated in advance with the survey data results. The model results presented in this section are based on a model application with three iterations of speed feedback.

The comparisons described in this section reflect model calibration adjustments. In some cases, model parameters were adjusted to produce more reasonable results, although there was not a universal attempt to match all results from the expanded household survey for all market segments by adjusting model constants or other parameters. This type of adjustment was only made when the uncalibrated model results did not appear reasonable and the survey data results were based on a substantial number of observations. The specific calibration adjustments are documented in the Excel files.

Because of the extensive number of comparisons, the spreadsheet files themselves are incorporated as appendices to this report. The remainder of this section summarizes the validation results as presented in the spreadsheet files.

8.1.1. Long Term Choice Models

8.1.1.1. Vehicle Availability Model

The vehicle availability model simulates the number of vehicles owned by each household in the synthetic population. The Excel file with the results of the vehicle availability model is *VehicleAvailability.xlsm*. Table 8-1 summarizes the regional results of the calibrated model. On a regional basis, the number of households by number of vehicles owned matches well.

Table 8-1. Vehicle Availability Model – Regional Validation

Vehicles	Expanded household survey data		Model Results		Percentage Point Difference	Percentage Difference
	Households	Percentage	Households	Percentage		
0	231,695	11.2%	208,160	10.0%	-1.2%	-10.3%
1	690,202	33.3%	652,543	31.4%	-1.9%	-5.6%
2	753,072	36.3%	806,458	38.8%	2.5%	6.9%
3+	398,131	19.2%	409,036	19.7%	0.5%	2.6%
Total	2,073,100		2,076,197			

The more detailed comparisons in the Excel files show the following results:

- The modeled percentage of households owning each number of vehicles matches the survey data well for each county in the model region. The model slightly underestimates vehicle ownership in Carroll County, the smallest county in the model region.
- Vehicle availability levels were compared for cross-classifications of household size (1, 2, 3, 4+) and by income level (<$15,000, $15,000-$29,999, $30,000-$49,999, $50,000-$99,999, >$100,000). The model results match the expanded survey data well.
- Vehicle availability levels were compared for cross-classifications of number of workers (0, 1, 2, 3, 4+) and by income level (<$15,000, $15,000-$29,999, $30,000-$49,999, $50,000-$99,999, >$100,000). The model results match the expanded survey data well, considering the relatively low number of households surveyed for many of the cells.
- Vehicle availability levels were compared for cross-classifications of number of workers (0, 1, 2, 3+) by number of children (0, 1, 2+). The model results match the expanded survey data well, again considering the relatively low number of households surveyed for many of the cells, especially those representing households with zero vehicles.

8.1.1.2. Regular Workplace Location

The regular workplace location model simulates whether each worker in the synthetic population has a regular workplace and the location of that workplace. The Excel file with the results of the regular workplace location model is *UsualWork.xlsm*. Table 8-2 summarizes the regional modeled and observed (from the survey data set) percentages of workers by type (full time, part time, and senior) with regular workplaces. The survey data closely matches the survey data.

Table 8-2. Percentage of Workers by Type with Regular Workplaces

Worker Status	Expanded household survey data			Model			Diff.
	No Usual Workplace	Total	Percentage	No Usual Workplace	Total	Percentage	

Full-Time	317,530	2,143,942	14.8%	175,681	2,317,432	7.6%	-7.2%
Part-Time	84,267	266,015	31.7%	72,338	339,193	21.3%	-10.4%
Total	401,796	2,409,957	16.7%	248,019	2,656,625	9.3%	-7.3%

Error! Reference source not found. and **Error! Reference source not found.** show the comparison between the observed (survey) and modeled tour length frequency distributions for distance and highway time, respectively. While there are some differences in the distributions, the fits are good; the coincidence ratios are 92% for both distance and time. The average tour times are 25.3 minutes (observed) and 25.5 minutes (modeled); the average tour distances are 13.1 miles (observed) and 12.7 miles (modeled).

The more detailed comparisons in the Excel files show the following results:

- Full time workers have longer tour lengths than part time workers. The model results match the survey results well in this case.
- The distance between home and the regular workplace increases with income. In the survey data, this increase is a little steeper than in the model results.
- The distance between home and the regular workplace increases as the home location becomes less urban; the survey data trend is well reflected in the model results.
- The modeled percentage of workers whose regular workplaces are in the same zone as their homes (the "intra-zonal percentage") is 1.8%, compared to 2.0% in the survey data.
- The modeled and observed intra-zonal percentages are slightly lower for full time workers than for part time workers.
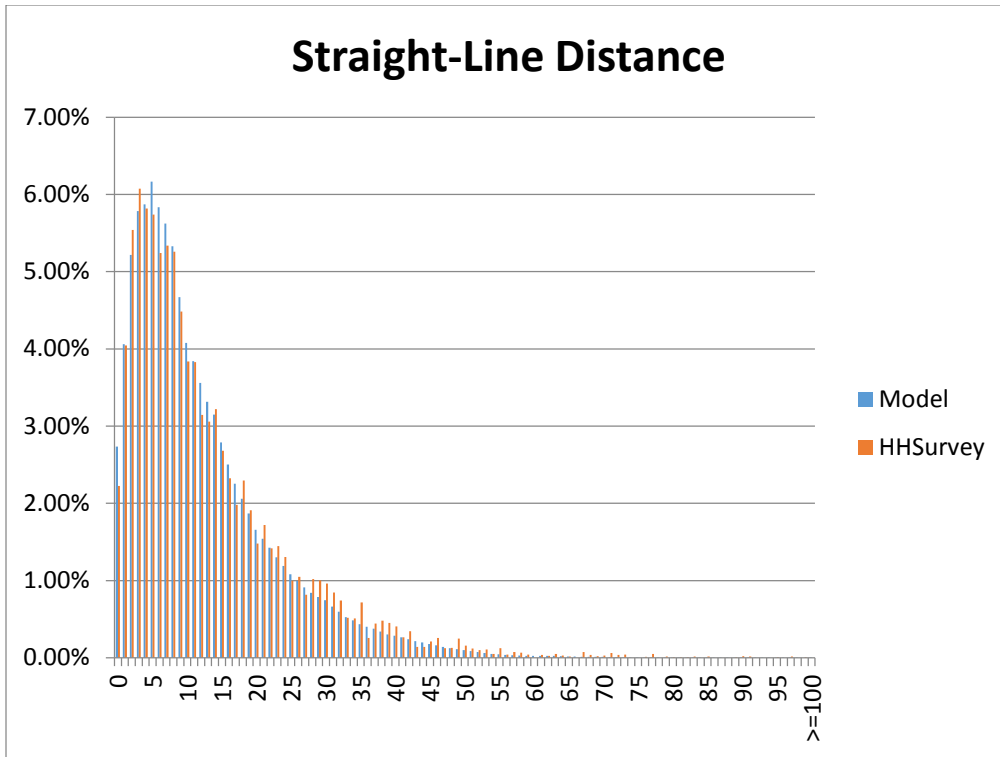
Figure 8-1. Home to regular workplace tour length frequency distribution (distance).
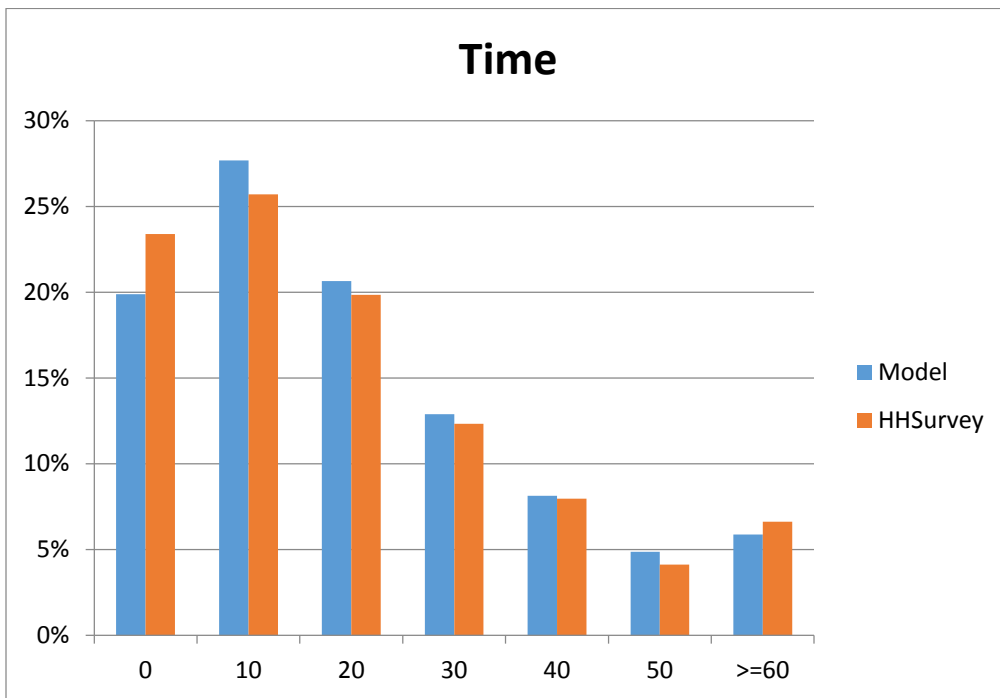

Figure 8-2. Home to regular workplace tour length frequency distribution (time).

8.1.1.3. School Location

The school location model simulates the school location for each child in the synthetic population. The Excel file with the results of the school location model is *SchLocation.xlsm*.

Error! Reference source not found. and **Error! Reference source not found.** show the comparison between the observed (survey) and modeled tour length frequency distributions for distance and time, respectively. While there are some differences in the distributions, the fits are good; the coincidence ratios are 82% for distance and 70% for time. The average tour times are 11.4 minutes (observed) and 11.5 minutes (modeled); the average tour distances are 6.0 miles (observed) and 6.0 miles (modeled).
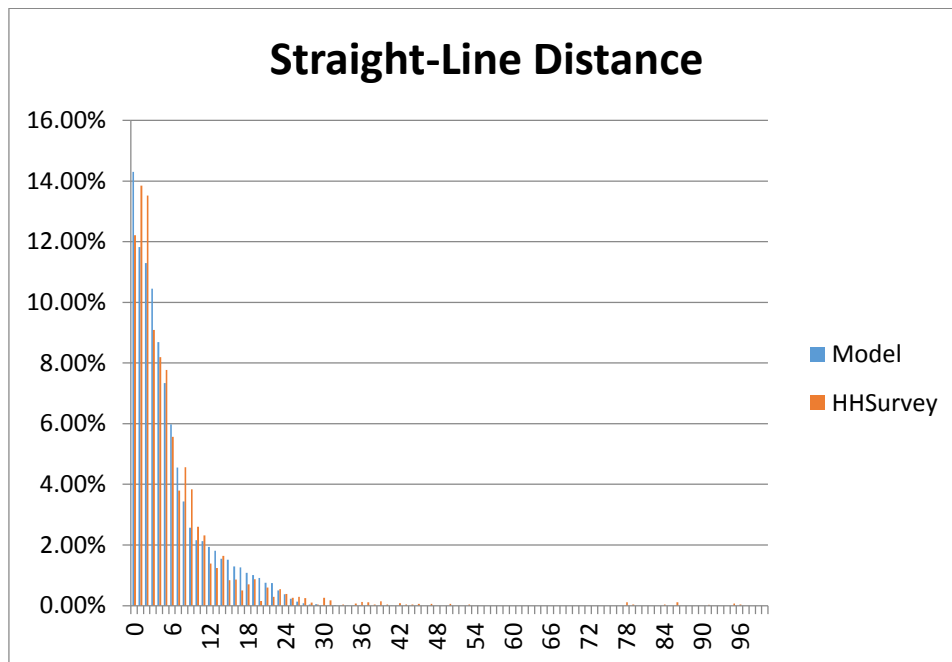

Figure 8-3. Home to school tour length frequency distribution (distance).
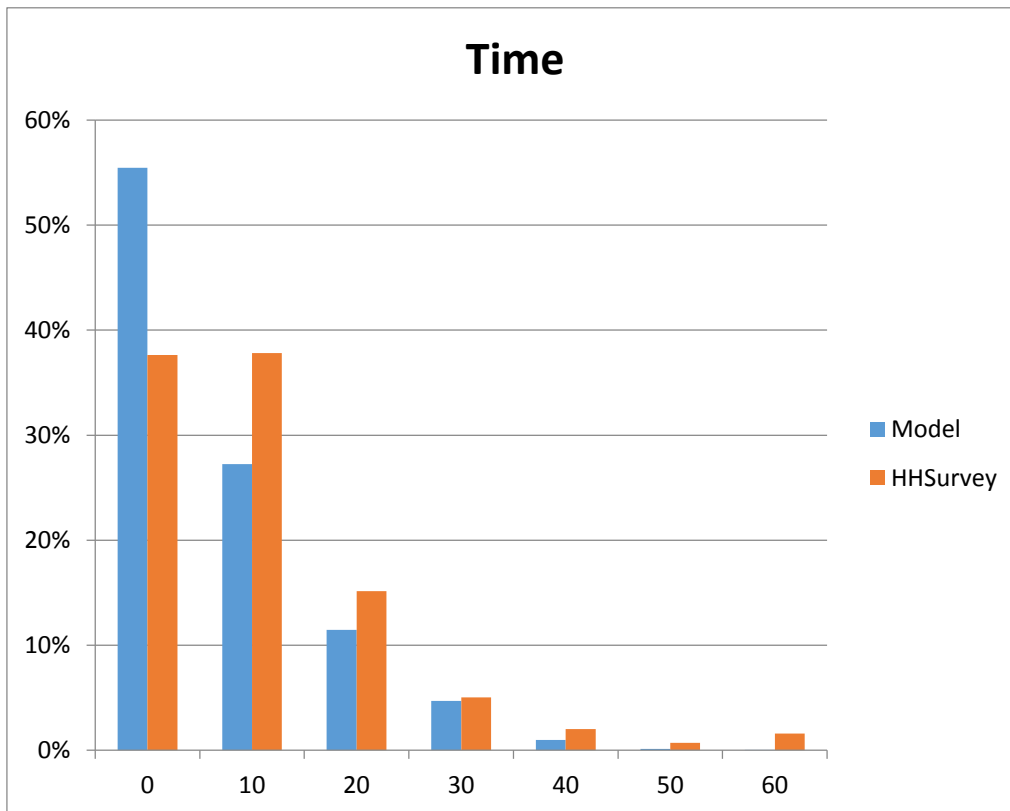
Figure 8-4. Home to school tour length frequency distribution (time).

The more detailed comparisons in the Excel files show the following results:

- Both the survey data set and the model results show that children between the ages of 5 and 15 have shorter school trips than younger children, who have shorter trip lengths than students age 16 and older. This reflects that high school students often travel longer distances to school than younger children, and that pre-school children may travel farther to day care than the distance to elementary school.
- The model results show a slight increase in trip length to school as income increases. This is generally true in the survey data although the trend is not consistent (and seems illogical).
- The highest percentage of students who go to school in their zone of residence is for students ages 6 to 15, in both the survey data and model results. The overall percentage of students who attend school in their residence zone is 12% (survey) and 11% (model).

8.1.1.4. Transit Pass Ownership

The transit pass ownership model simulates whether each household in the synthetic population has a transit pass. The Excel file with the results of the transit pass ownership model is *Transit Pass Ownership.xlsm*.

Table 8-3 compares the modeled and observed transit pass ownership by county. The modeled percentage of households with transit passes matches the observed percentages well within the BMC region. The modeled percentages are low in the MWCOG region, especially in Washington, D.C.

Table 8-3. Comparison of Modeled and Observed Transit Pass Ownership by County

County	Survey			Model Results			
	Yes	No	Survey	Yes	No	Model	Difference
Baltimore City	24,120	228,088	10%	20,705	216,077	9%	-0.8%
Baltimore County	10,528	236,218	4%	14,516	295,706	5%	0.4%
Anne Arundel	8,717	164,813	5%	7,727	190,810	4%	-1.1%
Howard	6,456	119,521	5%	4,729	100,455	4%	-0.6%
Carroll	1,415	96,741	1%	2,021	60,317	3%	1.8%
Harford	1,515	122,941	1%	4,387	85,272	5%	3.7%
Montgomery/Prince George's/Frederick	111,908	665,605	14%	45,925	696,530	6%	-8.2%
D.C.	71,095	203,421	26%	26,310	234,828	10%	-15.8%
Total	**235,754**	**1,837,348**	**11%**	**126,320**	**1,879,995**	**6%**	**-5.1%**
BMC region	**52,751**	**968,322**	**5%**	**54,085**	**948,637**	**5%**	**0.2%**

8.1.1.5.   E-ZPass Transponder Ownership

The E-ZPass transponder ownership model simulates whether each household in the synthetic population has a transponder. The Excel file with the results of this model is *E-ZPass Ownership.xlsm*.

It should be noted that the household survey data set did not include E-ZPass transponder ownership because the survey did not ask whether households owned transponders. The observed data for comparison therefore is obtained from an alternate source—a data set obtained by BMC from the Maryland Transportation Authority (MDTA) that provided the number of transponders owned by zip code. The number of "commercial" transponders (for example, the "standard business plan") was removed from the totals prior to comparison.

The use of this alternate observed data source means that the observed data does not exactly correspond to the number of households with transponders. Most notably, the MDTA data set counts the *number of transponders*, not the *number of households with transponders*. If a household has more than one transponder, the MDTA data set would count multiple transponders. In addition, some households may own E-ZPass transponders obtained from agencies in other states. Furthermore, some commercial vehicles may have transponders that do not fall into the excluded categories, while some personal vehicles may have transponders that are counted in the commercial categories. There are also some differences between the survey period (2007-2008) and the relatively recent MDTA data set.

Table 8-4 compares the modeled and observed transponder ownership by county. The model results show some noticeable differences from the observed data by county. While some model calibration was performed, it was decided not to adjust the model results too much given the different nature of the MDTA data set.

Table 8-4. Comparison of Modeled and Observed E-ZPass Transponder Ownership by County

County	Observed			Modeled			
	Households	Number of Passes	% Households with Passes	Households	Number of Passes	% Households with Passes	% Diff.
Baltimore City	252,208	66,654	26%	252,718	46,247	18%	-8%
Baltimore County	246,748	165,833	67%	318,820	151,812	48%	-20%
Anne Arundel	173,529	90,189	52%	202,188	148,250	73%	21%
Howard	125,976	45,981	36%	107,719	75,412	70%	34%
Carroll	98,156	9,816	10%	63,098	36,723	58%	48%
Harford	124,455	80,360	65%	91,762	68,410	75%	10%
Montgomery/Prince Georges/Frederick	777,512	177,348	23%	764,828	390,920	51%	28%
DC	274,517	23,852	9%	275,064	46,614	17%	8%
Total	2,073,100	660,033	32%	2,076,197	964,388	46%	15%
BMC Region	1,295,588	482,685	37%	1,311,369	573,468	44%	6%

8.1.2.   Daily Activity Pattern and Related Models

8.1.2.1.   Daily Activity Pattern Model

The daily activity pattern model simulates whether each person in the synthetic population has mandatory (work, university, or school) activities, has non-mandatory activities only, or makes no travel within the region (i.e., stays at home, is temporarily out of the model region, or has only external travel—travels only between home and locations outside the model region). If a mandatory activity pattern is chosen, the number of mandatory tours (zero, one, or two) is simulated, as well as whether any simulated work tours have stops.

Excel files summarize the results of the daily activity pattern model for each person type:

- Senior - DAP_Senior.xlsm
- Full time worker - DAP_FTW.xlsm
- Part time worker - DAP_PTW.xlsm
- Adult (university) student - DAP_Adult Student.xlsm
- Non-working adult - DAP_NWA.xlsm

- Child age less than 5 - DAP_Child1.xlsm
- Child age 5-15 - DAP_Child2.xlsm
- Child age 16 or older - DAP_Child3.xlsm

Table 8-5 through Table 8-12 summarize the regional results of the calibrated daily activity pattern model for each person type.

The Excel files show the results segmented by various variables of interest, including county of residence, household size, income level, vehicle availability, and gender. These comparisons show only minor differences between the survey data and the model results (though in many cases, the large number of alternatives in the daily activity pattern model means that the survey data has few observations for several of the alternatives for many of the market segments).

Table 8-5. Regional Comparison of Daily Activity Patterns: *Full Time Worker*

Daily Activity Pattern Type	Expanded Household Survey Data		Model Results	
	Count	Percentage	Count	Percentage
1 Work Tour, No Stops	986,563	42.6%	955,922	41.2%
1 Work Tour, with Stops	729,477	31.5%	783,134	33.8%
2 Work Tours, No Stops	47,491	2.0%	36,614	1.6%
2 Work Tours, Stops on One	36,882	1.6%	30,623	1.3%
2 Work Tours, Stops on Both	10,098	0.4%	9,168	0.4%
1 Univ. Tour/1 Work Tour, No Stops	5,966	0.3%	4,599	0.2%
1 Univ. Tour/1 Work Tour, Stops on Work Tour	3,892	0.2%	3,355	0.1%
1 School Tour/1 Work Tour, No Stops	0	0.0%	0	0.0%
1 School Tour/1 Work Tour, Stops on Work Tour	0	0.0%	0	0.0%
1 Univ. Tour	10,445	0.5%	10,477	0.5%
2 Univ. Tours	0	0.0%	0	0.0%
1 School Tour	0	0.0%	0	0.0%
2 School Tours	0	0.0%	0	0.0%
Non-Mandatory Travel Only	244,707	10.6%	267,483	11.5%
Stay at Home/Out of Area/ External Travel Only	241,912	10.4%	216,057	9.3%
Total	2,317,432	100%	2,317,432	100%

Table 8-6. Regional Comparison of Daily Activity Patterns: *Part Time Worker*

Daily Activity Pattern Type	Expanded Household Survey Data		Model Results	
	Count	Percentage	Count	Percentage
1 Work Tour, No Stops	72,243	21.3%	70,730	20.9%
1 Work Tour, with Stops	62,267	18.4%	69,263	20.4%
2 Work Tours, No Stops	3,353	1.0%	2,187	0.6%
2 Work Tours, Stops on One	4,477	1.3%	3,047	0.9%
2 Work Tours, Stops on Both	2,081	0.6%	1,532	0.5%
1 Univ. Tour/1 Work Tour, No Stops	0	0.0%	1,724	0.5%
1 Univ. Tour/1 Work Tour, Stops on Work Tour	108	0.0%	99	0.0%
1 School Tour/1 Work Tour, No Stops	0	0.0%	0	0.0%
1 School Tour/1 Work Tour, Stops on Work Tour	0	0.0%	0	0.0%
1 Univ. Tour	277	0.1%	405	0.1%
2 Univ. Tours	0	0.0%	0	0.0%
1 School Tour	0	0.0%	0	0.0%
2 School Tours	0	0.0%	0	0.0%
Non-Mandatory Travel Only	136,425	40.2%	138,126	40.7%
Stay at Home/Out of Area/ External Travel Only	57,963	17.1%	52,080	15.4%
Total	339,193	100%	339,193	100%

Table 8-7. Regional Comparison of Daily Activity Patterns: *Adult Student*

Daily Activity Pattern Type	Expanded Household Survey Data		Model Results	
	Count	Percentage	Count	Percentage
1 Work Tour, No Stops	19,694	7.6%	19,331	7.4%
1 Work Tour, with Stops	19,338	7.4%	19,091	7.3%
2 Work Tours, No Stops		0.0%	0	0.0%
2 Work Tours, Stops on One		0.0%	0	0.0%
2 Work Tours, Stops on Both	0	0.0%	0	0.0%
1 Univ. Tour/1 Work Tour, No Stops	4,111	1.6%	3,989	1.5%
1 Univ. Tour/1 Work Tour, Stops on Work Tour	2,076	0.8%	2,064	0.8%
1 School Tour/1 Work Tour, No Stops	0	0.0%	0	0.0%
1 School Tour/1 Work Tour, Stops on Work Tour	0	0.0%	0	0.0%
1 Univ. Tour	92,835	35.7%	91,869	35.2%
2 Univ. Tours	4,127	1.6%	4,131	1.6%
1 School Tour	0	0.0%	0	0.0%
2 School Tours	0	0.0%	0	0.0%
Non-Mandatory Travel Only	75,908	29.2%	76,026	29.1%
Stay at Home/Out of Area/ External Travel Only	41,699	16.1%	44,832	17.2%
Total	259,788	100%	261,333	100%

Table 8-8. Regional Comparison of Daily Activity Patterns: *Senior*

Daily Activity Pattern Type	Expanded Household Survey Data		Model Results	
	Count	Percentage	Count	Percentage
1 Work Tour, No Stops	8,322	1.8%	8,452	1.8%
1 Work Tour, with Stops	6,859	1.5%	8,067	1.8%
2 Work Tours, No Stops	838	0.2%	0	0.0%
2 Work Tours, Stops on One	570	0.1%	0	0.0%
2 Work Tours, Stops on Both	0	0.0%	0	0.0%
1 Univ. Tour/1 Work Tour, No Stops	0	0.0%	0	0.0%
1 Univ. Tour/1 Work Tour, Stops on Work Tour	0	0.0%	0	0.0%
1 School Tour/1 Work Tour, No Stops	0	0.0%	0	0.0%
1 School Tour/1 Work Tour, Stops on Work Tour	0	0.0%	0	0.0%
1 Univ. Tour	1,775	0.4%	791	0.2%
2 Univ. Tours	0	0.0%	0	0.0%
1 School Tour	0	0.0%	0	0.0%
2 School Tours	0	0.0%	0	0.0%
Non-Mandatory Travel Only	272,942	59.7%	263,859	57.7%
Stay at Home/Out of Area/ External Travel Only	165,985	36.3%	176,122	38.5%
Total	457,291	100%	457,291	100%

Table 8-9. Regional Comparison of Daily Activity Patterns: *Non-Working Adult*

Daily Activity Pattern Type	Expanded Household Survey Data		Model Results	
	Count	Percentage	Count	Percentage
1 Work Tour, No Stops	393	0.1%	467	0.1%
1 Work Tour, sith Stops	357	0.1%	472	0.1%
2 Work Tours, No Stops	0	0.0%	0	0.0%
2 Work Tours, Stops on One	0	0.0%	0	0.0%
2 Work Tours, Stops on Both	0	0.0%	0	0.0%
1 Univ. Tour/1 Work Tour, No Stops	0	0.0%	0	0.0%
1 Univ. Tour/1 Work Tour, Stops on Work Tour	0	0.0%	0	0.0%
1 School Tour/1 Work Tour, No Stops	0	0.0%	0	0.0%
1 School Tour/1 Work Tour, Stops on Work Tour	0	0.0%	0	0.0%
1 Univ. Tour	587	0.1%	369	0.1%
2 Univ. Tours	0	0.0%	0	0.0%
1 School Tour	0	0.0%	0	0.0%
2 School Tours	0	0.0%	0	0.0%
Non-Mandatory Travel Only	398,552	68.4%	397,174	68.2%
Stay at Home/Out of Area/ External Travel Only	182,503	31.3%	183,910	31.6%
Total	582,392	100%	582,392	100%

Table 8-10: Regional Comparison of Daily Activity Patterns: *Child Age Less than 5*

Daily Activity Pattern Type	Expanded Household Survey Data		Model Results	
	Count	Percentage	Count	Percentage
1 Work Tour, No Stops	0	0.0%	0	0.0%
1 Work Tour, with Stops	0	0.0%	0	0.0%
2 Work Tours, No Stops	0	0.0%	0	0.0%
2 Work Tours, Stops on One	0	0.0%	0	0.0%
2 Work Tours, Stops on Both	0	0.0%	0	0.0%
1 Univ. Tour/1 Work Tour, No Stops	0	0.0%	0	0.0%
1 Univ. Tour/1 Work Tour, Stops on Work Tour	0	0.0%	0	0.0%
1 School Tour/1 Work Tour, No Stops	0	0.0%	0	0.0%
1 School Tour/1 Work Tour, Stops on Work Tour	0	0.0%	0	0.0%
1 Univ. Tour	0	0.0%	0	0.0%
2 Univ. Tours	0	0.0%	0	0.0%
1 School Tour	125,930	37.3%	133,788	39.6%
2 School Tours	456	0.1%	0	0.0%
Non-Mandatory Travel Only	126,069	37.3%	132,190	39.1%
Stay at Home/Out of Area/ External Travel Only	85,476	25.3%	71,953	21.3%
Total	337,931	100%	337,931	100%

Table 8-11. Regional Comparison of Daily Activity Patterns: *Child Age 5-15*

Daily Activity Pattern Type	Expanded Household Survey Data		Model Results	
	Count	Percentage	Count	Percentage
1 Work Tour, No Stops	365	0.0%	0	0.0%
1 Work Tour, with Stops	167	0.0%	0	0.0%
2 Work Tours, No Stops	0	0.0%	0	0.0%
2 Work Tours, Stops on One	0	0.0%	0	0.0%
2 Work Tours, Stops on Both	0	0.0%	0	0.0%
1 Univ. Tour/1 Work Tour, No Stops	0	0.0%	0	0.0%
1 Univ. Tour/1 Work Tour, Stops on Work Tour	0	0.0%	0	0.0%
1 School Tour/1 Work Tour, No Stops	701	0.1%	0	0.0%
1 School Tour/1 Work Tour, Stops on Work Tour	457	0.1%	0	0.0%
1 Univ. Tour	0	0.0%	0	0.0%
2 Univ. Tours	0	0.0%	0	0.0%
1 School Tour	513,502	70.1%	531,673	72.6%
2 School Tours	10,047	1.4%	10,717	1.5%
Non-Mandatory Travel Only	125,639	17.2%	117,309	16.0%
Stay at Home/Out of Area/ External Travel Only	81,658	11.1%	72,837	9.9%
Total	732,536	100%	732,536	100%

Table 8-12. Regional Comparison of Daily Activity Patterns: *Child Age 16-17*

Daily Activity Pattern Type	Expanded Household Survey Data		Model Results	
	Count	Percentage	Count	Percentage
1 Work Tour, No Stops	4,783	3.6%	5,042	3.8%
1 Work Tour, with Stops	1,123	0.8%	1,654	1.2%
2 Work Tours, No Stops	0	0.0%	0	0.0%
2 Work Tours, Stops on One	0	0.0%	0	0.0%
2 Work Tours, Stops on Both	0	0.0%	0	0.0%
1 Univ. Tour/1 Work Tour, No Stops	0	0.0%	0	0.0%
1 Univ. Tour/1 Work Tour, Stops on Work Tour	0	0.0%	0	0.0%
1 School Tour/1 Work Tour, No Stops	4,699	3.5%	5,789	4.3%
1 School Tour/1 Work Tour, Stops on Work Tour	924	0.7%	934	0.7%
1 Univ. Tour	0	0.0%	0	0.0%
2 Univ. Tours	0	0.0%	0	0.0%
1 School Tour	87,307	64.9%	86,685	64.5%
2 School Tours	3,080	2.3%	2,601	1.9%
Non-Mandatory Travel Only	17,482	13.0%	17,152	12.8%
Stay at Home/Out of Area/ External Travel Only	15,052	11.2%	14,592	10.9%
Total	134,449	100%	134,449	100%

8.1.2.2. School Escorting Model

For each child traveling to school, the school escorting model determines whether he or she is escorted by another household member to school or from school, and, if so, which household member does the escorting, and whether that household member escorts the student as part of a mandatory tour (for example, on the way to or from work). The Excel file that summarizes the results of the school escorting model is SchoolEscort.xlsm.

Table 8-13 presents a summary of the comparison of the percentage of school escorting alternatives from the survey data set and the model results by child age group (0-4, 5-15, and 16+). In this table, the five alternatives for each student are:

- Outbound mandatory – escorting to school as part of a mandatory tour
- Outbound stand-alone – escorting to school as part of a stand-alone tour
- Return mandatory – escorting from school as part of a mandatory tour

- Return stand-alone – escorting from school as part of a stand-alone tour
- None – student is not escorted

Table 8-14 shows the comparison of escort person types between the survey data set and model results. Both Table 8-13 and Table 8-14 show relatively close agreement between the observed and model results.

The more detailed comparisons in the Excel files show the following results:

- The survey data shows that very little school escorting occurs in zero car households. The model results reflect this unsurprising result.
- Both the survey data and model results show that most escorts are full time workers or non-working adults.
- The household survey data show that 72% of escorts are female; the model results show a lower percentage of female escorts (62%).
- Generally, fewer children from higher income households are escorted, especially for the youngest children.

Table 8-13. Regional Comparison of School Escorting Alternatives

Escort Type	Child Age	Expanded Household Survey Data		Model Results		Percentage Point Difference (Model - Survey)
		Count	Percentage	Count	Percentage	
Outbound mandatory	< 5 Years	53,716	20.9%	54,958	20.5%	-0.4%
Outbound standalone	< 5 Years	39,172	15.2%	38,501	14.4%	-0.9%
Return mandatory	< 5 Years	49,848	19.4%	51,756	19.3%	-0.1%
Return standalone	< 5 Years	39,941	15.5%	41,492	15.5%	0.0%
None	< 5 Years	74,263	28.9%	80,869	30.2%	1.3%
Total	< 5 Years	256,940		267,576		
Outbound mandatory	5-15 Years	78,304	7.3%	115,881	10.5%	3.2%
Outbound standalone	5-15 Years	114,090	10.6%	84,261	7.6%	-3.0%
Return mandatory	5-15 Years	48,057	4.5%	78,485	7.1%	2.6%
Return standalone	5-15 Years	99,897	9.3%	75,770	6.8%	-2.4%
None	5-15 Years	734,989	68.3%	751,817	68.0%	-0.4%
Total	5-15 Years	1,075,337		1,106,214		
Outbound mandatory	16+ Years	8,528	4.3%	6,736	3.4%	-0.9%
Outbound standalone	16+ Years	11,927	6.0%	21,528	10.9%	4.9%
Return mandatory	16+ Years	3,840	1.9%	5,327	2.7%	0.8%
Return standalone	16+ Years	6,166	3.1%	3,982	2.0%	-1.1%
None	16+ Years	167,920	84.6%	159,647	80.9%	-3.7%
Total	16+ Years	198,381		197,220		
Outbound mandatory	All	140,548	9.2%	177,575	11.3%	2.1%
Outbound standalone	All	165,189	10.8%	144,290	9.2%	-1.6%
Return mandatory	All	101,746	6.6%	135,568	8.6%	2.0%
Return standalone	All	146,004	9.5%	121,244	7.7%	-1.8%
None	All	977,171	63.8%	992,333	63.2%	-0.7%
Total	**All**	**1,530,658**		**1,571,010**		

105

Table 8-14. Regional Comparison of School Escorting Types – Household Survey vs. Model Results

Escort Person Type	Expanded Household Survey Data		Model Results		Percentage Point Difference (Model - Survey)
	Count	Percentage	Count	Percentage	
Adult student	22,542	5.0%	6,629	1.5%	-3.6%
Full-time worker	274,394	61.4%	308,295	68.2%	6.8%
Part-time worker	53,548	12.0%	65,243	14.4%	2.5%
Non-working adult	90,952	20.3%	67,654	15.0%	-5.4%
Senior	5,603	1.3%	4,347	1.0%	-0.3%
Total	447,038		452,168		

8.1.2.3. Joint Travel Model

The fully joint tour models include two models: a generation model, which simulates the number (zero, one, or two) and purposes (meal, shopping, personal business, or social-recreation) of fully joint tours made by each household; and a participation model, which determines which household members participate in each simulated joint tour. The Excel file that summarizes the results of the fully joint tour models is JointTour Gen & Part.xlsm. The household survey data set shows an average of 0.255 fully joint tours per household while the model results show 0.262 joint tours per household.

The more detailed comparisons in the Excel files show the following results:

- The survey data set shows varying rates of joint tours per household by county, with the lowest rates in Baltimore City and Washington, D.C. This is not surprising since average household size is lower in these cities than in the rest of the model region. These two jurisdictions also have the lowest joint tour rates in the model results, though the model somewhat overestimates the joint tour rate in Baltimore City.
- Among households making joint tours, there is no discernable pattern of the number of joint tours made by income level. The model somewhat overestimates the number of joint tours for the lowest income group.
- Among households making joint tours, zero-vehicle households make fewer joint tours, though the model somewhat overestimates the joint tour rate for these households. The model also underestimates the joint tour rate for households with three or more vehicles.
- The distributions of joint tours by purpose and by party size (2, 3, or 3+) are similar for the survey data set and the model results. The cross-classifications of tour purpose by party size also match well, with the largest differences appearing for the combinations with the lowest incidence in the survey data set (generally, the 3+ person tours).

8.1.2.4. Individual Non-Mandatory Tour Generation Model

The individual non-mandatory tour generation model simulates the number (zero, one, two, or three) and purpose (meal, shopping, personal business, escorting, or social-recreation) of non-mandatory tours made by each person in the synthetic population for whom a mandatory or non-mandatory daily activity pattern has been simulated (at least one non-mandatory tour must be simulated for persons with non-mandatory patterns). The Excel file that summarizes the results of the non-mandatory tour generation model is *INMTourGeneration.xlsm*. The number of modeled non-mandatory tours per person is slightly lower (about 4%) than the number of such tours in the expanded household survey data set.

The more detailed comparisons in the Excel files show the following results:

- The modeled percentages of non-mandatory tours by purpose match the percentages from the survey data set almost exactly.
- Compared to the survey data set, the model slightly overestimates the number of non-mandatory tours for seniors and slightly underestimates for children, adult students, and adult non-workers. The model percentages of non-mandatory tours by purpose for each person type also match the percentages from the survey data set almost exactly.
- Most of the underestimation of non-mandatory tours by the model is for males. The expanded survey shows very similar percentages of tours by purpose for males and females. The slight differences between genders in the data (for example, slightly higher percentages of shopping and escort tours for females) are reflected in the model.
- There are a few relatively minor differences in the model results by income level, household size, and vehicle availability compared to the survey data set. Perhaps most notable is that the model does not pick up the differences by household size for individual meal tours—the survey data show that the larger the household, the lower the incidence of individual meal tours.

8.1.2.5. Work-Based Sub-Tour Generation Model

The work-based sub-tour generation model simulates the number (zero, one, or two) and purposes (work, meal, shopping, personal business, escorting, or social-recreation) of work-based sub-tours made by persons making work tours. The Excel file that summarizes the results of the work based sub-tour generation model is *WBTourGeneration.xlsm*. The number of modeled work-based sub-tours per work tour is about the same as the number of such sub-tours in the expanded household survey data set (0.154 observed versus 0.155 modeled).

The more detailed comparisons in the Excel files show the following results:

- The modeled percentages of work based sub-tours by purpose match the percentages from the survey data set well.

- The model results show that those who take non-auto modes to work make fewer sub-tours than those who take auto modes, and those who walk or bike to work making fewer sub-tours than those who took transit. This seems to make sense in that these people usually would not have access to cars for making work-based sub-tours. However, the survey data set generally shows the opposite pattern. The model results slightly overestimate the number of sub-tours by workers who take auto modes to work and slightly underestimate the number of sub-tours by workers who use non-auto modes to travel to work.
- The survey data set shows that males make more work-based sub-tours than females, and the model results match this result.
- The household survey data shows that the rate of making work-based sub-tours increases with income level. The model data show this pattern for all travelers, but at a much more moderate rate of increase.

8.1.3. Tour Level Choice Models

8.1.3.1.   Tour Destination Choice Models

The tour destination choice models simulate the location of the primary activity of each tour. There are Excel files with detailed results for various aggregate activity purposes:

- Work (not to regular workplace) – Tour Dest Work.xlsm
- University - Tour Dest Uni.xlsm
- Fully joint – Tour Dest Joint.xlsm
- Individual non-mandatory (except escort tours) - Tour Dest INM.xlsm
- Work based sub-tours – WB Tour Dest.xlsm

Each spreadsheet file includes histograms comparing the tour length frequency distributions, by both time and distance, for the corresponding activity purpose. Table 8-15 summarizes the coincidence ratios for these comparisons.

Table 8-15. Coincidence Ratios for Tour Length Frequency Distributions

	Coincidence Ratio	
Tour Purpose	**Time**	**Distance**
Work (including tours to regular workplace)	86%	80%
University	91%	76%
Joint	90%	85%
Individual non-mandatory	89%	89%
Work based sub-tours	98%	77%

For each tour purpose, the following comparisons between the observed (expanded household survey) data and model results are included in the Excel files:

- Average tour length (time and distance) by:

  o Tour activity (meal, shop, personal business, or social-recreation) - *for non-mandatory tours only*

  o Income level

  o Area type at home (or workplace for work-based sub-tours) and at the primary activity location

  o Person type – *except joint tours*

  o Number of household vehicles – *except work-based sub-tours*

  o Parent tour mode –*work-based sub-tours only*

- Percentage of intra-zonal tours (primary activity location zone is the same as the home zone (or work zone for work-based sub-tours)) by:

  o Area type

  o Person type – *work tours only*

  o Number of household vehicles – *except work-based sub-tours*

  o Parent tour mode –*work-based sub-tours only*

Generally, both the average tour lengths (see Table 8-16) and intra-zonal percentages (see Table 8-17) from the model matched those from the survey data well. In the model results, the average tour lengths show a logical progression with tour lengths increasing as income increases; the survey data do not show this pattern for all tour purposes. Another difference is that the model shows more intra-zonal tours for less dense areas, while the survey data do not show this pattern.

Table 8-16. Average Trip Length Comparisons by Tour Purpose

Tour Purpose	Observed		Model	
	Time (min)	Distance (miles)	Time (min)	Distance (miles)
Work	23.0	12.0	23.1	11.3
University	17.2	9.0	17.2	8.3
Joint	12.8	7.1	14.4	7.2
Individual non-mandatory	11.0	5.7	11.9	5.6
Work-based sub-tours	5.7	3.3	6.0	2.5

Table 8-17. Comparison of Intra-Zonal Percentages

Tour Purpose	Time	Distance
Work (including tours to regular workplace)	2%	3%
University	3%	4%
Joint	7%	8%
Individual non-mandatory	11%	13%
Work-based sub-tours	18%	18%

8.1.3.2.   Tour Time-of-day Choice Models

The tour time-of-day choice models simulate the start and end times, in half hour increments, of the primary activity of each tour. There are Excel files with detailed results for various aggregate activity purposes:

- Mandatory (work, school and university) - *TOD_Mand.xlsm*
- Joint - *TOD_Joint.xlsm*
- Individual non-mandatory - *TOD_NM.xlsm*
- Work-based sub-tours - *TOD_WB.xlsm*

Each spreadsheet presents histograms comparing the distributions of activity arrival and departures for the corresponding activity purpose. Table 8-18 summarizes the coincidence ratios for these comparisons.

Table 8-18. Coincidence Ratios for Time-of-Day Distributions

Tour Purpose	Coincidence Ratio	
	Arrival	Departure
Work	85%	83%
School	81%	86%
University	61%	51%
Joint	83%	78%
Individual non-mandatory	83%	83%
Work-based sub-tours	90%	91%

Table 8-19 presents the activity durations by purpose for the survey data set and the model results. These figures match well.

Table 8-19. Modeled and Observed Activity Durations by Purpose

Tour Purpose	Duration (hours)	
	Survey	**Model**
Work	7.1	7.4
School	7.0	7.2
University	4.5	5.0
Joint	1.9	1.7
Meal	0.9	1.5
Shopping	0.9	1.0
Personal Business	1.8	1.4
Social-recreation	1.0	1.8
Escort	0.1	0.2
Work-based sub-tours	0.8	0.7

Each spreadsheet also compares the average activity duration in hours by the following segmentations:

- Income level
- Person type – except joint tours
- Gender – except joint tours
- Specific activity purpose (e.g., meal, shopping) – joint, non-mandatory, and work based

In most cases, the modeled and survey activity durations are within 10% or within 10 minutes of one another when there are sufficient observations in the segment. Some exceptions include the following:

- Modeled activity durations for work tours are about 15% higher for part time workers and about 25%higher for adult students. Modeled activity durations for work tours are about 15% higher for the $15,000 to $30,000 income group and about 10% higher for the $30,000 to $50,000 income group.
- Modeled activity durations for non-mandatory tours are high for most segments since the average activity duration is high by about 15 minutes.
- Several of the university tour segments have greater differences between the observed data and model results; this is due to the relatively low numbers of these tours.

8.1.3.3. Tour Mode Choice Models

The tour mode choice models simulate the main mode of each tour. There are Excel files with detailed validation results for various aggregate activity purposes:

- Work - TourModeChoice_Work.xlsm
- School - TourModeChoice_Sch.xlsm
- University - TourModeChoice_Uni.xlsm
- Joint - TourModeChoice_Joint.xlsm

- Individual non-mandatory (except escort tours) - TourModeChoice_INM.xlsm
- Escort - TourModeChoice_Escort.xlsm
- Work based sub-tours - TourModeChoice_WB.xlsm

Error! Reference source not found. compares the regional observed and modeled mode shares by tour purpose.

Table 8-20. Regional Modeled and Observed Tour Mode Shares by Purpose

Tour Mode	Work		School		University		Individual Non-Mandatory	
	Survey	Model	Survey	Model	Survey	Model	Survey	Model
Drive Alone	59.0%	58.6%	2.0%	1.9%	50.9%	49.5%	46.9%	47.4%
Shared Ride 2	14.4%	14.9%	19.5%	7.1%	11.6%	11.5%	24.3%	23.5%
Shared Ride 3+	7.7%	8.2%	38.0%	53.3%	11.4%	11.3%	12.2%	11.7%
Transit-Walk Access	9.1%	8.6%	4.5%	4.4%	13.3%	14.3%	6.0%	5.8%
Transit-Auto Access	6.6%	6.4%	0.3%	0.3%	3.9%	3.9%	0.9%	0.9%
Walk	2.3%	2.5%	7.7%	6.3%	6.3%	7.6%	8.9%	10.1%
Bike	0.8%	0.8%	0.4%	0.4%	1.7%	1.9%	0.6%	0.6%
School Bus			27.7%	26.3%			0.2%	

Tour Mode	Escort		Joint		Work-Based Sub-tours		ALL TRIPS	
	Survey	Model	Survey	Model	Survey	Model	Survey	Model
Drive Alone					44.5%	46.1%	38.9%	38.4%
Shared Ride 2	45.7%	46.9%	46.7%	47.2%	12.9%	12.7%	22.4%	20.9%
Shared Ride 3+	42.7%	40.5%	43.0%	43.5%	7.2%	7.0%	17.9%	20.1%
Transit-Walk Access			2.9%	2.6%	2.0%	2.2%	6.1%	5.9%
Transit-Auto Access			0.4%	0.0%	0.2%	0.0%	2.8%	2.7%
Walk	11.7%	12.5%	6.7%	6.4%	32.8%	31.7%	7.8%	8.0%
Bike			0.3%	0.3%	0.2%		0.6%	0.6%
School Bus					0.2%		3.5%	3.5%

For each tour purpose, the spreadsheet files show the following comparisons between the survey and modeled tour mode shares:

- Area type at home (or workplace for work based sub-tours) and at the primary activity location
- Distance range
- Transit in-vehicle time ranges (walk and auto access)
- Household size and income level
- Vehicles less than, equal to, or greater than number of workers/drivers
- Age and gender

In nearly all cases, the mode shares from the model matched those from the survey data well. Some of the key results, which are true for both the observed and model data, include the following:

- Not surprisingly, transit and non-motorized mode shares increase as the area becomes more densely developed, while auto mode shares decrease. This trend noted in the expanded survey data is also seen in the model, although the rate of changes among area types is more moderate in the model (it should be noted that except for work and individual non-mandatory tours, the number of survey observations is fairly small for the most urban area types).
- Transit-walk access mode shares decrease with increased distance; the opposite holds for transit-auto access shares (nearly all transit-auto access tours are for work or university purposes). The model captures these trends better for the walk access tours. Non-motorized trips, naturally, are nearly all short-distance, and the model results reflect this.
- Transit mode shares to all counties are low—from 0-2%—for all tour purposes, with the exception of the three Maryland Counties in the MWCOG region, where the transit shares are a bit higher. Transit shares to the cities of Baltimore and Washington are substantially higher. The model results reflect these trends.
- Transit-walk access shares decrease with increasing income levels for all tour purposes, and the model results accurately reflect this trend. For work tours, transit-auto access shares increase with increasing income levels, and the model results accurately reflect this as well.
- Generally, transit shares decrease with increasing household size, and the model accurately reflects this trend.
- Not surprisingly, transit and non-motorized mode shares are much higher in households with fewer vehicles than workers, or fewer vehicles than drivers, and are even higher in households with zero vehicles. The model reflects these trends accurately.
- Auto shares, especially drive alone, decrease while transit shares decrease with increasing age.
- For work tours, transit and shared-ride mode shares are higher for females; drive alone and bike mode shares are higher for males.

8.1.3.4. Stop Generation Models

The stop generation models simulate the number and purposes of stops made on each tour. Separate models were estimated for each tour purpose. There are Excel files with detailed results for various aggregate activity purposes:

- Mandatory (work, school, and university) - Stops_Mand.xlsm
- Joint - Stops_Joint.xlsm
- Individual non-mandatory - Stops_INM.xlsm
- Work-based sub-tours - Stops_WB.xlsm

Table 8-21 compares the number of observed and modeled half tours with stops by tour purpose. Table 8-22 presents the observed and modeled daily stops per half tour in each direction by tour purpose. As these tables show, the model results are close to the observed results from the expanded household survey data set.

Table 8-21. Regional Modeled and Observed Shares of Half Tours by Number of Stops by Purpose

Stops	Work		School/ University		Individual Mandatory	Non-
	Survey	Model	Survey	Model	Survey	Model
0 stops	50%	48%	78%	80%	55%	60%
1 stop	35%	37%	15%	13%	30%	27%
2 stops	10%	9%	5%	5%	9%	8%
3 stops	6%	7%	2%	2%	6%	5%

Stops	Joint		Work-Based Sub-tours		All Half Tours	
	Survey	Model	Survey	Model	Survey	Model
0 stops	51%	58%	80%	86%	57%	60%
1 stop	34%	30%	17%	13%	29%	28%
2 stops	10%	9%	3%	2%	9%	8%
3 stops	4%	3%	0%	0%	5%	5%

**Note: Model is constrained to produce only 1 or 2 stops per half tour on work based sub-tours.

Table 8-22. Observed and Modeled Average Number of Stops per Half Tour

Tour Purpose	Outbound Half Tour		Return Half Tour	
	Survey	Model	Survey	Model
Work	0.24	0.28	0.47	0.46
School/University	0.08	0.08	0.24	0.22
Individual Mon-Mandatory	0.30	0.26	0.37	0.32
Joint	0.32	0.24	0.36	0.34
Work Based Sub-tours	0.10	0.07	0.17	0.13
Total – All Tours	**0.21**	**0.25**	**0.33**	**0.34**

For each tour purpose, the spreadsheet files show the following comparisons between survey and modeled stops:

- Income level
- Person type – except joint tours
- Age and gender – except joint tours

The more detailed comparisons in the Excel files show some differences between the model results and the expanded survey data (some due to small sample sizes for certain segments), but overall the model results reflect the observed data fairly well.

8.1.4.   Stop/Trip Level Choice Models

8.1.4.1.   Stop Destination Choice Models

The stop destination choice model simulates the locations of all intermediate stops between the home (or workplace, for work-based sub-tours) and primary activity location on tours. The Excel file that summarizes the results of this model is *StopDestChoice.xlsm*.

Error! Reference source not found. presents a comparison between the observed (survey) and modeled trip length distance distributions. While there are some differences in the distributions, the fits are good; the coincidence ratio is 85%. The average trip distances are 4.1 miles (observed) and 4.4 miles (modeled).
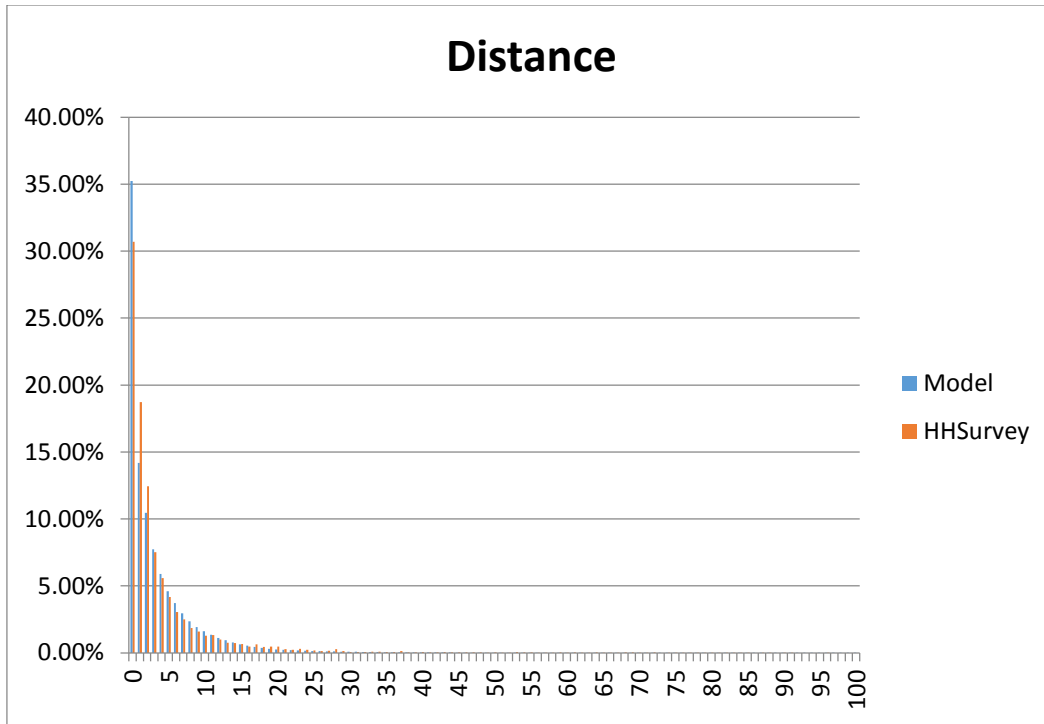
Figure 8-5. Trip length frequency distribution for stops (miles).

The spreadsheet file also provides comparisons between the observed data from the household survey and the model results for the average trip distances segmented by stop purpose, household income level, tour mode, area type at home and primary destination, and tour purpose. These comparisons show a good match, with 36 of the 54 segments having modeled trip lengths within one half mile of the observed trip length and 47 of the segments having modeled trip lengths within one mile of observed trip length. The largest overestimates by the model are for two of the mid-range area types and for school stops, while the largest underestimates are for the rural area type and for work and university stops.

8.1.4.2.   Stop Time-of-day Choice Models

The stop time-of-day choice model simulates the times (at the half hour level) of all intermediate stops between the home (or workplace, for work-based sub-tours) and primary activity location on tours. The Excel file that summarizes the results of this model (as well as the trip mode choice model, described in the next section) is *TripModeTODChoice.xlsm*.

Error! Reference source not found. presents a comparison between the observed (survey) and modeled stop time-of-day distributions. While there are some differences in the distributions, the fit is good. The spreadsheet file also compared the distributions by area type; these also show good fits.

116
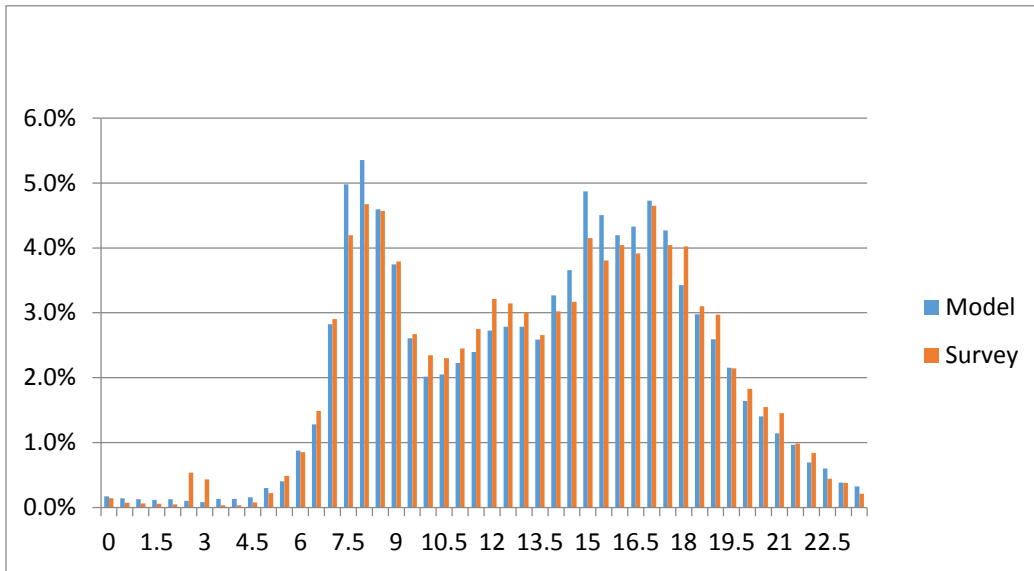
Figure 8-6. Trip length frequency distribution for stops (miles)

8.1.4.3.   Trip Mode Choice Model

The trip mode choice model simulates the mode for each trip that is part of a tour, conditional on the simulated tour mode. The Excel file *TripModeTODChoice.xlsm* also summarizes the results of the trip mode choice model. The following is a summary of trip mode shares by tour mode:

- Drive alone:

    o <u>Survey</u>: Drive alone – 99%, walk 1%
    o <u>Model</u>: Drive alone – 98%, walk 2%
- Tour mode shared ride 2:

    o <u>Survey</u>: Drive alone – 30%, shared ride 2 – 68%, walk 2%
    o <u>Model</u>: Drive alone – 31%, shared ride 2 – 67%, walk 2%
- Tour mode shared ride 3+:

    o <u>Survey</u>:   Drive alone – 18%, shared ride 2 – 18%, shared ride 3+ – 62%, walk – 2%
    o <u>Model</u>:   Drive alone – 15%, shared ride 2 – 21%, shared ride 3+ – 63%, walk – 1%
- Tour mode transit-walk access:

    o <u>Survey</u>:   Drive alone – 1%, shared ride 2 – 5%, shared ride 3+ – 3%, transit-walk access – 68%, walk – 23%, bike – 1%
    o <u>Model</u>:   Drive alone – 6%, shared ride 2 – 8%, shared ride 3+ – 4%, transit-walk access – 51%, walk – 29%, bike – 3%
- Tour mode transit-auto access:

- o <u>Survey</u>:   Drive alone – 9%, shared ride 2 – 9%, shared ride 3+ – 5%, transit-walk access – 5%, transit-auto access – 65%, walk – 7%
- o <u>Model</u>:   Drive alone – 6%, shared ride 2 – 29%, shared ride 3+ – 18%, transit-walk access – 2%, transit-auto access – 39%, walk – 6%
- Tour mode school bus:

  - o <u>Survey</u>: Shared ride 2 – 4%, shared ride 3+ – 6%, walk – 2%, school bus – 88%
  - o <u>Model</u>: Shared ride 2 – 5%, shared ride 3+ – 8%, walk – 3%, school bus – 84%

Note that by definition, all walking trips and bicycle tours have the same trip mode as the tour mode. The spreadsheet file also shows the following comparisons between survey and modeled trip mode shares:

- Area type at home (or workplace for work based sub-tours) and at the primary activity location
- Distance range
- Transit in-vehicle time ranges (walk access)
- Household size and income level
- Vehicles less than, equal to, or greater than number of workers/drivers
- Age and gender

In most cases, the mode shares from the model matched those from the survey data well.

8.2. Long-distance Model Calibration and Validation

8.2.1. Model Calibration

Travel demand model calibration is essential to accurately model people's travel. Model calibration is the process of adjusting the model parameter values until the simulated or estimated travel results closely match the observed travel for the base year.

In general, the process of system-wide calibration of parameters of a simulation-based model is understood to find the value of the parameters that minimize the error between observed outputs and simulated outputs. The observed outputs, usually, refer to trusted external data sets containing aggregate measures matching the base conditions (i.e. OD tables, traffic counts for base year). Thus, we want the simulation-based model to replicate these base conditions. The degree of trust in each measurement may be represented as a weight in the calibration process. The simulated output refers to the aggregation of the simulation results into measures that are equivalent to the observed output's measures.

The calibration process is formulated as a constrained minimization optimization problem as follows,

$$minimize_\theta \ w\left\|O_m - O_S(\theta)\right\|^2$$

$$s.t.$$

$$O_S = F(Z;\theta) \hspace{4cm} \text{(8-1)}$$

$$l \leq \theta \leq u$$

Where w is a vector of weights indicating the trust of the modeler on different observed outputs, and in our calibration, the w is set as 1; O_m is a vector representing the observed outputs and O_s is a vector representing the simulated outputs; θ is the vector of parameters from the model to be calibrated, and l and u are vectors of lower bounds and upper bounds for the parameters; $F(Z;\theta)$ represents the link between the simulated outputs and the simulation-based model, and Z are the inputs required to run the simulation-based model. Lastly, the ||.|| represent the Euclidian norm. The algorithm selected to solve this constrained minimization problem is *Simultaneous Perturbation Stochastic Approximation* (SPSA) (Spall, 2005). This algorithm has the following advantages: 1) it accounts for the simulation error in the simulation-based model output; 2) it can be applied in stochastic gradient setting or a gradient-free setting; 3) it only requires two function evaluations per iteration regardless of the length of the vector of parameters.

In essence, the SPSA algorithm works by perturbing the components of the vector of parameters and computing difference of these perturbations with respect to the objective function of the constrained minimization problem. More detailed information can be found in Spall's work (Spall, 2005). One example for calibrating models using a variation of SPSA is the work from Antoniou et al. (Antoniou, et al., 2015). The implementation step of the algorithm is briefly described as follows,

- *Step 0 Initialization and coefficient selection:* in this step, the SPSA algorithm is set up with the initial values for the vector of parameters and also the values for the vector of hyper-parameters. These hyper-parameters belong exclusively to the SPSA algorithm.
- *Step 1 Generation of simultaneous perturbation vector:* a perturbation vector is generated using Monte Carlo simulation. This perturbation vector uses a Bernoulli distribution centered at zero.
- *Step 2 Objective function evaluations:* evaluate the objective function twice using the perturbation vector.
- *Step 3 Gradient approximation:* compute the gradient approximation using the perturbation vector and two evaluations of the objective function computed from step 2.
- *Step 4 Update vector of parameters:* update the values of the vector of parameters based on the gradient descent using the approximated gradient computed from step 3.
- *Step 5 Iteration or termination:* return to step 1 to continue iterating or terminate if there is negligible change between iterations in the objective function and/or values of the vector of parameters.

For the passenger long-distance model, we initially calibrated the alternative specific constants of the time of year choice model, and the travel mode choice model using the Airline Origin and Destination Survey (DB1B) data. This survey is a 10% sample of airline tickets from reporting carriers collected by the Office of Airline Information of the Bureau of Transportation Statistics. Data includes origin, destination and other itinerary details of passengers transported. This database is used to determine air traffic patterns, air carrier market shares and passenger flows. The model is calibrated on the base year model.

For the purpose of this calibration effort, airline OD data is summarized to the following 4 values for each quarter:

- Number of flights departing from Maryland
- Number of flights landing in Maryland
- Number of flights departing from all other states
- Number of flights landing in all other states

Consequently, airline OD data was summarized into 16 variables. The model outputs were also summarized to capture these 16 variables.

The sum of the squared differences between model simulation and airline OD values for these 16 variables was used as the objective function of the calibration. SPSA seeks to minimize the objective function by changing model parameters. For that, some of the model parameters should be selected as the inputs of SPSA to be changed. Mode choice and time of year choice models were selected to be calibrated against airline OD data. Alternative specific constants for these models were selected as the variables to be calibrated. The followings are the variables selected for calibration:

- Business trips time of year model alternative specific constant for quarter 2
- Business trips time of year model alternative specific constant for quarter 3
- Business trips time of year model alternative specific constant for quarter 4
- Pleasure trips simple time of year model alternative specific constant for quarter 2
- Pleasure trips simple time of year model alternative specific constant for quarter 3
- Pleasure trips simple time of year model alternative specific constant for quarter 4
- Pleasure trips full time of year model alternative specific constant for quarter 2
- Pleasure trips full time of year model alternative specific constant for quarter 3
- Pleasure trips full time of year model alternative specific constant for quarter 4
- Business trips mode choice model alternative specific constant for Air
- Business trips mode choice model alternative specific constant for Rail
- Pleasure trips mode choice model alternative specific constant for Air
- Pleasure trips mode choice model alternative specific constant for Rail
- Personal Business trips mode choice model alternative specific constant for Air
- Personal Business trips mode choice model alternative specific constant for Rail

Another important aspect of the calibration is hyper-parameter selection for SPSA algorithm. SPSA includes five hyper-parameters: a, c, α, γ, and A. More information about these parameters can be found in the literature of Spall's (Spall, 2005). These parameters were selected based on the suggestions in the optimization literature using variance of objective function, and average of gradients.

We conducted several rounds of calibration, and finally chose the calibrated parameters under 60 iterations of calibration, as it shows better performance in trip distribution by time of year and other trip distributions. SPSA algorithm was coded in Java to find the calibrated variables. The initial values for calibration variables were set to their estimated values from the estimation process.

The 60-iteration calibration results are shown as below:

Table 8-23. Calibration Hyper-Parameter

Hyper-Parameter	Value
A	100
a	1.37245315551126038E-18
c	0.1
α	0.6
γ	0.1

Table 8-24. Calibration Properties

Number of iterations	60
Initial objective function value	5.3036402075560817E18
Final objective function value	4.692863010673278E18

Table 8-25. Calibration Results

Variable	Calibrated Value	Initial Value
toyBusiness_ASC1	-2.199	-1.469
toyBusiness_ASC2	-1.695	-0.744
toyBusiness_ASC3	-0.427	-0.370
toySimplePleasure_ASC1	-2.017	-1.228
toySimplePleasure_ASC2	-1.239	-1.285
toySimplePleasure_ASC3	-0.384	-0.541
toyFullPleasure_ASC1	-1.415	-1.120

toyFullPleasure_ASC2	-1.049	-0.653
toyFullPleasure_ASC3	-1.011	-0.403
mc_BUSINESS_Air	-1.622	-0.44
mc_BUSINESS_Train	-2.350	-2.93
mc_PLEASURE_Air	-3.960	-2.95
mc_PLEASURE_Train	-2.676	-3.56
mc_PERSONAL_BUSINESS_Air	-1.626	-1.49
mc_PERSONAL_BUSINESS_Train	-4.110	-3.75

8.2.2. Model Validation

The validation of the long-distance model components is done by cross-validation using 80% of the sample size for model estimation, and 20% of the sample size for model validation.

8.2.2.1. Travel Mode Choice Model

Figure 8-7, Figure 8-8, and Figure 8-9 present the aggregate share of the observed tour travel mode choice and the estimated one for business, pleasure, and personal business purposes. Results show that the mode choice models can estimate the mode choice market share in the right trend but with errors. Compared to business and personal business mode choice models, pleasure mode choice model has a weaker performance. The air mode for pleasure purposes is underestimated by almost 40%, and the large percentage is also caused by the small sample size of the air pleasure travel in the validation data set.
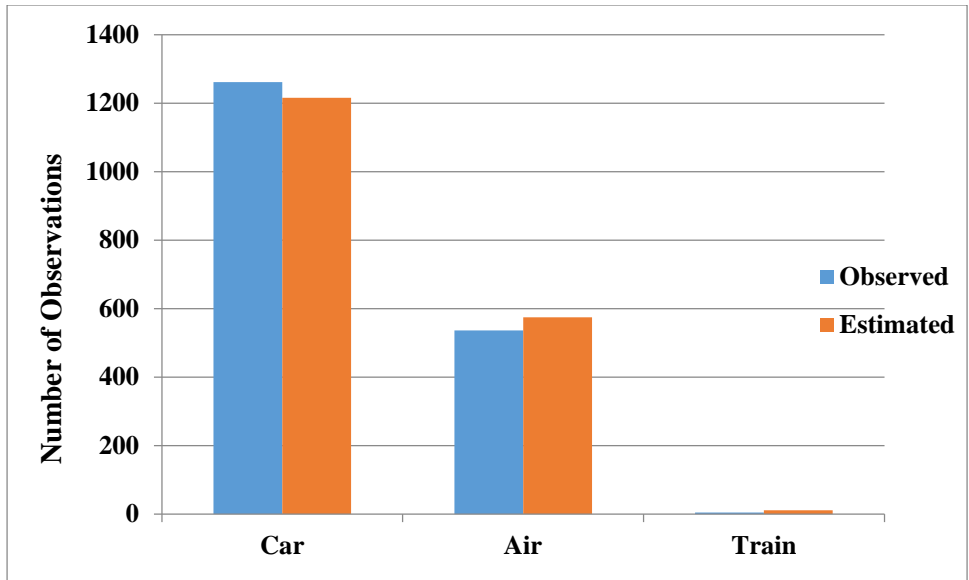
Figure 8-7. Tour mode choice validation for business purpose.
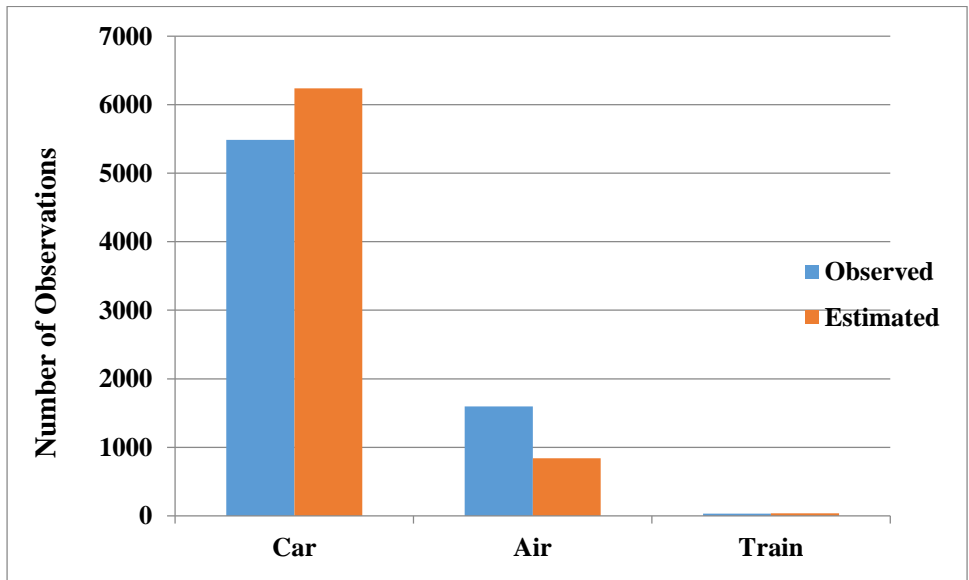
Figure 8-8. Tour mode choice validation for pleasure purpose.
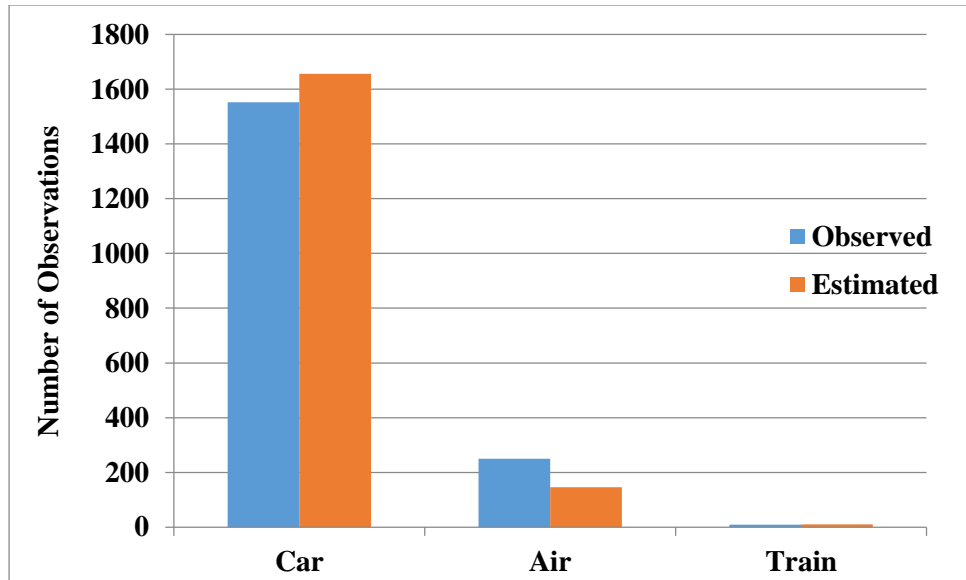
Figure 8-9. Tour mode choice validation for personal business purpose.

8.2.2.2. Time-of-Year Choice

Good estimation can be observed from the without sample validation results (Figure 8-10, Figure 8-11, and Figure 8-12), possessing a very small difference between the observed time-of-year distributions and the estimated ones.
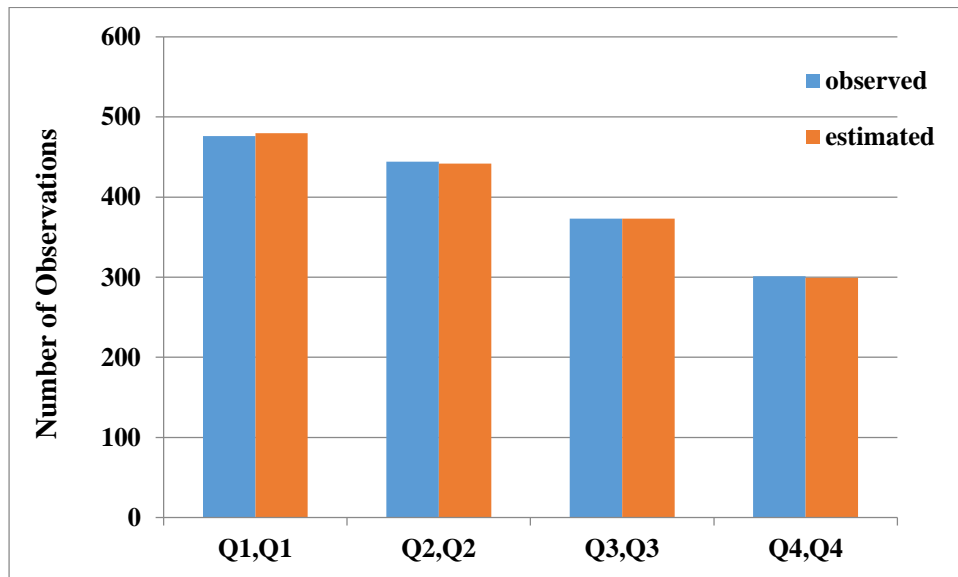
Figure 8-10. Time-of-year choice validation for business purpose.
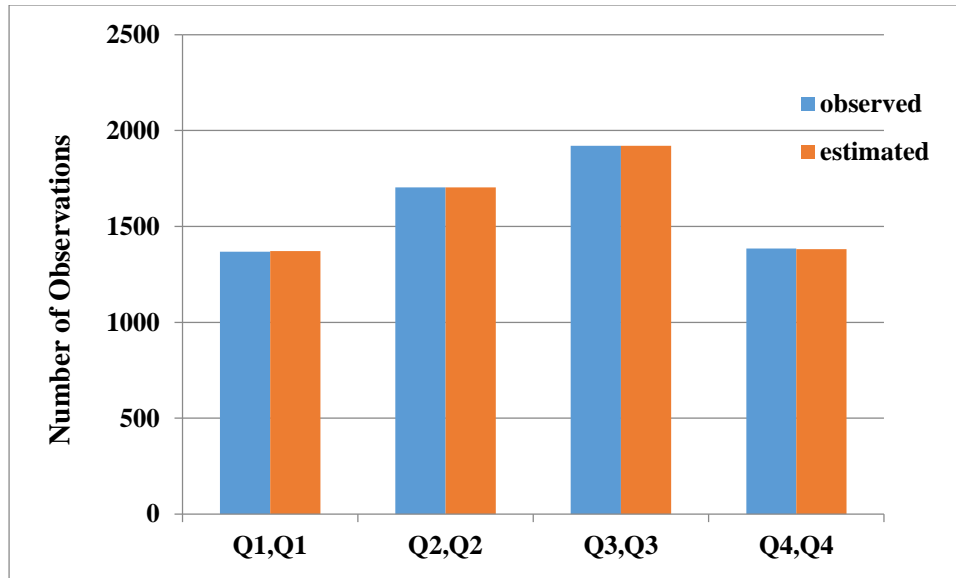
Figure 8-11. Simple time-of-year choice validation for pleasure purpose.
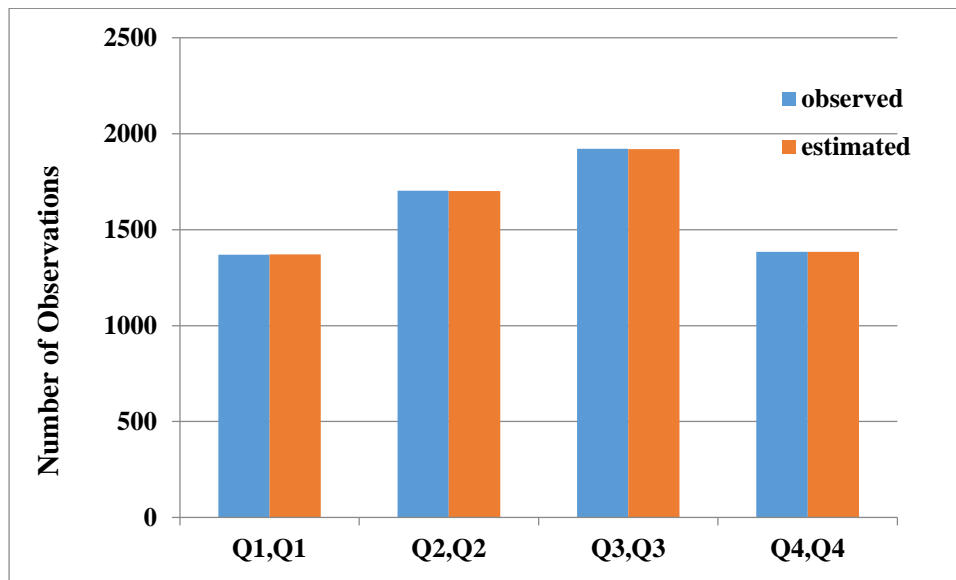
Figure 8-12. Full time of year choice validation for pleasure purpose.

8.2.2.3.   Tour Duration Choice Model

The validation results between the observed duration distribution and estimated duration distribution are shown in Figure 8-13 and Figure 8-14 for business and personal business purposes separately. It is observed that except for the duration of two days, the estimated duration distributions follow the same pattern as the observed distribution for both long-distance purposes.
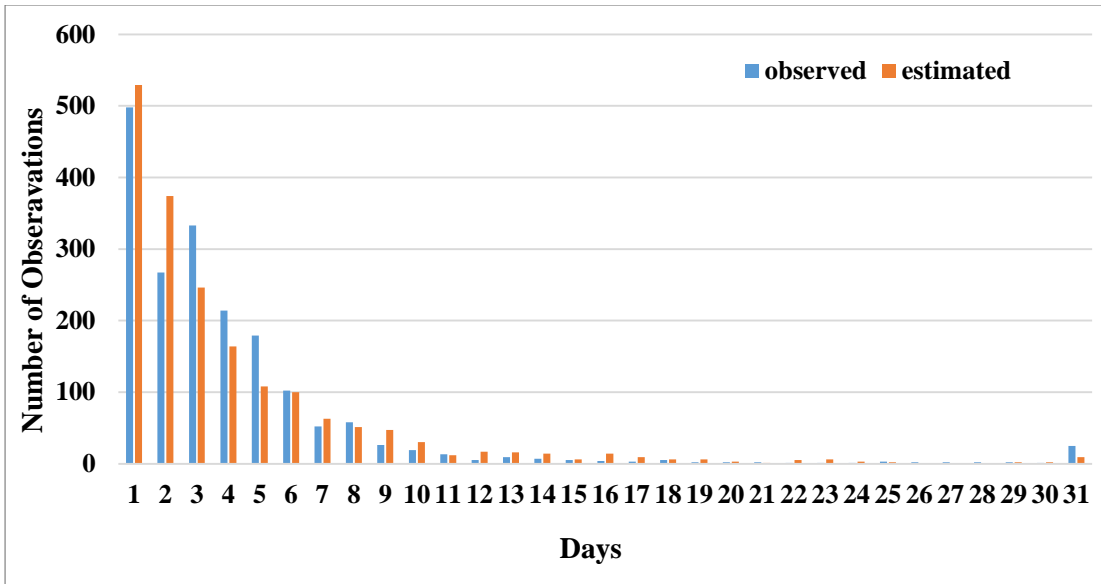
Figure 8-13: Validation results for Business Duration Model
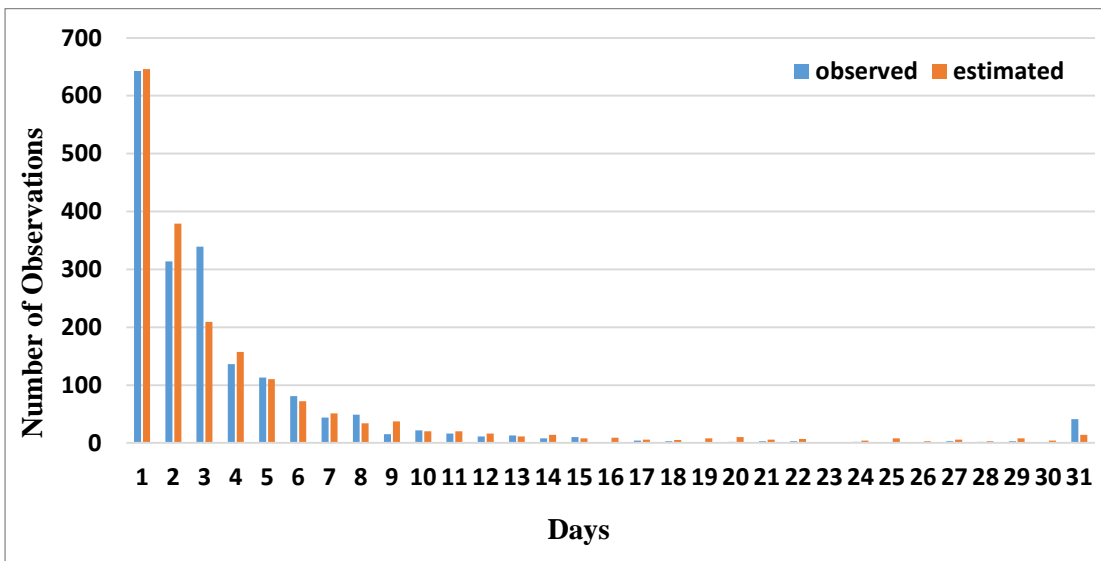
Figure 8-14. Validation results for Personal Business Duration Model.

8.2.2.4. Travel Party Size Choice Model

The without sample validation results (Figure 8-15, Figure 8-16, and Figure 8-17) illustrate that the travel party size choice models of all three purposes can estimate the number of persons during the entire tour with small errors.
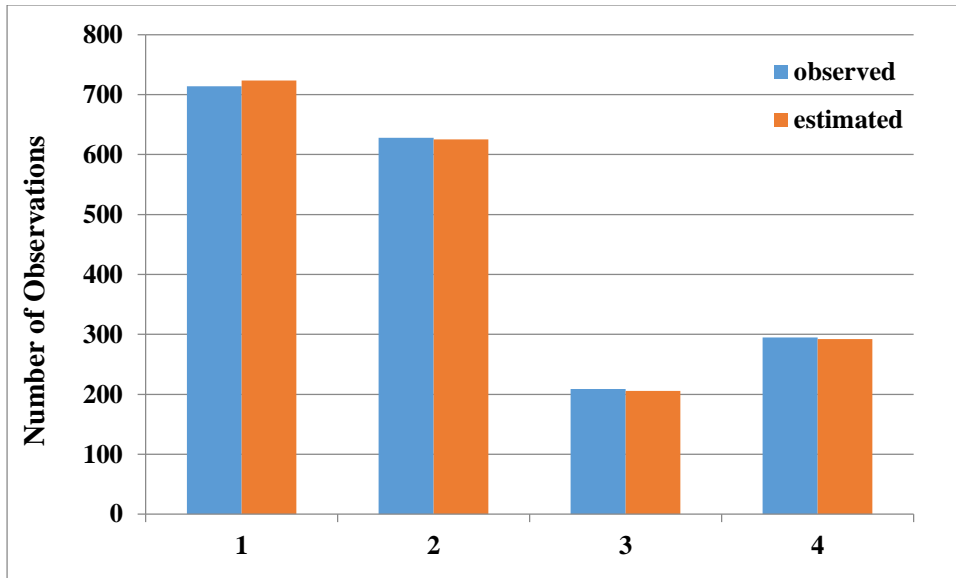
Figure 8-15. Travel party size choice validation for business purpose.
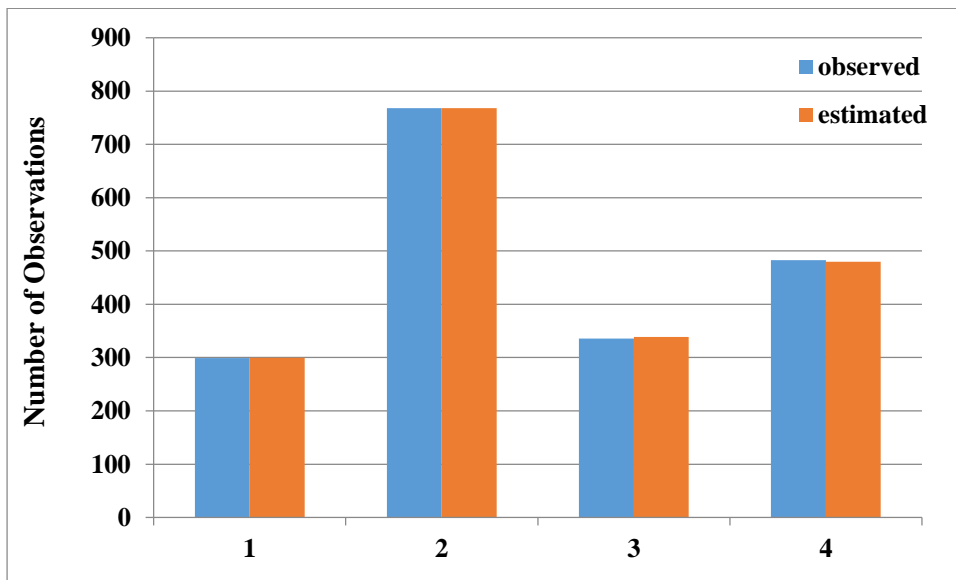
Figure 8-16. Travel party size choice validation for personal business.
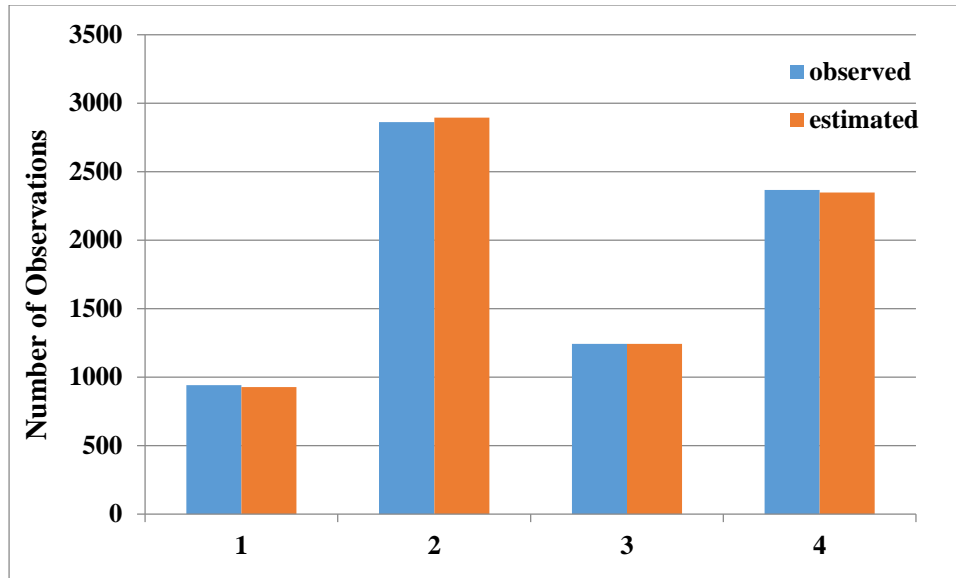
Figure 8-17. Travel party size choice validation for pleasure purpose.

8.2.2.5. Tour Destination Choice Model

Figure 8-18, Figure 8-19, and Figure 8-20 show the validation results for only 206 zones that exist in the dataset for destination model estimations for the three long-distance travel primary purposes. The figures show that most of the destinations can be estimated with small difference from the observed, except for a few with large errors due to the relatively small sample size. Among all three purposes, the destination model for pleasure presents the overall largest error.
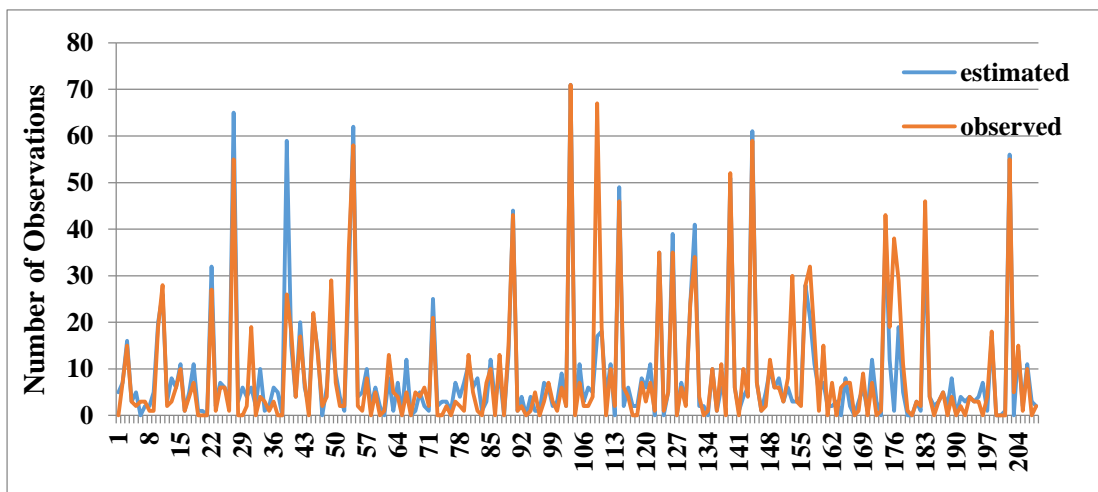
Figure 8-18. Destination choice validation for business purpose.
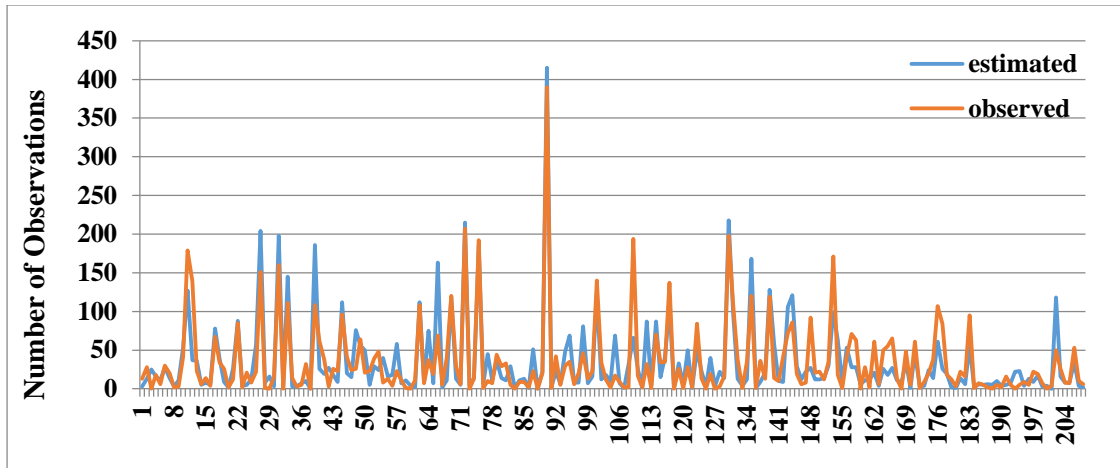
Figure 8-19. Destination choice validation for pleasure purpose.
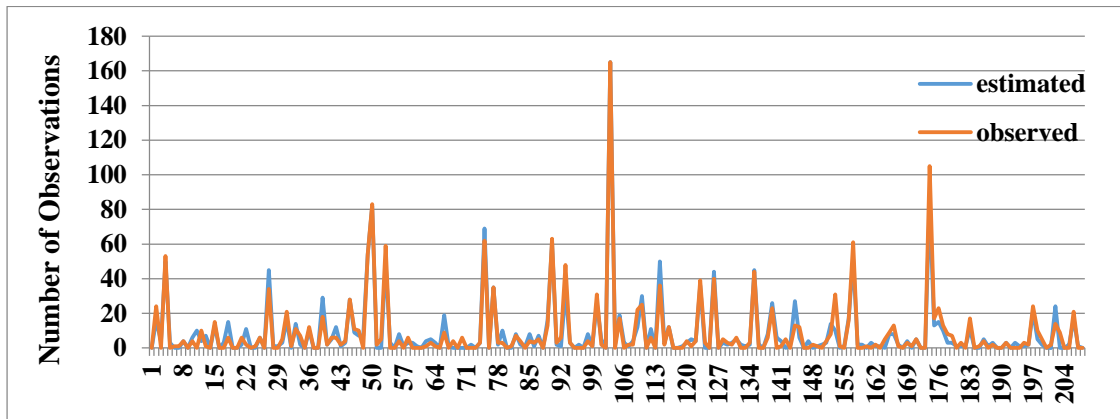
Figure 8-20. Destination choice validation for personal business purpose.

8.2.2.6.   Stop Frequency Choice Model

The without sample validation results in Figure 8-21 and Figure 8-22 show that the stop frequency models perform well in estimating the number of stops during either inbound or outbound legs of the tour.
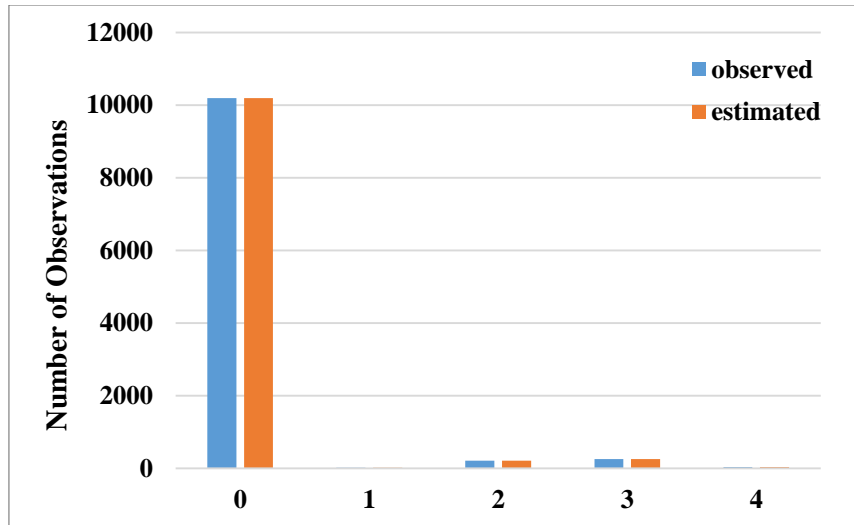
Figure 8-21. Inbound Stop Frequency Model Validation.
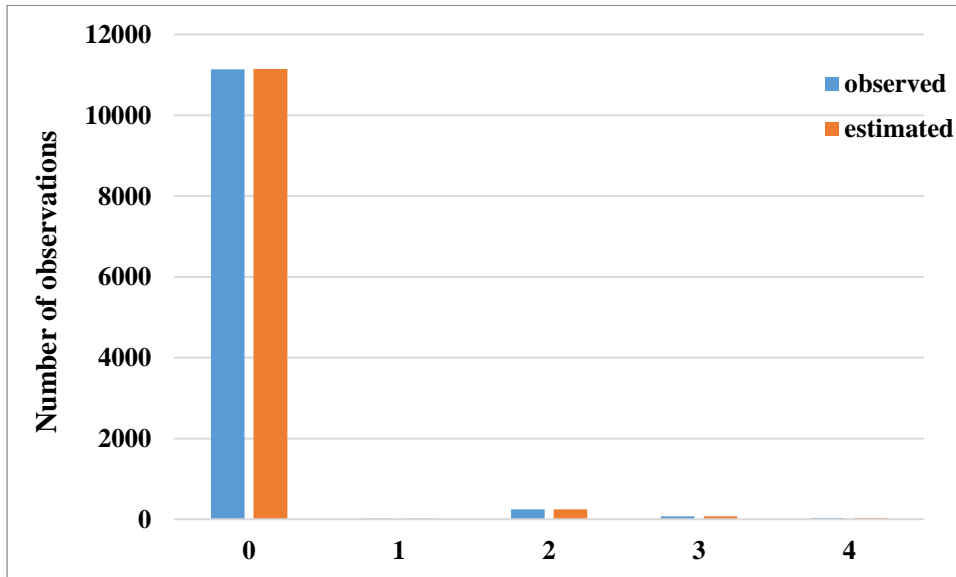
Figure 8-22. Outbound stop frequency model validation.
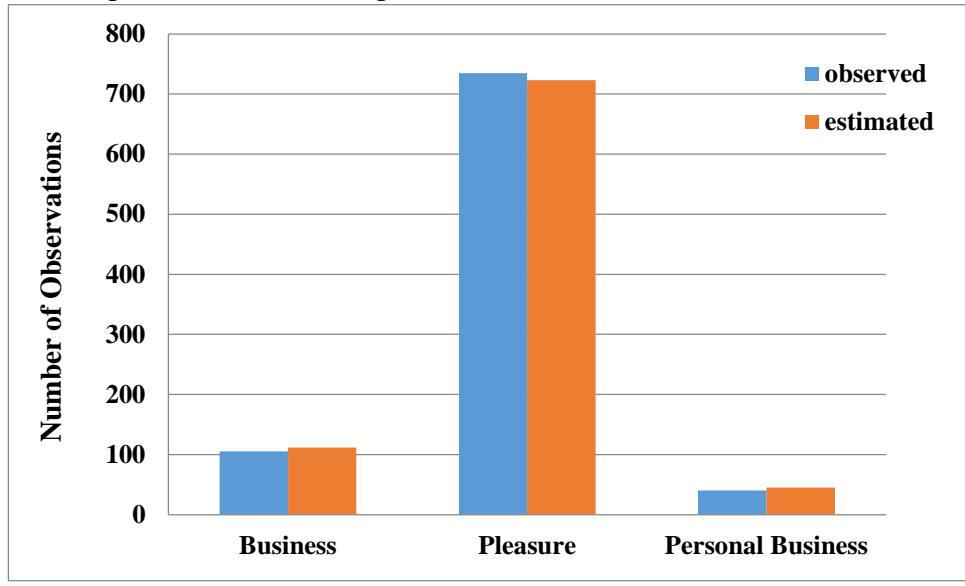
Figure 8-23 and Figure 8-24 illustrate the good performance of the stop purpose models in estimating the purposes of each stop during each half leg of the long-distance tour.
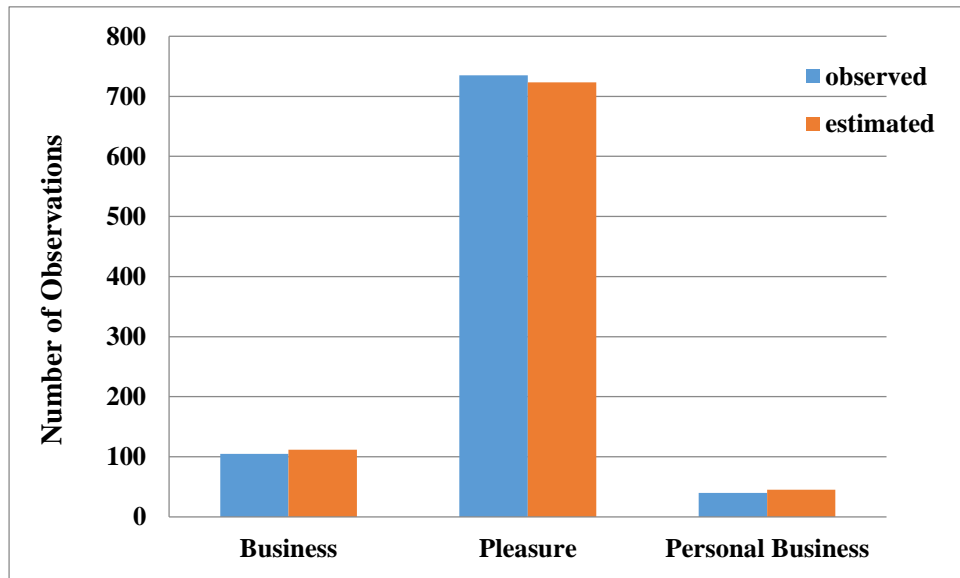
Figure 8-23. Outbound stop purpose model validation.
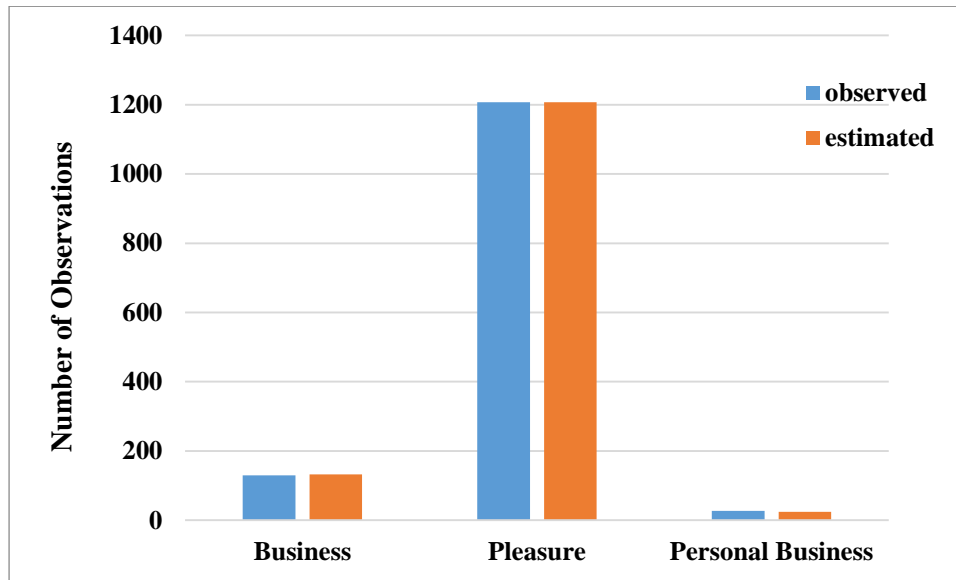
Figure 8-24. Inbound Stop Purpose Model Validation.

8.2.2.8. Stop Location Choice Model

Figure 8-25 and Figure 8-26 shows the without sample validation results for the two models. From the figures, we can see that the size of the validation sample is very small for each zone, with the largest number of trips less than 30. Based on the very few validation data records, we can say that both models can estimate the stop location well for a number of zones. Furthermore, the outbound location choice model shows a better performance than the inbound location model, with a smaller number of trips under/over-estimated.
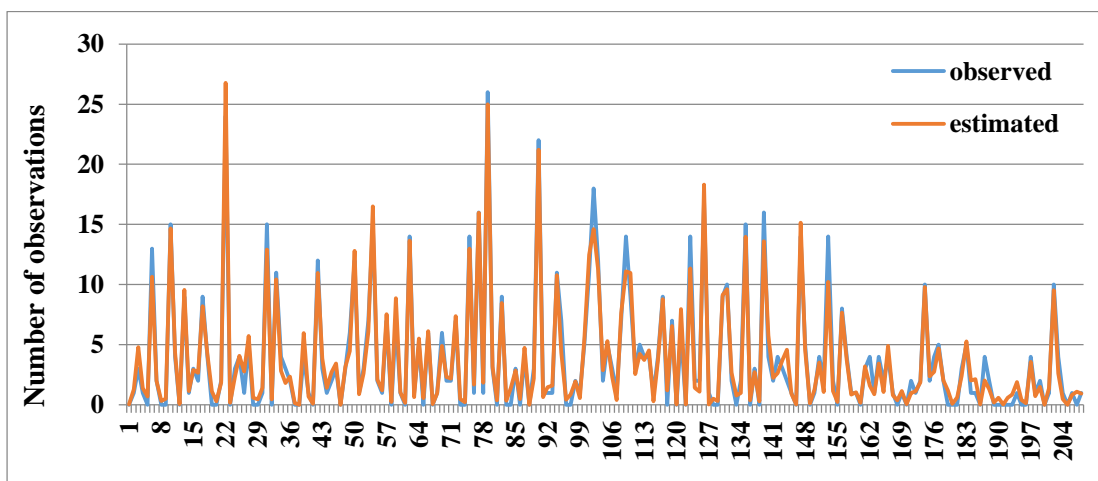
Figure 8-25. Outbound stop location choice validation.
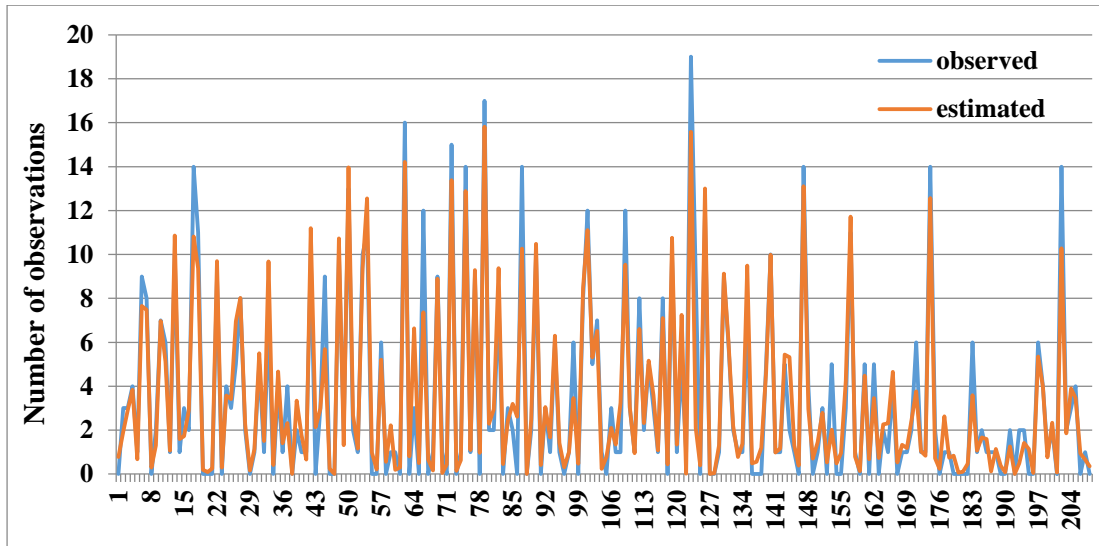
Figure 8-26. Inbound stop location choice validation.

8.2.2.9. Comparison with Other Long-distance Models

In order to check the accuracy of the model, results are validated with other data sources, which are mainly the output of other demand models. The total number of long-distance trips to Maryland and total number of long-distance trips generated from Maryland are between three different sources for validation. Figure 8-27 shows the comparison.

As shown, the number of trips to Maryland is equal to the number of trips generated from Maryland in all datasets, as they are all obtained from balanced models. MSTM1 is the first version of the Maryland state-wide transportation model that was developed based on a traditional four-step model. CDM-Smith data are derived from Multimodal Interregional Passenger Travel Origin Destination Data for FHWA, which was conducted by the C.D. Smith team. RSG data are the long-distance passenger OD estimated by RSG company.

Definition of long-distance trips are different among some of these datasets. For instance, DM-Smith dataset defines long-distance trips as trips longer than 100 miles, while other datasets define long-distance trips as trips longer than 50 miles. However, the number of predicted trips is not significantly different among these datasets.

133
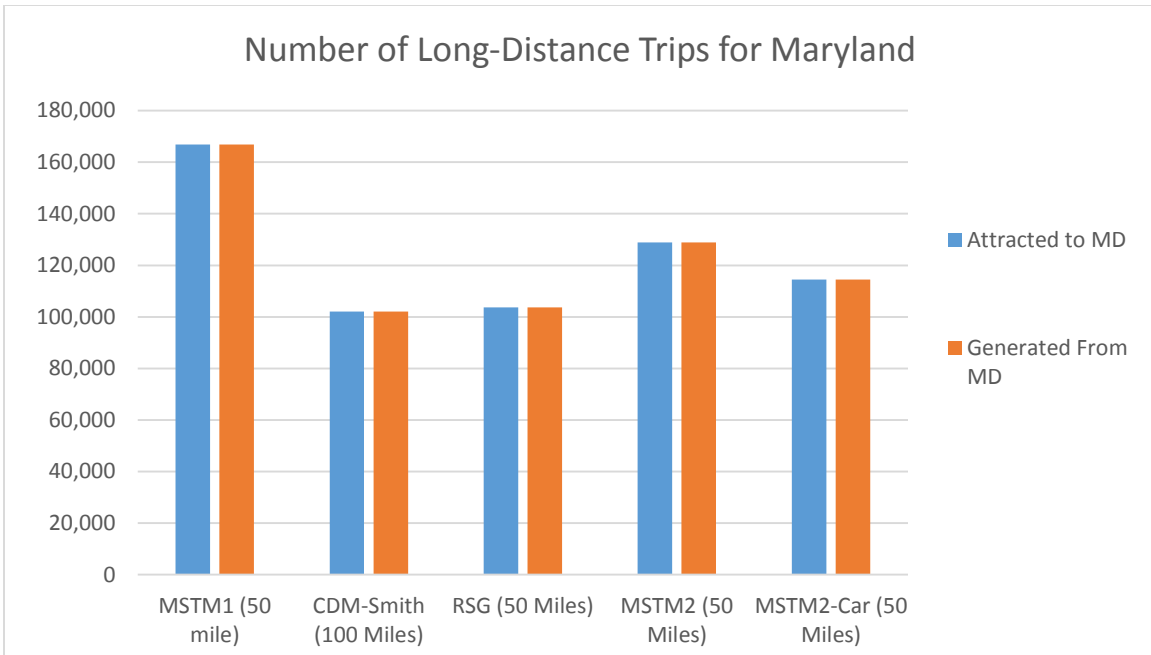
Figure 8-27. Total generation/attraction validation for Maryland.

8.3. Truck Model Validation

The model validation was conducted by comparing observed truck count and modeled truck volume. During the project duration, truck VMT data was not available and validation related to VMT is proposed to be conducted later. Figure 8-28 shows various major corridors and the comparison of observed and modeled volume. The comparison suggests that both observed and modeled volumes compare well for all corridors.
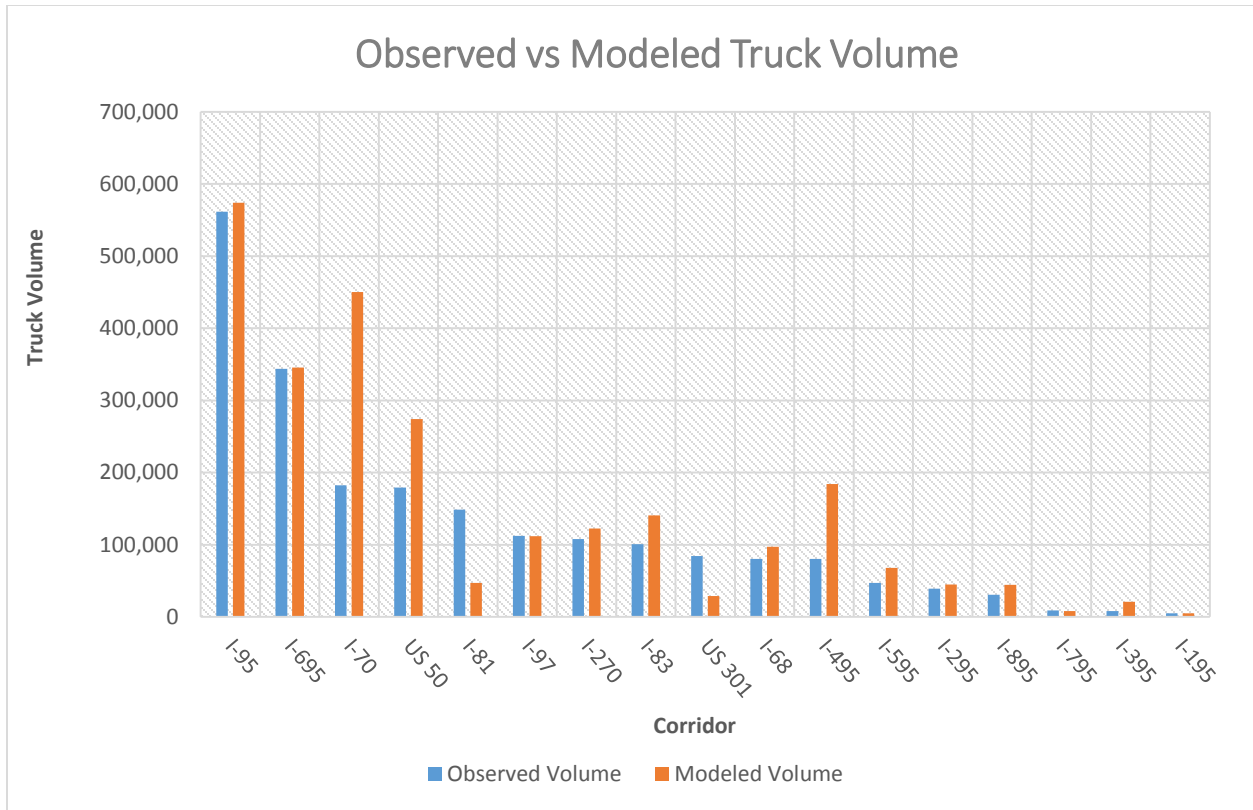
Figure 8-28. Truck count comparison at major corridors.

Figure 8-29 shows observed and modeled truck volume for various functional classes. There is not a pattern of modeled volumes, either lower or higher, across functional classes. However, there is a reasonable match between both data sets. Figure 8-30 shows the comparison of observed and modeled volume at various volume classes. It is expected for the model volumes to be more precise in higher volume classes and vice versa. However, for all volume classes, both data sets suggest a reasonable comparison. Figure 8-31 shows a county-level comparison. Figure 8-32 shows screenline validation. At each screenline, both observed and modeled volumes are close to the 45-degree slope line. R-square of 0.679 shows reasonable match between observed and modeled volume at screenlines.
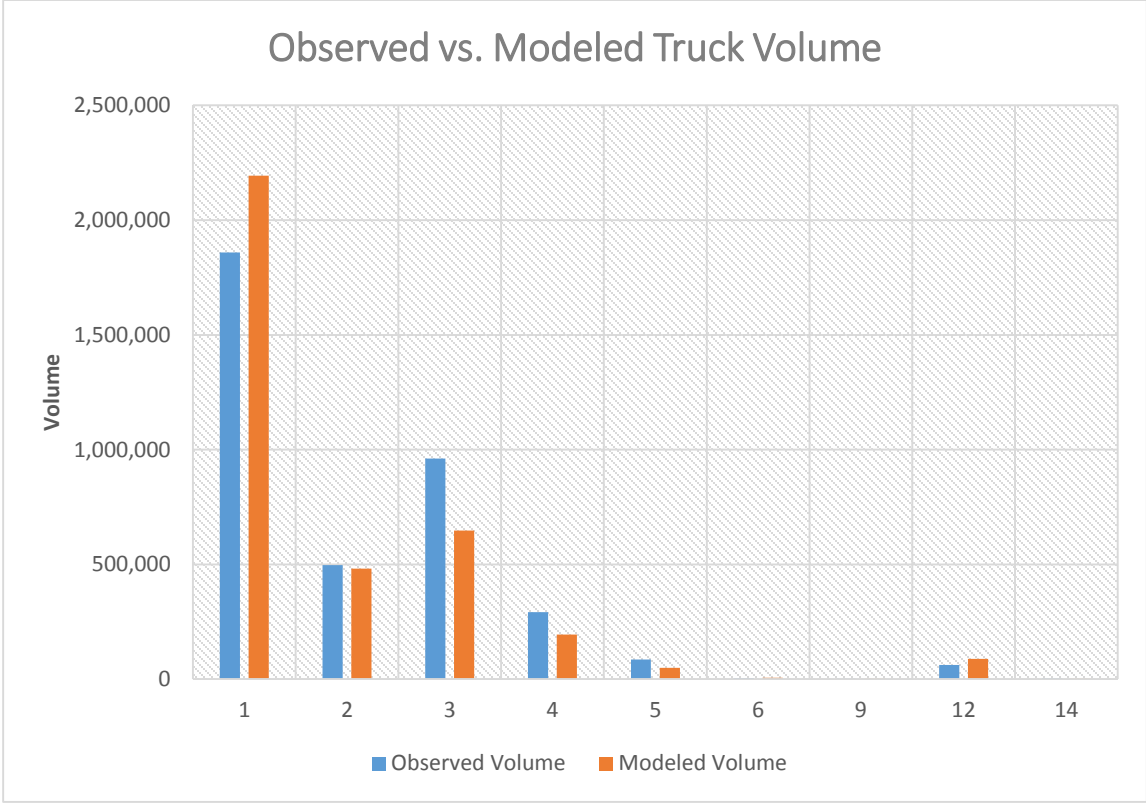
Figure 8-29. Truck count comparison at major facility types.
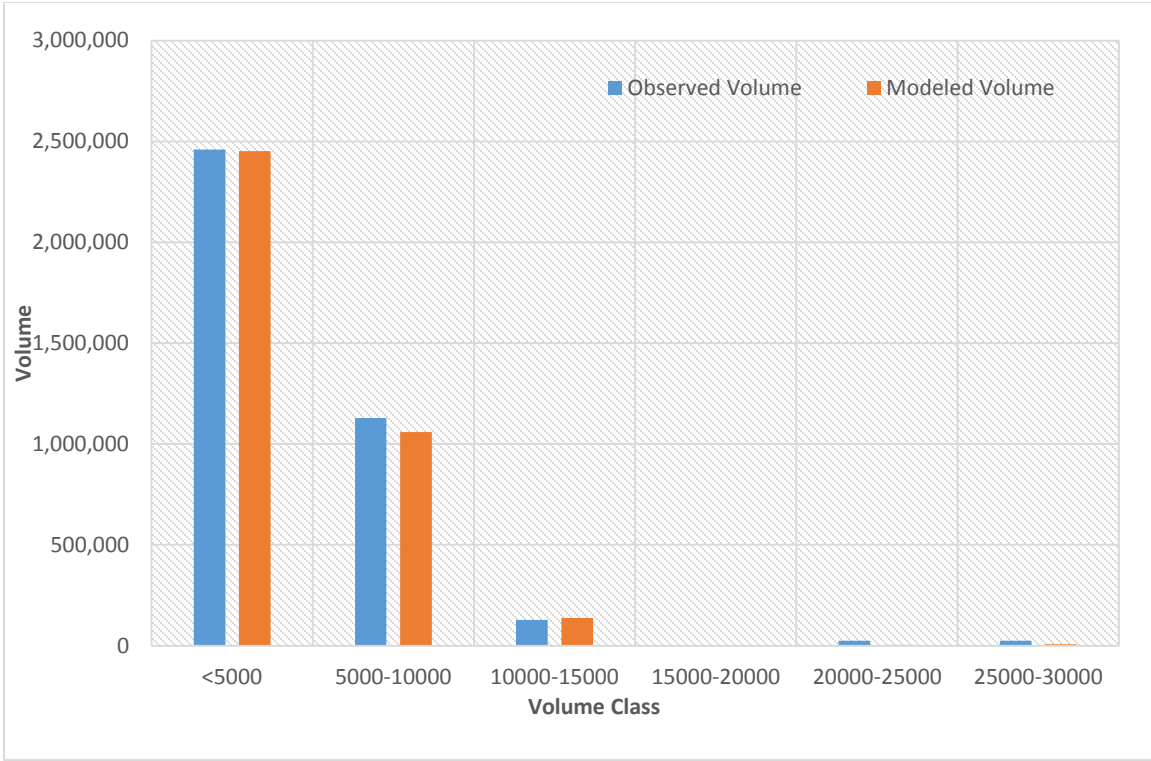
Figure 8-30. Truck count comparison for volume classes.
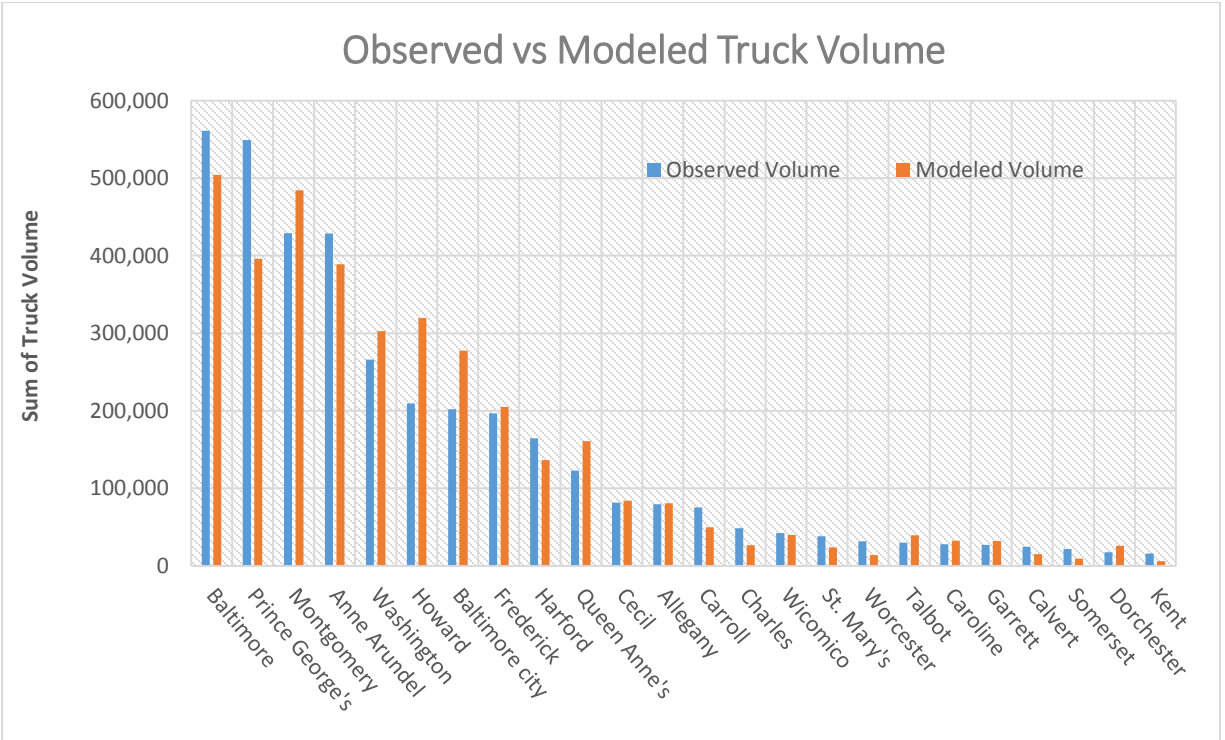
Figure 8-31. Truck count comparison for different counties.
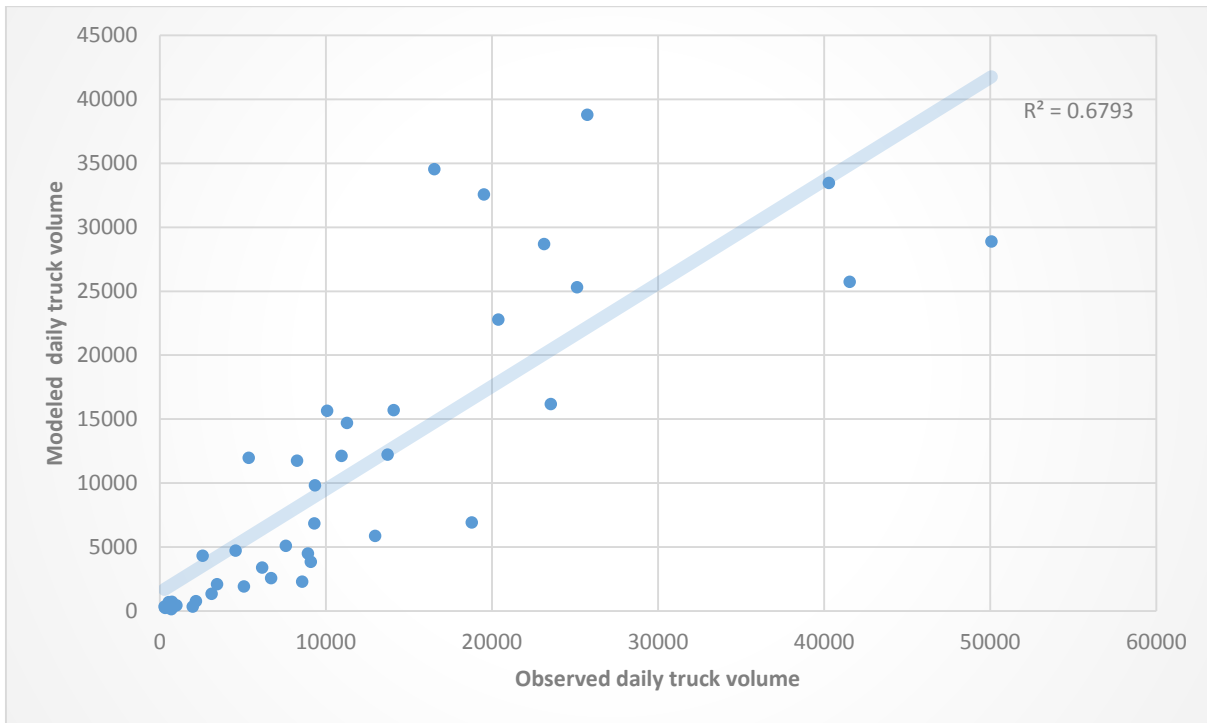
Figure 8-32. Truck count comparison at major facility types.

8.4. Integrated Model Calibration and Validation

8.4.1. Model Calibration for Non-BMC Area

The first step for calibrating the model for the non-BMC area is to obtain information such as mode share, time-of day distribution, and trip length distribution by region. This information helps identifying possible needs for parameter changes for certain regions.

The following information was summarized by region:

- Mode share
- Trip length for work trips
- Time-of-day distribution for mandatory tours
- Time-of-day distribution for non-mandatory tours
- Vehicle availability
- Toll transponder ownership
- Transit pass ownership

Figure 8-33 and Figure 8-34 show the mode share and work-trip length distribution by region. The unexpected observation in these two figures is the high walk mode share and high percentage of

zero-length trips in some Halo zones. These observations indicated a need for calibration of the destination choice and the mode choice models in the Halo area.
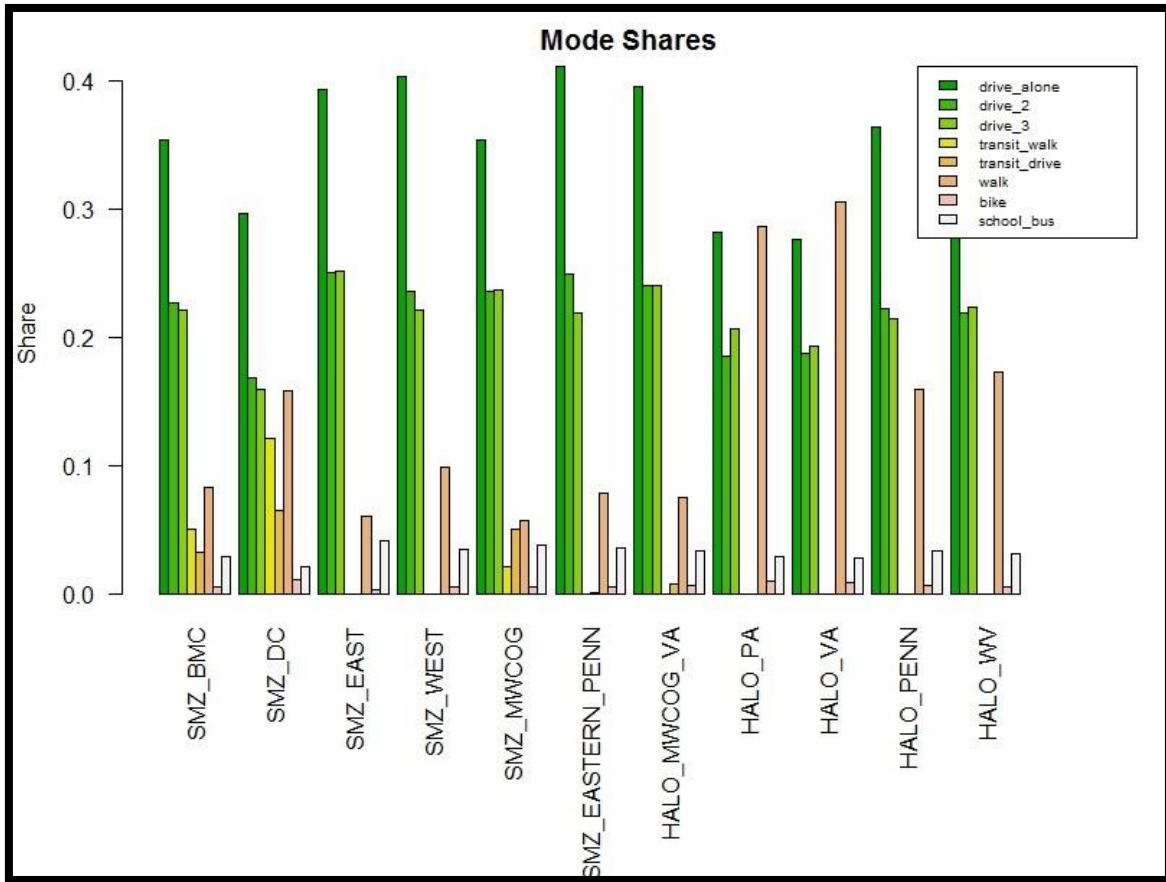
Figure 8-33. Mode shares by region.
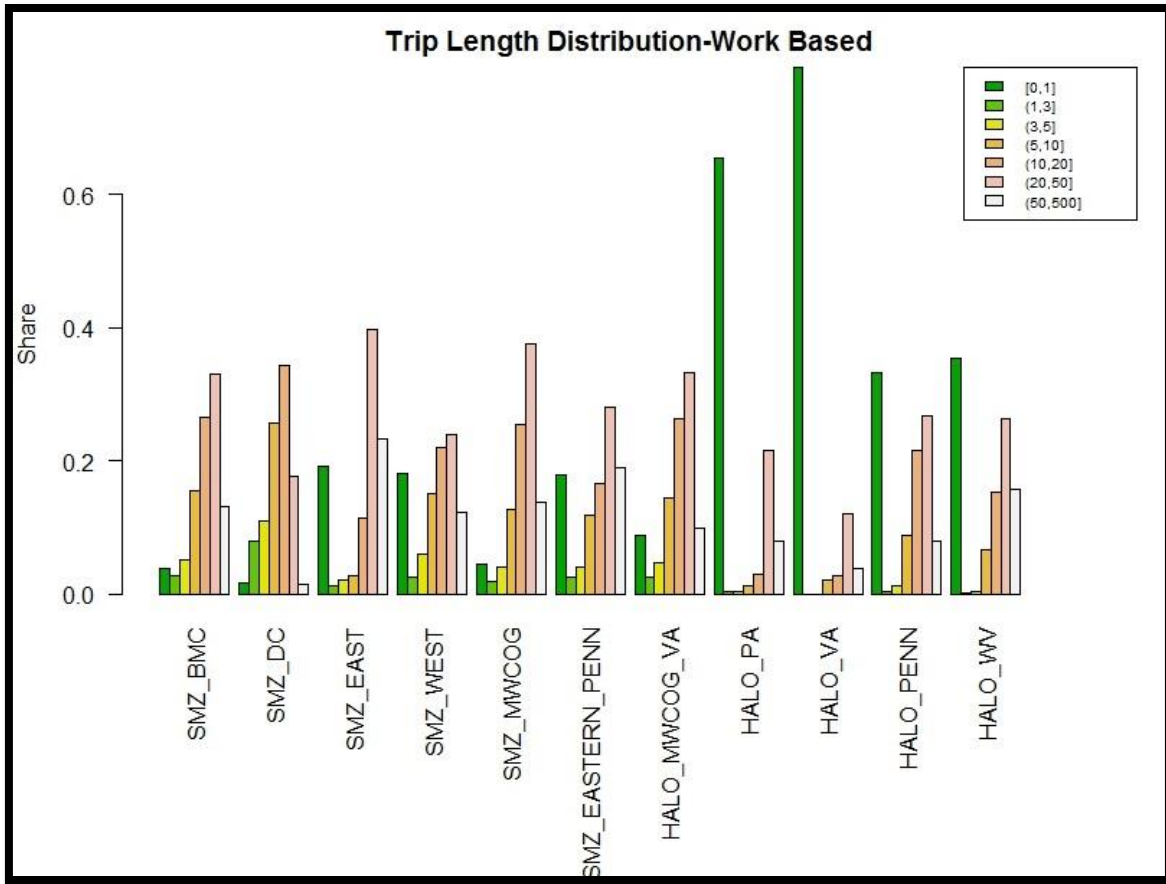
Figure 8-34. Work-trip length distribution by region.

The calibration of model parameters and skims was performed for the Halo area to decrease the walk mode share and shift the trips to destinations further away from the origin. Figure 8-35 and Figure 8-36 show the calibrated distributions that better match the expectation.

140
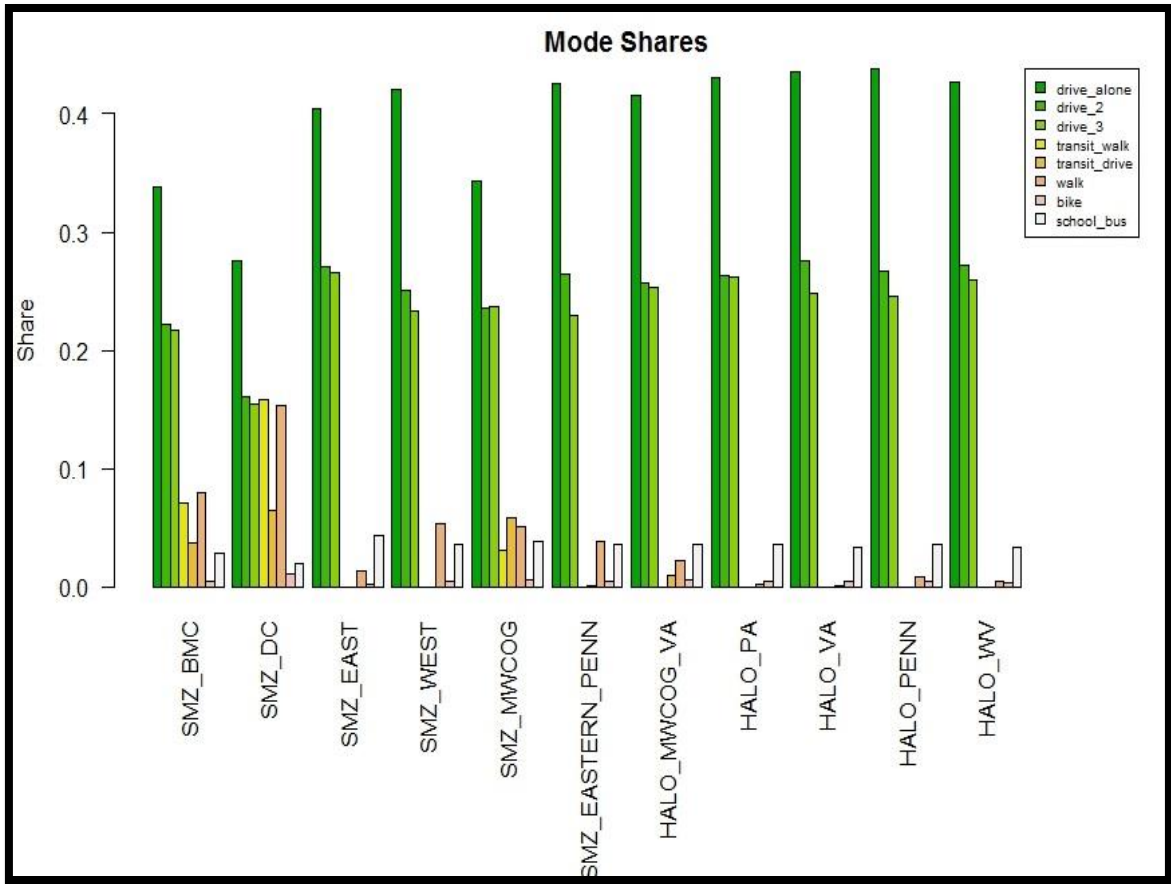
Figure 8-35. Calibrated mode share distribution by region.
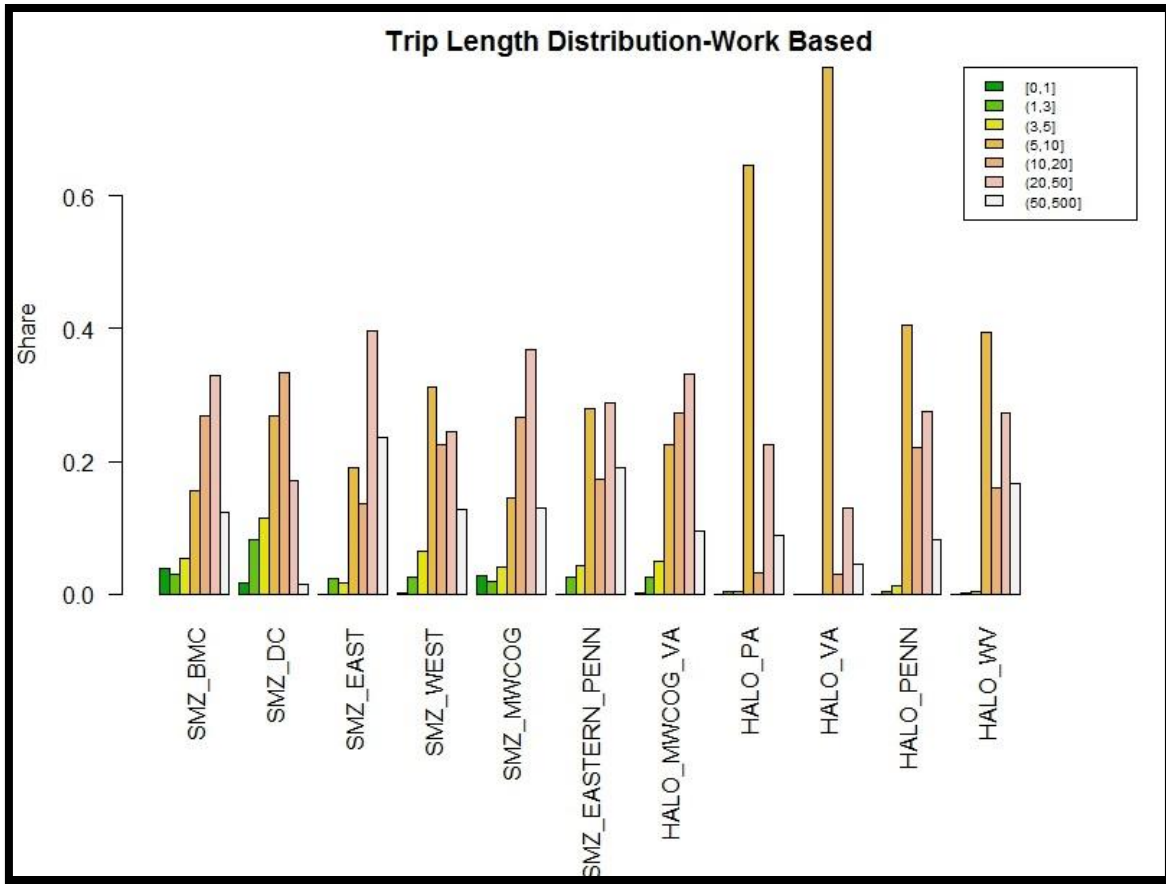
Figure 8-36. Calibrated work-trip length distribution by region.

The other model components did not show any signs that indicate a need for the calibration. Figure 8-37 through Figure 8-41 show the summaries for other model components by region.
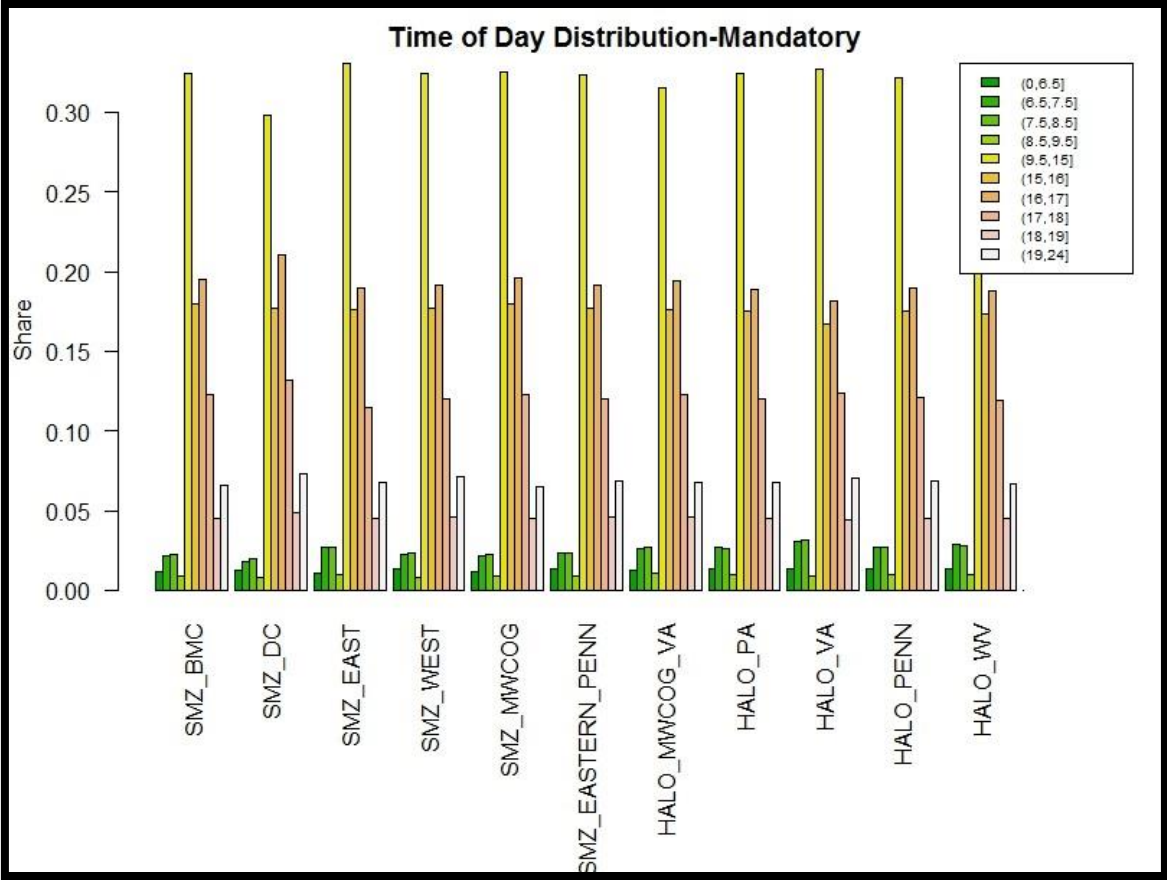
Figure 8-37. Time-of-day distribution for mandatory tours by region.
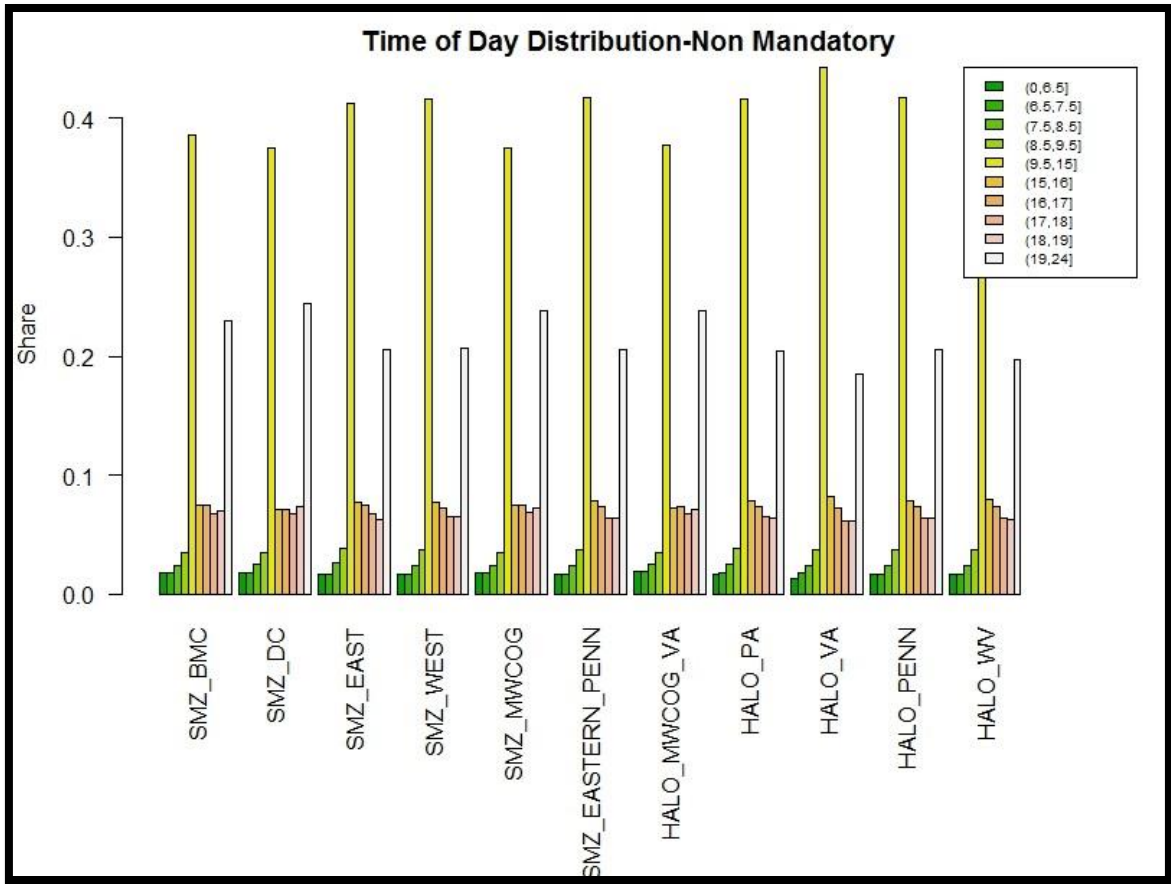
Figure 8-38. Time-of-day distribution for non-mandatory tours by region.
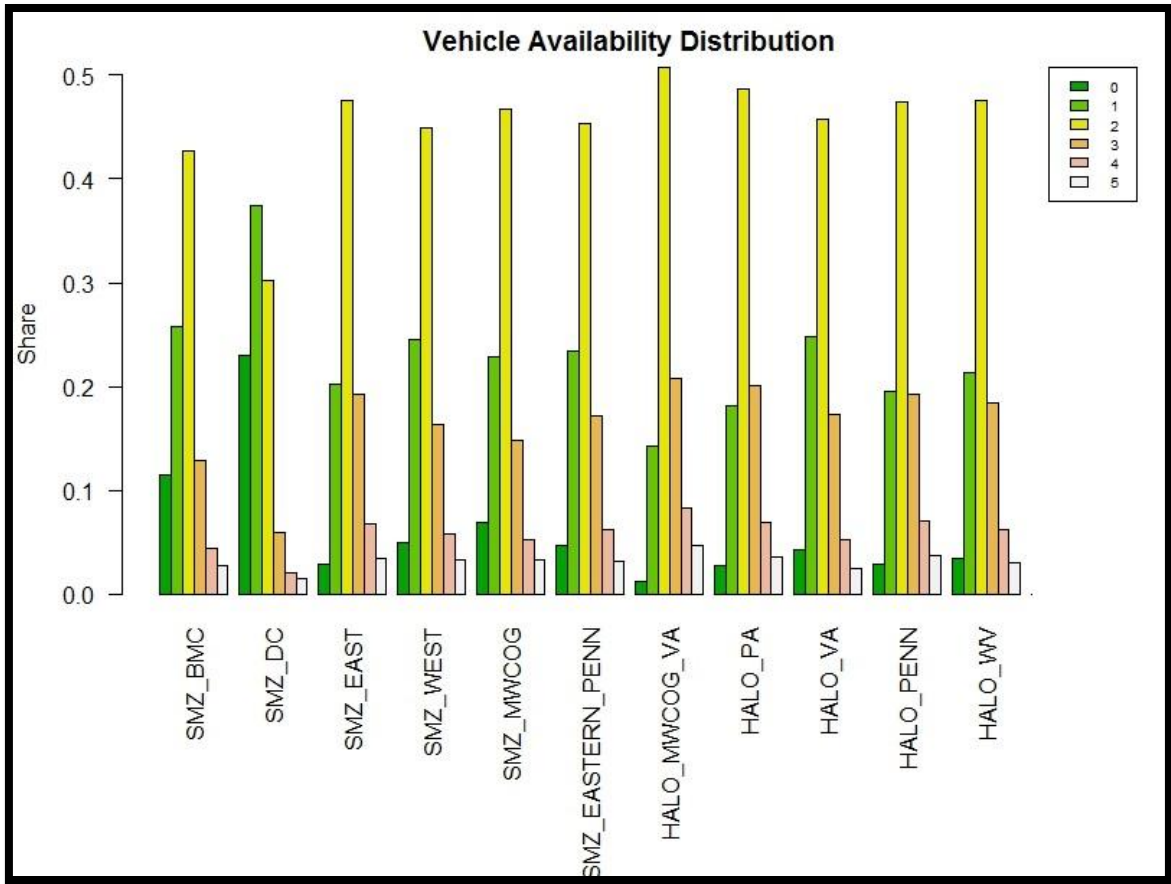
Figure 8-39. Vehicle availability distribution by region.
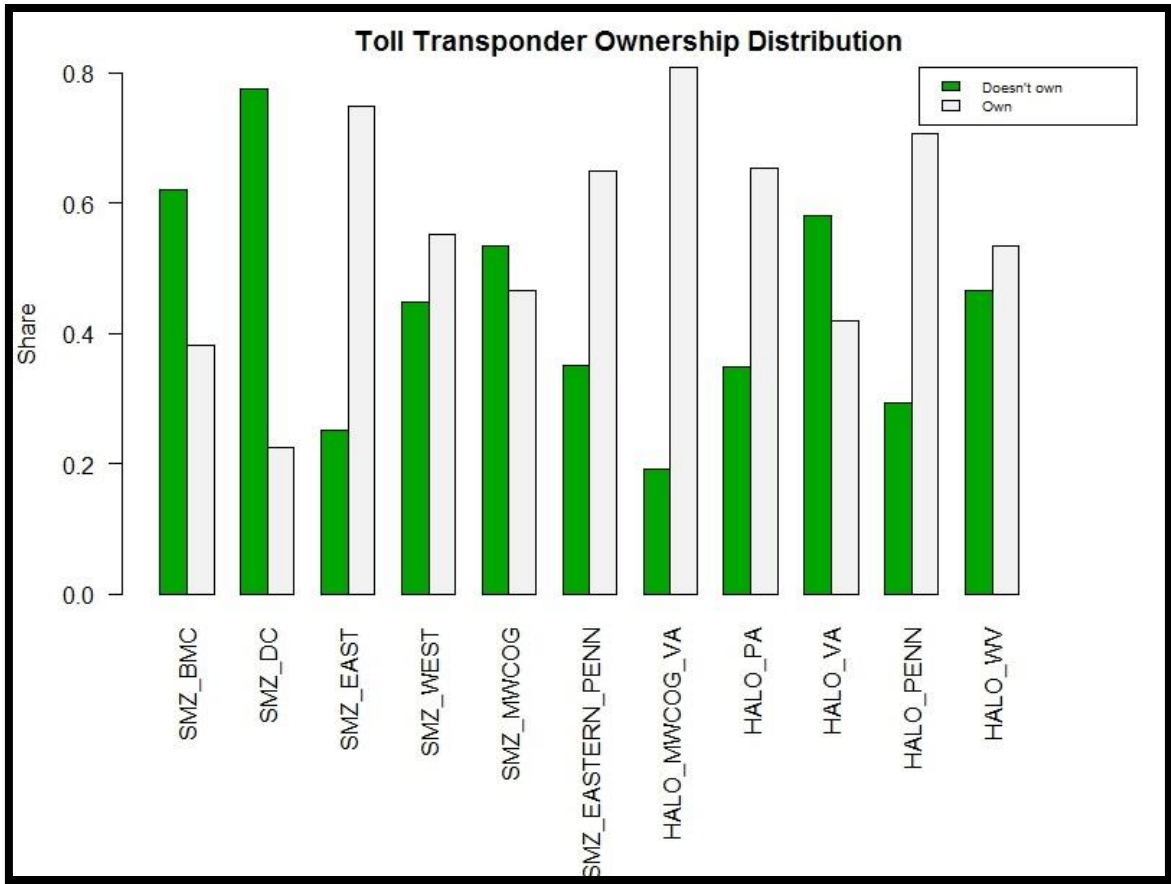
Figure 8-40. Toll transponder ownership distribution by region.
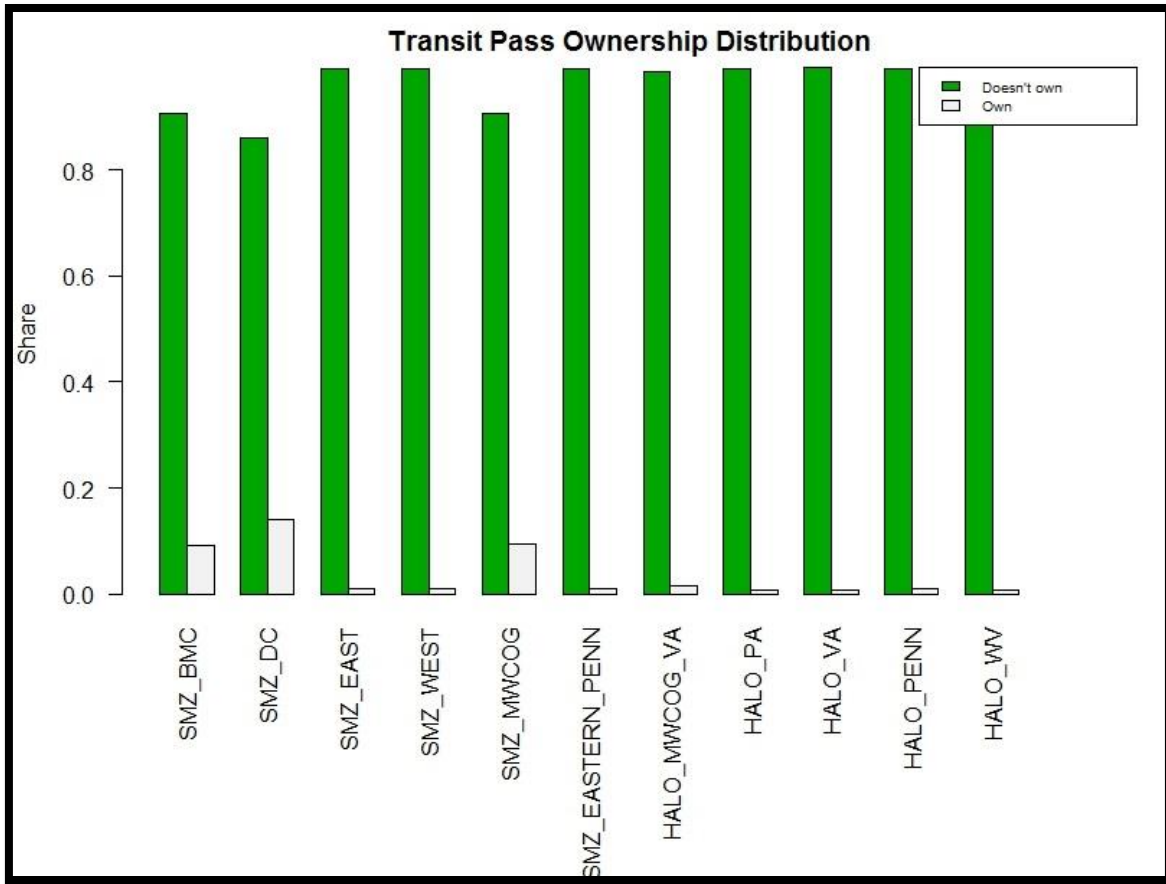
Figure 8-41. Transit pass ownership distribution by region.

8.4.2. Traffic Count Validation

The system level highway validation is an overall validation of the travel modeling process with a very specific focus on the reproduction of observed traffic volumes in a region. For highway assignment, validation tests focus on vehicular volumes and VMT across screenlines, along corridors, and any other geographic representation of the network.

To validate the produced traffic volumes, two classes of measures are considered here:

- Individual link traffic volumes;
- Intra-regional traffic flows as defined by screenlines

The screen-line validation is presented in the next sub-section. Table 8-26 summarizes the system level highway validation measures for the auto mode.

Table 8-26: System Level Highway Validation Tests

VALIDATION FOCUS	VALIDATION MEASURES	EXPECTED OUTCOMES	PRIORITY
Individual link volumes	• Plot of absolute and percent difference between modeled and observed volumes for links with counts	• Visual inspection for large errors in modeled link volumes or for general trends in errors	Level 1 (Basic Test)
	• Scatterplot of modeled versus observed daily traffic volumes by link	• RMSE and %RMSE by functional class should be within targets shown in Table; %RMSE by volume group should be within the targets shown in figure	
	• Root mean squared error (RMSE) and percent RMSE by:	• Coefficient of variation by volume group should be within the targets shown in figure	
	• state		
	• functional class (with particular attention to toll roads, HOT, and HOV facilities)		
	• volume group		
	• Coefficient of variation by volume group		
	• Anomalous links		
	• Links with zero volumes		
	• Links with very high v/c ratios		
Screenlines	• Percent deviation by screenline	• Percent deviation should be within the targets shown in figure	Level 1 (Basic test)

The hourly count data is available for 9021 of the links in Maryland, which is about the 2.8% of total links in the study area. MSTM2 assign trips into four time periods. To compare the volume for each time period, hourly count data should be aggregated. The precision of count data is one hour, however model time periods needs half-counts. Therefore, in the event it is needed, we allocate the hourly count data evenly into half-hour periods. For instance, AM period is 6:30-9:30 a.m. for volume data, and its corresponding count data are calculated based on the following equation:

$$\text{Count (AM Period)} = 0.5 * \text{Count (7)} + \text{Count (8)} + \text{Count (9)} + 0.5 * \text{Count (10)} \quad (8\text{-}2)$$

After each model iteration, link volumes are compared with count data to better understand the improvement in model performance. Figure 8-42 to Figure 8-45 show the scatterplot of modeled versus observed traffic volumes by link in four different time periods. In each figure there are three plots from three iterations of the integrated model.

As it was expected from iteration 1 to iteration 3, volumes are getting closer to count data and points are less scattered. Iteration 3 results for all time periods look reasonable. The other observation is that links with higher volume dominate the observations, whereas links with lower volume are dominant in the night period.
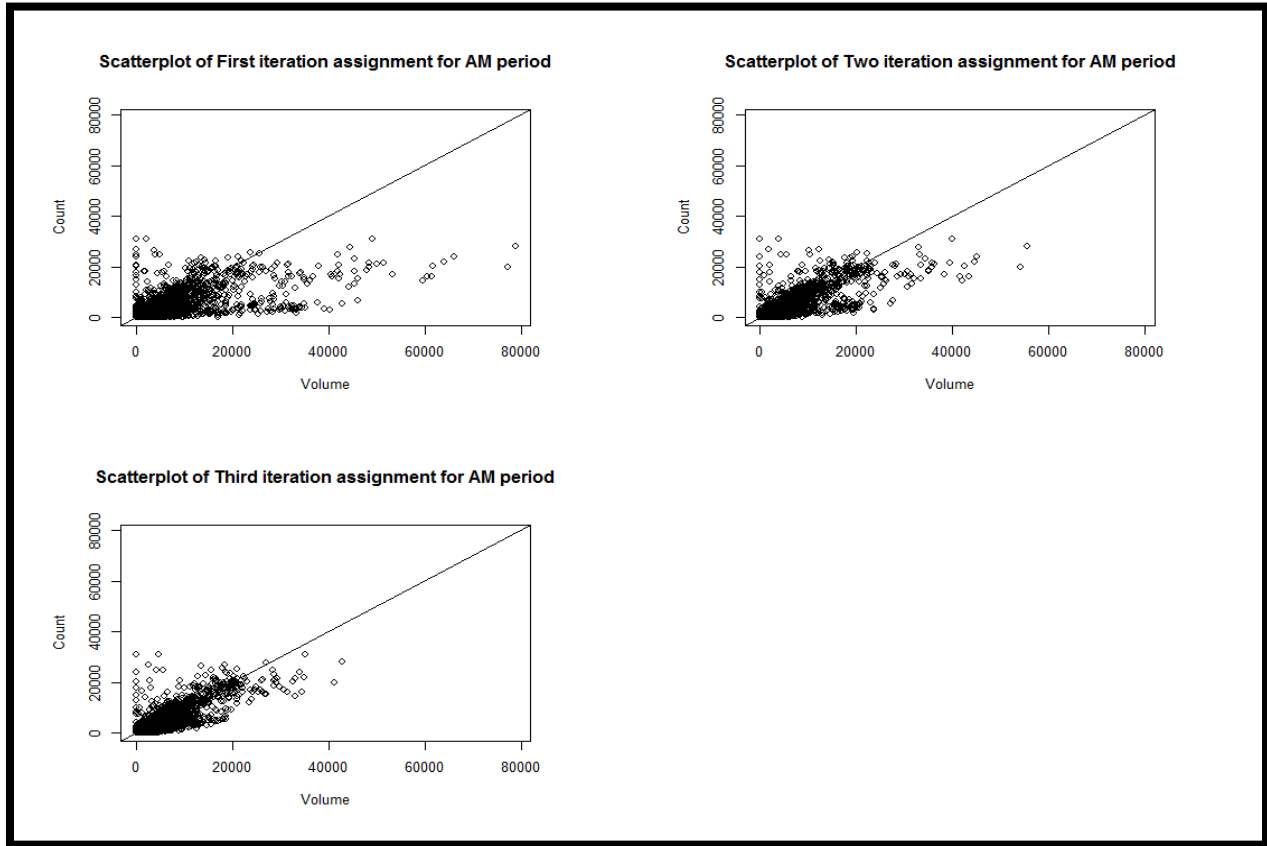
Figure 8-42. Scatterplot of volumes versus count data for am period.
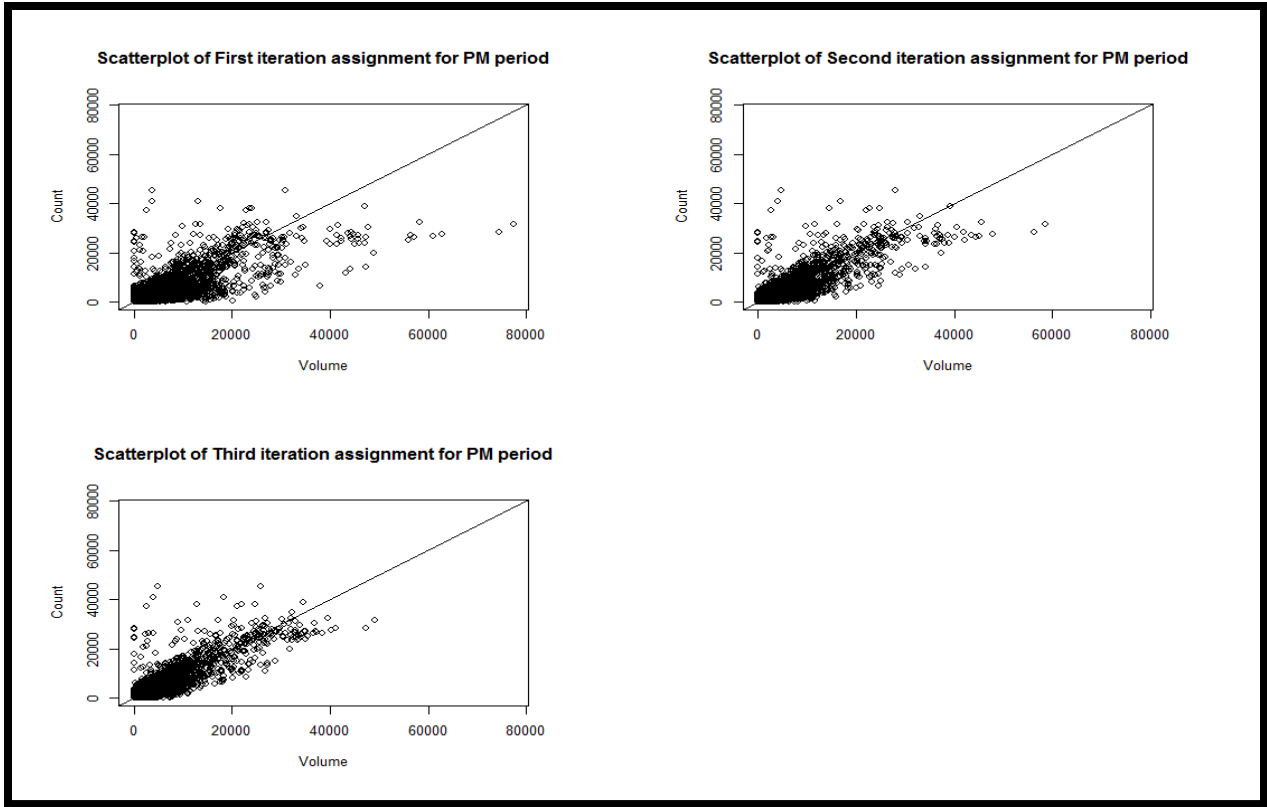
Figure 8-43. Scatterplot of volumes versus count data for p.m. period.
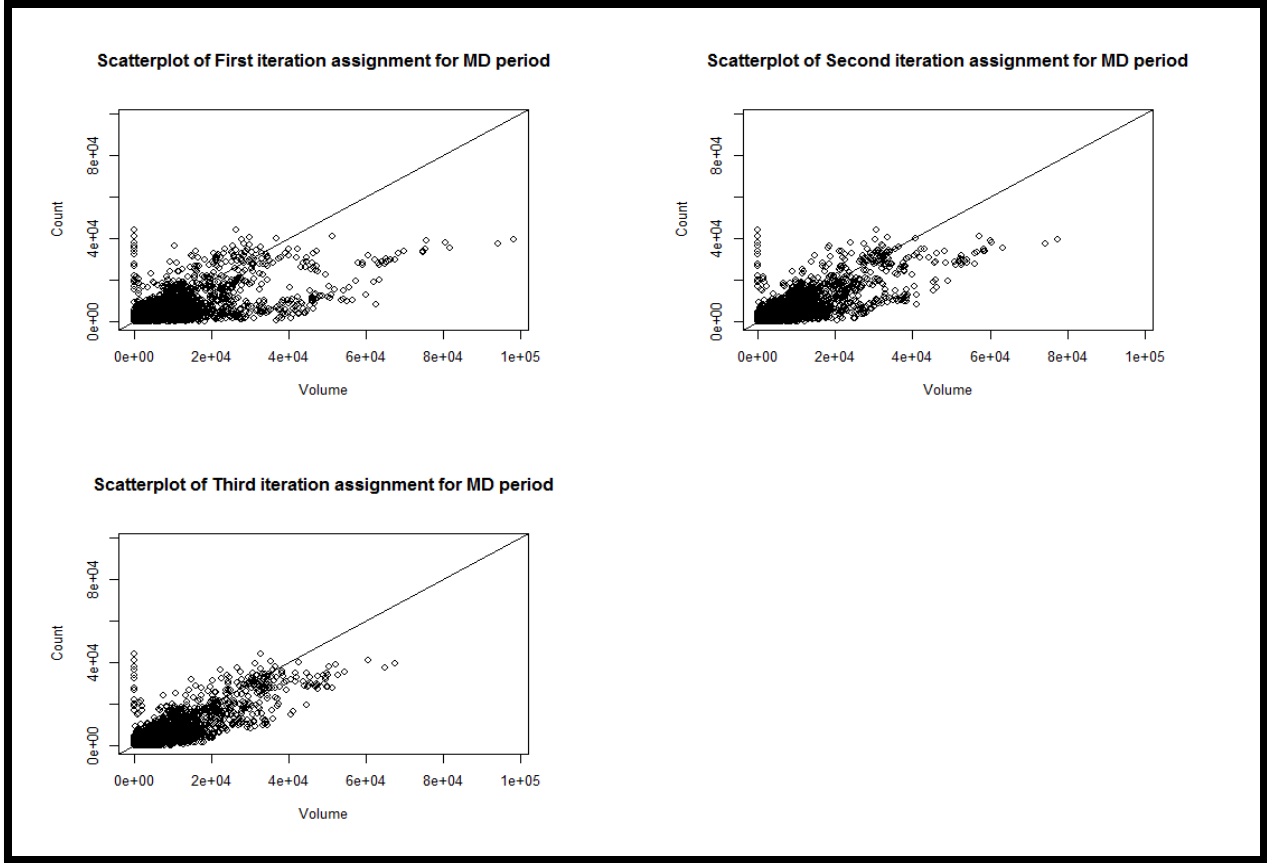
Figure 8-44. Scatterplot of volumes versus count data for mid-day period.
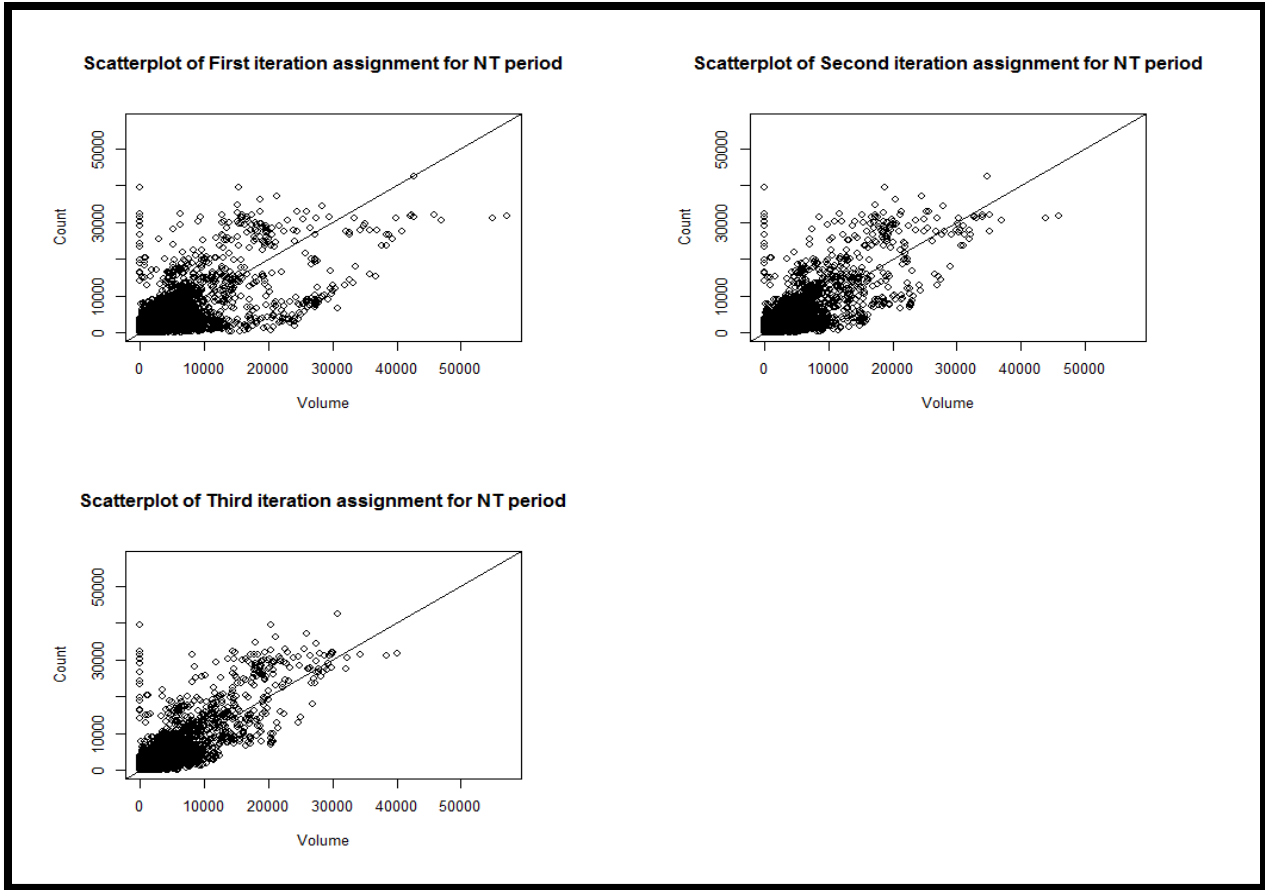
Figure 8-45. Scatterplot of volumes versus count data for the night period.

Table 8-27 shows the targets and results for RMSE and RMSE% by functional class. The results are not in the range of targeted values.

Figure 8-46 presents the target and results for assignment by volume group using %RMSE. The RMSE percentages are higher for most of the groups in the results.

Figure 8-47 presents the target and results for assignment by volume group using percent difference. Percent difference is measured based on following equation.

$$Percent\ Difference = \frac{|Assigned - Count|}{Count} \qquad (8\text{-}3)$$

Overall, results are better using %RMSE measure. For the higher volume groups the error ranges are beyond the targeted thresholds.

Table 8-27. RMSE and %RMSE Targets by Functional Class[1]

Functional Class	Target Values		Model Values	
	RMSE[2]	%RMSE[3]	RMSE[2]	%RMSE[3]
Freeways	<12,500	<20%	13164	46%
Principal Arterials	<3,750	<30%	8548	42%
Minor Arterials	<3,000	<40%	4618	55%
Collectors	<2,250	<70%	3264	78%
All Links	n/a	<40%	-	73%

[1] $$RMSE = \sqrt{\frac{\sum(Count-Assigned)^2}{(Number\ of\ Observations-1)}}$$

$$\%RMSE = \frac{100 \times RMSE}{Avg.\ Count}$$

[2] Based on one-half lane of capacity and assumption of 8% peak hour factors for interstates, freeways, system ramps, expressways, and external connectors; based on 10% peak hour factor for other functional classes

[3] Rules of thumb from model validation efforts for several regions.

153
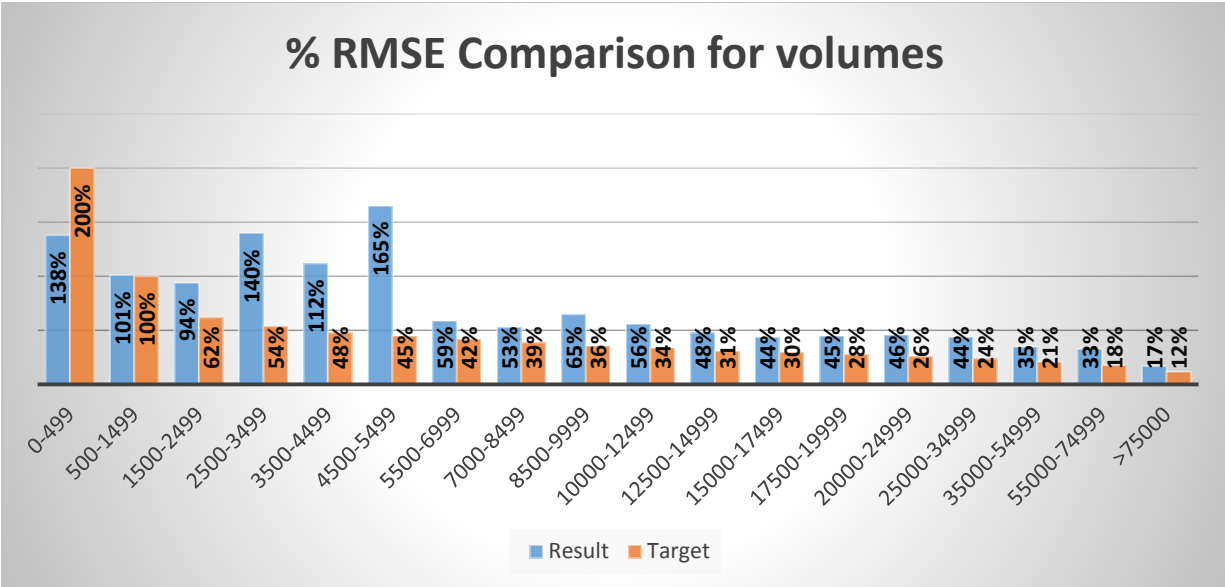
Figure 8-46. Traffic assignment %RMSE comparison by volume group.
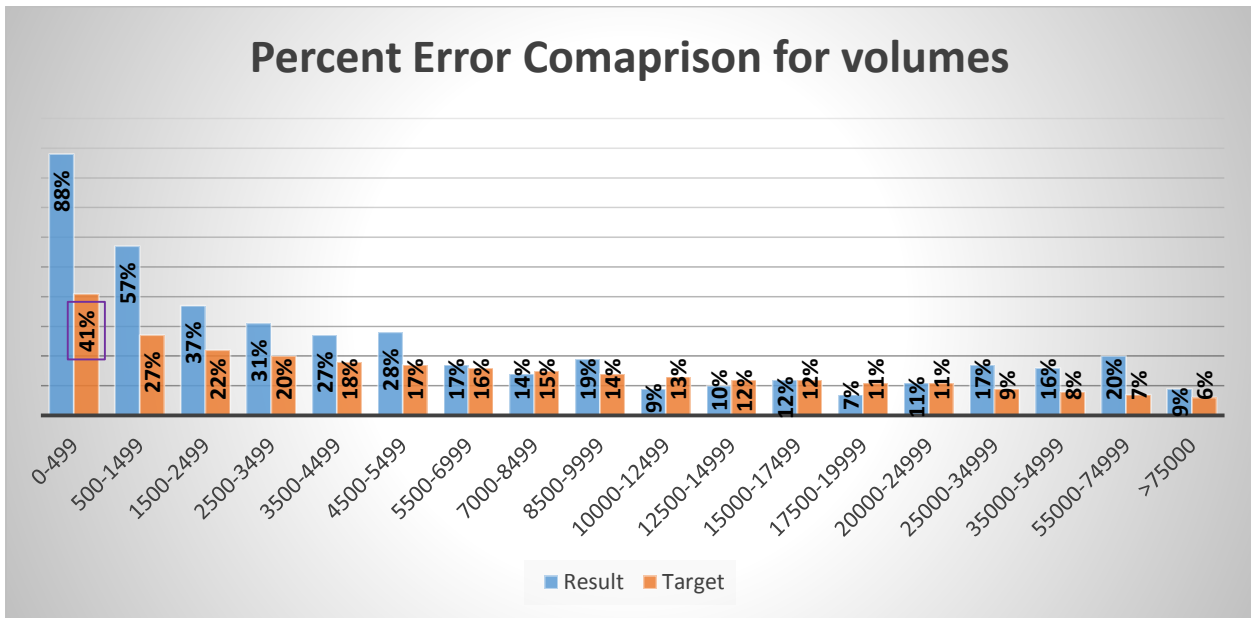
Figure 8-47. Traffic assignment percent difference targets by volume group.

154

The region level summaries are important in determining the validity of the model parameters estimated within the BMC model region for the rest of the state of Maryland. These differences can help guide the adjustment of model parameters specific to particular regions. Table 8-28 shows the comparison between estimated VMT for different regions in Maryland and SHA year 2015 mobility report estimates. These results can be used as a guide for further calibration of models for different regions.

Table 8-28. VMT comparison for Maryland regions.

Region	Annual VMT(billions)		
	SHA 2015 estimates	MSTM 2.0	% Difference
BMC	25.6	21.13	-17.5%
Washington Metropolitan	19.5	17.04	-12.6%
Western Maryland	3.0	2.7	-10%
Southern Maryland	5.9	2.33	-61%
Eastern Shore	3.3	4.85	+46%
Total	57.3	48.14	-16%

There are also a couple of other measures of validation that can give us better insight about model performance. Up until now, all comparisons were based on absolute value of differences, but another important factor to consider is the ratio of volumes that are being over-predicted. Here, out of 9021 links, 4872 (55%) of the links have higher volumes and the rest has higher counts. It seems the model does not suffer from a systematic bias for predicting volumes.

The number of links with zero volumes is another important measure that can represent a source of possible error. There are 921 (~2%) of links with zero volume (but positive counts) that need further investigation. There is not any specific pattern for their facility type or region and they should be considered case by case.

8.4.3. Screen-line Validation

SHA has identified 31 screenlines for Maryland. Figure shows the validation of assigned volumes against counts for these screenlines. At each screenline, both volumes and counts are close to the 45-degree slope line (Figure 8-48). R-square of 0.91 shows reasonable match between observed and modeled volume at screenlines.
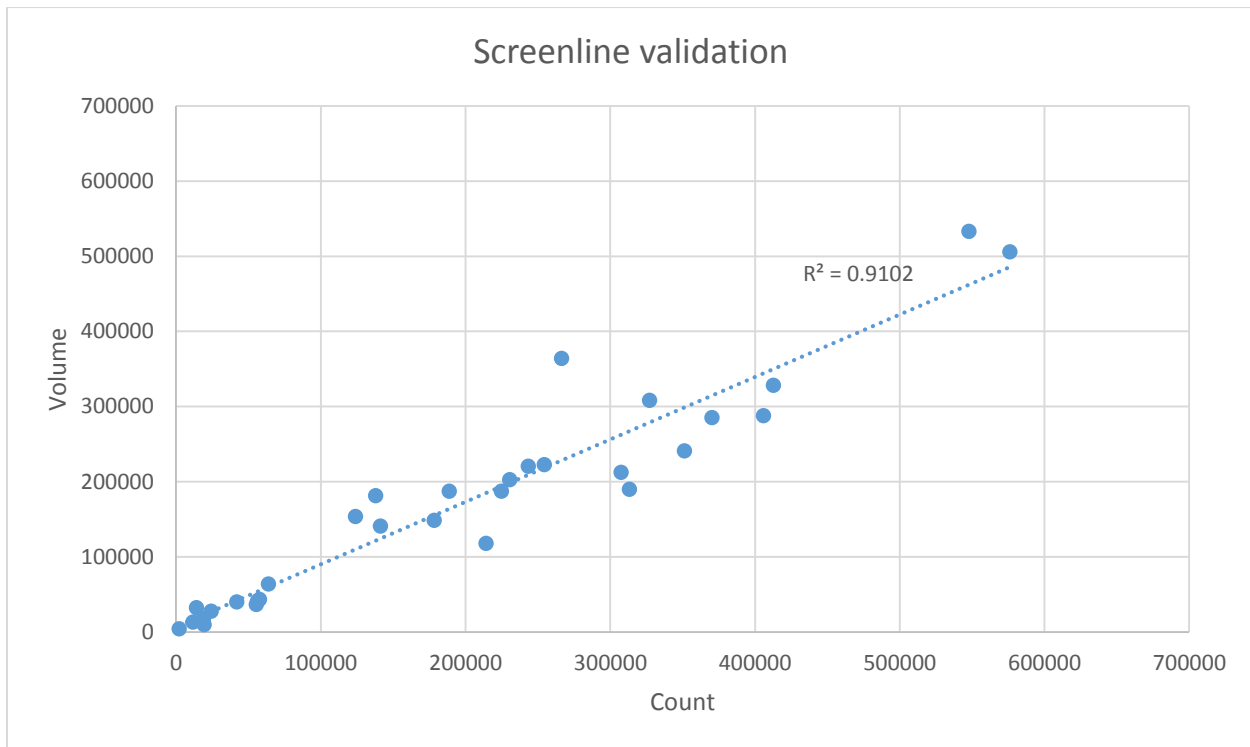
Figure 8-48. Screenline validation results

8.4.4. *Validation against Travel Surveys*

In this section, validation of the integrated model result has been completed using three independent data sources; i.e., data that have not been used in model development. The MSTM1 outputs, the 2009 National Household Travel Survey (NHTS), the 2007-2008 TPB Household Travel Survey (HHTS), and the American Community Survey (ACS) have been used to validate the integrated model result.

MSTM1 outputs are used as a benchmark to analyze the results of MSTM2 integrated model. Results of short-distance model, long-distance model, and freight model of both MSTM1 and MSTM2 has been compared separately. Comparison is shown in Table 8-29.

Table 8-29. Comparison of Number of Trips Generated between MSTM1 and MSTM2

Demand Comparison	Entire Zones		
	MSTM1	MSTM2	Difference (%)
Short-Distance Model	AM 6681024.96	AM 7246138.62	8.46
	PM 9962562.21	PM 10903384.36	9.44
	MD 16416633.96	MD 9578618.89	-41.65
	NT 7337137.7	NT 5514494.12	-24.84
Long-Distance Model	AM 2566401.21	AM 1441730.23	-43.82
	PM 4030855.24	PM 1603392.04	-60.22
	MD 5928761.4	MD 2418737.18	-59.20
	NT 3469168.2	NT 1844456.44	-46.83
Freight Model	AM 1031272.32	AM 264614	-74.34
	PM 959388.04	PM 266726	-72.20
	MD 2596003.45	MD 882243	-66.02
	NT 1261414.62	NT 165760	-86.86

It should be noted that the comparison of the number of generated trips between MSTM1 and MSTM2 is not an apples-to-apples comparison, since not only does base year differ in MSTM1 and MSTM2, but also the time period definition. Time period definition for both MSTM1 and MSTM2 is shown in Table 8-30.

Table 8-30. MSTM1 and MSTM2 Time Period Definitions

Time Period	MSTM1	MSTM2
AM Peak	6:30 am - 9:30 am	6:30-9:30
Mid-Day	9:30 AM - 3:30PM	9:30:00 AM- 3:00PM
PM Peak	3:30 PM - 6:30 PM	3:00 PM- 7:00PM
Night	6:30 PM - 6:30 AM	7:00 PM- 6:30AM

Although there are some differences in base year and time period of these two models, this comparison can still provide some insight regarding the overall performance of MSTM2 model.

For further investigations, a more detailed comparison of short distance model results has been conducted. Table 8-31 and Table 8-32 show the demand comparison of short-distance models for trips for which both origin and destination zones are located in Maryland, and for trips for which either origin zone or destination zone is located in Maryland, respectively.

Table 8-31. Short-Distance Models' Number of Trips Comparison (both Origin and Destination Located in Maryland)

Short-Distance Model Comparison		
MSTM1	**MSTM2**	**Difference (%)**
AM 2956540	AM 3089169	4.485953175
PM 4426183	PM 4426879	0.01572461
MD 7356356	MD 3754640	-48.96059951
NT 3243770	NT 2237638	-31.0173656
Total 17982849	Total 13508326	-24.88216967

Table 8-32. Short-Distance Models' Number of Trips Comparison (either Origin or Destination Located in Maryland)

Short-Distance Model Comparison (From MD or to MD)		
MSTM1	**MSTM2**	**Difference (%)**
AM 3388684	AM 3581974	5.703984202
PM 4978283	PM 5209251	4.639511253
MD 8131125	MD 4397203	-45.92134545
NT 3645680	NT 2624270	-28.01699546
Total 20143772	Total 15812698	-21.50080928

2009 NHTS survey and 2007-2008 TPB HHTS can also be used to compare the generated trip rates, number of trips generated by purpose, time-of-day choices, and mode shares. Since NHTS 2009 did not provide detailed origin and destination location of each trip, an apples-to-apples comparison of generated trips based on their origin and destination location is impossible. On the other hand, 2007-2008 TPB HHTS has the origin and destination location of each trip based on BMC area zone structure, which makes it possible to compare the integrated model results against this survey data in more details. Table 8-33 shows the comparison of number of trips generated

for MSTM2 model and the 2007-2008 TPB HHTS survey results. HHTS samples are expanded to population using provided weights.

Table 8-33. MSTM2 Number of Trips Generated Comparison to TPB HHTS Survey

Number of generated trip					
Time Period	MSTM2 results for HHTS area	Trip Ratio	HHTS results	Trip Ratio	Difference (%)
AM	3200584	22.43	3212755	19.54	-0.38
PM	4581328	32.11	5566065	33.85	-17.69
MD	4134866	28.98	5110478	31.08	-19.09
NT	2352439	16.49	2554356	15.53	-7.90
Total	14269217	100	16443654	100.00	-13.22

Comparison of time-of-day distribution is shown in Table 8-34 and Figure 8-49. The number of generated trips is calculated based on MSTM2 time period definition for both HHTS and NHTS data survey. Time-of-day for MSTM2 is calculated based on three different scenarios, including time-of-day choice for trips for which either origin or destination is located in Maryland, trips for which both origin and destination are located in Maryland, and trips generated in HHTS survey study area.

Table 8-34. Comparison of Time-of-Day Choice between Integrated Model Result and Surveys

Time-of-Day					
Time Period	MSTM2 (From MD or to MD)	MSTM2 (From MD and to MD)	MSTM2 (HHTS Study Area)	2009 NHTS	TPB HHTS Survey
AM	22.65	22.87	23.22	18.52	19.54
PM	32.94	32.77	33.31	27.82	33.50
MD	27.81	27.80	26.96	38.59	31.13
NT	16.60	16.56	16.51	15.08	15.83
Total	100.00	100.00	100.00	100.00	100.00
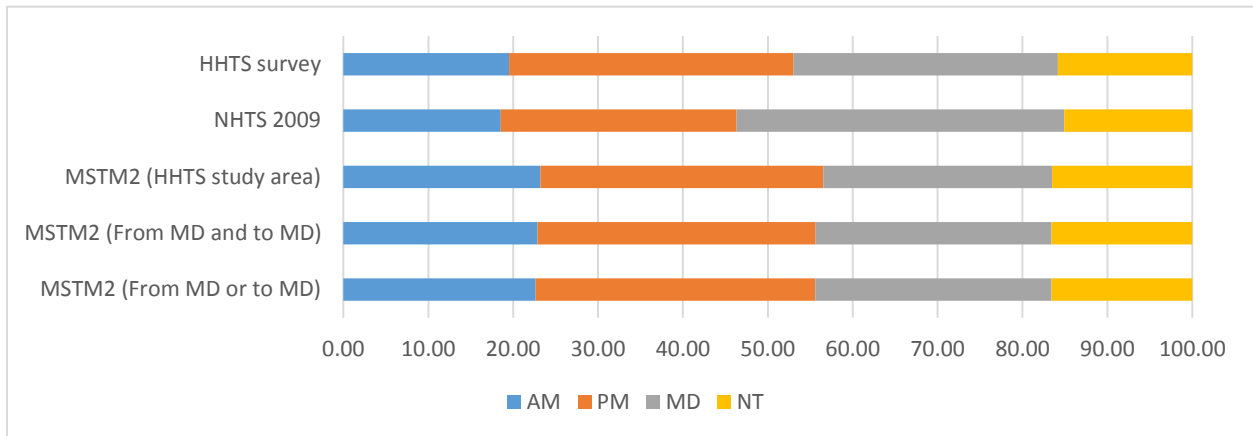
Figure 8-49. Comparison of time-of-day choice between integrated model result and surveys.

Trip length distribution between the model and NHTS survey was analyzed (Figure 8-50). In order to have a geographically consistent comparison, the comparison is focused on trips made by households in Maryland. The trip length is segmented into 5 intervals from 0 to 50 miles.
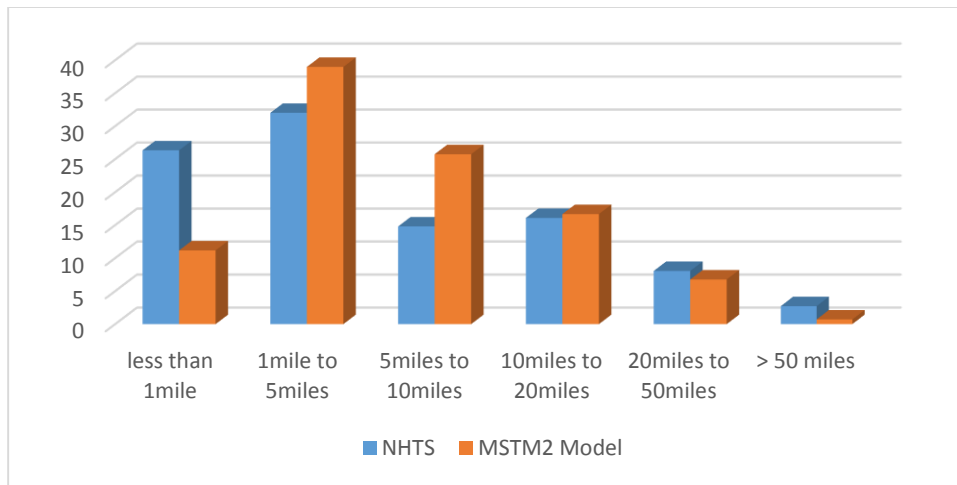
Figure 8-50. Comparison of trip length distribution for Maryland residents.

MSTM2 trip mode choice results were validated against the ACS 2015 dataset that devoted a section to work trip mode choice. The work trip mode choice was compared for Maryland and Washington, D.C. (Figure 8-51) and non-Maryland (Figure 8-52) separately. ACS 2015 contains work trip mode choice information for 10 Maryland counties, the District of Columbia, and 10 counties outside of Maryland that were considered in MSTM2 short distance model.
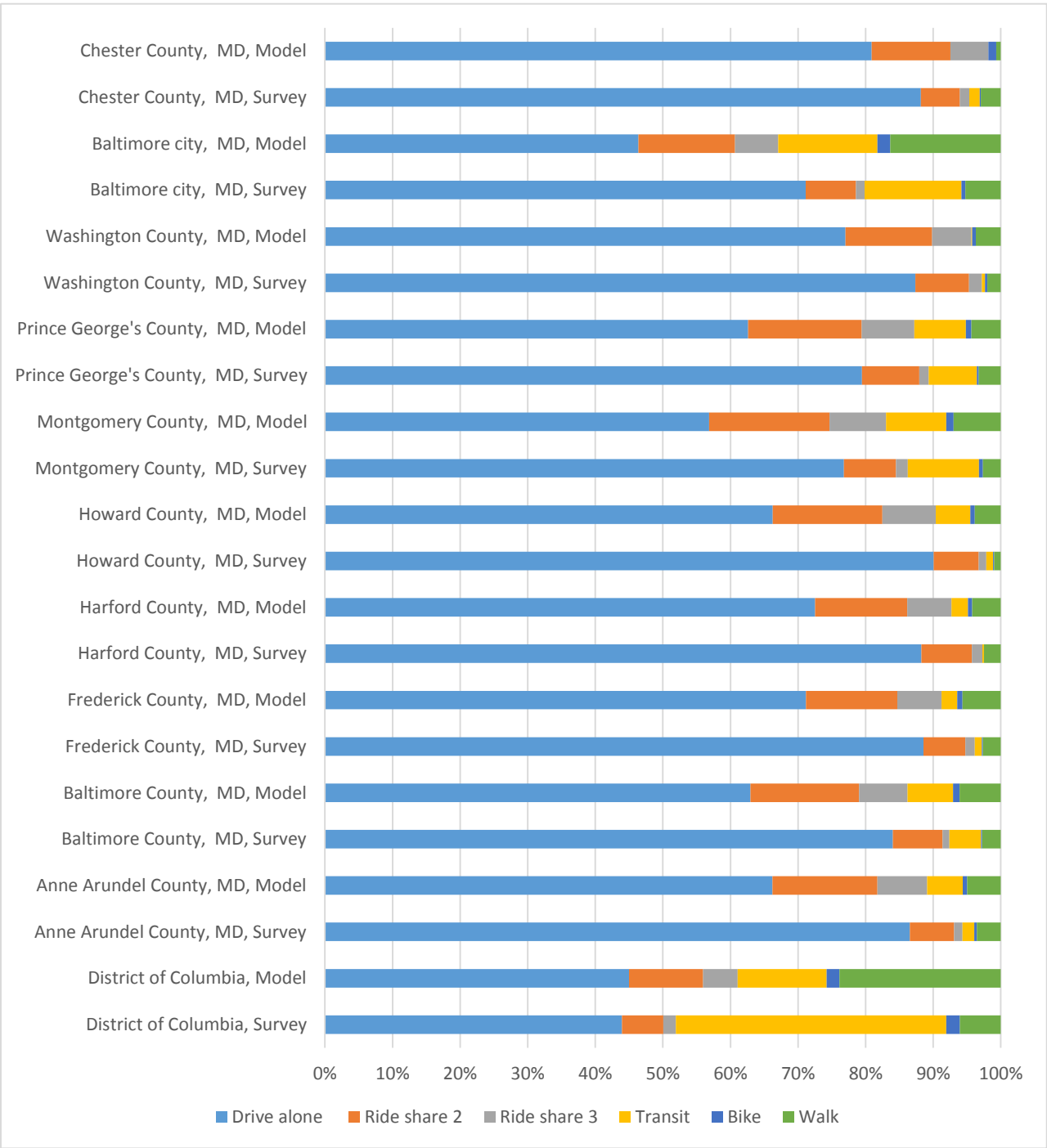
Figure 8-51. Validation of work trip's mode choice for Maryland and District of Columbia.
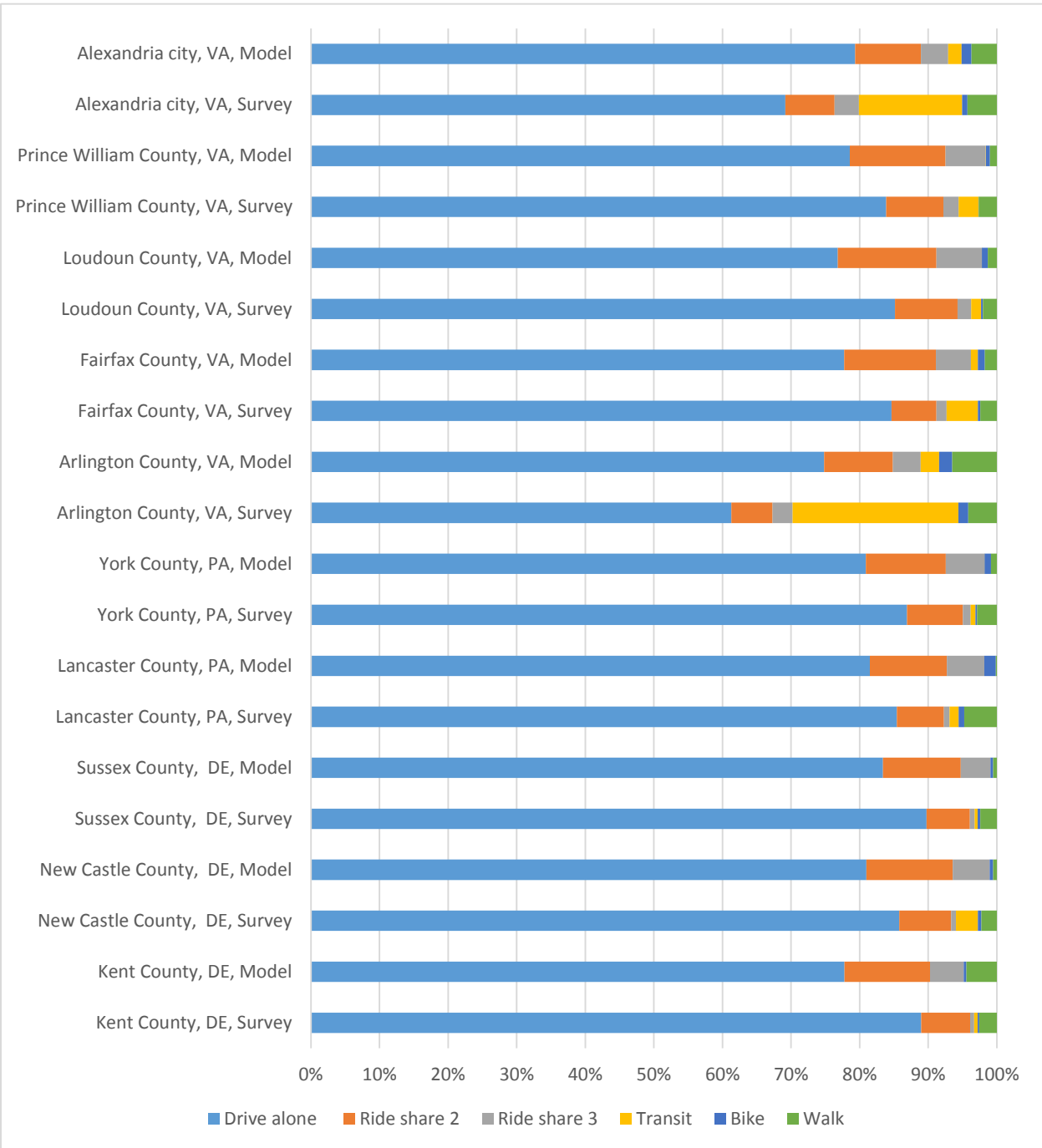
Figure 8-52. Validation of work trips mode share for non-Maryland area.

9. OUTREACH AND TECHNOLOGY TRANSFER

This chapter presents the outreach plan and technology transfer activities proposed as part of this project. The main objective of the outreach and technology transfer activities is to provide effective training of agency users for practice-ready integrated models.

9.1. Engage stakeholders

The project team works closely with MD SHA and BMC to help showcase the capabilities of MSTM2 as a unique and comprehensive decision-making tool that helps agencies plan solutions for congestion, sustainability, transit, pollution, and others. The MSTM2 model is an indispensable tool for both MD SHA and BMC to highlight the impacts of policies related to passenger and freight travel to key citizen groups, the private sector, and other agencies in the state of Maryland.

9.2. Visualization of Model Results for Decision-making Support

9.3. Supporting Documents

The MSTM2 team has provided a brief user's guide as a part of this document. The team will provide a more comprehensive user's guide covering each of the separate modeling tools and the integrated model to offer step-by-step guidance to the agency staff. The goal is to provide a go-to resource that, together with the continuous support of the project team, will support the agencies in integrating these tools as part of their internal procedures.

10. USER GUIDE: HOW TO RUN THE MODEL

10.1. Installation and Implementation Overview

The Maryland Statewide Transportation Model Version 2 activity-based model (ABM) is powered by TourCast and built to take advantage of multi-core processors within a single machine running the Windows Server or Windows 7 operating system.

10.1.1. Hardware Requirements

Below are the requirements for an efficient ABM operation. The program will work on systems with less RAM and processors (e.g., a laptop with 8GB RAM and four cores), although it will perform better on the configuration defined below:

- Hard disk capacity:

  - ~200GB per scenario
- RAM requirements

  - 16GB
- CPU requirements

  - 8-16 core processors
  - Processor speed matters nearly as much as number of cores, e.g. a 3.4GHz i7 processor with 4 cores may outperform a 2.4GHz Xeon processor set with 24 cores.

10.1.2. Software Requirements

These are the minimum requirements to run the ABM with TourCast. Please note that PopGen2 is only needed when new synthetic population is necessary.

- PopGen2 (optional)
- Cube 6.4.1 with Cube Cluster

  - Confirm the presence of the license file. Depending on the installation, there may need to be a cubevoyager.lci in the Program Files\Cube\Voyager directory contains the file. If this file is present in the 32-bit program files folder, copy it to the 64-bit folder.
  - Confirm that tpptools.dll, tppdlibx.dll, tputlib.dll are present in the Program Files\Cube\Voyager directory.
- Python 2.7

  - Ensure that Python.exe is included in the PATH variable.
- Visual Studio Redistributable or Visual Studio 2010 or later

  - http://www.microsoft.com/en-us/download/confirmation.aspx?id=30679

- Microsoft .NET Framework 4.0
- TourCast binaries and configuration scripts (included as part of model zip file)

10.1.3. Installation Process

This section describes the steps to install the model.

1. Ensure that the required software is present.
2. Choose a location on the computer to unzip the model. The intermediate and final data will be generated within the model folder, so select a location that has sufficient storage (~50GB per speed feedback).
3. NOTE: do not install the model to a directory that includes an "@" symbol. This symbol is interpreted by Cube to represent variables and will prevent Cube from following the file path. Note that folders under the "Users" directory, such as My Documents, include an "@" symbol in the file path.
4. Using Cube, open the InSITE_Ver**.cat security catalog file. Click "Yes" when prompted to update the Application structure path (
5. Figure 10-1).
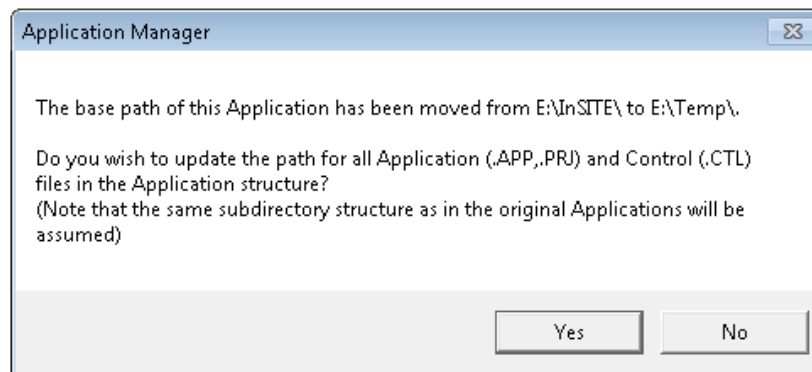
Figure 10-1. Update application structure path.

You may also see the following warning message (

Figure 10-2) appear several times as the path is updated; they are benign and you can just click "OK" to each one:
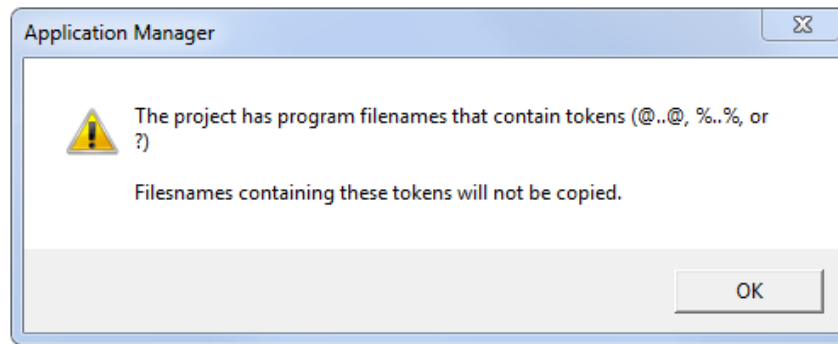
Figure 10-2. Update path warning.

Once the catalog is open, you should see the root application group. Details of each application group are provided in the following catalog overview section.

10.1.4. Folder Structure

Within the catalog directory that has been expanded there are the following folders:

- Cube: Cube application group (*.APP) and script files (*.S)
  - The Cube files are organized into subdirectories according to the application group.
  - The Cube script file names are set by the application group name and the program type. For example, a Pilot program in the Root application will have a file name RTPIL***.S
- Input: synthesized population, highway factors, and external/special generator parameters
  - **accessibilities:** accessibility logsum parameters
  - **network**: preprocessed highway networks
  - **skims**: loaded highway and transit skims to prime first iteration (optional)
  - **transit**: transit factor, line, fare, and system files.
  - **Popgen:** preprocessed PopGen outputs
  - **InputsFromPrevModel**: external, truck, and special generator files from trip based model
  - **sedata**: zonal and parcel data
- Doc: user documentation and validation reports
- TourCast:
  - **bin:** TourCast executables and libraries
  - **logs:** log output from TourCast
  - **script:** TourCast model component configuration files
  - **support:** visual studio redistributable and Cube DLLs
- Root: Cube scenario directory – all model outputs will be written to subdirectories of this folder.

10.1.5. Catalog Overview

At a high level, the InSITE ABM catalog is organized with initial skimming and non-ABM components, and a speed feedback loop (Figure 10-3).
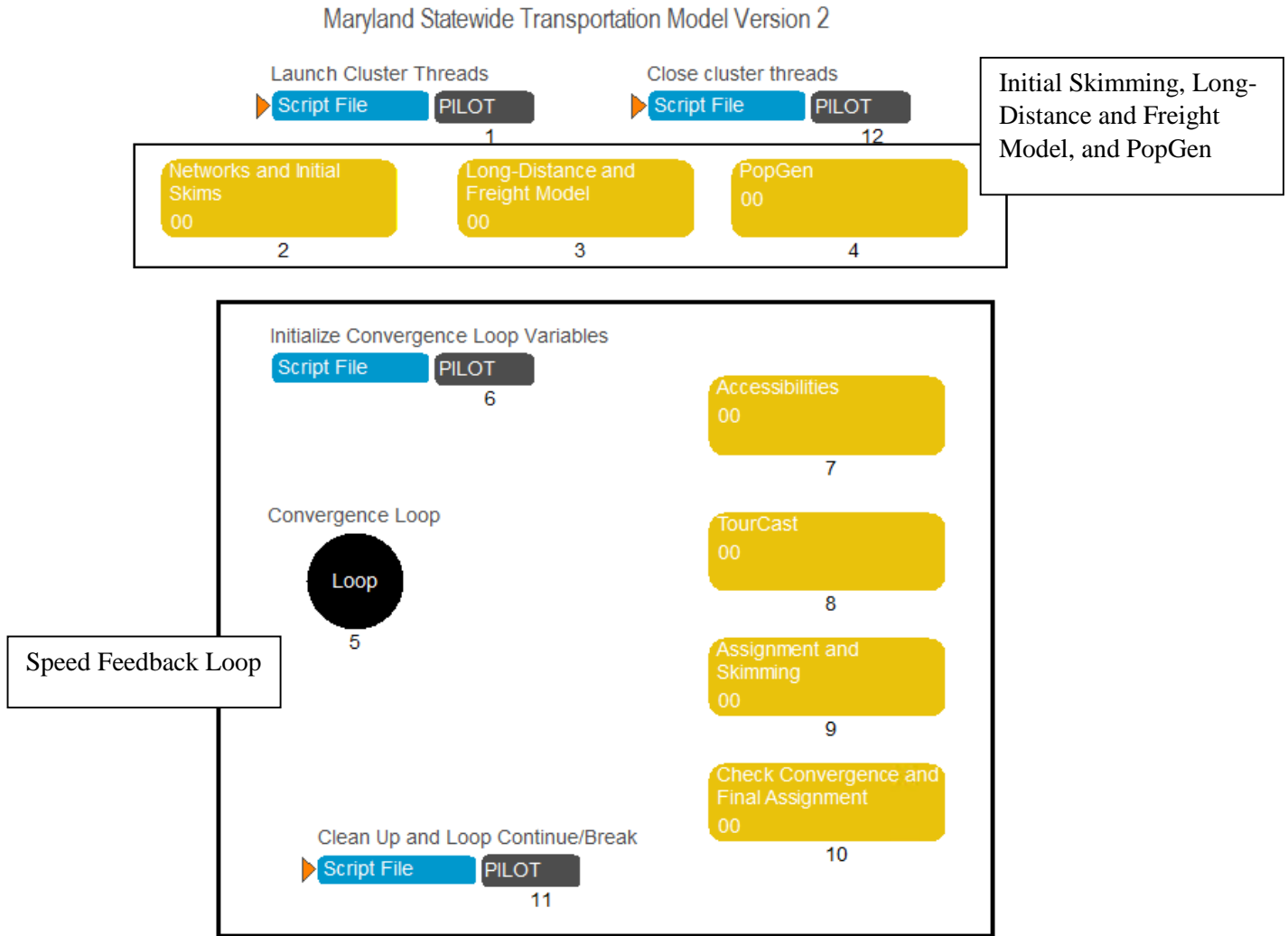
Figure 10-3. Catalog overview.

10.1.5.1. Initial Skimming, Long-Distance and Freight Model, PopGen

The initial skimming and long-distance and freight model includes model operations that do not change across speed feedback iterations, such as the population generator and external demand process. Also, the skim inputs are prepared in these application groups for the first iteration.

Networks and Initial Skims – Group 2
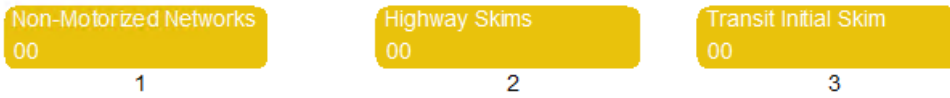
Figure 10-4. Network and initial skims.

This group takes the highway network, with speeds and capacities and the transit line files, as input. Free-flow skims are generated for the non-motorized modes at a daily level, highway for four time periods, and transit for a peak and off-peak time period (**Error! Reference source not found.**).

Non-Motorized Networks – Group 2.1
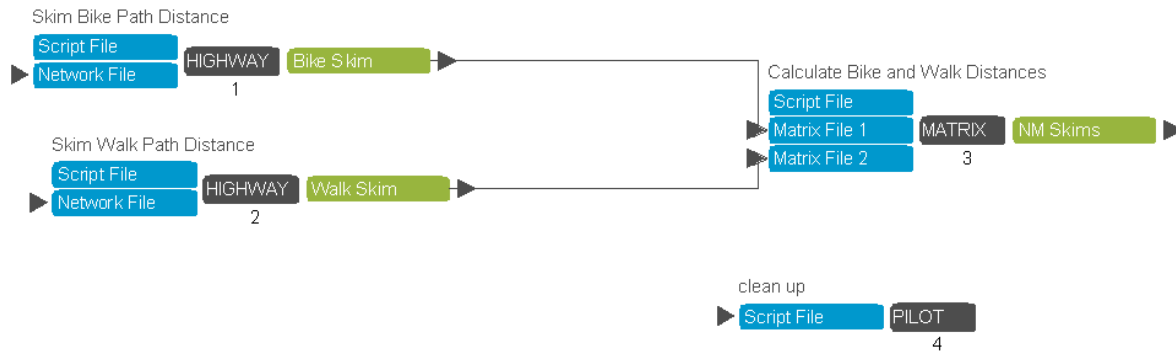
Figure 10-5. Non-motorized network.

The current model uses the highway distances for both bike and walk skims. For non-motorized mode, a separate structure and process are maintained in the model (**Error! Reference source not found.**).

Highway Free Flow Skims – Group 2.2
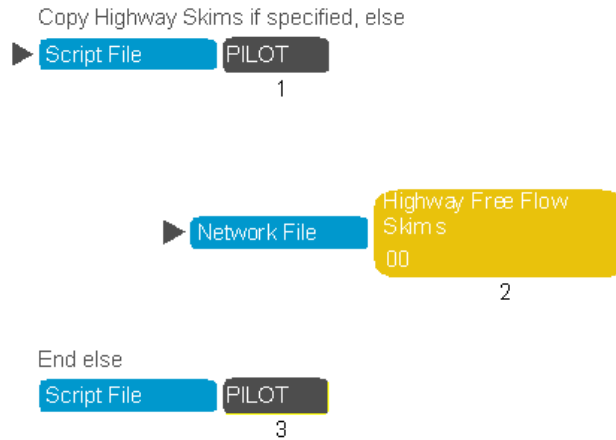
Figure 10-6. Highway free flow skims.

If generating free-flow skims is not specified in the scenario manager, highway skims will be copied from the specified location (Figure 10-6).

Transit Free Flow Skims – Group 2.3
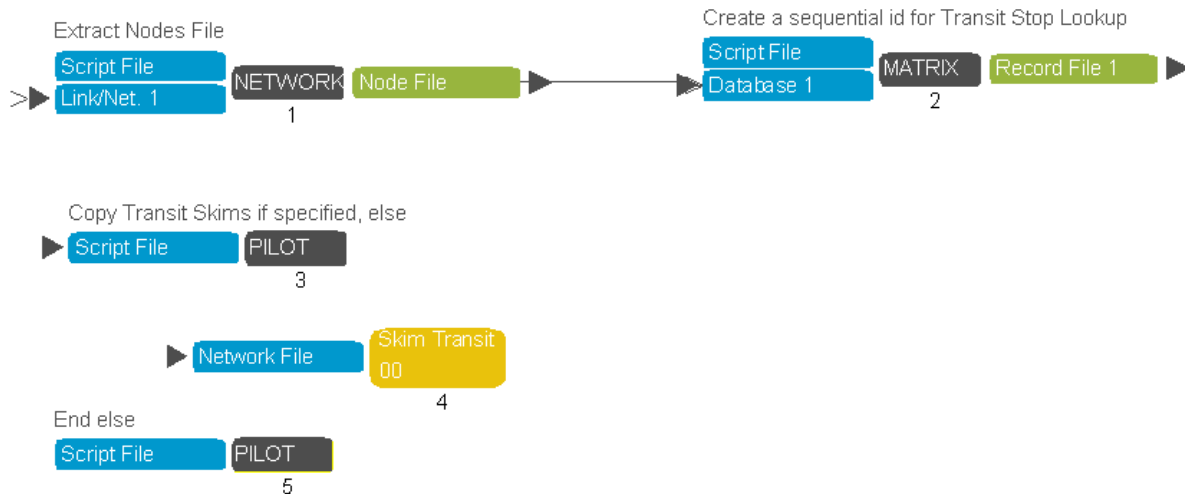
Figure 10-7. Transit free flow skims.

A transit stop lookup file is generated from the network to calculate the parcel-transit stop access and egress times. The transit skimming procedure is identical to that done after highway assignment. Please refer to Group 2.3 for more information (Figure 10-7).

Long-Distance and Freight – Group 3

169
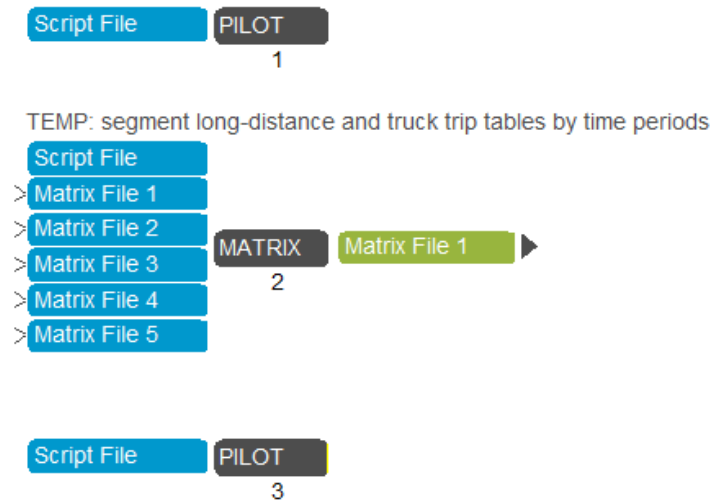
Figure 10-8. Long-distance and truck tables.

This group models the long-distance and truck trip tables that are constant across speed-feedback iterations (**Error! Reference source not found.**). The current implementation copies the outputs from the long-distance model and the truck model.

Run PopGen – Group 4
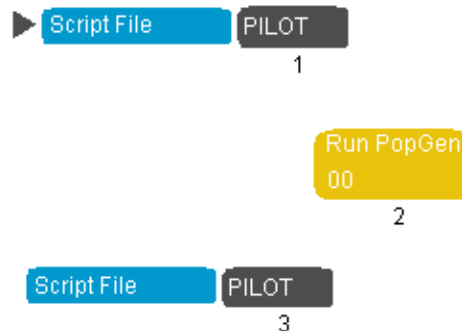
Figure 10-9. Run PopGen.

Running PopGen is controlled by a scenario key. If the scenario key is de-selected, the selected households and persons tables are copied to the scenario directory (Figure 10-9). The PopGen application group is a placeholder and will eventually be updated to run PopGen natively from Cube.

Each speed feedback loop begins with calculating the aggregate logsums. Then, the disaggregate demand models powered by TourCast are run. The trip tables from the disaggregate and aggregate models are then combined for highway and transit assignment. After the first iteration, the assigned link volume is compared to the convergence threshold. If the current iteration results are within the convergence threshold, then a final set of highway assignments and transit assignments are completed and the loop exits.

Accessibilities – Group 7
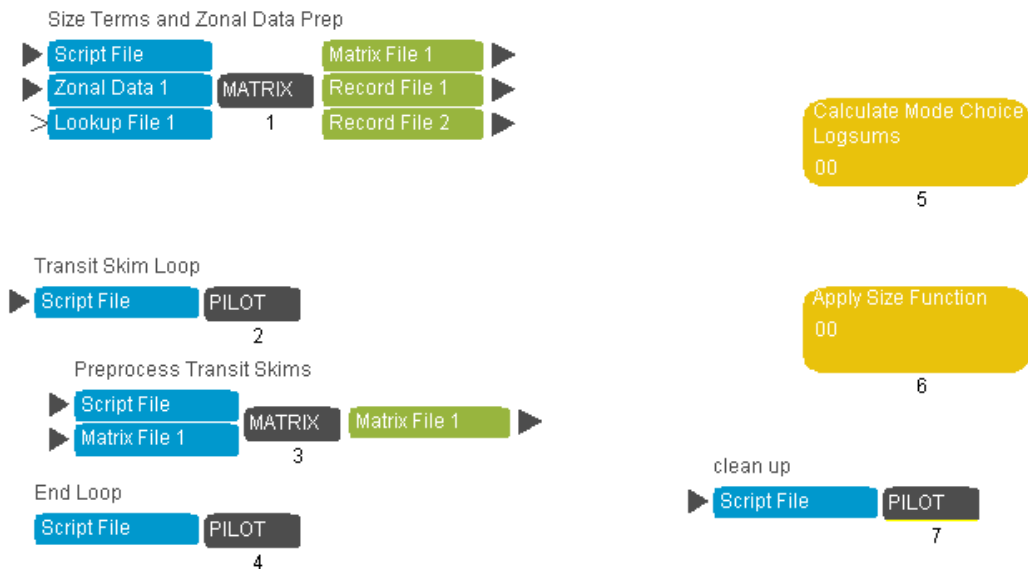
Figure 10-10. Accessibilities in speed feedback loops.

The accessibilities group calculates the aggregate logsums used in the early activity-based model components. The initial steps in the calculation prepare the size terms and process the transit skims. The mode choice logsum is calculated and size terms are applied to the logsum prior to aggregation for across all destination zones for each origin zone (**Error! Reference source not found.**).

TourCast – Group 8
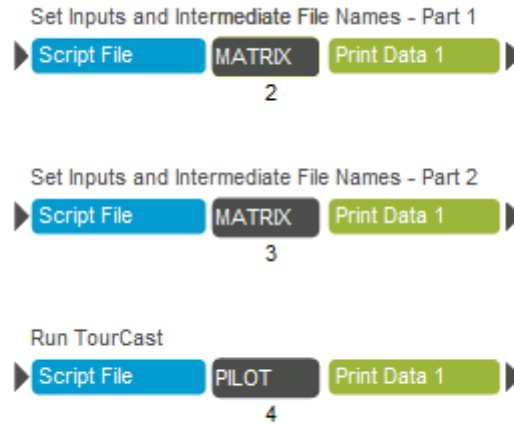
Figure 10-11. Tourcast in speed feedback loops.

TourCast runs by calling the ModelEngine.exe file in the TourCast\bin folder of the catalog and passing the component configuration file(s) as arguments. In order to call an external program from within Cube Catalog and capture error codes from the external program, a batch program needs to be dynamically created and called. The first two Matrix programs within this application group set the input and output file names for this speed feedback iteration; the pilot program assembles the TourCast calls according to the components selected through the scenario manager (Figure 10-11).

Highway Assignment – Group 9.2
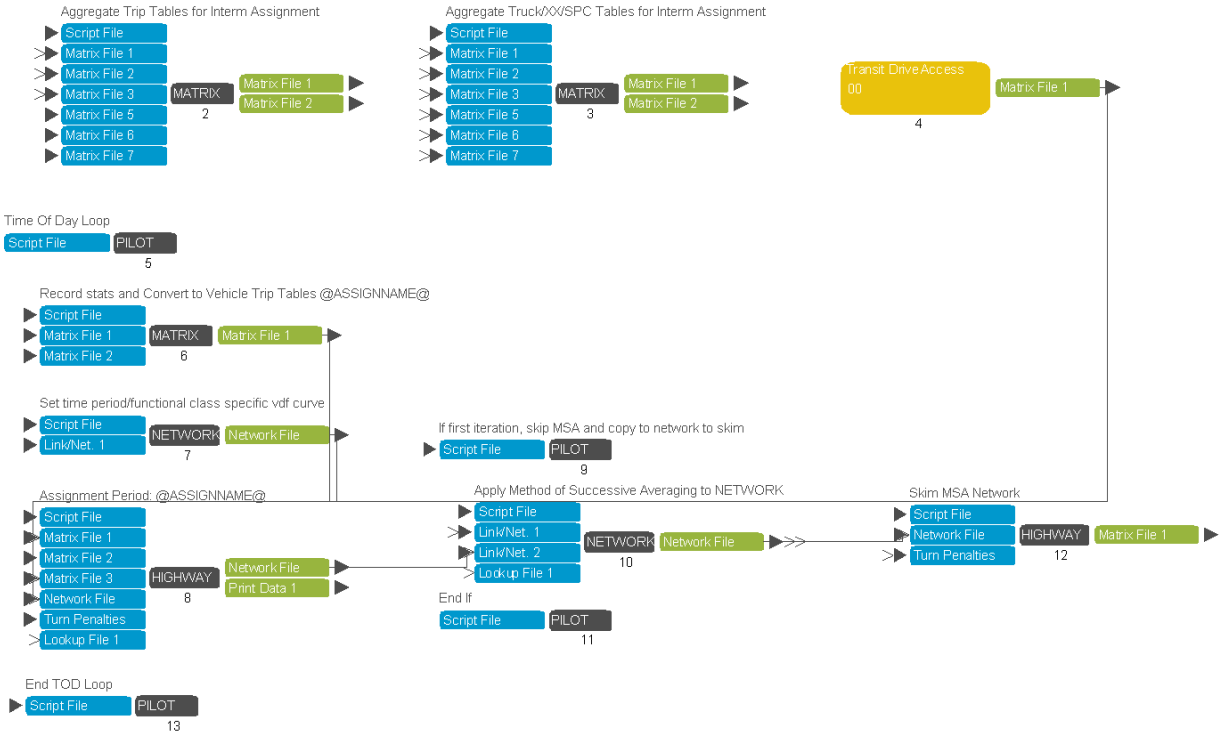
Figure 10-12. Highway assignment in speed feedback loops.

This highway assignment is the convergence loop assignment that only assigns four-time periods. Therefore, the first step in this group is to aggregate the eight time-period tables into four time periods (MATRIX 2 – MATRIX 3).

The programs in application group 4 derive the auto trip component from the transit drive-transit person trips for each time period. The group then loops through each time period, converts from person to vehicle trip tables (MATRIX 6), runs the assignment, and conducts an MSA process on the loaded network before collecting skims (HIGHWAY 8 – HIGHWAY 12) (Figure 10-12).

Transit Skims – Group 9.3
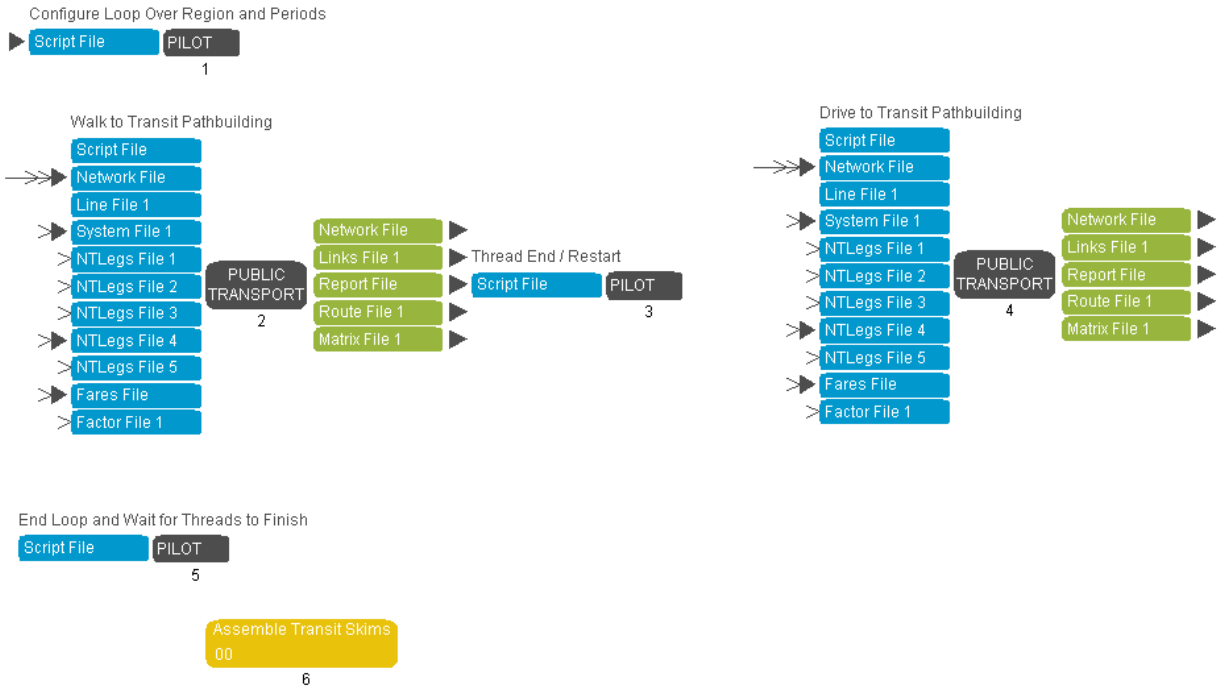
Figure 10-13. Transit skims in speed feedback loops.

The transit skim group uses the loaded highway network to update the transit speeds and generate a new set of transit skims. The skim loop is done by time period (peak/off-peak), access mode (walk/drive), and region (BMC, COG) (Figure 10-13).

Transit Skims Assembly – Group 9.3.6
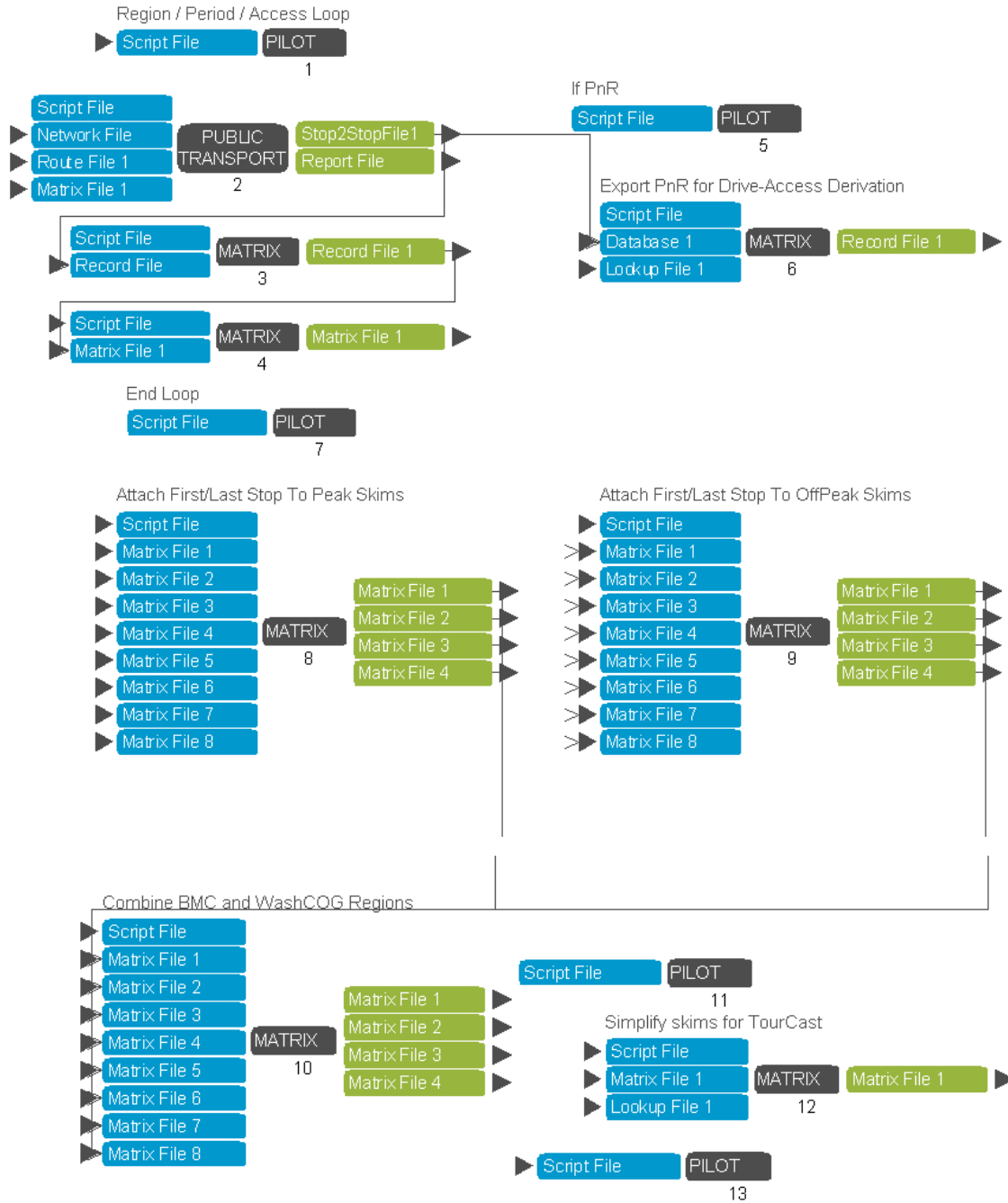
Figure 10-14. Transit skims assembly in speed feedback loops.

The transit skim assembly combines the transit skims between the two regions (MWCOG and BMC) and converts the skims into the necessary cores for TourCast (Figure 10-14).

Convergence Check – Group 10.2
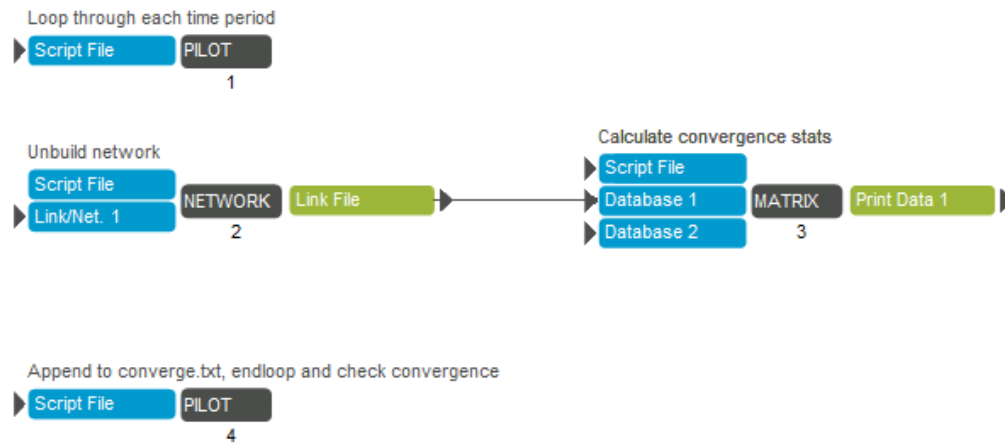
Figure 10-15. Convergence check in speed feedback loops.

After the second iteration, the highway assignments are compared to the previous iteration to see if the highway volume has changed in a manner greater than the convergence thresholds. Convergence is checked for each time period and the speed feedback loop continues if convergence is not reached for all time periods. The convergence status is appended to the **converge.txt** file in the scenario directory (Figure 10-15).

Assign Highway and Transit – Group 10.4
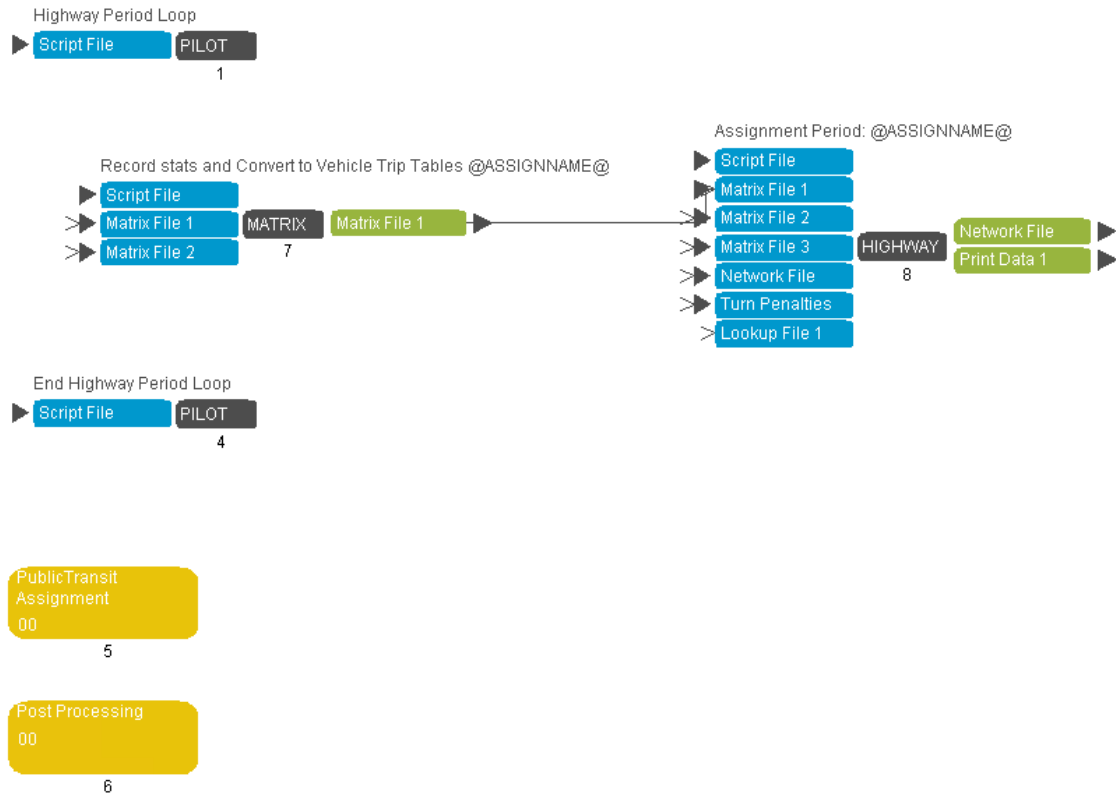
Figure 10-16. Assign highway and transit in speed feedback loops.

If convergence has been reached, then the eight time-period highway assignment and the transit assignment are run. The post-processing step at the end of the eight time-period assignment combines the results of each time period as well as the results from the four time-period assignment into a single network for analysis (Figure 10-16).

Transit Assignment – Group 10.4.5
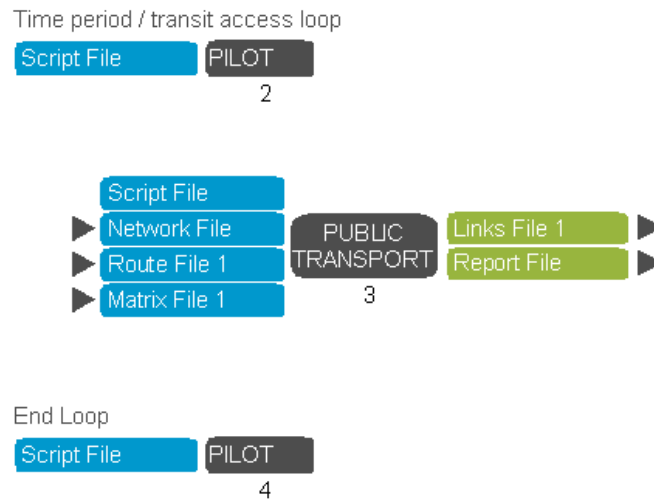
Figure 10-17. Transit assignment in speed feedback loops.

This group assigns the walk-transit and drive-transit trip tables for peak and off-peak periods (Figure 10-17).

10.2.   Preparing a Model Run

10.2.1. Input Data and Parameter Files

10.2.1.1.   TourCast component configuration files

Each disaggregate demand model includes a main configuration file and may require several other configuration files to be fully defined. The configuration files define the following for each model component:

- Component type
- Alternatives
- Variables and coefficients
- Nest structure (if the model is nested)
- Model parameters
- Segmentation

10.2.2. Cube Catalog Scenario Manager

The scenario manager has a set of user-defined keys. The purpose and default value of each key is described in this section.

The catalog scenario manager can be used to specify the scenario input files and parameters. To access the scenario manager, open the security catalog file in Cube and double click the desired scenario in the Scenario window (Figure 10-18).
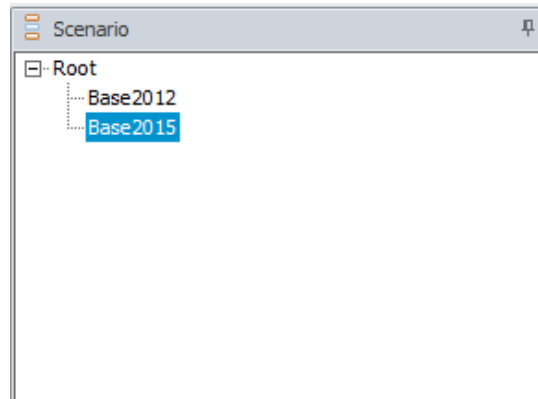
Figure 10-18. Catalog scenario manager.

The scenario manager contains two pages, organized by:

- Base 2012
- Base 2015

10.2.2.1. <u>Major Inputs and Run Parameters</u>

Figure 10-19. Major inputs and run parameters.

The Major Inputs include the highway network file, zonal data file, parcel data file, and the synthesized population files (Figure 10-19). The Catalog Run Parameters specify the catalog parameters and initial skim inputs:

- *Maximum number of Cluster Sessions:* this should be set to a value less than or equal to the number of CPU cores in the computer.
- *Maximum Full Feedback Iterations:* if convergence is not reached within this number of iterations, the model will complete and run the final assignment as if convergence was reached.
- *Previous iteration assignment weight:* this factor is applied to the previous weighted assignment results and combines them with one minus this factor applied to the current assignment results. A factor of 0 would use only the current assignment results and a factor of 1 would use only the previous results.
- *Convergence Criteria:* the catalog will run speed feedback iterations until either the convergence thresholds are reached, or the maximum full feedback iterations is met. The

convergence criteria are applied by time period and are met if fewer than 1% of links differ in volume by more than 10% (numbers from configuration shown in figure above).

- *Remove temporary files:* when selected, intermediate group files are deleted.
- Convergence can be hastened by using initial highway skim inputs from a previous model run. These skim files are specified in the four lines under the highway skim checkbox. These files will be used as the initial skim inputs unless the Generate Free-Flow Skims box is checked.
- Model computational time can be reduced by using initial transit skim inputs from a previous model run. These skim files are specified in the eight lines under the transit skim checkbox. These files will be used as the initial skim inputs unless the Reskim transit networks box is checked.
- Select to run PopGen; once PopGen is integrated into the catalog, its operation will be controlled through this checkbox.

10.2.2.2. TourCast Operational Control

TourCast Operational Control
- ☑ Run TourCast Post-PopGen component
- ☑ Vehicle Availability
- ☑ School Location
- ☑ Usual Work Location
- ☑ Pass Ownership Model
- ☑ Daily Activity Pattern
- ☑ Mandatory Tour Destination
- ☑ Mandatory Tour Time Of Day
- ☑ School Escort
- ☑ Fully Joint
- ☑ Individual Non-Mandatory
- ☑ Home-Based Tour Stop Generation
- ☑ Home-Based Tour Mode Choice
- ☑ Work-Based Subtour
- ☑ Stop-Level Models
- ☑ Trip Mode Choice

Figure 10-20. Tourcast operational control.

The second tab of the Scenario Manager interface contains a series of check-boxes that are associated with the TourCast component groups in their order of execution. By default, all TourCast components are selected. NOTE: the TourCast components are loosely coupled, but deselecting component(s) means that some input data to downstream components will not be generated and must be made available in the scenario directory manually (Figure 10-20).

10.2.2.3. Highway Parameters

Highway Parameters

HOV2 Average Occupancy	2
HOV3 Average Occupancy	3.2
Commercial Vehicle PCE	1
Medium Truck PCE	1.5
Heavy Truck PCE	2
Highway Turn Penalty File	F:\Di Yang\InSITE_VerB04_MSTM2initialsteps_Final\Input\network\Base2015\turn.pen
AM Period Capacity Factor	2.34
MD Period Capacity Factor	4.965
PM Period Capacity Factor	3.65
NT Period Capacity Factor	3.111
Maximum Highway Assignment Iterations	30

Figure 10-21. Highway parameters.

This section contains:

- Average occupancy for HOV2 and HOV3 modes, which is used to convert from TourCast person trip outputs to vehicle trips for assignment
- Truck PCE: passenger car equivalents of each truck by truck category
- Turn penalty file
- Maximum iterations to run for the highway static assignment

This section specifies the public transit inputs. The transit input files are stored in the Inputs\Transit folder. The scenario keys specify (Figure 10-21):

- Transit line file that describes the path, operating time, and headway of each route
- Access/egress links
- Transit fare and system files that describe parameters for each transit mode
- Factor files by time-of-day and region

10.3.　　　Running the Entire Model

Once all of the input files are created and mapped using the scenario manager, running the entire model is as simple as clicking "Run" from the scenario manager.

10.3.1. Task Monitor Window

While the catalog is running, the progress can be monitored through the task monitor window, which will display the current application group being executed. Even when Cube Cluster is in operation, the task monitor window will still be present (Figure 10-22).
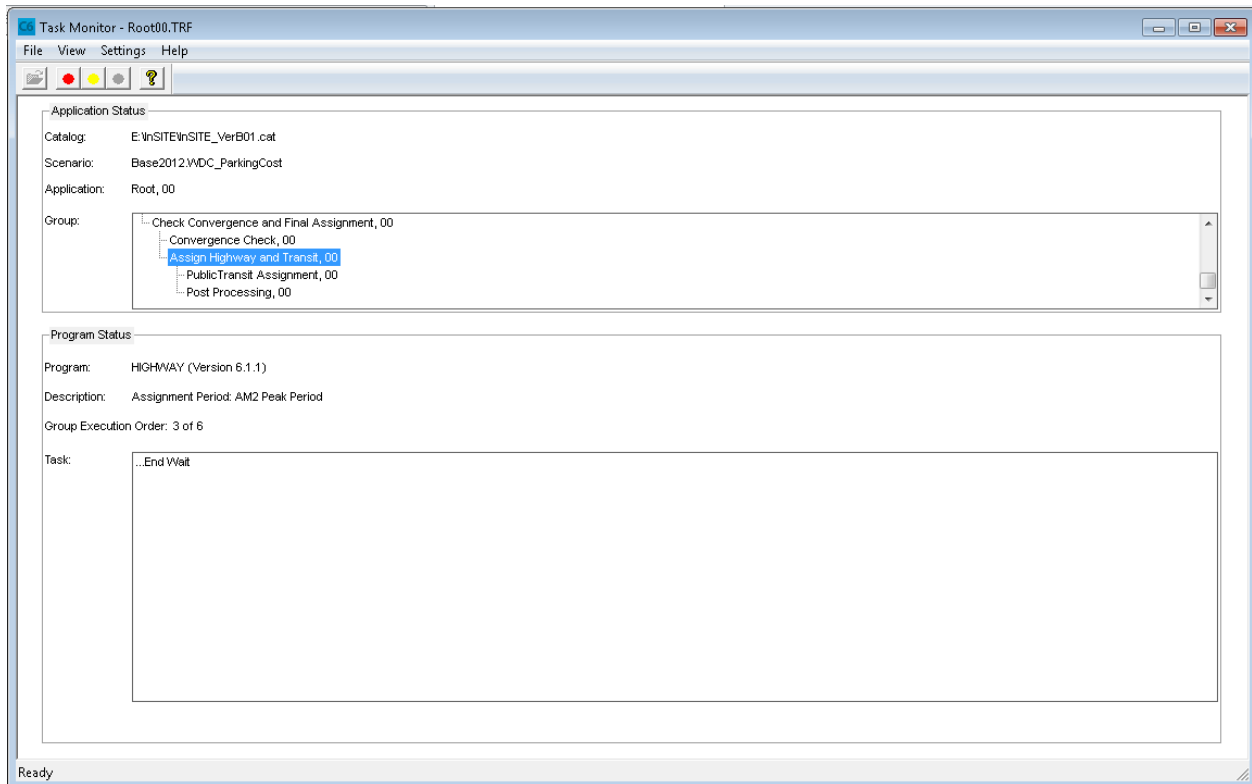

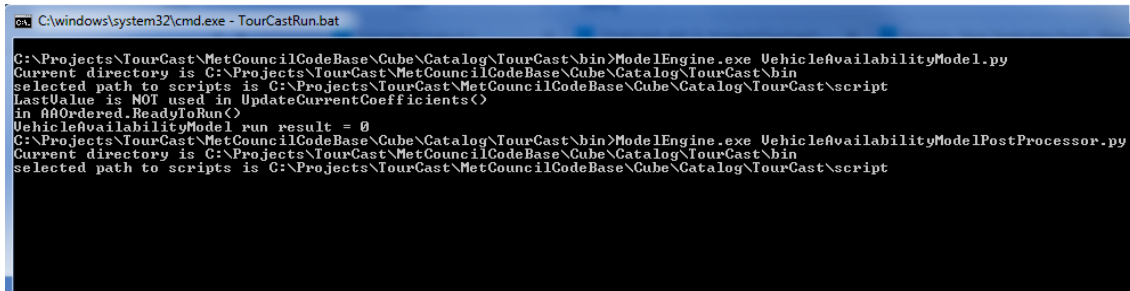**Figure 10-22: Task monitor window.**

10.3.2. TourCast Operating Window

TourCast is executed through a dynamically built batch command **TourCastRun.bat** in the Cube folder. When TourCast is called, the task monitor will display (Figure 10-23):

Executing: "start /wait TourCastRun.bat"

and two command prompt windows will open. One window is present to carry error codes back to Cube and will not display any output. The other window shows the status of each component execution. Once the TourCast components are complete, control will return to Cube to run the next

set of components. The call to each component is indicated by a new ModelEngine.exe call. Each component will end with a run result. A run result of "0" indicates success.

Figure 10-23. Tourcast operating window.

If there is an error, this window will be paused. It is good practice to copy the full text of this window if there is an error to assist in debugging. Once the text is copied, press any key and close the windows. This will return control to Cube and the catalog run will end with an error.

10.3.3. Model outputs

A reliable way to monitor the model status is to examine the output files. Each file has the speed feedback iteration number included.

Catalog outputs files are defined in Appendix A.

TourCast output files are defined in Appendix A.

10.3.4. Running a single application group

A single application group can be run in the catalog as long as all variables are defined. Within the speed feedback loop, at least the iteration variable ITER needs to be defined. This can be done by creating a pilot program in the application group with the following text:

ITER = 1

For example, to run iteration 1.

10.3.4.1.　Running TourCast Only

TourCast can be run individually by running within the TourCast application group. Note that the scenario manager configuration settings are active if TourCast is run using the Run Application control. Alternatively, TourCast can be run by calling ModelEngine.exe directly with the desired component configuration file. Please see the TourCastRun.bat file in the Cube folder for examples.

10.4. Preparing a New Scenario

To build a new scenario, the following inputs potentially need to be updated:

- Highway Network
- Transit Line File
- Zonal data (sociodemographics)
- External station volumes
- Synthesized population

11. CONCLUSION

REFERENCES

(Caltrans) California Department of Transportation State-wide Travel Demand Model [Report]. - [s.l.] : http://www.dot.ca.gov/hq/tsip/ smb/swm_documentation.html, 2011.

Beckman RJ, Baggerly KA and McKay MD Creating Synthetic Baseline Population [Journal]. - [s.l.] : Transportation Research Part A: Policy and Practice, 1996.

Chase KM, Anater P and Phelan TJ Freight Demand Modelin and Data Improvement [Book]. - [s.l.] : Transportation Research Board, 2013.

ConnDOT Modeling Procedure [Report]. - [s.l.] : Connecticut Department of Transportation, 1997.

Erhardt G. Evaluating Travel Markets in a Statewide Model: An Arizona Case Study [Conference]. - [s.l.] : Presented at 91st Transportation Research Board Annual Meeting, Washington DC., 2012.

FDOT The Integrated Florida Transportation Model [Report]. - [s.l.] : Florida Department of Transportation, 2008.

Konduri K and Pendyala R Updates to the BMC Population Synthesis Model: Incorporating Controls at Multiple Geographic Resolutions, Final Report [Report]. - 2015.

KTC Kentucky Statewide Traffic Model Final Calibration Report [Report]. - [s.l.] : Report for Kentucky Transportation Cabinet (KTC) , prepared by Wilbur Smith Associates., 1997.

MDOT State Long-Range Transportation Plan 2005-2030: Land Use [Report]. - [s.l.] : Report for Michigan Department of Transportation, prepared by Wilbur Smith Associates, Inc., 2006.

mishra S [et al.] Development of MAryland Statewide Transportation Model and Its Application in Scenario Planning [Conference]. - [s.l.] : Transportation Research Board 92nd Annual Meeting, 2013.

Model Design Plan for BMC Activity-Based Model [Report]. - [s.l.] : Cambridge Systematics, 2013.

NCHRP Travel Estimation Techniques for Urban Planning [Report]. - [s.l.] : Barton-Aschman Associates for Transportation Research Board, National Research Council, 1998.

ODOT Ohio Statewide Model [Report]. - [s.l.] : Report for Ohio Department of Transportation prepared by Parsons Brinkerhoff, 2010.

ODOT Travel Demand Model Development and Application Guidelines [Report]. - [s.l.] : Report for Oregon Department of Transportation (ODOT), Prepared by Parsons Brinckerhoff Quade & Douglas, Inc., and Kittelson & Associates, Inc., 2010.

Ryan J, Maoh H and Kanaroglou P Comparing the Major Techniques Using a Small Complete Population Of Firms [Journal]. - [s.l.] : Geographical Analysis, 2009.

THE MARYLAND INTEGRATED TRAVEL ANALYSIS MODELING SYSTEM [Online]. - http://www.mitams.org/.

VAT Vermont Statewide Travel Demand Model Development. [Report]. - [s.l.] : Report prepared by Vermont Agency of Transportation, 2010.

Zhang L and Lu Y U.S. National and Inter-Regional Travel Demand Analysis: Person-Level Microsimulation Model and Application to High-Speed Rail Demand Forecasting: Final Report [Report]. - [s.l.] : National Transportation Center at Maryland (NTC@Maryland), 2014.