THE MARYLAND STATEWIDE TRANSPORTATION MODEL

This report documents the data sources, development, validation and execution of the Maryland Statewide Transportation Model (MSTM). This planning tool was developed to provide analytical support in SHAs decision-making and to help implement transportation policies, programs and initiatives throughout the State of Maryland.

Model
cumentation

Title: The Maryland Statewide Transportation Model. Model Documentation (version 1.0)	**Date:**	October, 2013
	No. of Pages:	187
	Publication No.:	

Description:	This report documents the data sources, development, validation and execution of the Maryland Statewide Transportation Model (MSTM). This planning tool was developed to provide analytical support in SHAs decision-making and to help implement transportation policies, programs and initiatives throughout the State of Maryland.
Project Administration:	Morteza Tadayon Data Services Engineering Division Office of Planning and Preliminary Engineering Maryland State Highway Administration Lisa Shemer Data Services Engineering Division Office of Planning and Preliminary Engineering Maryland State Highway Administration
Project Management:	Subrat Mahapatra Data Services Engineering Division Office of Planning and Preliminary Engineering Maryland State Highway Administration
Project Staff:	Rick Donnelly, Parsons Brinkerhoff Leta Huntsinger, Parsons Brinkerhoff Amar Sarvepalli, Parsons Brinkerhoff Fred Ducca, NCSG/University of Maryland Rolf Moeckel, NCSG/University of Maryland Sabyasachee Mishra, University of Memphis Tim Welch, Georgia Institute of Technology Mark Radovic, Gannett Fleming, Inc.

Table of Contents

List of Figures

List of Tables

1 Model Overview

The Maryland State Highway Administration (SHA) has developed a statewide transportation model that (1) will allow consistent and defensible estimates of how different patterns of future development change key measures of transportation performance, and (2) can contribute to discussion and other evaluation tools that address how future transportation improvements may affect development patterns.

The Maryland Statewide Travel Model (MSTM) is by design a multi-layer model working at a Regional, Statewide and Urban level (Figure 1-1). The Regional Model covers North America, the Statewide Model includes Maryland, Washington DC, Delaware and selected areas in Pennsylvania, Virginia and West Virginia, and the Urban Model which serves to link for comparison purposes only, the urban travel models where they exist within the statewide model study area, for instance by connecting MSTM with the Baltimore Metropolitan Council (BMC) Model or the Metro Washington Council of Governments (MWCOG) Model.

This documentation is a User's Guide focusing on the implementation of the Regional and the Statewide Model components. Past and future efforts strive to compare MSTM model results to MPO models and data at the Urban level. Every level is simulated to study travel behavior at an appropriate level of detail. The interaction of the three levels potentially improves every level by providing simulation results between upper and lower levels. All MSTM assignment of the travel demand occurs at the Statewide level.

At the Statewide Level, there are The 1588 Statewide Model level Zones (SMZs) that cover Maryland, Delaware, Washington DC, and parts of New Jersey, Pennsylvania, Virginia and West Virginia (Figure 1-2). The 151 Regional Model Zones (RMZs) cover the full US, Canada, and Mexico. RMZs are used for the multi-state commodity flow model and the long distance passenger model only and are eventually translated into flows assigned to networks and zones at the Maryland-focused (SMZ) level.

summarizes the MSTM model components within the Statewide and Regional levels. Economic and Land Use assumptions drive the model. On the person travel side, the Regional model includes a person long-distance travel model for all resident and visitor trips over 50 miles, reflecting only travel between their local trip end and their point of entry/exit (highway, airport, train station or bus terminal). These trips are combined with Statewide level short-distance person trips by study area residents, produced using a trip generation, trip distribution, and mode choice components. On the freight side, the Regional model includes a long-distance commodity-flow based freight model of truck trips into/out of and through the study area (EI/IE/EE trips). These flows are originally estimated for the entire US and disaggregated to the study area zonal system. These trips are combined with short distance truck trips (II trips) generated at the Statewide level using a trip generation and trip distribution method. The passenger and truck trips from both the Regional (long-distance) and Statewide (short-distance) model components provide traffic flows allocated to a time period (AM peak, PM peak or off-peak) are input to a single Multiclass Assignment [1], [2], [3], [4].

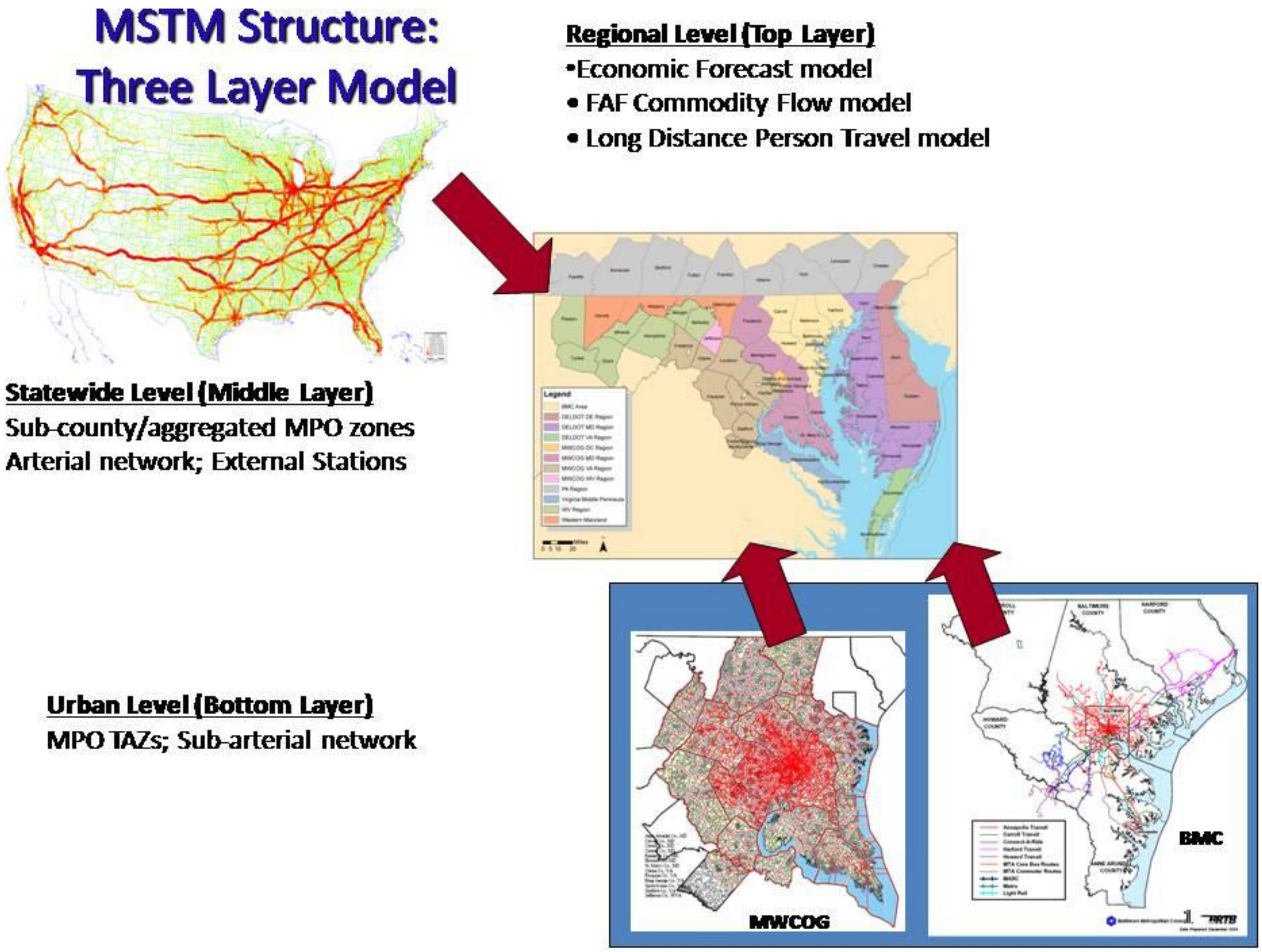


Figure 1-1: MSTM three level model

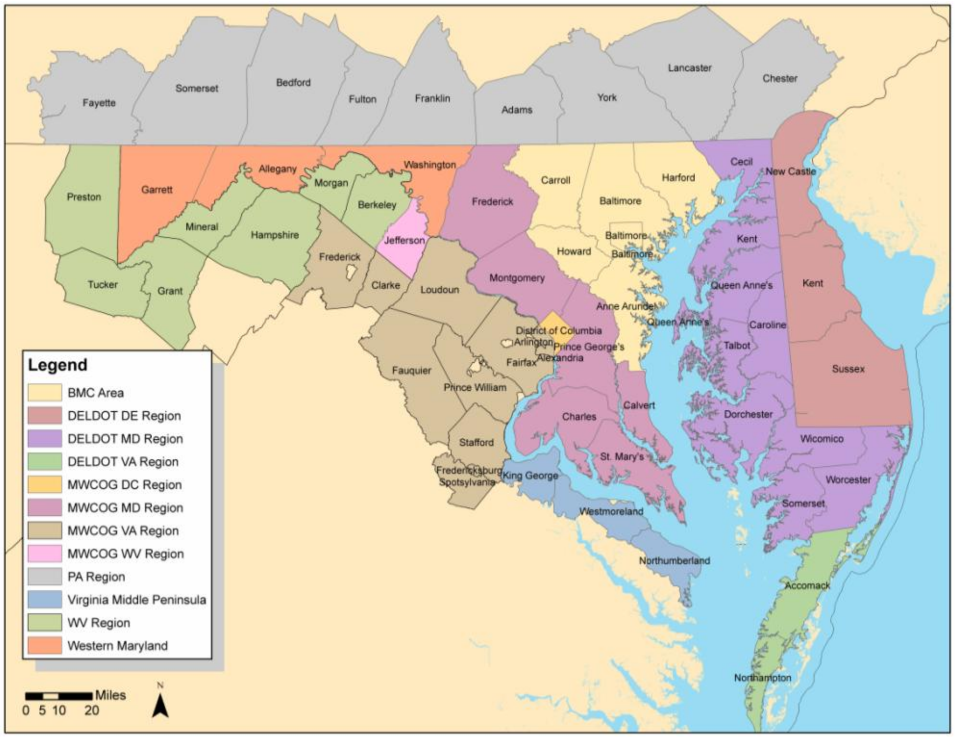


Figure 1-2: MSTM statewide level map

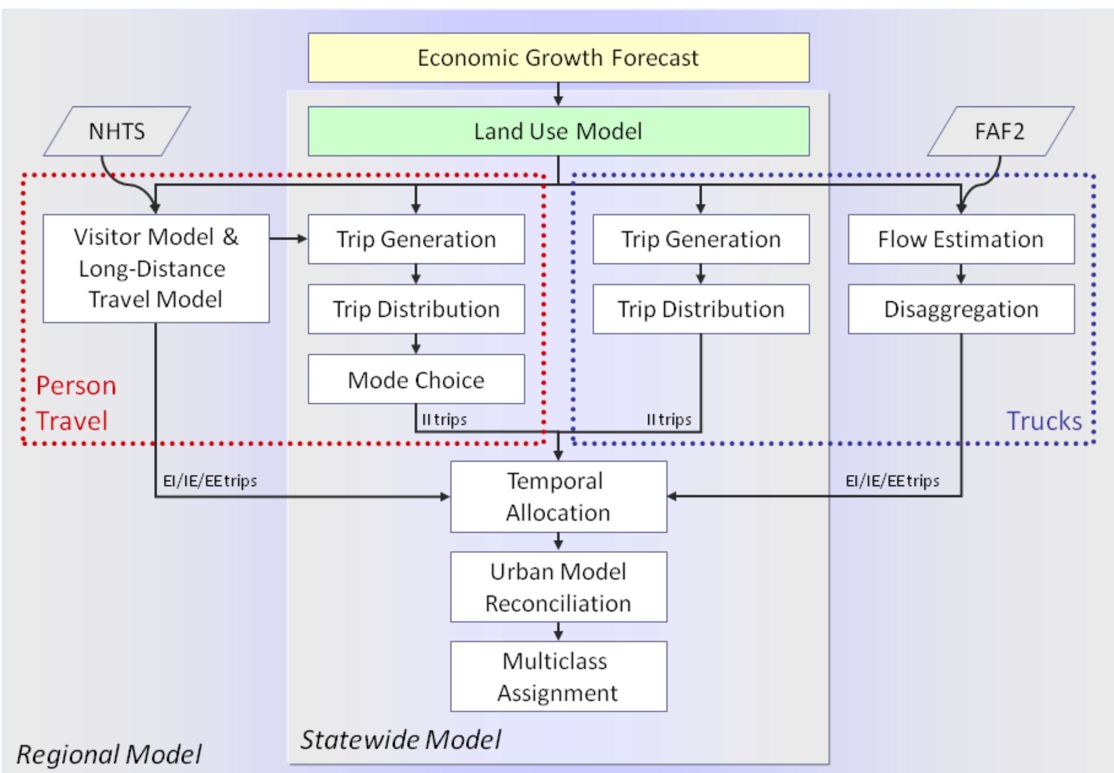


Figure 1-3: Overview of the MSTM model components

2 Model Inputs

2.1 Zone System

Regional Level: 151 Regional Model Zones (RMZs) in the MSTM Regional model cover the entire US, Canada, and Mexico. These zones are used for the Regional long distance models only. Flows from these model zones are eventually translated into flows assigned to networks and zones at the Statewide Model Zone (SMZ) level, discussed below.

Statewide Level: 1,588 Statewide Model Zones (SMZs) in the MSTM Statewide level cover all of Maryland and selected counties in adjacent states. SMZs are the basis for MSTM transportation assignment and input land use assumptions. They nest within counties and are aggregations of MPO TAZs where they exist.

Urban Level: 3,056 Urban Model Zones (UMZs) in the MSTM urban level are taken directly from the Traffic Analysis Zones (TAZs) in the Baltimore Metropolitan Council (BMC) and Metro Washington Council of Governments (MWCOG) MPO models.[1]

The numbering of the MSTM zones reflects this three-level hierarchy. At the Urban Level, TAZ numbers are retained directly preceded by a 1 for BMC and a 2 for MWCOG. At the Statewide and Regional levels, two zone numbering systems are used. The "SMZ_GeoRef" system includes FIPS codes that enable the zone to be located by state and county, while "SMZ_CUBE" is a sequential numbering system for use in CUBE traffic assignment (some blank zones between major geographic coverages were left in for future flexibility). The "RMZ_GeoRef" also uses state and county FIPS codes, but is preceded by a coverage area code (1-6), as shown in Figure 2-1. The numbering system is summarized below, with actual numbers by region noted in Table 2-1.

[1]The reviewed BMC zone system has as a total of 1,421 TAZs numbered 1-2,928 (98 RPDs). The reviewed MWCOG zone system has 1,972 TAZs numbered 1-2,141 (333 TADs).Where the models overlapped, BMC TAZs were used in Anne Arundel County, Baltimore County, and Carroll County, and MWCOG TAZs were used in Frederick County, Montgomery County, Prince George's County, and District of Columbia.

Table 2-1: MSTM zone numbering

Model Area	Coverage	Count	CUBE Zone Number Start	CUBE Zone Number End
Maryland				
MD-BMC	6 counties/cities	599	1	599
MD-MWCOG	6 counties/cities	401	609	1009
MD West	3 counties	65	1019	1083
MD Eastern Shore	9 counties	86	1093	1178
District of Columbia				
District of Columbia	All	84	1188	1271
Virginia				
VA-MWCOG	15 counties/cities	148	1281	1428
VA-Frederick County	2 county/city	5	1438	1442
VA-Mid Pen	2 counties	7	1443	1449
VA-Eastern Shore	2 counties	11	1450	1460
West Virginia				
WV-MWCOG	1 county	4	1470	1473
WV	7 counties	26	1474	1499
Delaware				
DelDOT	3 counties	97	1509	1605
Pennsylvania				
PennDOT	5 counties	31	1615	1645
PennDOT	4 counties	24	1651	1674
SMZ Total		1588	1	1674
RMZ (Regional Model Zones)				
NJ	3 counties	19	1850	1873
NJ, PA, VA, WV	Counties and aggregation of counties	85	1701	1785
Rest of USA	States	44	1786	1829
Canada	Aggregation of Provinces	2	1830	1831
Mexico	Nation	1	1832	1832
RMZ Total		151	1701	1873

[1] In Virginia, the independent cities of Fairfax City, Falls Church, Manassas, Manassas Park, and Winchester were assigned to surrounding/adjacent counties

2.1.1 Statewide Model Zones (SMZs)

The MSTM SMZs were developed through an iterative process. The outer study area was identified from analysis of 2000 Census Transportation Package (CTPP) data to encompass the bulk of labor flows in/out of Maryland. Within this larger boundary, six regions were identified for SMZ formation, treating each region as a separate entity with its own datasets and issues. These regions are shown in Figure 2-1.

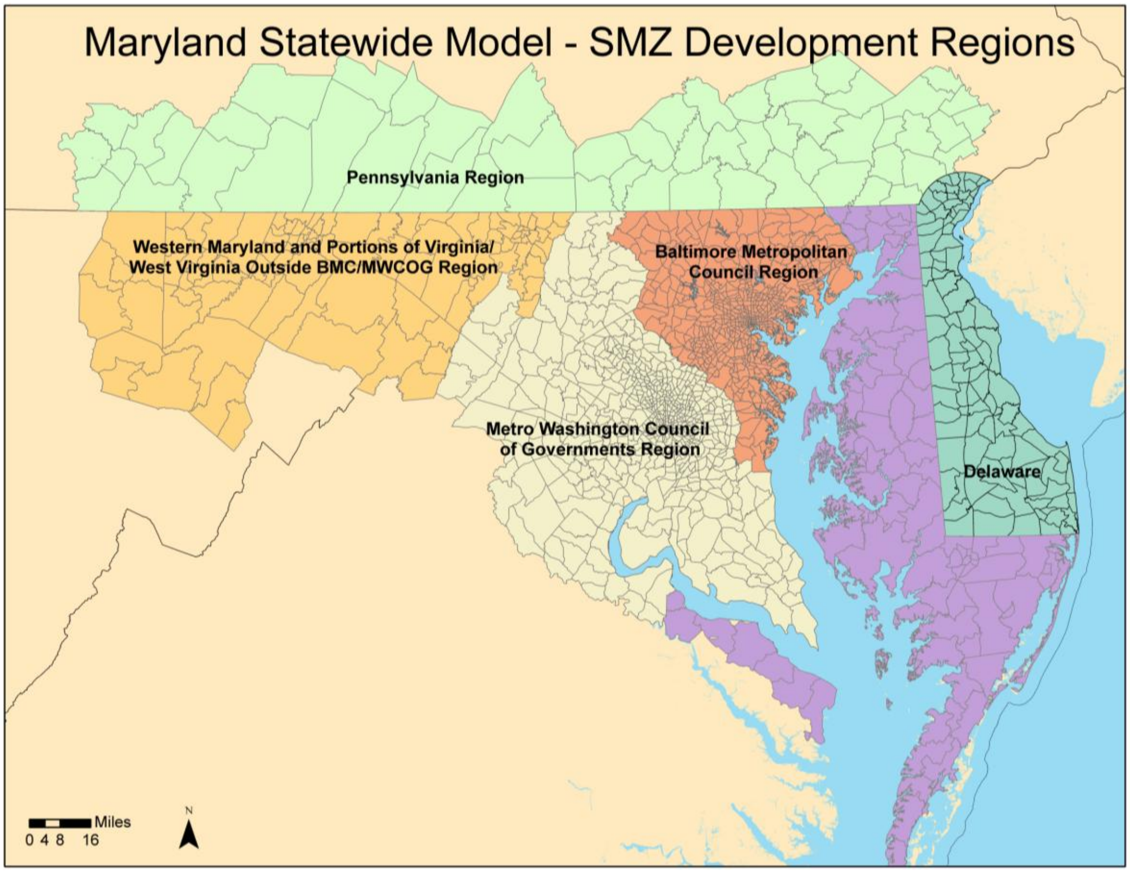


Figure 2-1: Regions used to develop SMZs

The remainder of this section discusses the process and assumptions made in developing SMZs for each of these sub-regions and overall. The goal was to adhere to the following major factors in the development of the SMZs.

- To the extent possible, SMZs conform to census geography to best utilize census data products in model development/updates and model calibration/validation. However, MWCOG MPO TAZs[2]are retained, and do not follow census geography.
- SMZs must nest within Counties and conform to County boundaries.
- Aggregations of MPO zones, to facilitate linkages between MPO and statewide models.
    - Within Washington and Baltimore MPO areas, SMZs should be equal to or aggregations of MPO TAZs and nest within the MPO's TADs/RPDs.
    - SMZs should be more uniform in size than TAZs. In general, SMZ should be greater than 0.25 and less than 10 square miles. There should be greater aggregation in central areas where MPO TAZs are smaller (often individual street blocks) and little to no aggregation of larger MPO TAZs.
- SMZs should not straddle freeways, major rivers or other natural barriers.
- SMZs should separate the traffic sheds of major roads. MPO TAZs on opposite sides of a major road can be combined to define a traffic shed or corridor.

[2] Metropolitan Washington Council of Governments (MWCOGs) version 2.2 Travel Demand Model

- SMZs should separate activity centers from surrounding areas and, where the activity center has been subdivided into multiple MPO TAZs, group adjacent TAZs into a single SMZ.

In each region, SMZs were developed with reference to various GIS overlays.

- MPO or other TAZ GIS shape file (where available) with activity density (ActDen) symbology (where TAZ data available) and Labels = TAZ number.
- Activity Density maps, calculated from historic/forecast demographic and acreage in areas of Maryland where TAZ demographic data is not available;
- Where TAZ shape files and related data are not available, use statewide land use or zoning coverage instead of Activity Density.
- Major roads coverage, from MPO networks where available, with Freeways and Major Arterials highlighted.
- MPO analysis districts (i.e., TAD or RPD) boundaries, where relevant.
- County boundaries.

The process for developing the zones consisted of a first cut based on the criteria above followed by review by SHA and other team members. Comments were addressed and conflicting comments resolved. During a final review the following additional changes were made:

- Isolate protected or restricted development lands for the land use model.
- Baltimore and District central business district aggregation to provide somewhat more uniform SMZ size and accentuate downtown activity levels on par with suburban centers.
- Distinctions were made to delineate areas with good accessibility to Metrorail stations.
- To the extent possible, the SMZ boundaries outside the MPOs and Eastern Maryland were made to distinguish rural from urban/suburban development zoning boundaries, with zones centered upon activity/town centers and major crossroads.

2.1.2 Regional Model Zones (RMZs)

The MSTM Regional model, primarily used in multi-state freight modeling, has its own zone system of RMZs. In Maryland and adjacent areas where MSTM RMZs and SMZs overlap, SMZs nest within RMZs, i.e., RMZs are aggregations of smaller SMZs. The following approach was followed.

- In Maryland, District of Columbia, and Delaware, counties were used to form RMZs.
- In four adjacent states, counties were used near the Maryland border with aggregations of counties in outer areas. Aggregation were based on the following sources:
    - Pennsylvania commodity flow districts per Pennsylvania DOT Statewide Freight Model User’s Guide v2.1 (August 2006).
    - West Virginia Department of Motor Vehicles (DMV) Districts.
    - Virginia DOT Construction districts, with some adjustments.
- In the remainder of the US, states were used, including Alaska and Hawaii.
- In the remainder of North America, three zones were as follows:

- o Canada East: Ontario, Quebec, New Brunswick, Nova Scotia, Prince Edward Island, Newfoundland and Labrador.
- o Canada West: Manitoba, Saskatchewan, Alberta, British Columbia, Yukon, Northwest Territories, and Nunavut.
- o Mexico.

The resulting RMZs are shown in Figure 2-2.

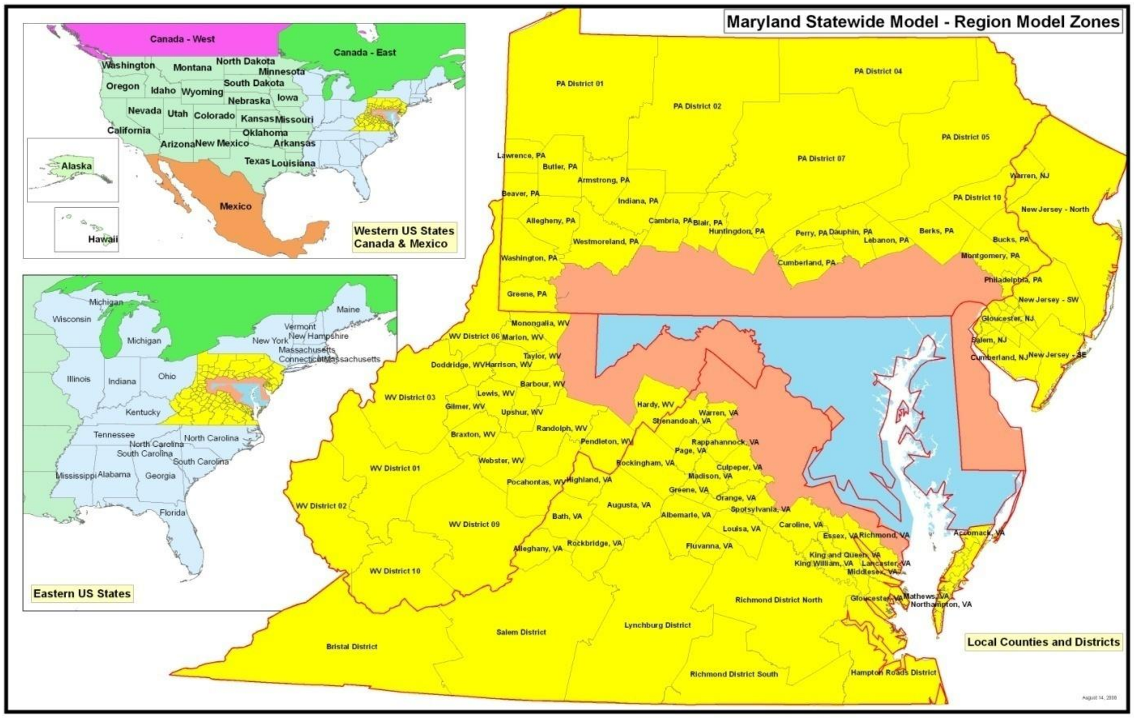


Figure 2-2: *Map of RMZ zones*

2.2 Socioeconomic Data Development

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 to the extent practical given the availability of source data. SMZ data was developed initially for 2000 and then used to develop 2007 (for validation) and 2030 (future year) model inputs.

For 2000 SMZ socio-economic data, household data were drawn from Census 2000 which provides consistent data throughout the model area. Consistent employment data was produced for the entire model area at a county level[3], but more spatially detailed employment, developed later, had to drawn from a variety of sources including MPO TAZ data, Quarterly Census Employment and Wages (QCEW) data for Maryland and TAZ data from statewide modeling efforts in adjacent states. Following is an outline of the primary data used from Census and QCEW sources.

Census 2000 Based Data. The following is Census 2000 data used at SMZ level for the MSTM Statewide model. Portions of this data are used in the Trip Generation model (Section 5.1), to provide a pattern that can disaggregate data to the detail required in that module.

1. Population (SF1)
    a. Population by age group (0-4, 5-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+)
    b. Population in households
    c. Population in Group Quarters
        i. Institutionalized by type
        ii. Non-Institutionalized by type
2. Housing Units (SF1)
    a. Occupied
    b. Vacant
3. Households by income quintile in 1999 dollars) (SF3)
    a. Lower quintile (<$20,000)
    b. Lower-middle quintile ($20,000 to $39,999)
    c. Middle quintile ($40,000 to $59,999)
    d. Upper-middle quintile ($60,000 to $99,999)
    e. Upper quintile ($100,000 or more)
4. Households by number of persons in household (SF3) (1, 2, 3, 4, 5 or more)
5. Households by number of workers in household (CTPP) (0, 1, 2, 3 or more)
6. Average household income (SF3)
7. Median household income (SF3) (optional)
8. Total Workers (CTPP)

2000 Census Transportation Planning Package (CTPP) data was also utilized.

Employment Security Based Employment.[4] The MSTM employment categories for tabulating the QCEW dataset are indicated in Table 2-2. Two levels of detail are specified. The more detailed categories are subject to extensive masking at SMZ level due to confidentiality requirements. In addition to SMZ level summaries, independent summaries by county (based on county codes in the QCEW records that do not depend on geocoding) for each set of categories provided a check on SMZ tabulations and a basis for developing county level expansion factors. The county summaries minimize masking of data and provide a direct comparison to the more detailed county employment estimates.

QCEW data for the year 2000 is not available. The closest QCEW data is for 2003, therefore it was necessary to devise procedures for developing SMZ level employment estimates using a combination of 2003 QCEW data, 2000 MPO TAZ employment data, 2000 county employment and other data and GIS coverages as appropriate. Parsons Brinckerhoff and National Center for Smart Growth (NCSG) staff collaborated on developing the necessary procedures.

[4]A federal-state program summarizing employment, wage and contribution data from employers subject to state unemployment laws, as well as workers covered by unemployment compensation for federal employees (UCFE). The QCEW program is also called Covered Employment and Payrolls (CEP) program and involves the Bureau of Labor Statistics (BLS) of the U.S. Department of Labor and the State Employment Security Agencies (SESAs).

Table 2-2: Aggregate categories for QCEW Data

NAICS CODE	DESCRIPTION	MSTM designation	Intermediate Categories	SMZ Categories
111,112	Farm	01_Farm	OthBasic	Other
113-115,21	Mining, forestry, fish. & ag. supt.	02_OtherAg&Mining	OthBasic	Other
23	Construction	03_Construction	Other	Other
31,32,33	Manufacturing	04_Manufacturing	Industrial	Industrial
42	Wholesale trade	05_Wholesale	Trade	Retail
44	Retail trade	06_Retail	Trade	Retail
484,493	Trucking & warehousing	07_Trucking&Wrhsg	Industrial	Industrial
22,48x,49x	Utilities & other transportation	08_UtilitiesOtherTransp	Industrial	Industrial
51	Information	09_Information	Office	Office
52,531,533	FIRE excluding rental	10_FIRE(excl rental)	Office	Office
54,55	Prof & tech serv plus mgmt off.	11_ProfTechServ&Mgmt	Office	Office
56	Administration & waste services	12_Admin&WasteMgmt	Office	Office
61	Educational services	13_Educational services	Office	Office
62	Health & social services	14_Health&SocSrvcS	Other	Other
71	Arts, entertainment & recreation	15_ArtsEntertmnt&Rec	Other	Other
721	Accommodations	16_Accommodations	Other	Other
722	Food services	17_FoodServices	Other	Other
81,532	Other services incl rental	18_OtherServices	Other	Other
92 (pt)	Federal government incl military	19_FederalGovernment	FedGovMil	Office
92 (pt)	State government	20_StateGovernment	StaLocGov	Office
92 (pt)	Local government	21_Local Government	StaLocGov	Office

In addition to preparation of data received from other states and from the BMC and MWCOG it is necessary to develop employment data for the areas of Maryland not covered by BMC or MWCOG. To do this the QCEW data was used. The QCEW dataset was created by the Maryland Department of Labor, Licensing and Regulation (DLLR) to comply with federal unemployment insurance regulations. The data are collected quarterly and provide monthly summaries of employment by workplace. Appendix A provides more detail on the data and its processing, including the time period of the data, how master account records were treated, and how workplaces with zero employment were treated.

MPO Base Year and Collaborative Forecasts. The primary source for socio-economic data in the Baltimore and Washington DC regions are the MPO model base year and forecast data used in the BMC and MWCOG models. Similar data was obtained from the VDOT, PennDOT and DelDOT models. These data were adjusted in the reconciliation process to account for definitional definitions, etc. The key data used from these other models includes the following:

BMC:	2000, 2010, and 2030 (7.0)	(Release Year: 2010)
MWCOG:	2000, 2010, and 2030 (7.2a)	(Release Year: 2010)
PennDOT:	2002 and 2030	(Release Year: 2005)
VDOT:	2000 and 2030	(Release Year: 2005)
DELDOT:	2000 and 2030	(Release Year: 2005)

Multi-State Forecasts A multi-state base and forecast year estimate of socio-economic data was developed to provide a common method across the MSTM study area [7]. This provided a consistent method and control totals for comparison with MPOs in the data reconciliation process (Appendix A), as well as a forecast for model regions outside the cooperative forecasts.

For the base year 2005, this data are derived from Census data (for households) and the REIS database from the BEA (for employment). For 2040 and 5-year intermediate years (ending in 0 and 5) a top-down allocation of economic and demographic magnitudes from the nation to “regions” to individual counties was employed, shown schematically in Figure 2-3. The regions comprise an intermediate level spanning labor markets and metropolitan districts.

As shown by the two upper sections of the schematic diagram, the national forecasting process works from population to employment. National employment totals for future years are derived from Census Bureau population projections. Breakdowns of these totals by industry then become the basis for forecasting employment in regional industries, given ratios of regional to national employment projected from historical data. Cohort-survival analysis is then used for each region to derive a population profile consistent with its employment level in each future year. An allocation model disaggregates the regional totals to jurisdictions, calibrated to 1995-2005 historical data. Because all predictors in a recursive forecasting framework must themselves be predicted, the explanatory variables are limited almost entirely to past changes, initial conditions and current changes in the target variables themselves. “Proximity” variables integrate across employment or households in all of a region’s counties, weighted negatively by distance to the subject county and positively by a measure of that county’s available land. In the testing and retention of explanatory variables, all sectors are eligible as predictors of all other sectors, with no overall direction of causality imposed as in the regional forecasting case.

The calibrated model involves 40 equations (in each of two versions) because the five household variables are estimated four times using progressively more inclusive sets of predictors. The economic descriptors throughout the process consist of wage-and-salary employment by industry, with the number of industries varying from 20 to 22 at different stages.

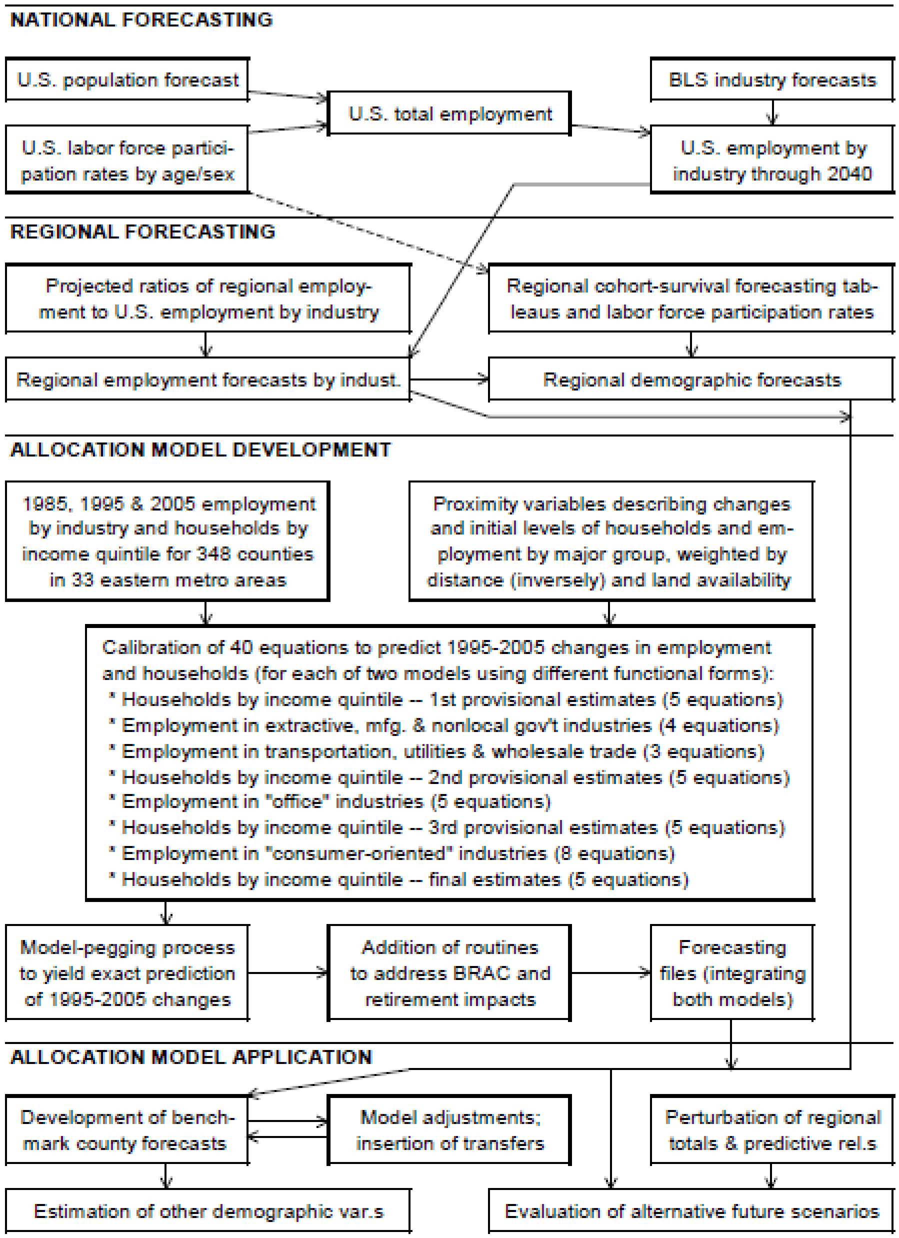


Figure 2-3: Schematic of top-down forecast allocation process

2.2.1 Socio-Economic (SE) Data Reconciliation

The Socio-Economic (SE) data reconciliation is an important part of establishing the inputs to the MSTM. As the modeling region in MSTM consists of Maryland, and six other neighboring states, the SE data is collected from numerous sources such as MPOs, state DOTs and local agencies. The data sources do not follow the same definition and are not in the same format. The SE data reconciliation integrated all the data sources to provide a unified set of inputs to the MSTM. The methods used for the year 2000, the future year 2030 and the validation year 2007 is summarized in Table 2-3 and described in the following section. Further details and formulae for data reconciliation methods are found in Appendix A-C.

Table 2-3: Summary of source data for MSTM socio-economic inputs

	HH		Emp	
	county control totals	SMZ+sector distribution	county control totals	SMZ+sector distribution
2000 Baseyear				
BMC	NA	2000 Census	2000 BEA	2000 BMC (7.0) [1]
MWCOG-MD	NA	2000 Census	2000 MWCOG (7.2a)	2000 MWCOG (7.2a) sector factors, 2000 CTPP [2]
MWCOG-VA	NA	2000 Census	2000 MWCOG (7.2a)	2000 MWCOG (7.2a) sector factors, 2000 CTPP
rest of MD	NA	2000 Census	2000 BEA	2007 QCEW
non-MD	NA	2000 Census	2000 BEA	DL: 2000 DELDOT PA/VA: 2000 PENNDOT/VDOT [3] NJ/WV: 2000 CTPP
2030 Consolidated forecast				
BMC	NA	2030 BMC (7.0)	2030 BMC	2030 BMC (7.0)
MWCOG-MD	NA	2030 MWCOG (7.2a)	2030 MWCOG (7.2a)	2000 MWCOG (7.2a) sector factors, 2000 CTPP
MWCOG-VA	NA	2030 MWCOG (7.2a)	2030 MWCOG (7.2a)	2000 MWCOG (7.2a) sector factors, 2000 CTPP
rest of MD	2030 TH	2000 Census=TH	2030 TH	2007 QCEW
non-MD	2030 TH	DL: 2030 DELDOT PA/VA: 2030 PENNDOT/VDOT NJ/WV: 2000 Census	2030 TH	DL: 2030 DELDOT PA/VA: 2030 PENNDOT/VDOT NJ/WV: 2000 CTPP
2007 Validation Year				
BMC	2005-2010 BMC (7.0)	2010 BMC (7.0)	2005-2010 BEA	2005-2010 BMC (7.0)
MWCOG-MD	2005-2010 MWCOG (7.2a)	2010 MWCOG (7.2a)	2005-2010 MWCOG (7.2a)	2030 MWCOG (7.2a)
MWCOG-VA	2005-2010 MWCOG (7.2a)	2010 MWCOG (7.2a)	2005-2010 MWCOG (7.2a)	2030 MWCOG (7.2a)
rest of MD	2007 Census	2000 Census	2005-2010 BEA	2007 QCEW
non-MD	2007 Census	2000 Census	2005-2010 BEA	DL: 2000 DELDOT PA/VA: 2000 PENNDOT/VDOT NJ/WV: 2000 CTPP

[1] In future if there is not much difference between the employment categorization between BMC and ES-202 at SMZ level, ES-202 can be used in BMC region.
[2] In future if there is not much difference between the employment categorization between CTPP 2000 and ES-202 at SMZ level, ES-202 can be used in MWCOG region.
[3] For Industrial and Other category, CTPP 2000 data is used at SMZ level for employment proportions, to avoid definition problems from PennDOT and VDOT data
TH = Tommy Hammer BEA/Census-based forecast

2.2.2 General Methodology

County level information is the basic source of input for all employment data. County level employment is then allocated to individual SMZs based on the proportion of employment as determined by MPO estimates, CTPP or QCEW data. For households, 2000 Census allocations were used directly, with future year data taken directly from MPOs or forecast county household allocated to SMZs based on 2000 Census, MPO or State DOT model projections.

2.2.2.1 HorizonYear -2000

- **Employment**
  - BMC: BEA 2000 control totals are used to estimate County level employment. County level employment is then proportionally disaggregated to SMZ based on employment estimates from the BMC 2000 (round 7.0) employment estimates. The pro-

portion used for allocation is BMC employment in the SMZ divided by total BMC employment in the County.

- MWCOG-within Maryland: At the county level, MWCOG 2000 control totals (adjusted to BEA definitions after multiplying factors provided by MWCOG) are used. At the SMZ level, total employment is the proportion of jurisdiction and SMZ total employment from the MWCOG 2000 (round 7.0) multiplied with the adjusted MWCOG 2000 county control totals. At the SMZ level, the distribution of employment category is based on CTPP[5] 2000.

- MWCOG-outside Maryland: At the county level, MWCOG 2000 control totals (adjusted to BEA definitions after multiplying factors provided by MWCOG) are used. At the SMZ level, total employment is the proportion of jurisdiction and SMZ total employment from the MWCOG 2000 (round 7.0) multiplied with the adjusted MWCOG 2000 county control totals. At the SMZ level, the employment categories are based on CTPP 2000.

- Non-MPO Region Maryland: At the county level, BEA 2000 control totals are used. At the SMZ level, the total employment is the proportion of jurisdiction and SMZ total employment from the QCEW 2007 (ES-202) multiplied with the BEA 2000 county control totals. At the SMZ level, the distribution of employment category is based on QCEW 2007.

- Regions outside Maryland: At the county level, BEA 2000 control totals are used and at the SMZ level, the following is used in different regions:

  - New Jersey and Remainder of West Virginia: At the county level, BEA 2000 control totals are used. The allocation to SMZs is based on the distribution of employment by category in the CTPP 2000

  - Delaware: At the County level BEA control totals are used. To allocate to the SMZ level the proportions of DELDOT 2000 employment was used.

  - Pennsylvania and Virginia: At the county level the BEA 2000 controls were used. Employment was then sub-allocated to SMZs based on PennDOT/VDOT 2000.

- **Household (Population)**

  - For Households, Census 2000 data are used throughout the modeling region.

[5]In the future if there is not much difference between the employment categorization between CTPP 2000 and QCEW at SMZ level, ES-202 can be used in MWCOG region.

2.2.2.2 FutureYear-2030

- **Employment**
  - BMC: At the county level, 2030 control totals are used. At the SMZ level, the total employment is the proportion of jurisdiction and SMZ total employment from the BMC 2030 (round 7.0) multiplied with the BMC 2030 county control totals. At the SMZ level, the distribution of employment category is based on BMC 2030.
  - MWCOG-within Maryland: At the county level, MWCOG 2030 control totals (adjusted to BEA definitions after multiplying factors provided by MWCOG) are used. At the SMZ level, each employment category is multiplied with growth the jurisdiction has received between 2000 and 2030 (round 7.2a). Then the revised total employment (at jurisdiction level) is compared with the 2030 total employment. Then employment categories at SMZ level is multiplied with the proportion of 2030 total employment (round 7.2a adjusted with factors provided by MWCOG) and revised total employment at the jurisdictional level.
  - MWCOG-outside Maryland: At the county level, MWCOG 2030 control totals (adjusted to BEA definitions after multiplying factors provided by MWCOG) are used. At the SMZ level, each employment category is multiplied with growth the jurisdiction has received between 2000 and 2030 (round 7.2a). Then the revised total employment (at jurisdiction level) is compared with the 2030 total employment. Then employment categories at SMZ level is multiplied with the proportion of 2030 total employment (round 7.2a adjusted with factors provided by MWCOG) and revised total employment at the jurisdictional level.
  - Non-MPO Region Maryland: At the county level,2030 control totals are used. At the SMZ level, the total employment is the proportion of jurisdiction and SMZ total employment from the QCEW 2007 QCEW multiplied with the 2030 county control totals. At the SMZ level, the distribution of employment category is based on QCEW 2007.
  - Region outside Maryland: At the county level, 2030 control totals are used, and at the SMZ level, the following is used in different regions:
    - New Jersey and Reminder of West Virginia: At the SMZ level, the total employment is the proportion of jurisdiction and SMZ total employment from the CTPP 2000 multiplied with the 2030 county control totals. At the SMZ level, the distribution of employment category is based on CTPP 2000.
    - Delaware: At the SMZ level, the total employment is the proportion of jurisdiction and SMZ total employment from the DELDOT 2030 multiplied with

the 2030 county control totals. At the SMZ level, the distribution of employment category is based on DELDOT 2030.

    - Pennsylvania and Virginia: At the SMZ level, the total employment is the proportion of jurisdiction and SMZ total employment from the PennDOT/VDOT[6] 2030 multiplied with the 2030 county control totals. At the SMZ level, the distribution of employment category is based on PennDOT/VDOT 2030.

- **Household (Population)**
  - BMC: The households from the BMC 2030 (round 7.0) TAZ level is summed to the SMZ level.
  - MWCOG-within Maryland: The households from the MWCOG 2030 (round 7.2a) TAZ level is summed to the SMZ level.
  - MWCOG-outside Maryland: The households from the MWCOG 2030 (round 7.2a) TAZ level is summed to the SMZ level.
  - Non-MPO Region Maryland: At the county level, 2030 control totals are used and at the SMZ level Census 2000 household proportions are used.
  - Region outside Maryland: At the county level, 2030 control totals are used and at the SMZ level jurisdiction proportions are used (except in New Jersey and remainder of West Virginia, where census 2000 household proportions are used).
    - New Jersey and Reminder of West Virginia: At the SMZ level Census 2000 household proportions are multiplied with 2030 county control totals.
    - Delaware: At the SMZ level, the total household is the proportion of jurisdiction and SMZ total household from the DELDOT 2030 multiplied with the 2030 county control totals are used.
    - Pennsylvania and Virginia: At the SMZ level, the total household is the proportion of jurisdiction and SMZ total household from the PennDOT/VDOT 2030 multiplied with the 2030 county control totals are used.

2.2.2.3 2007 SE Data Estimation Procedure

1. The estimation procedure is conducted in two steps: Linear interpolation from 2005 to 2007 - The first step is to determine the population, household, and employment (total

[6] VDOT 2025 SE data is converted to year 2030 based on the past growth.

and by industry) for the year 2007, when the SE data for year 2005 and 2010 is given[7]. This assumes a linear growth of SE variables over time.

2. Adjust to BEA and Census controls - The second step is to adjust the SE control totals[8] obtained from step 1 with the official data from census (population and household) and BEA (employment).

General Formula

The 2005 and 2010 SE data is available from official sources at SMZ level. The formula for obtaining 2007 SE data is

$SE_{2007} = SE_{2005} + [(SE_{2010} - SE_{2005}) / (2010-2005)] \times (2007-2005)$

Population and Household

The control total for population is obtained from http://www.census.gov/popest/counties/

The control total for household is not available from the same source[9]. Household at the SMZ level is obtained as:

$HH_{2007} = POP_{2007\text{-adjusted}} / (POP / HH)_{2007\text{-unadjusted}}$

Employment

The control total for employment is obtained from (CA25): http://www.bea.gov/regional/reis/

Employment control total data by industry for all the counties is not available for year 2007 from BEA. Hence, total employment is used in the second step to adjust control totals.

2.2.3 Statewide Layer

The statewide person trip component of the MSTM travel model is very similar in structure to the BMC MPO model but covers the entire SMZ area. Parameters vary by SMZ and by Region. A complete description of the development of various Statewide Model parameters from the 2007-2008 combined Baltimore and Washington Household Travel Survey (HTS) data can be found in Appendix D [8].

2.3 Network and Skim Development

MSTM uses a multi-modal network at the Statewide level, including highway and transit networks and associated assumptions on link attributes and model-wide intercity and urban transit service. The networks were compiled from various existing models, including MPO, DOT, and other sources, and standardized. Extensive efforts were made to map the highway network to the SHA CenterLine network to enable sharing of data. Initial network and skim development is discussed in greater extent in the *MSTM Model Networks Development* document listed in the reference section [6].

[7] When 2005 and 2010 data are not available the nearest time slices are considered.

[8] Control totals at county level

[9] Household (and population) data is available from the American Community Survey (ACS) at: http://www.census.gov/acs/www/. But the 2007 data is not available for all the counties. Example, for Maryland household data is available for 16 counties (as opposed to all 24 counties).

2.3.1 Consolidated Roadway Network

This section describes the link attributes provided in each regional, state, and national source used to develop the MSTM network and the key adjustments made to form a unified set of network link attributes. This includes the re-numbering of nodes to establish unique values for modeling processing. Several sources were used to develop an initial set of network attributes for the MSTM. The attributes provided in the BMC network were used as the main source. Model networks from MWCOG, DelDOT, and a network prepared by Caliper for a previous regional project were reviewed to identify attributes that matched or nearly matched those provided by the BMC.

2.3.1.1 MSTM Network Attributes

Table 2-4 provides a summary of the attributes that have been developed for the MSTM. Other attributes from the various networks may be adopted in the future if deemed necessary. Since several of the coding conventions used in the various networks are not the same, a hybrid set of codes had to be developed for the MSTM.

Table 2-4: MSTM network metadata for links

Field	Description
A	A node
B	B node
AMLIMIT	AM peak link usage restriction code
PMLIMIT	PM peak link usage restriction code
OFFLIMIT	Off-peak link usage restriction code
FT	Facility type
DISTANCE	Distance in miles
SPDP	Posted speed limit, mph
CAPCLASS	Maximum daily lane capacity divided by 50 (Service level 'E')
CNTID	Regional count database identification
CNT00	Year 2000 daily count
CNTWKD00	Year 2000 weekday count
HTCNT00	Year 2000 heavy truck count
MTCNT00	Year 2000 medium truck count
COMCNT00	Year 2000 commercial vehicle count (not presently coded)
AMLANE	AM peak number of lanes
PMLANE	PM peak number of lanes
OFFLANE	Off-peak number of lanes
FFSPEED	Free-flow speed, mph
CONGSPD	Initial congested speed, mph
CAPE	Maximum daily lane capacity (Service level 'E')
TOLLCOSTOF	Off-peak toll, cents (year 2000 $)
TOLLCOSTPK	Peak toll, cents (year 2000 $)
FROM_TO_ID	Local network link identifier
MODEL	Local model identifier
PB_DIST	PB calculated distance in feet

Field	Description
RECID	Temporary ID number for links used to stitch networks
FROM_X	From Node X Coordinate
FROM_Y	From Node Y Coordinate
TO_X	To Node X Coordinate
TO_Y	To Node Y Coordinate
SWFT	Statewide Model facility type
DIR	One-way directional code
RMZ_NAME	RMZ name
JUR_NAME	Jurisdiction Name
JUR_FIPS	Jurisdiction FIPS Code
SMZRMZ	SMZ or RMZ number
RT_ID	Route ID number
RT_NAME	Route Name
ACRES	Acres
PBAREATYPE	PB defined area type
AREATYPE	Local network defined area type
FT_ORIG	Original FT

Table 2-5: MSTM limits codes

Code	Description
0	No restriction/GeneralUse
1	General Use
2	HOV2+ only
3	HOV3+ only
4	no Medium or Heavy Trucks allowed
5	Non-Airport Vehicles Prohibited
6	Transit Only
9	no vehicles (used in order to allow a link to physically remain in the network, but be closed to all traffic during certain periods; certain HOV lanes operate in this manner)

The various roadway functional classifications used in the MSTM are shown in Table 2-6. As discussed previously, the original MPO functional class is used to determine statewide functional class, link speeds, capacities, and VDFs.

Table 2-6: MSTM functional type

Functional Type Code	Description
1	Interstate
2	Freeway
3	Expressway
4	Major Arterial
5	Minor Arterial
6	Collector
7	Not Used
8	Medium Speed Ramps
9	High Speed Ramps
10	Local Roads
11	Centroid connector
13	Drive Access Link (Hwy - PNR)
15	Rail Links
19	Drive Access Links to IntercityBus
20	Drive Access links to IntercityRail
21	PNR - Hwy walk link
22	Not Used
23	PNR - rail walk link
24	Rail - Rail walk link, Hwy – Hwy walk link
26	Amtrak

Other look-up tables from the BMC and MWCOG model documentation were used to help complete the initial set of MSTM attributes. The codes used as variables for these look-up tables will be maintained in the MSTM attribute table. A more generic set of look-up tables may be created at a later stage in the model development. For now, the values from the individual model look-up tables will be used.

Within Maryland roadway tolls are configured as link attributes and peak and off-peak tolls have been added (in 2000$). Tolls on a link basis apply to all vehicle types. Tolls on the Delaware Memorial Bridge have also been included. Other toll roads outside Maryland have also been identified but the tolls have not been included in the MSTM.

2.3.1.2 Area Type Attribute Update

MSTM calculates its own area type, consistent across the model area. The area type attribute indices are used in the mode choice models and to assist in estimating capacity on certain highway links. When a new network is created or the SMZ data updated (Section 2), the area type attribute must also be updated. It then serves as a lookup table for additional attributes on the network. The MPO models use measures of zonal activity, combined with area size, to develop indices of area type. In the MSTM and BMC model the households and employment are used to measure activity whereas in the MWCOG model population and employment are used.

For the MSTM, area types are classified into nine categories. The identification of an area type in the MSTM consists of four steps:

1. A measure of activity is calculated for each SMZ equal to households plus retail employment plus total employment.
2. The activity measure is then divided by SMZ total area in acres to obtain activity density.
3. Third, SMZ's are then sorted by activity density
4. SMZ's are then assigned an area type code from 9 to 1 according to the following:
    a. Using the measure of density and the total activity, starting from the most dense SMZ, the SMZs which include one ninth of the total activity have area type 9 assigned.
    b. Area type 8 is then assigned to the next group of SMZs which also contains one ninth of total activity.
    c. This process is repeated until each SMZ has been assigned an area type (9 to 1).
5. These initial area type breaks listed below are then held fixed in all other model years and alternate scenarios:
    a. 1 – Less than 0.3914 activity density measure (step 1)
    b. 2- 0. 3915 to 0.9446 activity density
    c. 3- 0.9447 to 2.7507 activity density
    d. 4- 2.7508 to 3.6032 activity density
    e. 5- 3.6033 to 5.3648 activity density
    f. 6- 5.3649 to 7.7239 activity density
    g. 7- 7.7240 to 12.0503 activity density
    h. 8- 12.0504 to 31.2705 activity density
    i. 9- Higher than 31.2705 activity density

This distribution of area types is shown in Figure 2-4.

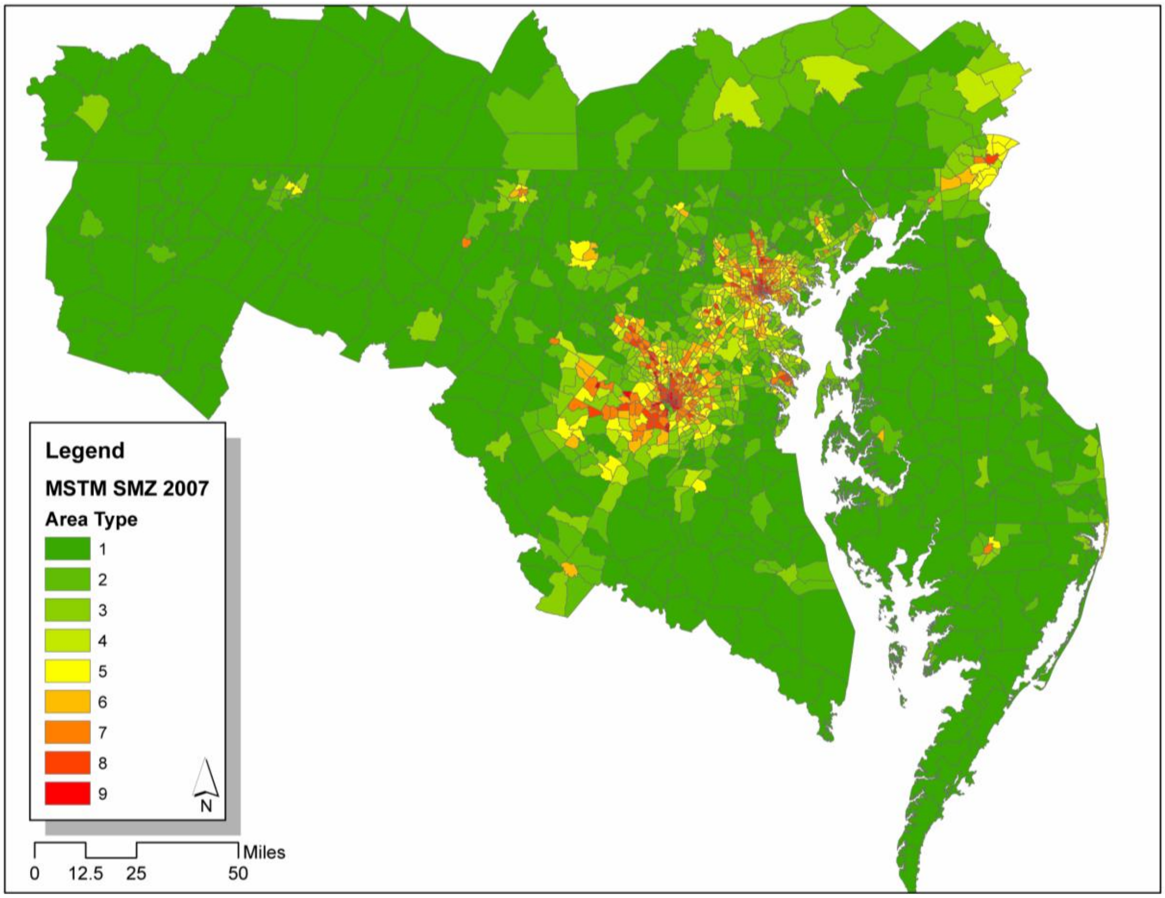


Figure 2-4: Area types for MSTM SMZs

2.3.1.3 Node Numbering

Since several sources were used to develop the MSTM network, the node numbering sequence had to be revised to eliminate duplications. The revised numbering sequence for the MSTM network was designed so that the values could be cross-referenced to the original network node numbers. This will allow for updates to the MSTM network based on changes to the original networks used and facilitate in the creation of a future year 2030 network. Table 2-7 summarizes the numbering sequence developed for the MSTM network.

Table 2-7: Node numbering system

Model System	Original Node Numbers	New Node Numbers	Comments
BMC	3002 to 39283	Unchanged	Unchanged
MWCOG	2358 to 19064	42358 to 59064	60000
DE	331 to 242037	80001 to 83165	Re-numbered 80K +
EastC	Null	83166 to 108772	Continued from DE
US	Null	108773 to 130952	Continued from EastC
SMZs	None	1 to 1588	Gaps (1607 total)
RMZs	None	1701 to 1873	Gaps (151RMZs)
Rail Nodes	None	4000 series	

2.3.2 Consolidated Transit Network

The MSTM network includes both MPO and intercity transit systems in Maryland and selected counties of adjacent states. As the transit focus of alternative scenarios will be on intercity transit facilities, ways to simplify local bus services in the transit networks were explored to expedite network coding. This includes the following transit systems and their system miles (2-way distance).

2.3.2.1 Transit Network Development

The objective of transit coding is to provide service to the zones that have service in the real world, not to serve as an exact representation of the route system. For example, streets that are too insignificant to be in the highway network are not added to the transit route. This would not result in a detailed description of transit service but would provide connectivity to the respective zones.

Unlike the MPO models where the non-transit links are added during the model run, in MSTM these have to be a part of the Transportation Network which is input to the model. Hence, the Park-N-Ride (PnR) node information was extracted from the MPO model files, and then those nodes were re-numbered and added to the MSTM network. PnR lots serve some specific stations which have to be coded along with the PnR information during the model run to facilitate the generation of Zonal Drive access legs described in the last section. These legs allow people to park their vehicle at the PnR lots and board the services at the stations being served.

Transit route files from the respective BMC and MWCOG models were combined and mode numbers were edited appropriately to reflect the new system. The node numbers that each route serves had to be re-numbered if they lie in MWCOG model area or if they were modified during the creation of MSTM roadway network so that they can fit on the new roadway network. This was a time consuming task as there is no automated procedure for such a conversion. It has been verified that all the transit stop nodes are highway nodes that are well connected to the network. Segments of the transit network had to be re-done to make them use the new more detailed network that came from the other MPO model. Some of the links in the present transit network may have only one link connecting two nodes while underlying highway network may have two links to establish the same connectivity, these do not cause a significant change in the results hence they were corrected to the extent possible given the scope of the project. A default speed called XYSPEED has been coded for each route to be used to calculate the time required to traverse such links using the XY distance.

The transit line descriptions follow the standard CUBE coding convention. The time periods are the same as the highway network assignment. Coded headways reflect the headway that is generally implied by the published timetable and are coded to the nearest whole minute. If the timetable suggests “clock” headways, that is what is coded (rather than the more intricate calculation used in some models, dividing the number of trips into the minutes in each time period).

2.3.2.2 Urban Transit

MSTM contains Baltimore and Metro Washington urban transit networks. These networks are taken directly from the BMC and MWCOG MPO model network files. There are two separate files, one for the Peak and one for the Off-Peak periods. These files consist of the route information for the Urban Transit Service. Bus Lines and Rail Lines are also present in separate files. The route files have been modified to reflect the re-numbered nodes in the MWCOG area. Since MSTM network derives parts of its network from different MPO networks, the transit lines had to be modified to fit the new network that came in from other MPO model. For example, parts of transit lines from BMC MPO area lying in the MWCOG's network had to be altered to fit the new network.

Modes from BMC and MWCOG models have been reorganized to form the MSTM mode system. Mode numbers 9 and 10 are not used. All modes are accessible via walk and Park-n-Ride (PnR). Below is a brief summary of the urban transit modes used in MSTM:

MODE 1. Local Bus- includes the following Bus Systems:

- BMC Buses: MTA Local Bus, MTA Premium Bus, Harford County Bus, HATS/Howard Transit/Connect-a-Ride (Howard County Bus), Carroll County Bus, Annapolis Transit Bus.
- MWCOG Buses: Local Metrobus, Other Primary - Local Bus, Other Secondary - Local Bus.

MODE 2. Express Bus- includes the following Bus Systems:

- BMC Buses: MTA Express Bus, MTA Premium Bus
- MWCOG Buses: Express Metrobus, Other Primary - Express Bus, Other Secondary - Express Bus.

MODE 3. Premium Bus: Includes BMC's MTA premium bus.

MODE 4. Light Rail: includes Baltimore light rail, Georgetown Branch, Anacostia and Montgomery Co. Corridor Cities Light Rail Lines.

MODE 5. Metro Rail: includes Baltimore Metro rail and DC Metro Subway.

MODE 6. Commuter Rail: includes MARC and Virginia Rail Express' Frederick and Manassas Lines.

2.3.2.3 Urban Transit Fares, Routes, and Schedules

Fare matrices were imported from the BMC (Version 3.3) and MWCOG (Version 2.2) models and combined to obtain the Fare matrix for the MSTM model (in 2000$). The weighted average of the trip matrix and fare matrix were used to convert the matrix from the earlier format to the newer one. Some other additional parameters like the HEADWAY for the lines is imported from the MPO models. HEADWAY 1 is for Peak period and HEADWAY 2 is for the Off-Peak Period.

2.3.2.4 Intercity Transit

Intercity transit includes Greyhound Bus and Amtrak Rail Lines in the model area, which covers six states. It may be noted that some of the routes described in the Urban Transit section also serve multiple MPOs within the State. These may also be used to commute between DC and Baltimore. Below are brief summaries of the Intercity Transit modes.

MODE 7. Amtrak Rail: Includes those routes that run regularly between DC and Baltimore. Only parts of the routes lying inside or close to the model area are coded and headways are also based on the coded segments of these routes. The following Amtrak stations are included:

- Wilmington, DE (WIL)
- Baltimore - Penn Station, MD (BAL)
- BWI Airport - Thurgood Marshall Airport, MD (BWI)
- Washington - Union Station, DC (WAS)
- Rockville, MD (RKV)
- Alexandria, VA (ALX)
- Newark, DE (NRK)
- Aberdeen, MD (ABE)
- New Carrollton, MD (NCR)

MODE 8. Greyhound Buses: Some of these routes are coded in the same way as Amtrak lines. Intercity Bus includes the following major stations:

- Annapolis
- Baltimore Downtown
- Baltimore Travel Plaza
- Easton
- Frederick
- Hagerstown
- New Carrollton
- Ocean City
- Salisbury
- Silver Spring
- Univ Of Md Eastern Shore
- Washington DC
- Wilmington DE

2.3.2.5 Intercity Transit Fares, Routes, and Schedules

Fare and scheduling data was collected for intercity transit including Greyhound Bus and Amtrak Rail line systems (in 2000$). The Amtrak data and some Greyhound data were collected using online resources from the transit providers in 2008. Web pages were used to find the data for city pairs that are included in the model area, and one stop into the halo. This allowed the modeling team to approximate the frequency of service for the transit modes. Greyhound does not have an online schedule information so a Greyhound schedule book was obtained for the route and headway information.

2.3.2.6 Non-Transit Modes

Some of the mode numbers are reserved for Non-transit modes that connect Transit services to the Highway links. A Non-transit leg is an imaginary entity representing a series of links required to establish the connection between transit and highway. The costs, such as distance and time, needed to traverse the leg are derived from the sum of the links traversed. In the following diagrams, roadway and non-transit links are combined to form the following links for three non-transit modes:

W2R = C1 + L1 + W1

W2B = C1 + L1 + L2

D2R = C1 + L1 + D1 (drive segment) and W3 (walk segment)

D2B = C1 + L1 + D1 (drive segment) and W2 + L2 (walk segment)

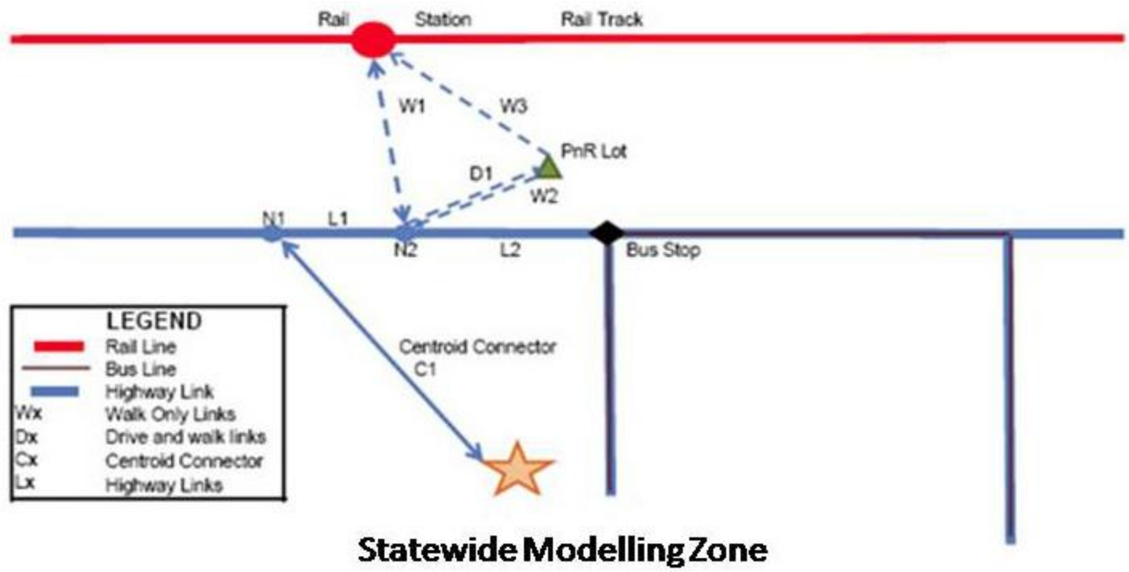


Figure 2-5: Transit coding diagram, transit and non-transit links

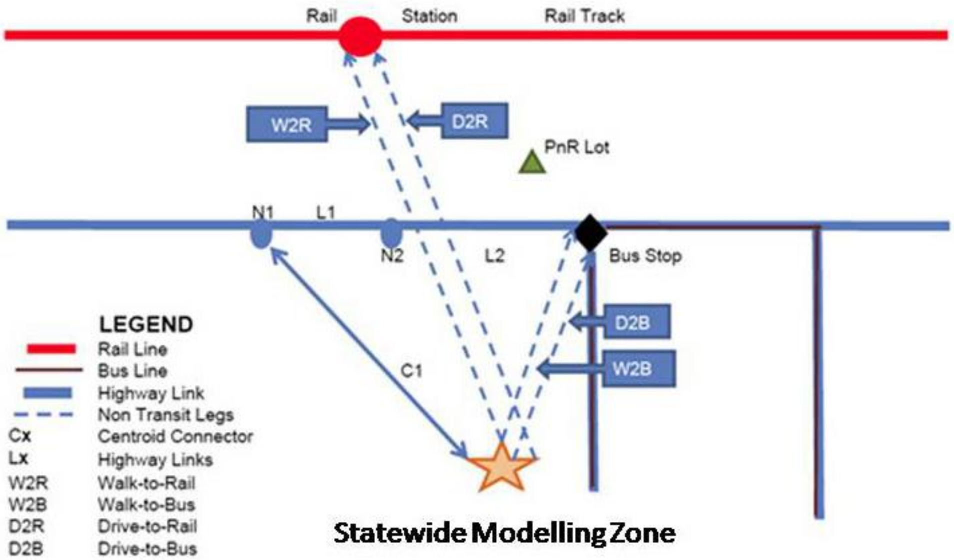


Figure 2-6: Transit coding diagram, transit and non-transit legs

MODE 11. Zonal Drive Access Legs: Connect the Zone Centroids with the nearby Park-n-Ride Lots. Unlike the Drive access Links whose purpose is to allow traffic to get on/off the roadway; legs connect a zone centroid to all the Park-n-Ride Lots within 10 mile distance. These PnR lots are then connected to the nearby stations/highway nodes via walk links.

MODE 12. Walk Transfer Legs: Hypothetical links that connect each line with nearby lines so that passengers can make transfers. These links derive their attribute values from the physical links that need to be traversed to establish connectivity.

MODE 13. Zonal Walk Access Legs: Similar to zonal drive access except they allow people to walk from the Zone Centroids to any of the nearby transit stop (within a mile of walking distance). These also derive their attribute values from the underlying network links.

2.3.3 Network Checking and Validation

Correct coding of the network and its attributes are critical for the model to produce reasonable outputs. To facilitate this process, tools have been developed to assist in network validation.

Some network coding errors are detected by Cube, but several definitional errors are not. A number of network validation checks were coded into a tool called NEVA[10] to ensure that the network is defined correctly. This tool should be run every time the roadway network is modified, covering the following checks. Use of this script is described at the end of this section.

- Links with differences between coded length and Euclidean distance
- Asymmetry of two-way link characteristics, such as length, functional classification (link type), area type, number of lanes, or capacity
- Dead-end or “dangling” links that do not connect to a downstream link or centroid connector

[10] NEVA: Network Validation

After the network passes these tests an assignment is carried out using a demand of 1 trip for each zone interchange in the trip matrix. The output of the assignment is checked for further problems with network coding:

- Traffic analysis zones that cannot be reached (i.e., have very large interzonal travel times associated with them, or the assignment fails)
- Links with zero flow after assignment (especially one-way links, which might have directionality coded improperly)

To run the NEtwork VAlidation (NEVA) tool, the Cube network is exported into a shapefile. The tool is started by opening a command prompt, navigating to the location where the NEVA tool is saved, and typing: NEVA <Name of shapefile>. The tool reads the shapefile and the corresponding attribute table and generates plots on the screen showing the links that potentially have problems. In addition, a file called <nevaReport.txt> is written that lists all links that should be checked for consistency.

2.3.4 Development of 2007 and 2030 networks

After a Year 2000 multi-modal network was developed as noted above, Year 2007 and Year 2030 networks were also developed. The 2007 network was used in model updates based on the 2007 joint Baltimore-Washington Household Travel Survey.

2007 Network
To create the 2007 network, modifications were made to the 2000 network, either through changes to specific links or the addition or deletion of links.

<u>Modifying Link Attributes</u>
Node IDs of MSTM network exactly match both the BMC and MWCOG networks. Node IDs of MSTM in BMC region are the same as those in BMC network. Node IDs of MSTM correspond to the MWCOG node IDs, however the MSTM node IDs have been incremented by 40,000 to avoid confusion. To create the 2007 network, the MSTM 2000 network's link attributes are updated according to the comparison of common links between 2000 and 2010 MPO networks. Network updates are made only with changes occurring before 2007.

<u>Adding and Deleting Links and Nodes</u>
Links that are part of the 2000 MPO networks and the MSTM, but not found in 2010 MPO networks are recorded as "to be deleted". If a link exists and is functional prior to 2007 in the 2010 MPO network but cannot be found in the 2000 network and MSTM 2000 network, the link will be tentatively added into the MSTM 2000 link set and denoted as "to be added". If those links are relevant to a new node, the node and its coordinates are also added into MSTM's new node set. In the new node set, 20,000 is added onto IDs of nodes whose original ID is greater than 60,000 in order to avoid node ID conflict with MPO node numbering systems. All the relevant links' node IDs are revised in the MSTM network, accordingly.

Extensive visual checks were conducted in the GIS interface by overlaying updated MSTM networks onto MPO networks. Redundant links, marked as "to be deleted", are confirmed and deleted in the GIS interface. New links to be added are retained if they match those in MPO network. In case that some links do not accurately represent links in MPO network, manual modification is conducted to guarantee newly added links correspond to links in MPO network. In this procedure, some nodes are either moved, added or deleted.

Modifying the Transit Network
Due to minor modifications to the highway network, the original bus lines do not always match the updated highway network. The function of "public transport" in Cube Voyager is applied to examine the integrity of transit lines. In cases that a transit line does not match highway links, the line input file is updated to match the new highway network.

2.3.5 Linkage to Centerline Data

The MSTM to Centerline transfer task involved an effort to establish a one-to-one correspondence between the generalized MSTM network (also called a stick network) and the true-shape centerline network. This task is a major effort due to the size of each network and the complexity of establishing a correspondence. Much of the process required the manual transfer of link node pairs from the MSTM network to the centerline network. On a few occasions the process was automated to check for errors and make corrections. This document describes the transfer process.

2.3.5.1 Objectives of task

The objective of this task is to make a one-to-one transfer of MSTM and Centerline network data possible. This means that results from a transportation model run can easily be transferred to the centerline file from the MSTM stick network. Secondly, establishing a one-to-one linkage between the two networks allows observed traffic count data and other highway attributes from the centerline network to be transferred to the MSTM network for model validation.

This task required manually comparing the two networks and transferring unique node pairs from the MSTM network to the matching link in the centerline file. This task could not be fully automated for several reasons. The first most significant reason is that in many cases, the generalization of the MSTM stick network results network geography that does not closely match that of the true shape centerline files. Figure 2-7 shows the typical case of non-matching geography. The top center two links illustrate the somewhat similar geography of a major highway section in both the MSTM and centerline networks. The generalized MSTM network is in red and the centerline file is in green. Even though the angle and distance between the line segments for these two roadways are similar a precise match could not be solved automatically.

Secondly, the ramps represented on this highway in true shape form are curved clover leafs. The MSTM as a stick network is not capable of representing fully rounded geometries. As a result, finding the proper link between all possible patterns of ramps between the two networks represents a complex problem.

The next issue involves roadways that require some judgment in finding matching links. The MSTM link below the one previously mentioned highway looks like it cuts through a subdivision and has no centerline network match. Through a manual process the matching centerline link was found below the area shown and a one-to-one link was established. The complexity of finding such a link means that an automated matching process was not feasible for this task.

The issues discussed above, while representing just three cases here, are repeated at various levels of complexity throughout the statewide network. The MSTM network has over 30,000 individual links that had to be matched against nearly 325,000 centerline links. Thus the process of creating a link between the two networks represented a significant amount of personnel hours.

Figure 2-7: Small area MSTM and Centerline network comparison

2.3.5.2 Transferability Issues

Completion of this task was not without some significant challenges. Once a significant amount of progress was made to manually complete the one-to-one linkage, several processes were devised to automate the error correction and validation process in order to save time and effort. The following sections briefly outline some of the challenges faced in completing the network linkage and the solutions that were created to solve the problems.

2.3.5.3 General Geometry

As described in the previous section the MSTM lacks the complex geometry of the centerline file. As a result of these geometric differences, the task of checking the network linkage for errors was made more difficult.

The first step in the error correction and validation process was to determine if all of the links in the MSTM had directions, slopes and angles similar to their associated centerline links within a certain amount of tolerance. This was a simple calculation for the MSTM since all links are in A_B node form already. Using the coordinates of these A_B nodes the geometry of each link was calculated. The Centerline file does not have explicit A_B nodes for each link, so the first task was to create these nodes. A tool was used to formulate the location of each node and based on connectivity and location of each line segment an A and B node for each link was determined. Once the node locations were determined, the direction, angle and slope of each line were calculated. In addition to end points, the mid points of each line were calculated and the mean distance of each segment from their matched line was compared. Figure 2-8 shows an example of two network segments that correspond to each other but do not share a close geometry, angle or direction. In these cases that segments were flagged and subjected to further validation.

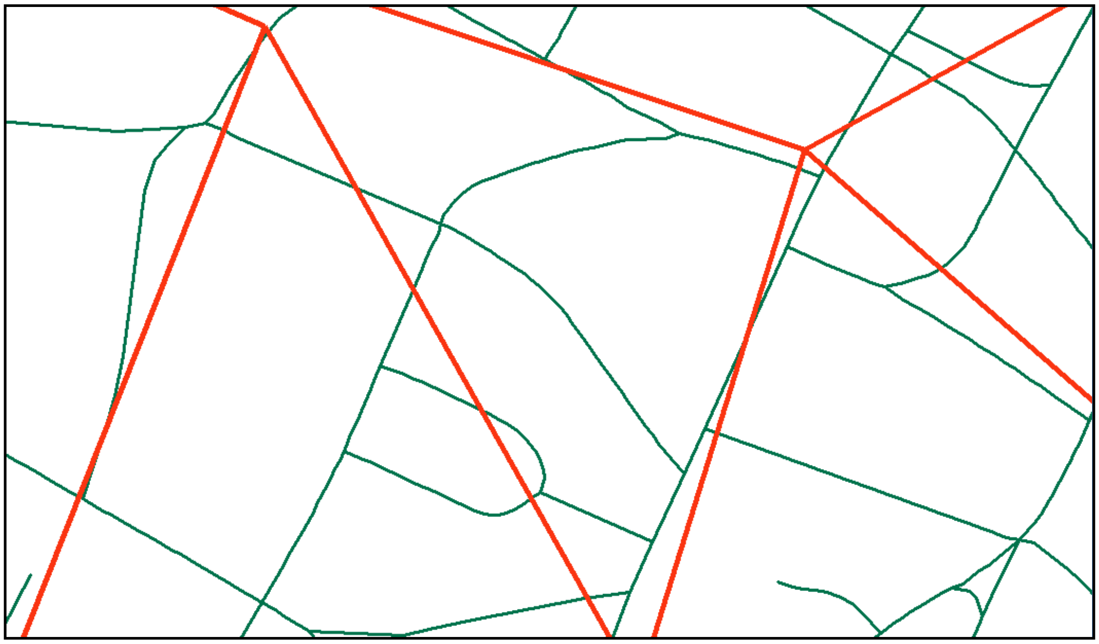


Figure 2-8: Network geometry differences

2.3.5.4 Multiple Correspondences

Another challenge in creating a one-to-on correspondence between the MSTM network and the centerline network was that line segments were not the same length between the two networks. Both networks created with a new line segment when the geometry of the road changes. Since the MSTM network is heavily generalized and the centerline network is very close to the real shape of the existing road network, on most occasions a link in the MSTM had a single segment while the centerline link had several segments. Since an automated process could solve for this issue, a manual transfer had to be used in order to fully transfer the attributes of the MSTM network to the centerline network. Figure 2-9 illustrates how many centerline segments correspond to a single MSTM line segment.

As part of the validation effort, an automated process was created at highlight links in the centerline network with corresponding MSTM links to determine if the attributes of the MSTM links where fully transferred to the multiple centerline links. Further validation was used to determine network connectivity to ensure the continuity of the transfer process by checking for gaps in the resulting MSTM attributed centerline network.

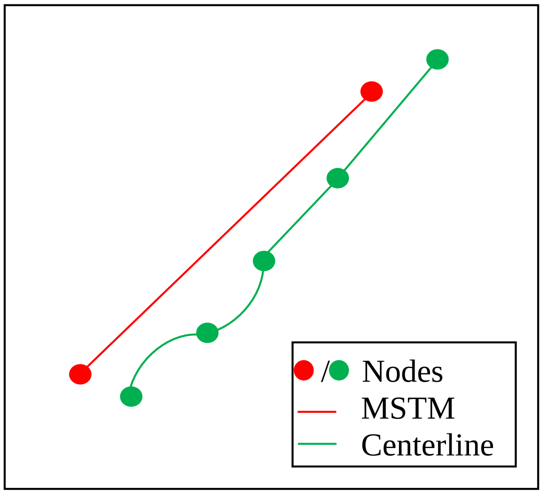


Figure 2-9: Network segments

2.3.5.5 Roadway Representation

An issue with creating an MSTM to centerline correspondence is differences in how lanes are represented in the two networks. In the MSTM network, major highways have two distinct lines, one for each direction. In the case of the Centerline network, roads are represented approximately how they exist on the ground. As a result, not just major highways but some minor highways, major arterials and boulevards that have medians between the two lanes are represented as two distinct lines in the network. A second issue is that many roadways in the MSTM network are represented by a single physical line but are coded with bi-directional data. However, the centerline file is a representation of the physical network so roads with more than a single direction are coded only as one direction.

2.3.5.6 Median Separated Road Segments

The complexity of a single MSTM bi-directional line segment to be coded to two centerline segments required the centerline line segments to be manually coded in each direction with the corresponding single bidirectional link in the MSTM network. Figure 2-10 shows two parallel line segments that represent a median separated road in the centerline network. Highlighted in yellow is the single MSTM line segment that corresponds to both of the centerline roads.

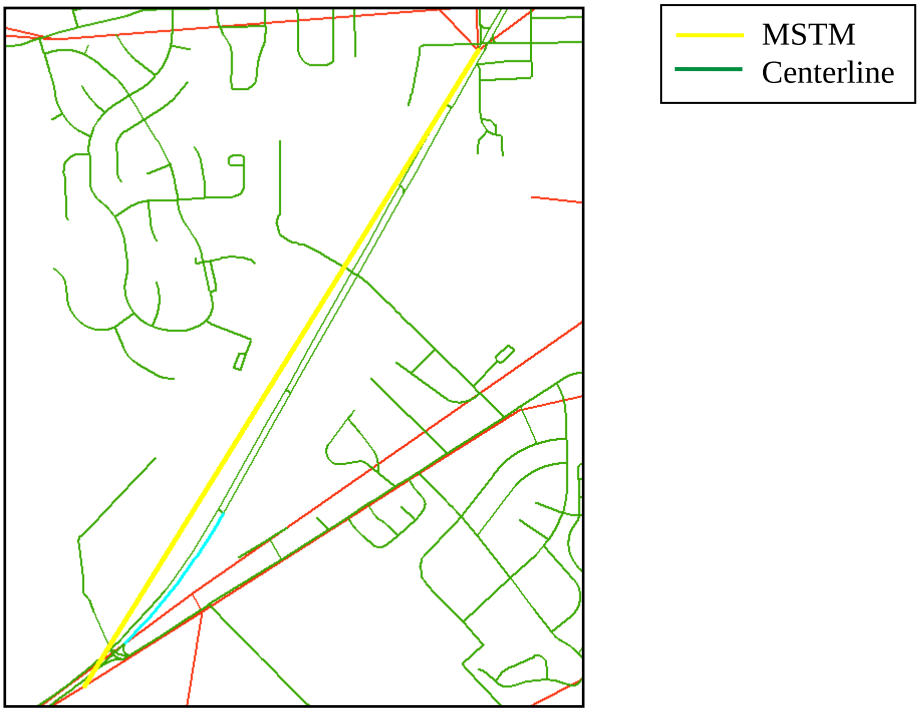


Figure 2-10: Median separation issues

When the current team received the partially completed correspondence network, many of the links that had this issue were coded only in one direction and in some cases with the wrong direction. To quickly solve this issue a method was created to detect the existence of these links, compare the attributes to the corresponding MSMT link and determine if all the links were correctly and fully coded. This process was able to repair all of the incorrectly coded links and served as a strong tool to check the validity of the completed one-to-one network correspondence.

2.3.5.7 Bi-directional Road Segments

The final major issue in transferring all of the MSTM attributes to the centerline network was that many of the centerline and MSTM links are rendered as a single physical line segment yet represents several lanes in opposite directions. To speed the process of creating a complete one-to-one correspondence, centerline segments were coded with just one direction from the MSTM network. Once each corresponding link in the centerline file had at least one direction from the MSTM coded to it, an automated process was created to find bidirectional links and add the additional MSTM directional attributes to the centerline network.

2.3.5.8 Observed Traffic Volume Transferability

The final objective of the MSTM and centerline correspondence task was to create a backwards link between the centerline network and the MSTM network to rapidly transfer observed traffic data from the centerline network to the MSTM network. The first step in this task was to attach all 3,883 AADT monitoring stations to the centerline network (Figure 2-11). Completing this task allows for a greater level of validation for the transportation model results.

Many of the stations are located in roadway medians and observe both sides of a divided roadway. Much like with the MSTM task, the stations were first located for one segment of the roadway then a process was created to locate the parallel roadway and attach station data for the opposite direction to that link. As a result of this effort over 7,000 traffic data points across the state can be transferred quickly to the centerline network and the MSTM network at anytime.

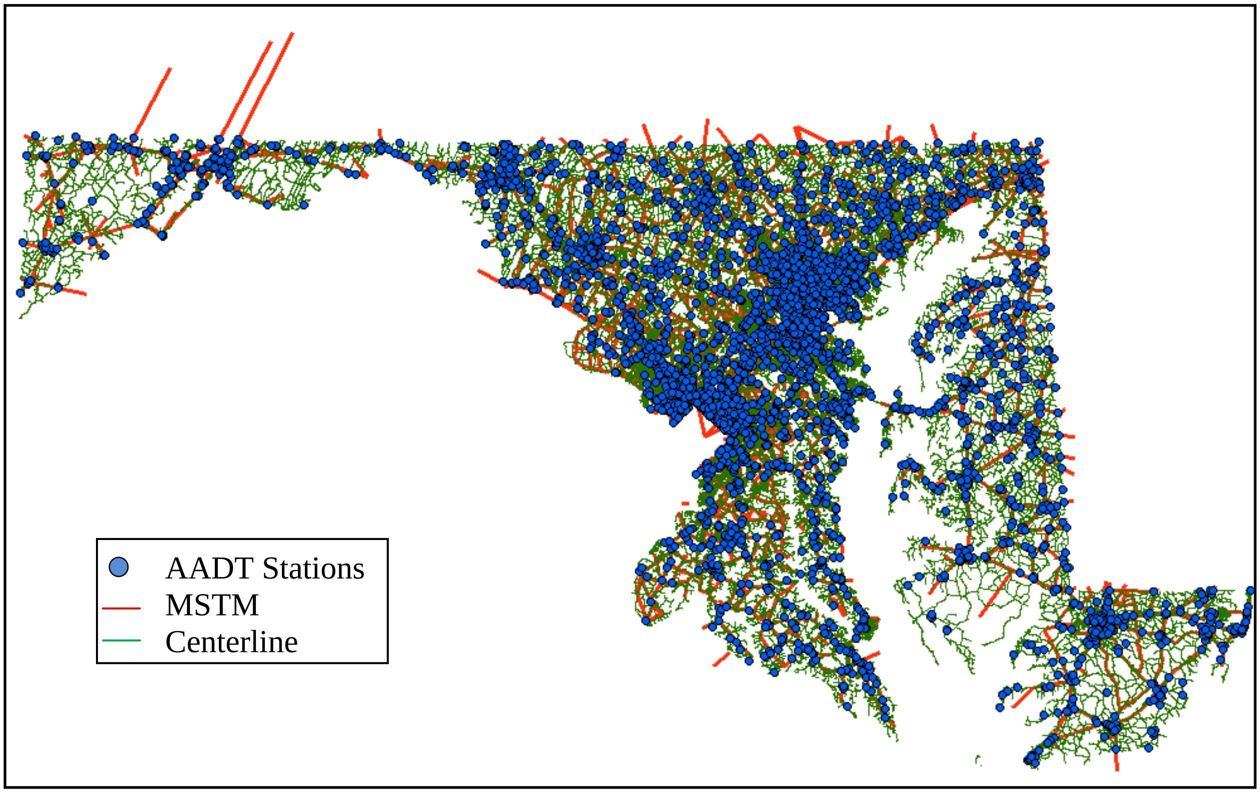


Figure 2-11: AADT Stations on the Centerline Network

3 Trip Generation

3.1 Statewide Layer

Person trip generation follows the same basic approach as the BMC model and encompasses the same trip purposes. The trip production component was updated to use household characteristics and trip rates derived from 2007-2008 HTS data and more recent Census data. The trip attraction component is based on linear regression equations derived from the same household survey data. Development of the independent household and employment variables required for each SMZ was described previously in Section 4.

3.1.1 Iterative Proportional Fitting:

MSTM person trip generation model uses trip production and attraction rates by household size (SIZ) by income (INC) and households workers (WRK) by income (INC). Since the SMZ data only provides households by income (see Section 4), a pre-generation step is applied to generate these joint distributions for the scenario year. An iterative proportional fitting (IPF) process combines the SMZ household data for the scenario year as marginals with joint-distribution seeds (from 2000 Census PUMS) to create households by SIZ and INC and households by WRK and INC at the SMZ level for a specified scenario year.

3.1.2 Trip Productions

The trip generation model produces trip productions by trip purpose for each SMZ based on joint distributions of households and trip production rates cross-classified by household category. The following trip purposes were identified:

- HBW = Home Based Work
- HBS=Home Based Shop
- HBO=Home Based Other
- HBSCH = Home Based School
- NHBW = Non Home Based Work
- NHBO = Non Home Based Other

Trip productions for work-related purposes are based on trip rates cross-classified by income and number of workers. The work related trips rates are slightly adjusted (reduced) to reflect the trips attracted to cities outside the MSTM region such as Philadelphia. Trip productions for non-work-related purposes are based on trip rates cross-classified by income and number of persons. Differences from the BMC approach are related to the income classification of households and the way motorized shares are derived and trip rates represent only trips within 50 miles. The long distance trips greater than 50 miles are modeled with the long distance travel model. Trip generation rates by household category and region are taken directly from the 2007-2008 HTS survey data. Rates are adjusted to the MSTM income categories (quintiles). The HTS regional rates used for the various MSTM regions are show in Table 3-1.

Table 3-1: Trip production rates by region and trip purpose

	HBW1				HBS1					HBO1				
	Wrks0	Wrks1	Wrks2	Wrks3	Size1	Size2	Size3	Size4	Size5	Size1	Size2	Size3	Size4	Size5
Urban	0.03194	1.11594	2.21429	2.7381	0.6754	0.9286	1.2676	1.1212	1.8913	0.984	1.7296	2.1831	3.3636	4.0435
Suburban	0.02715	1.12707	2.7381	2.7381	0.625	1.0874	1.8	1.3902	1.8913	0.965	2.1093	2.5867	4.1707	4.0435
Rural	0.02674	1.08602	2.7381	2.7381	0.6467	1.2737	1.8	1.3902	1.8913	0.8922	1.4526	2.5867	4.1707	4.0435
	HBW2				**HBS2**					**HBO2**				
	Wrks0	Wrks1	Wrks2	Wrks3	Size1	Size2	Size3	Size4	Size5	Size1	Size2	Size3	Size4	Size5
Urban	0.10963	1.23205	2.6	4.08696	0.6212	0.9676	1.3333	1.098	1.8354	1.0291	1.8866	2.6061	2.9608	5.5063
Suburban	0.05584	1.27261	2.35433	4.08696	0.6969	1.2694	1.3864	1.6444	1.8354	1.0857	2.0531	3.0568	3.4667	5.5063
Rural	0.13793	1.22697	2.5	4.08696	0.6293	1.2034	1.2063	1.3158	2.1316	0.9768	1.9186	3.2381	3.3158	5.2895
	HBW3				**HBS3**					**HBO3**				
	Wrks0	Wrks1	Wrks2	Wrks3	Size1	Size2	Size3	Size4	Size5	Size1	Size2	Size3	Size4	Size5
Urban	0.0719	1.30427	2.47699	3.98701	0.6472	1.0985	1.5	1.9756	1.902	0.8629	2.0925	3.7308	7.8293	7.1078
Suburban	0.05706	1.24526	2.41887	3.98701	0.6492	1.2407	1.5649	1.9949	1.902	0.959	2.0725	3.3789	5.1173	7.1078
Rural	0.11392	1.12834	2.28571	3.71642	0.5614	1.5013	1.7421	1.8027	2.1667	0.7602	1.9215	3.1006	4.3673	7.4881
	HBW4				**HBS4**					**HBO4**				
	Wrks0	Wrks1	Wrks2	Wrks3	Size1	Size2	Size3	Size4	Size5	Size1	Size2	Size3	Size4	Size5
Urban	0.03797	1.31975	2.43103	3.5974	0.627	1.2314	1.9	1.6111	2.472	0.9016	1.6829	3.11	7	7.4161
Suburban	0.09406	1.23503	2.36114	3.5974	0.657	1.2935	1.552	1.9966	2.472	0.9126	2.0064	3.2514	4.8537	7.4161
Rural	0.2	1.06993	2.12554	3.35443	0.6061	1.1296	1.3967	1.8358	3.0374	0.6212	1.6698	2.7554	4.3781	6.3645
	HBW5				**HBS5**					**HBO5**				
	Wrks0	Wrks1	Wrks2	Wrks3	Size1	Size2	Size3	Size4	Size5	Size1	Size2	Size3	Size4	Size5
Urban	0.1	1.24832	2.41411	3.92727	0.5889	1.259	1.7215	1.6232	2.1695	0.8333	1.8237	3.8101	6.0145	7.0678
Suburban	0.07692	1.27925	2.34343	3.92727	0.6782	1.165	1.3969	1.7742	2.1695	0.7931	1.8595	3.0825	5.2043	7.0678
Rural	0.07692	0.91667	2.30348	3.92857	0.6782	1.0063	1.4531	1.5625	2.1695	0.7931	1.4125	2.5625	4.6562	7.0678
	NHBW				**NHBO**					**HBSCH**				
	Wrks0	Wrks1	Wrks2	Wrks3	Size1	Size2	Size3	Size4	Size5	Size1	Size2	Size3	Size4	Size5
Urban	0.02716	0.81807	1.57447	1.29056	0.6667	1.1323	1.6267	1.6703	2.7386	0.0326	0.139486	0.71297	1.756256	2.690329
Suburban	0.02586	0.73898	1.23537	1.62068	0.7607	1.2917	1.56	1.9418	2.4039	0.016762	0.095771	0.787744	1.683333	2.890661
Rural	0.05386	0.69022	1.20296	1.71001	0.8789	1.4065	1.791	2.1243	2.7306	0.003852	0.048097	0.653769	1.647994	2.560217

The MSTM does not include school bus mode and HBSCH trip rates are adjusted during the trip generation to reflect only non-school bus mode trips.

3.1.3 Trip Attractions

Trip attractions by SMZ are calculated based on regression-type equations applied to SMZ socioeconomic variables for the non-home end of trips.

The attraction rates were derived from the combined HTS survey data. The rates were calculated for the entire survey area, not distinguishing urban, suburban and rural regions. For production rates, the objects that generate trips are households. The survey is large enough to calculate region-specific production rates by households. For attraction rates, however, the objects that attract trips are zones with their employment and household numbers. As few trips in the survey had the same zone as destination, it was impossible to create region-specific attractions that were statistically significant. Therefore, the entire survey area was treated as one region to increase the number of records used to estimate attraction rates for each trip purpose.

Table 3-2: Trip attraction rates

	Purpose					
Independent variable	**HBWork**	**HBShop**	**HBOther**	**HBSchool**	**NHBWork**	**NHBOther**
Households			3.158			0.82
Total employment	1.0286					
Retail employment		6.667				
Office employment					0.79	
Other employment			0.785		0.57	0.85
School enrollment				1.902		

3.1.4 HBW adjustment

An analysis was done to identify the number of residents who worked outside the model area. This was of particular concern in the Philadelphia area, where MSTM contains suburbs, but not the city. An analysis of 2000 Census CTPP data was done to identify by county, the number of worker flows that originated within the model area and destined outside the worker area. These county-level adjustment factors were applied to the HBW trip table.

4 Non-Motorized Share

The Maryland Statewide Transportation Model (MSTM) generates motorized trips only. Walk and bike trips are generated by trip generation, but shall not be included in trip tables for subsequent modules. A certain share of trips is dropped before trip productions and attractions are fed into the destination choice model. Previously, the MSTM model applied Weibull functions to estimate the non-motorized shares by area type and purpose. Plotting these shares showed unexpected patterns, which affect trip origins, mode choice and the assignment results. To mitigate the impact, non-motorized shares were averaged across counties. This resulted into reasonable patterns non-motorized shares, however, the was a steep border effect were two neighboring zones in different counties may have very different non-motorized shares, while all zones within one counties were treated as being equal in terms of non-motorized shares. **Error! Reference source not found.** shows the motorized share, which is the inverse of the non-motorized share, used in MSTM for Home-based Work trips up to phase 3.

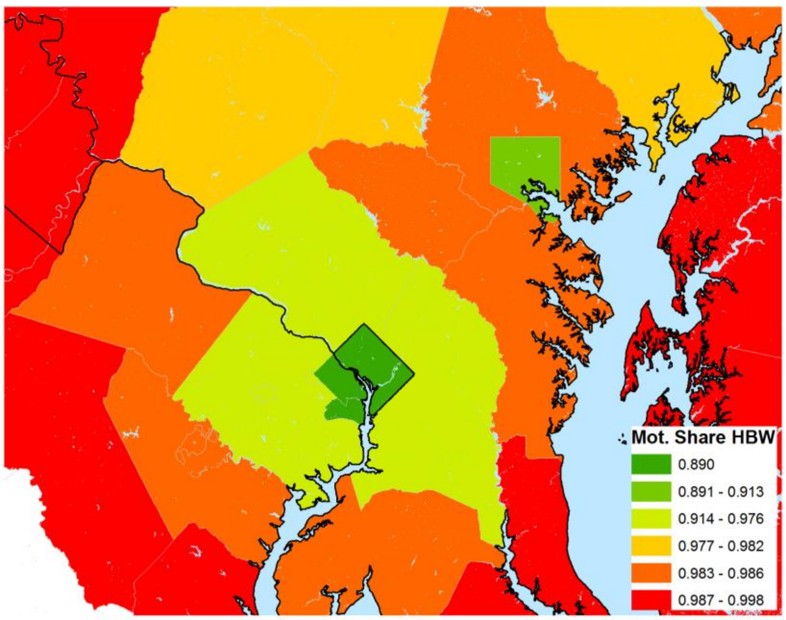


Figure 4-1: Previously assumed motorized share for HBW

In this phase, the 2007 Household Travel Survey was used to estimate the non-motorized share by zone. A multiple regression was set up to analyze the impact of various measures of densities and accessibilities on non-motorized shares at the zonal level.

4.1 Observed Data

The 2007 household travel survey was used to calculate the observed non-motorized shares. The primary travel modes designated in the survey are shown in Table 4-1. Each mode has been categorized as motorized or non-motorized. The survey trips data was aggregated by SMZ, purpose, and travel mode. The non-motorized shares are then calculated by SMZ for each of the 18 purposes using equation 1.

Table 4-1: Primary travel modes in the household travel survey

Travel Mode	Motorized	Non-Motorized
Transit	√	
Auto D	√	
Auto P	√	
Walk		√
Bike		√
Other	√	

$$Non-Motorized\ Share\ per\ SMZ = \frac{Non-Motorized\ \ Trips}{(Motorized\ \ Trips + Non-Motorized\ \ Trips)} \quad (1)$$

The socioeconomic data (Activities.csv) is used to calculate the SMZ density per acre for three different densities: household, employment, and activity density. These densities were used as independent variables in the stepwise multiple regression. Table 4-2 shows how each of the densities were calculated.

Table 4-2: Density equations

Density	Equation
Household	HH/Acres
Employment	TotalEmp/Acres
Activity	(HH + TotalEmp + RetailEmp)/Acres

4.2 Accessibility

Besides various measures of density, accessibility was tested as an additional independent variable. Accessibility is a relative measure describing for a given zone how easily all other zones can be reached.

A large number of accessibilities have been defined over the last five decades (compare Schürmann et al. 1997[11]). The Hansen accessibility, also called potential accessibility, is probably the version that is used most commonly in transportation and land-use analyses, because it takes both the size of potential destinations as well as their distance into account. A larger size of a destination zones (measured in, for example, population or employment) increases the accessibility, while the distance to destination zones is inversely proportional accessibility:

$$acc_i = \sum_j s_j^{\alpha} \cdot \exp\left(\beta \cdot d_{i,j}\right) \quad (2)$$

acc_i Accessibility of zone i
s_j Size term of zone j (for example, population or employment)

[11] Schürmann, C., K. Spiekermann, M. Wegener (1997) Accessibility Indicators. Report 39. Institute of Spatial Planning, University of Dortmund.

$d_{i,j}$ Distance from zone i to zone j (measured in travel time)
α, β Parameters

The parameter α serves to increase or decrease the relative importance of particularly large centers accounting for agglomeration effects. The parameter β is a negative value increasing the disutility with larger distances. The exponential function makes the effect of distance non-linear, i.e. the difference between 1 mile and 2 miles is perceived to be larger than between 11 miles and 12 miles. After a few iterations of testing the impact of different parameters, α was set to 1.0 and β was set to -0.3.

Twelve different accessibility measures were calculated and tested as independent variables in the stepwise multiple regression (Table 4-3).

Table 4-3: Analyzed accessibility measures

	Accessibility by auto	Accessibility by transit
Accessibility to households	1	7
Accessibility to university enrollment	2	8
Accessibility to retail employment	3	9
Accessibility to office employment	4	10
Accessibility to other employment	5	11
Accessibility to total employment	6	12

To calculate transit accessibilities, only walk access (and not drive access) to transit was considered, as the goal of this task was to explain non-motorized trip shares. Accessibility to transit with walk access was expected to work as a proxy for walkability. All four transit modes (bus, express bus, rail and commuter rail) were taken into account, using the output files of the skimming process WBusPK.skm, WCRailPK.skm, WExpBusPK.skm and WRailPK.skm. Of the 22 tables given in every skim file, the table 11_BestJrnyTime was used. This table provides a combined travel time including initial wait time, transfer time, walk time and a penalty for every transfer. Out of the four transit modes, the one mode with the shortest travel time for a given origin-destination pair was used when calculating the accessibility, as travelers are assumed to select the fastest transit mode. Zones with no walk-access to transit received a transit accessibility value of 0.

As accessibilities are dimensionless, calculated values were normalized to values between 0 and 100.

$$acc_i' = \frac{acc_i \cdot sc}{\max(acc)} \tag{3}$$

acc_i' Scaled accessibility of zone i
acc_i Accessibility of zone i
sc Scaler, set to 100

This ensures that the impact of accessibility remains unchanged across different scenarios and model years. As accessibility is a relative measure (zone A is more accessible than zone B), the absolute growth in accessibility between two years is irrelevant. For example, if the population grows by ten percent, and the accessibilities across the region grow accordingly, the share of non-motorized trips is not expected to be affected. Accessibility is only used to spatially distinguish non-motorized shares.

4.3 Stepwise Multiple Regression

An R-Statistical script was written to run a stepwise multiple regression using the calculated non-motorized shares as the dependent variable and densities and accessibilities as the independent variables.

First the observed trips, accessibilities and the socioeconomic data are read in. The non-motorized shares and the densities are . Then the stepwise multiple regression is run on each purpose. The regression function produces coefficients for each of the independent variables that are being used. These coefficients are then run through a check to determine if they are usable or not. If the coefficients are less than zero the corresponding independent variables are removed and the stepwise multiple regression is run again using the remaining variables. This check was implemented as it is hypothesized that the non-motorized share is going to be larger if densities or accessibilities are greater. Due to this theoretical concept, only coefficients greater than zero are accepted The check and re-run is looped over until the output coefficients are all greater than zero.

Once the stepwise multiple regressions for each purpose cleared the check the resulting coefficients are applied to the corresponding densities and accessibilities in each SMZ to estimate the non-motorized shares, as seen in the equation below.

$$Estimated\ Shares = hhDC * hhD + empDC * empD + actDC * actD + hhCAC * hhCA + retCAC * retCA + othCAC * othCA + hhTAC * hhTA + offTAC * offTA + othTAC * othTA \quad (4)$$

Where:

hhDC	= Household density coefficient
hhD	= Household density
empDC	= Employment density coefficient
empD	= Employment density
actDC	= Activity density coefficient
actD	= Activity density
hhCAC	= Household car accessibility coefficient
hhCA	= Household car accessibility
retCAC	= Retail employment car accessibility coefficient
retCA	= Retail employment car accessibility
othCAC	= Other employment car accessibility coefficient
othCA	= Other employment car accessibility
hhTAC	= Household transit accessibility coefficient
hhTA	= Household transit accessibility

offTAC = Office employment transit accessibility coefficient
offTA = Office employment transit accessibility
othTAC = Other employment transit accessibility coefficient
othTA = Other employment transit accessibility

The estimated non-motorized shares are then compared to the observed non-motorized shares and plotted and the R2, RMSE, and percent RMSE are calculated for each of the purposes.

4.4 Interpolation

When comparing observed non-motorized with estimated non-motorized share, a poor match was found. **Error! Reference source not found.** compares observed and estimated motorized shares for the purpose Home-based Work, Income Group 1. Every dot shows one zone that was included in the survey. Most noteworthy is the large number of dots on the right side showing an observed motorized share of 100 percent. Some investigation revealed that the relatively small number of survey records by zone limited calculating the full picture. If a zone has only three records, and all three were motorized, the motorized share in the survey is 100 percent, even though in reality a couple of walk and bike trips that were not captured by the survey may have occurred.

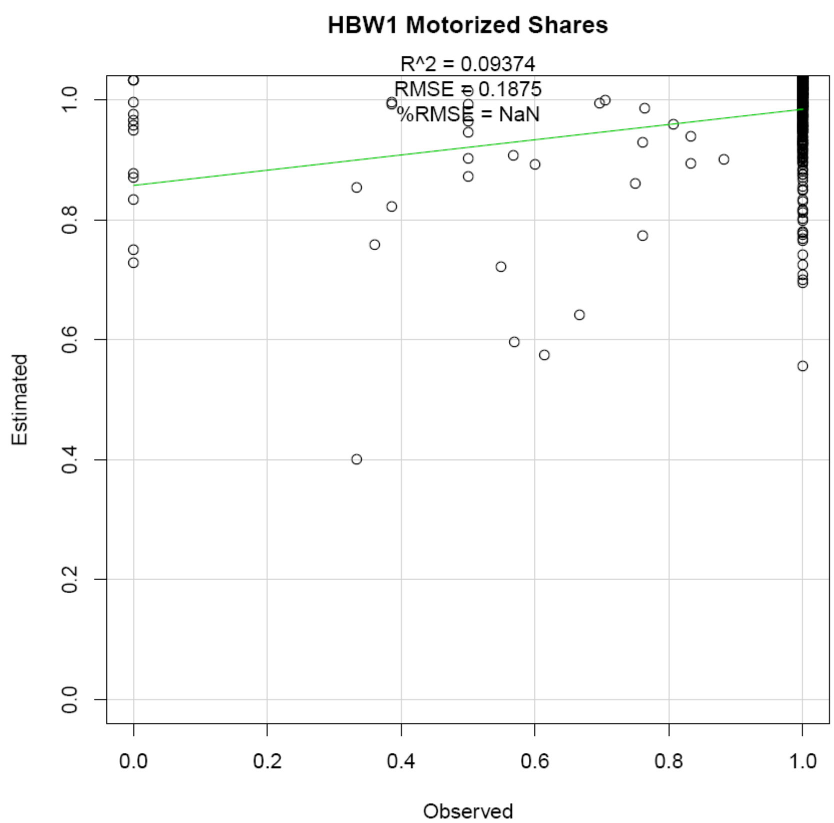


Figure 4-2: Observed and estimated motorized share for HBW1 by zone

Error! Reference source not found. shows the approximate location of survey records with blue dots for motorized and red dots for non-motorized trips for the city of Baltimore. Note that the precise location of survey records was unknown, and a GIS function was used to randomly distribute dots for every record across their zone. The color of the zones shows the percent share of motorized trips. White zones did not have any survey records. Zones with very few records are likely to either have a motorized share of 100 percent or 0 percent. The intermittent shape of the survey data does not allow calculating the motorized share at the zonal level.

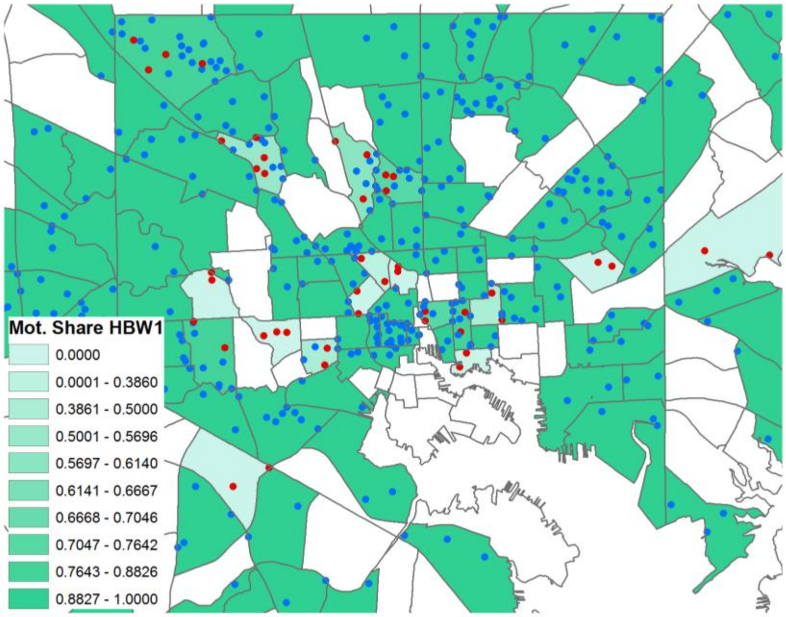


Figure 4-3: Location of motorized (blue) and non-motorized (red) HBW1 survey records

To overcome the intermittent pattern of the survey data, records were interpolated spatially to calculate a more reasonable observed motorized share. Inverse Distance Weighting was used to spatially interpolate across zones. For every zone, records of the nearest neighboring zones were taken into account until a minimum number of 500 survey records (raw record count) was reached. The motorized share of close zones (calculated using expanded survey records) was given a higher weight than the motorized share of more distant zones.

$$ms_i = \frac{\sum_j \frac{er_{mot,j}}{er_{tot,j}} \cdot d_{i,j}{}^{\beta}}{\sum_j d_{i,j}{}^{\beta}} \tag{5}$$

ms_i Motorized share in zone *i*
$er_{mot,j}$ Expanded records of motorized trips in zone *j*
$er_{tot,j}$ Expanded records of all trips in zone *j*
$d_{i,j}$ Distance from zone *i* to zone *j*
β Parameter, set to -1

Figure 4-4 shows the interpolated motorized share for zones in Baltimore. There is no boundary effect, but the motorized share smoothly changes from one zone to another.

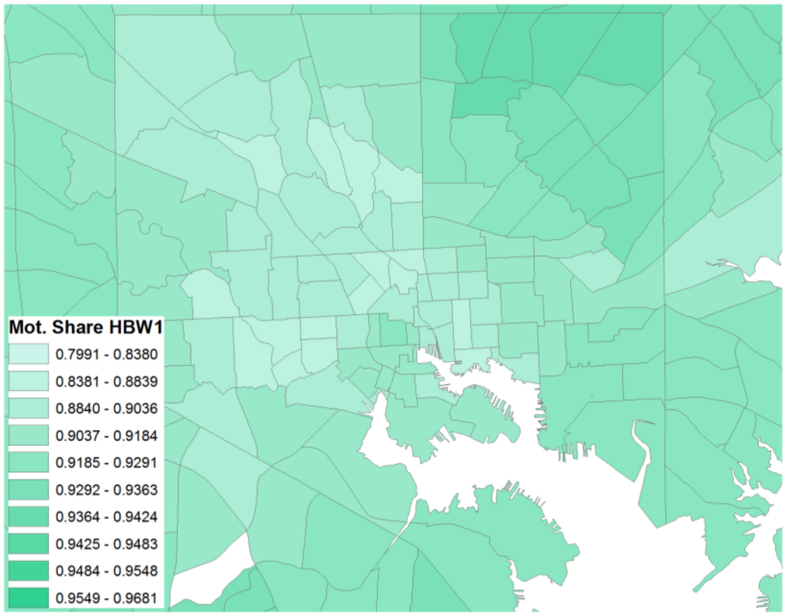


Figure 4-4: Interpolated motorized share for HBW1

4.5 Estimation Results

Using the interpolated non-motorized shares as the dependent variable and households, employment, and activity densities as well as accessibilities by auto or transit of households, retail employment, office employment, and other employment as independent variables the stepwise multiple regression script yielded to the coefficients shown in Table 4-4.

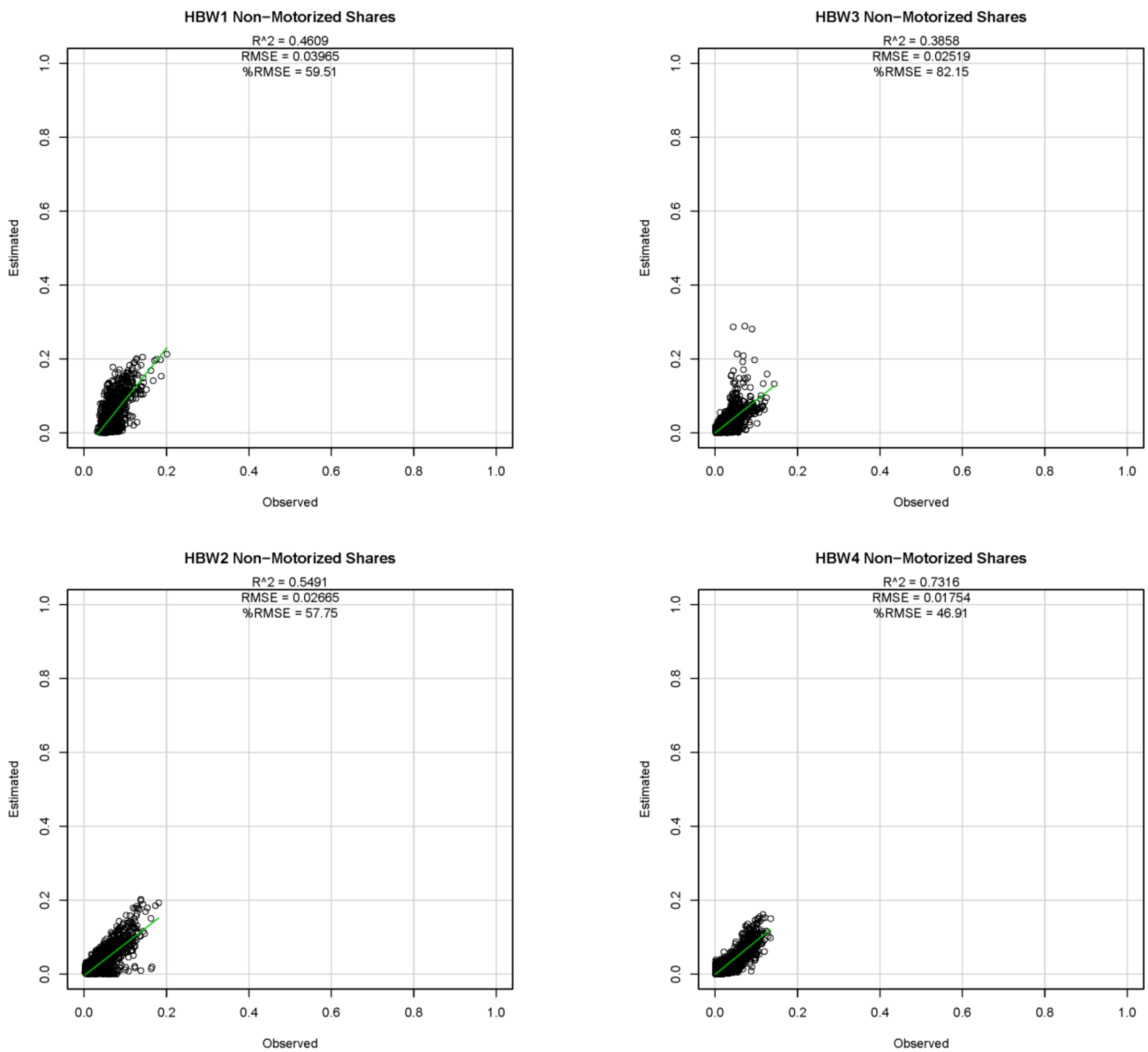
Table 4-4: Final independent variable coefficients

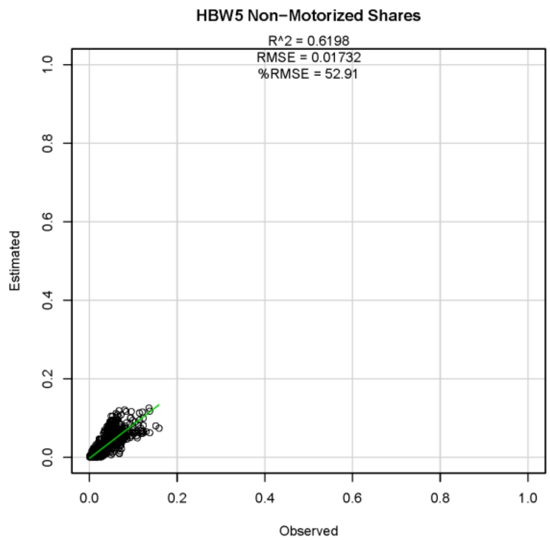
Purpose	hhDensity	actDensity	carAccHH	carAccRetailEmp	carAccOtherEmp	trnAccOtherEmp
HBO1	0.000000	0.000000	0.004424	0.000000	0.000000	0.000000
HBO2	0.000000	0.000000	0.002794	0.000645	0.000000	0.000000
HBO3	0.000000	0.000000	0.002721	0.000854	0.000000	0.000000
HBO4	0.000000	0.000000	0.002250	0.001837	0.000000	0.000000
HBO5	0.000000	0.000000	0.003816	0.000000	0.000000	0.002461
HBS1	0.000000	0.000000	0.006644	0.000000	0.000000	0.000000
HBS2	0.000000	0.000000	0.002246	0.004050	0.000000	0.000000
HBS3	0.000000	0.000000	0.000928	0.005008	0.000000	0.000000
HBS4	0.001081	0.000000	0.002040	0.002734	0.000000	0.000000
HBS5	0.000000	0.000000	0.003353	0.000000	0.000914	0.002194
HBSCH	0.000000	0.000000	0.004713	0.000000	0.000000	0.000000
HBW1	0.000000	0.000000	0.002127	0.000000	0.000000	0.000000
HBW2	0.000000	0.000000	0.001456	0.000631	0.000000	0.000000
HBW3	0.000000	0.000350	0.001030	0.000267	0.000000	0.000000
HBW4	0.000000	0.000000	0.001615	0.000000	0.000000	0.000000

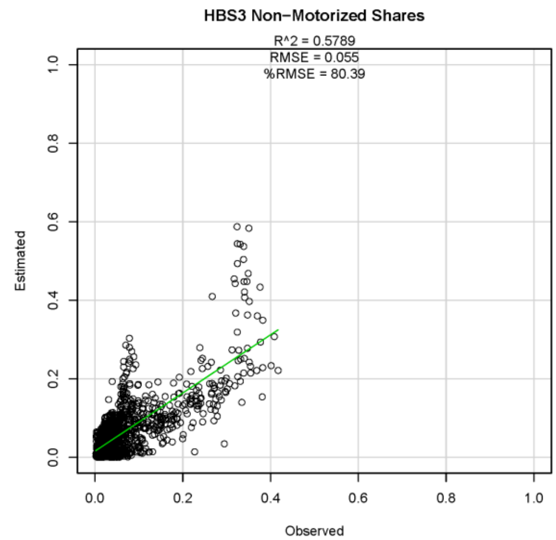
Purpose	hhDensity	actDensity	carAccHH	carAccRetailEmp	carAccOtherEmp	trnAccOtherEmp
HBW5	0.000000	0.000000	0.001254	0.000000	0.000000	0.000000
NHBO	0.000000	0.000000	0.002191	0.002258	0.001904	0.000000
NHBW	0.001161	0.000000	0.003140	0.002622	0.001532	0.000000

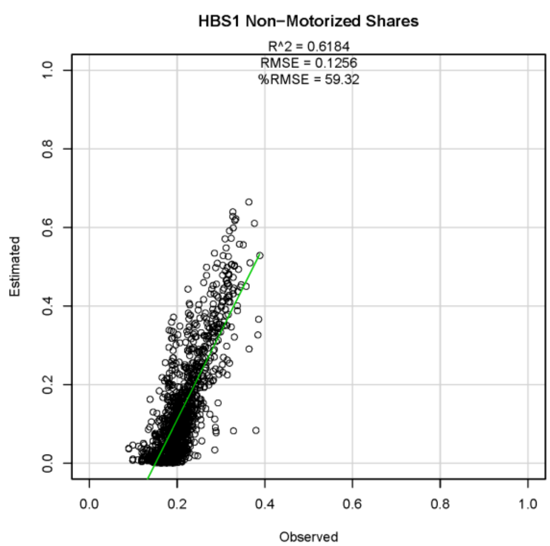
Using the above coefficients in equation 4, reasonable non-motorized shares were estimated. The following plots show a comparison of the estimated and observed non-motorized shares. The average R2 for all purposes is 0.5981.

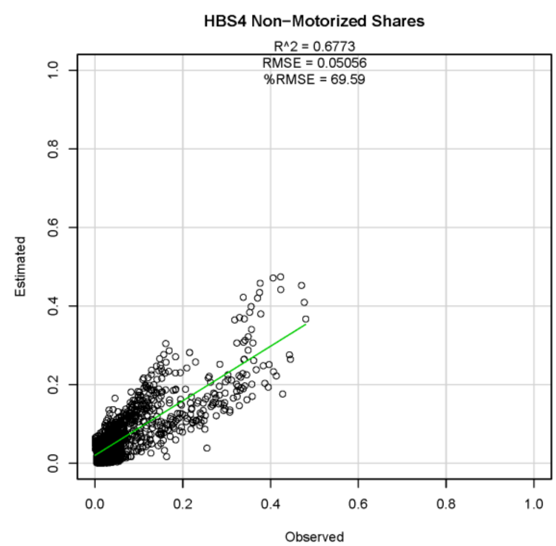
Figure 4-5: Comparison of observed and estimated shares of non-motorized trips by SMZ

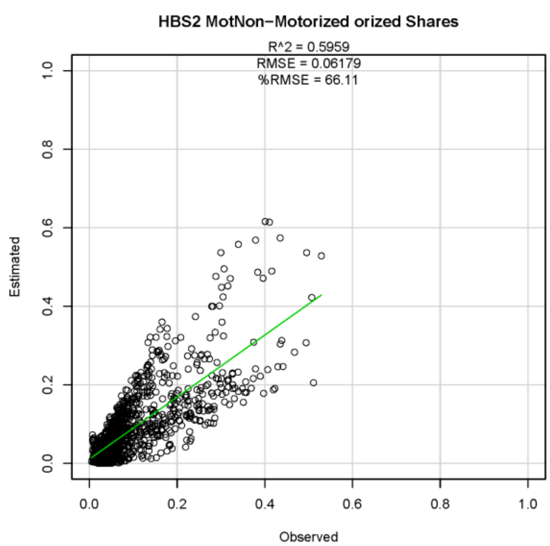


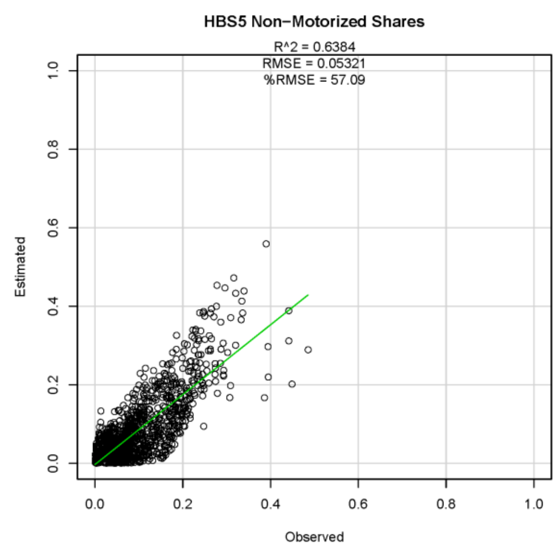

HBW5 Non−Motorized Shares
R^2 = 0.6198
RMSE = 0.01732
%RMSE = 52.91
Estimated
Observed

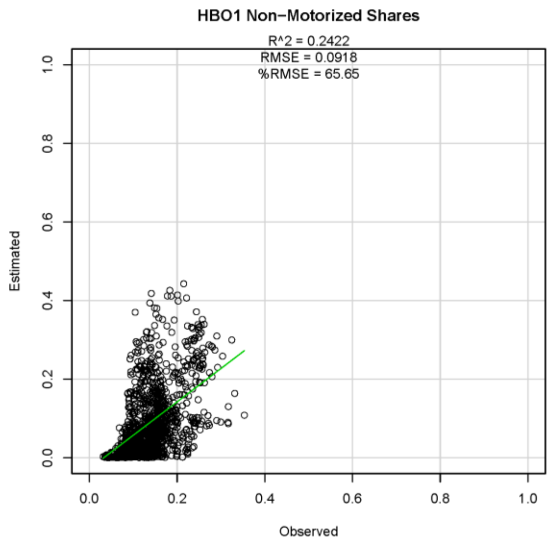

HBS3 Non−Motorized Shares
R^2 = 0.5789
RMSE = 0.055
%RMSE = 80.39
Estimated
Observed

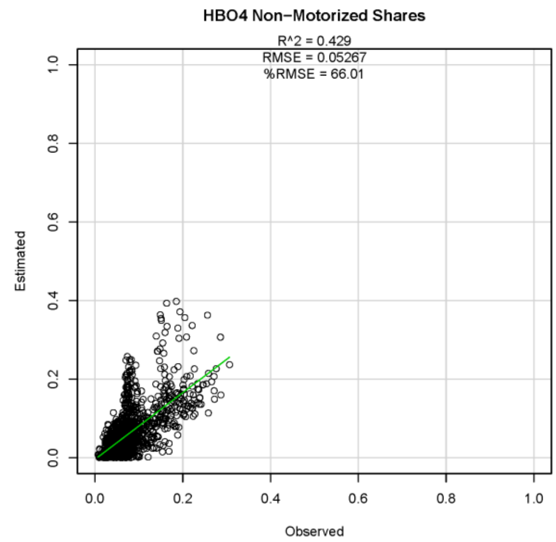

HBS1 Non−Motorized Shares
R^2 = 0.6184
RMSE = 0.1256
%RMSE = 59.32
Estimated
Observed

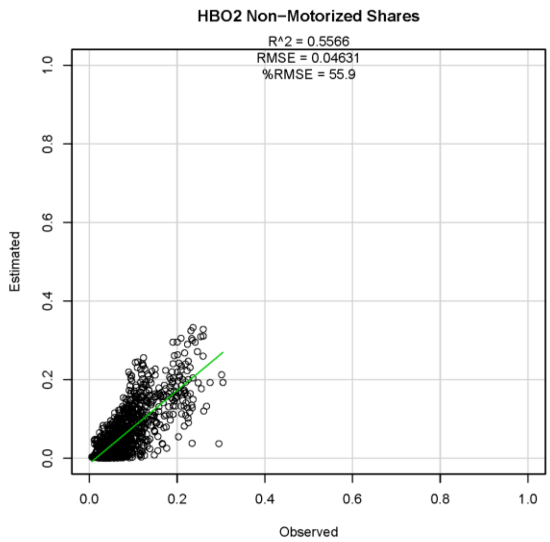

HBS4 Non−Motorized Shares
R^2 = 0.6773
RMSE = 0.05056
%RMSE = 69.59
Estimated
Observed

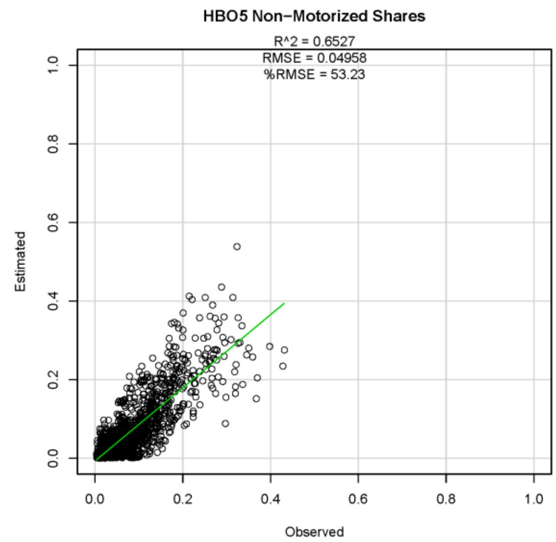

HBS2 MotNon−Motorized orized Shares
R^2 = 0.5959
RMSE = 0.06179
%RMSE = 66.11
Estimated
Observed

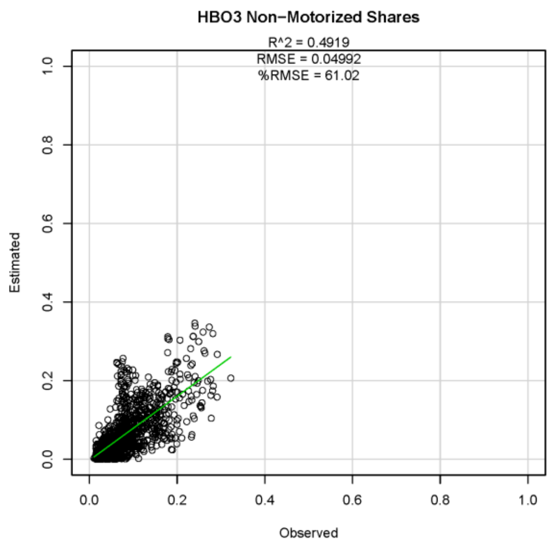

HBS5 Non−Motorized Shares
R^2 = 0.6384
RMSE = 0.05321
%RMSE = 57.09
Estimated
Observed

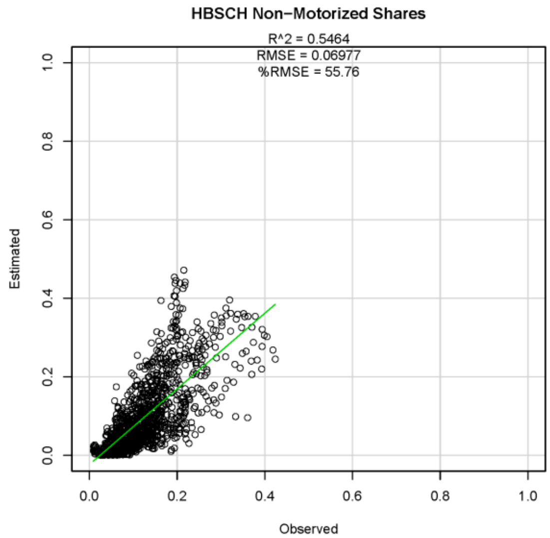

HBO1 Non−Motorized Shares
R^2 = 0.2422
RMSE = 0.0918
%RMSE = 65.65
Estimated
Observed

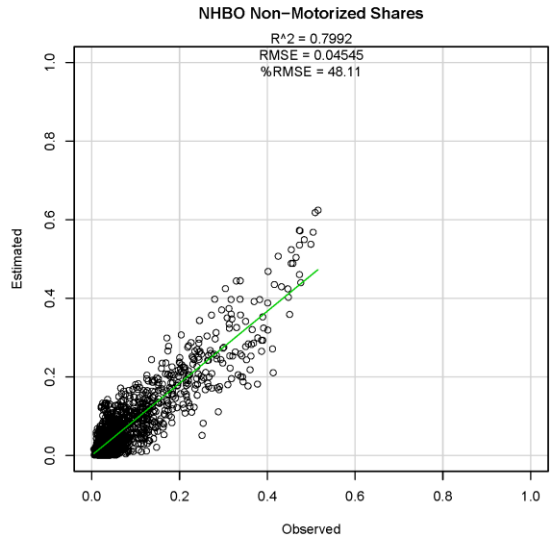

HBO4 Non−Motorized Shares
R^2 = 0.429
RMSE = 0.05267
%RMSE = 66.01
Estimated
Observed


HBO2 Non−Motorized Shares
R^2 = 0.5566
RMSE = 0.04631
%RMSE = 55.9
Estimated
Observed


HBO5 Non−Motorized Shares
R^2 = 0.6527
RMSE = 0.04958
%RMSE = 53.23
Estimated
Observed


HBO3 Non−Motorized Shares
R^2 = 0.4919
RMSE = 0.04992
%RMSE = 61.02
Estimated
Observed


HBSCH Non−Motorized Shares
R^2 = 0.5464
RMSE = 0.06977
%RMSE = 55.76
Estimated
Observed
0.0
0.2
0.4
0.6
0.8
1.0


NHBO Non−Motorized Shares
R^2 = 0.7992
RMSE = 0.04545
%RMSE = 48.11
Estimated
Observed
0.0
0.2
0.4
0.6
0.8
1.0

The variables were tested on multicollinearity. Multicollinearity analyses whether independent variables strongly correlate with each other. While multicollinearity is irrelevant when analyzing patterns, it may become problematic when parameters are used to forecast estimates (as done in MSTM). The *Variance Inflation Factor* (or VIF) was calculated for every set of independent variables used to estimate the motorized share. Values of 10 or greater are considered to be problematic.

Testing the results of the stepwise multiple regression applied here revealed only one case of multicollinearity. The purpose NHBO has a VIF value of more than 10 on the independent variable carAccOfficeEmployment. As this independent variable was found to be significant only for NHBO, this variable was eliminated from the estimation. All other variables found to be significant did not show any multicollinearity with a VIF value above 10.

Finally, non-motorized shares have to be scaled to match the average non-motorized share given by the survey. This step becomes necessary as the interpolation of observed values does not respect average non-motorized shares, but rather smoothes shares across zone. Figure 4-6 shows the estimated non-motorized share by purpose in blue, which is consistently lower than the observed non-motorized share, shown in green. The orange bars show how non-motorized shares were increased proportionally across zones to match the observed non-motorized shares. Those values are fed into the MSTM model in all future runs.

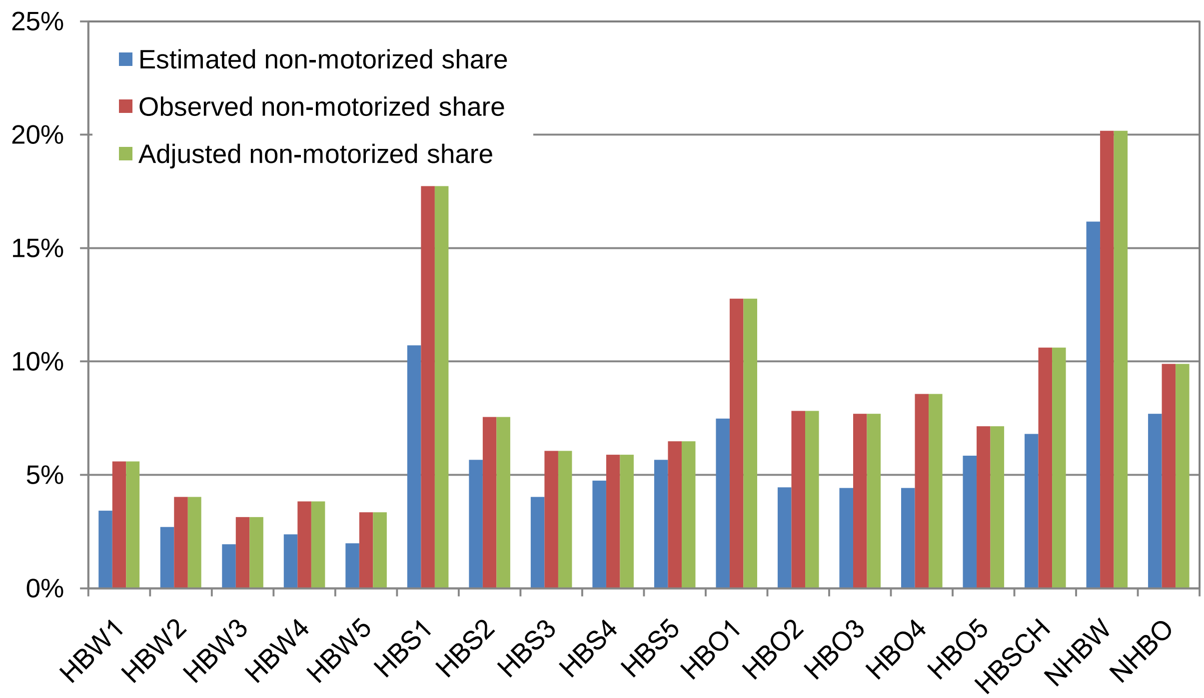


Figure 4-6: Non-motorized share by purpose

As an example, Figure 4-7 shows the estimated non-motorized share for Home-Based Work, Income Group 1 trips across the entire MSTM study area.

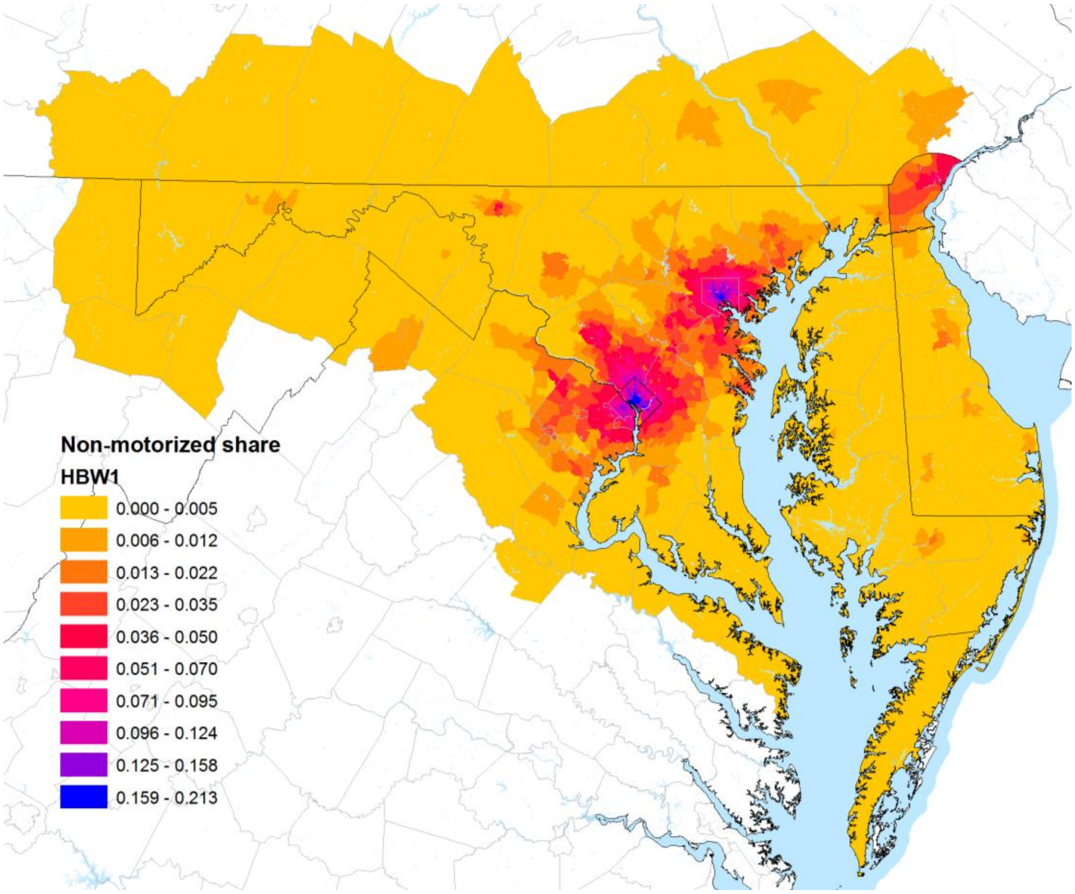


Figure 4-7: Estimated share of non-motorized trips for HBW1

5 Trip Distribution

5.1 Statewide Layer

The destination choice model predicts the probability of choosing any given zone as the trip attraction end. The model was estimated in a multinomial logit form using the ALOGIT software. These models are preceded by the trip production models, which forecast the number of productions by zone for different trip markets, chiefly identified by purpose and household income level. The destination choice models include mode choice logsums, distance terms, zonal employment, household characteristics and region geographic characteristics. The destination choice formulation is used for all purposes except for Home Based School (HBSCH), which uses a gravity formulation (see Section 4.1).

5.1.1 Estimation Dataset

The combined household travel surveys (HTS) in the MWCOG and BMC regions constitute the backbone of the estimation dataset. No travel behavior data is available for people residing outside of these two metropolitan areas. Information about trip characteristics obtained from the household survey includes trip production and attraction location, purpose, household income and auto ownership and departure time. While the surveys provide considerably more detail about trip-makers and their households, the models are limited to the attributes forecasted by the trip production models. Mode choice logsums and distance skims from the current version of the statewide model provide the trip impedance information. In addition, various terms identifying the region where the trip starts or ends were developed. These terms identify the metropolitan area (Washington DC or Baltimore) and the area type (CBD, Urban, Suburban, Other), as well as whether a bridge crossing is required.

Since there are a large number of destination alternatives, it is not possible to include all alternatives in the estimation dataset. A sampling-by-importance approach was used to choose alternatives sets for each trip. Each trip record was duplicated 10 times and different choice sets with 30 alternatives each were selected based on the size term and distance. This approach is nearly statistically equivalent to selecting 300 alternatives as the choice set of each trip, once a sampling correction term is applied in estimation.

5.1.2 Main Explanatory Variables

The following variables were examined and proved to be significant on many different purposes. By allowing for the inclusion of multi-modal accessibilities and several other region and trip market terms, the destination choice framework helps explain variation in travel across the state that was difficult to explain with a single gravity model impedance function (adopted in MSTM Phase II effort):

- Mode Choice Logsum
- Distance between the home and potential work destinations
  - Linear distance
  - Distance square root
  - Distance squared

  - Distance cubed
- Household income group interacted with distance terms:
  - Low income (less than $30,000)
  - Medium-Low income ($30,000-$60,000)
  - Medium income ($60,000-$90,000)
  - Medium-High income ($90,000-$150,000)
  - High income ($150,000 and more)
- Zero-car household interacted with distance terms (not found to be significant so not used)
- Production region interacted with distance terms:
  - Washington DC CBD
  - Washington semi-urban
  - Washington suburban
  - Baltimore CBD
  - Baltimore semi-urban
  - Baltimore suburban
- Intra-zonal indicator
- Attraction zone indicators:
  - Washington DC CBD
  - Baltimore CBD
- Employment:
  - Total employment
  - Office employment
  - Retail employment
  - Industrial employment
  - Other employment

The utility (U_{ijn}) of choosing a trip attraction destination (j) for a trip (n) produced in zone (i) is given by:

$$U_{ijn} = S_j + \alpha \times L_{ij} + \sum \beta^k \times D_{ij}^k + \sum \beta^k \times D_{ij}^k N_n^k + \sum \beta^k \times Z_j^k + C_{jn}$$

Where:

S_j is the size variable for destination zone *j*,

L_{ij} is the mode choice logsum between zone pair *ij*,

D_{ij}^k represents the various distance terms (linear, log, squared, cubed and square root),

N_n^k represent person, household or production zone characteristics for trip n and is used for creating interaction variables with distance terms,

Z_j^k represents attraction zone characteristics (other than the size term), and

C_{jn} is a correction term to compensate for the sampling error in the model estimation (i.e., it represents the difference between the sampling probability and final estimated probability for each alternative).

Appendix D explains how this correction factor is calculated.

The size variable may consist of several different terms; up to four categories of employment in addition to households. Weights (β^k) for each term in the size variable were estimated along with all other model parameters as follows, where E_j^k is employment of type k in zone j:

$$S_j = log(\sum \beta^k \times E_j^k)$$

Since the scale of the size term is arbitrary, one of the β^k coefficients is always set to 1.0. An alternative and equivalent specification of the size variable, implemented in ALOGIT is

$$S_j = log(\sum exp(\lambda^k) \times E_j^k)$$

ALOGIT reports the value of λ^k , instead of reporting directly the value of β^k . For this reason, the estimated size term coefficients may be negative; the actual coefficients are of course always positive, consistent with theory.

A combination of distance terms is used in the utility such that the composite distance utility function is monotonically decreasing. These distance terms are used to closely approximate the shape of the trip length frequency distribution. The distance-related disutility may be capped at a chosen maximum value, to maintain a reasonable probability of selecting far away destinations. The distance cap was established during model estimation at 30 miles, and may be adjusted during model calibration to ensure that the model reproduces the tail of the trip length frequency distributions. Note that even with a distance cap, the utility of a more distant zone decreases, all else equal, because of the mode choice logsum term.

Table 5-1 shows the trip length frequency for each purpose in the dataset. **Error! Reference source not found.** shows the trip length frequency in a diagram.

Table 5-1: Observed frequency of distance to chosen attraction zone

Miles	HBWork	HBShop	HBSchool	HBOther	NHBWork	NHBOther	Total
0 to 5	1,385,636	2,688,283	1,505,727	5,054,414	1,466,157	2,852,756	14,952,973
5 to 10	1,035,131	652,603	288,498	1,402,598	409,427	619,060	4,407,317
10 to 15	728,215	237,769	98,815	540,246	222,782	262,061	2,089,888
15 to 20	495,038	103,085	38,729	303,962	137,517	137,774	1,216,105
20 to 25	338,011	47,322	12,759	135,930	83,299	70,021	687,342
25 to 30	223,495	30,885	6,226	87,834	56,244	39,579	444,263
30 to 35	148,581	15,915	7,939	48,830	38,341	26,291	285,897
35 to 40	103,875	8,916	3,500	33,577	27,250	12,742	189,860
40 to 45	74,319	9,774	2,891	28,855	23,595	13,027	152,461
45 and up	127,528	18,223	5,491	48,048	30,788	21,358	251,436
Total	4,659,829	3,812,775	1,970,575	7,684,294	2,495,400	4,054,669	24,677,542

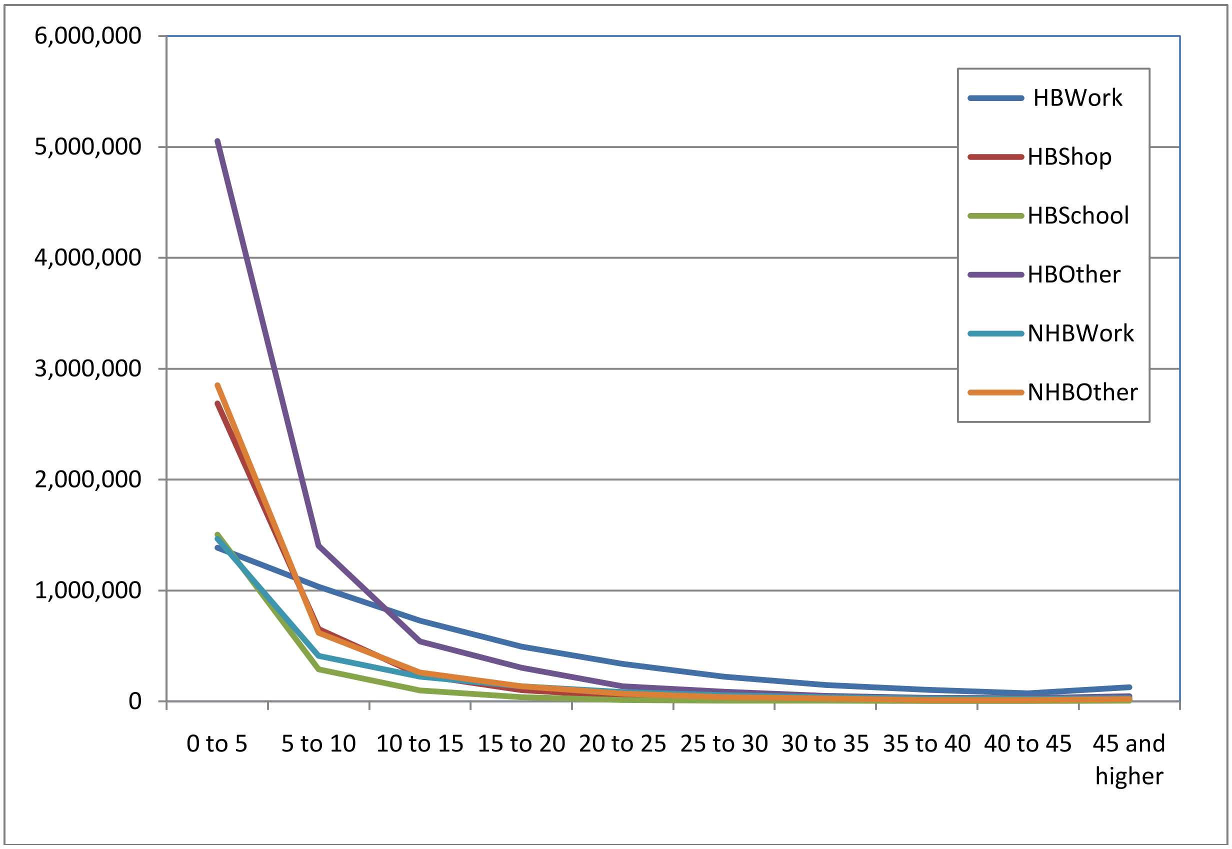


Figure 5-1: Observed trip length frequency

5.1.3 Home Based Work (HBW) Model Estimation

The first model estimated was the home-based work purpose. There are 20,626 HBW trip records in the survey file. The model specification was built incrementally, starting with a utility function that included only the mode choice logsum, and adding distance terms, size terms, and other trip attributes one at a time. Various specifications with a capped distance disutility were explored. The purpose of the cap in estimation is to reduce the influence of very long but infrequent trips on the distance polynomial coefficient estimates. The effect is similar to that of removing outliers. Note however that these trip observations were not removed from the estimation. The distance cap also helps to obtain a monotonically decreasing utility with respect to distance over the entire trip distance range comprised by the model area. The cap was set high enough to include a large majority of the trip records. Approximately 90% of the HBW trip observations exhibit a distance shorter than 30 miles. The final HBW model was estimated with a 25 mile distance cap. Estimation runs also tested trip length differences among socio-economic variables and home residential location, and the attractiveness of the two CBD areas.

Model Estimation Findings:

- The **mode choice logsum** coefficient is 0.58, consistent with theory and with the expectation of relatively elastic demand.
- The coefficients for the distance polynomial are all significant, and the combined total utility with respect to distance decreases monotonically with distance, as expected.

- **Household income** was interacted with the linear distance term. Following the trip generation segmentation, trips are stratified into five household income groups. The income coefficients are expressed relative to the lowest income (<$30K), which was given an income coefficient of zero. The coefficients on the other income categories were positive and significant, with a steadily increasing coefficient value as the income level increases. This result shows that persons from higher income households are likely to make longer work trips.
- The marginal utility of an **intrazonal trip** was captured with an indicator variable. The intrazonal coefficient is positive and significant, indicating a preference for destinations in the home zone, all other things equal.
- The effect of auto **ownership** on destination choice was examined by interacting a zero-car household indicator with distance, under the hypothesis that these households would travel shorter distances, on average, than other households. The estimated coefficient showed the opposite effect, possibly because it is highly correlated with low income households. No auto ownership effects were kept in the final estimated model.
- **CBD indicator** variables were used to explore the attractiveness of a destination in either the Baltimore or Washington DC CBDs. Both CBD variables exhibited negative coefficients. On its own, this result is unintuitive because the CBDs are major attractors, which would lead one to expect a positive coefficient. However the attractiveness of the CBDs may already be captured in the mode choice logsums or size terms. The CBD variables were dropped from the final model.
- Due to significant differences in trip lengths between the Baltimore and Washington regions observed in the household survey for the HBW purpose (**Error! Reference source not found.**), **region-specific indicator variables** were interacted with distance. The estimated production region coefficients were significant and exhibited small, negative magnitude. The negative sign indicates a preference for shorter distances in the non-rural locations. Ideally the underlying variables leading to these regional differences should be used instead of these geographic specific variables, which are somewhat akin to k-factors. However, these variables are included with the goal that they be further explored as part of future model improvement efforts.
- The effect that the Potomac River has on discouraging trips across was examined with a **bridge crossing** variable (**Error! Reference source not found.**, at least one crossing is assumed when the origin and destination regions differ). As expected, its coefficient is negative and significant, indicating that bridge crossings are not desirable. For HBW trips, a bridge crossing is equivalent to 12 additional minutes of travel time.
- **The size term** comprises retail, office, industrial and other services employment. All employment categories exhibited significant coefficients.

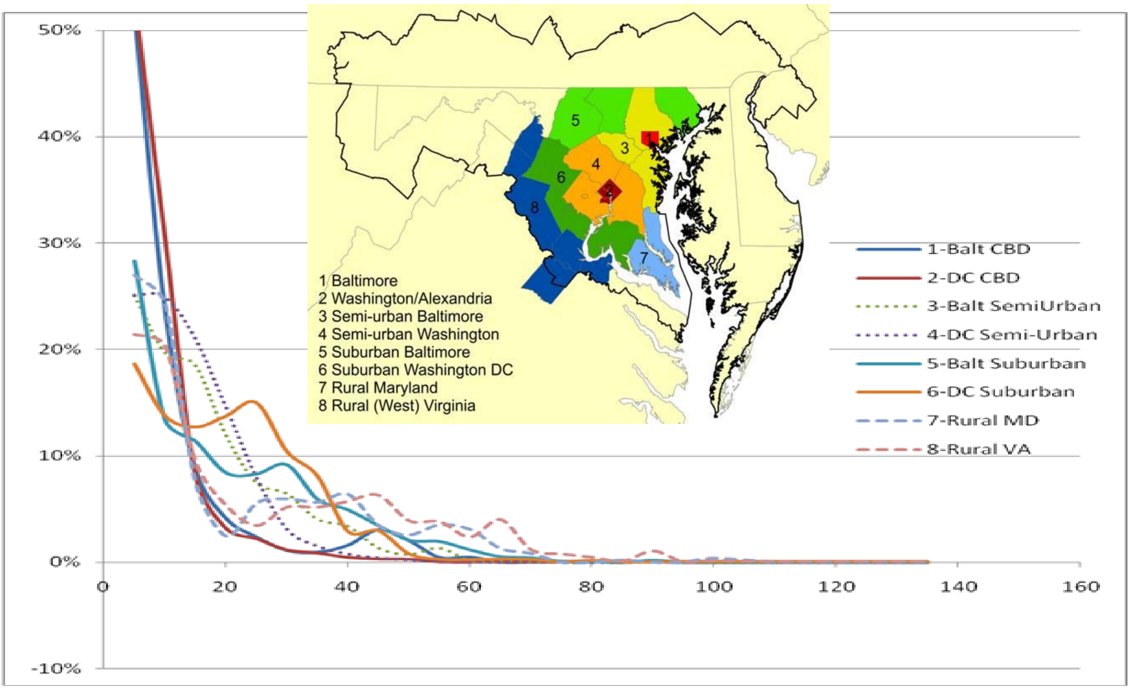


Figure 5-2: HBW observed trip length frequency variation by region

Note: HTS regions defined as: 1=1-110; 2=1188-1307, 3=111-405, 525-599, 1509-1543, 1634-1650, 1684-1697; 4=609-943, 1308-1355, 5=406-524,944-966, 1615-1633, 1651-1674, 6=992-1009, 1356-1397, 7=967-991, 1093-1178, 1443-1460, 1544-1605, 8=1019-1083, 1398-1442, 1470-1499.

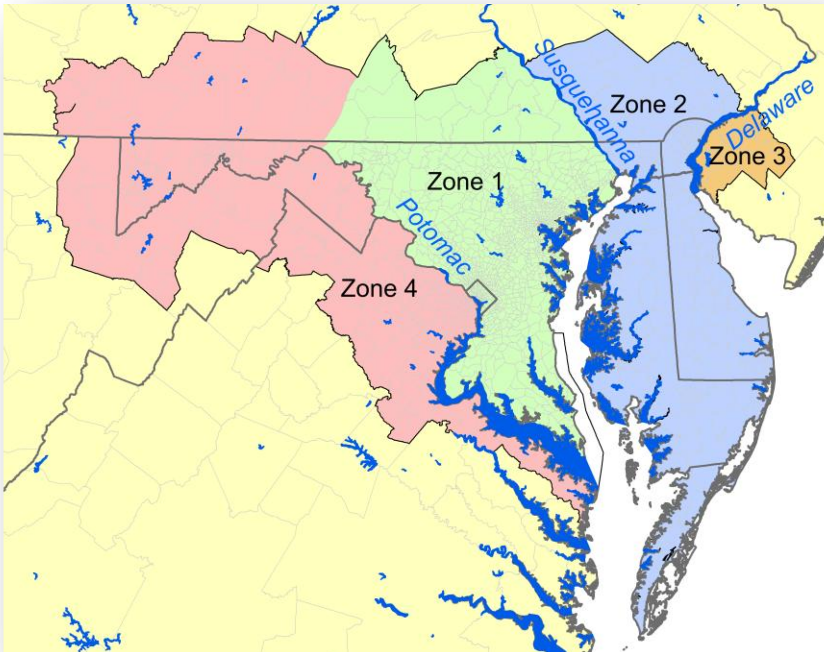


Figure 5-3: River crossing regions

Note: River Crossing regions defined as: 1 = 1-599, 609-1009,1054-1083,1188-1271,1625-1633,1667-1674, 2=1093-1178,1450-1460,1509-1605,1658-1666, 3 = 1684-1697 , 4 = 1019-1053,1281-1428,1438-1449,1470-1499,1615-1624,1651-1657, 5 = 600-608,1010-1018,1084-1092,1179-1187,1272-1280,1429-1437,1461-1469,1500-1508,1606-1614,1634-1650,1675-1683,1698-1832.

5.1.4 Home Based Shop (HBS) Model Estimation

The sample size for home based shop trips is 3,812,775 observations. The best model estimated for HBW trips was used as the starting point for HBS, without the inclusion of the regions interacted with distance. The disutility of distance was capped at 30 miles.

Model Estimation Findings:

- The **mode choice logsum** coefficient was consistently estimated at a value greater than 1.0, which is outside the theoretically acceptable range. The coefficient was therefore constrained to a value of 0.8.
- The **distance**, **distance cubed**, and **log of distance** coefficients were all negative and significant. The **distance squared** term was positive and significant. Combined, the total disutility with respect to distance decreases monotonically.
- The **household income** coefficients were positive and significant, but did not steadily increase with higher incomes. The two highest income categories were combined into one to obtain a monotonic progression.
- The **intrazonal coefficient** was negative and became insignificant when the logsum coefficient was constrained. Therefore it was dropped from the final run.
- The **CBD indicator** variables for Washington DC and Baltimore were negative. Thus, as was the case for HBW, these variables were excluded from the final model.

- The **bridge crossing** coefficient was negative and significant.
- Retail was the only employment category used for the HBS **size term** for HBS.

5.1.5 Home Based Other (HBO) Model Estimation

The sample size for home based other trips is 7,684,294. The best model estimated for the HBW trips was used as the starting point, without the inclusion of the regions interacted by distance. The disutility of distance was capped at 30 miles.

Model Estimation Findings:

- The **mode choice logsum** coefficient was estimated at a value of 0.8, which is a reasonable result.
- The **distance**, **distance cubed**, and **log of distance** coefficients were all negative and significant. The **distance squared** term was positive and significant. The total disutility of distance decreases monotonically with distance, as expected.
- The **household income** coefficients were positive and significant, but did not steadily increase with higher incomes over the five income groups. Therefore, income was collapsed into three categories: less than $30K, $30-60K and $60K or higher.
- The **intrazonal** coefficient was positive and significant.
- The **CBD indicator** coefficients for Washington DC and Baltimore were negative and therefore dropped from the final model.
- The **bridge crossing** coefficient was negative and significant.
- The **size term** consists of number of households, retail employment, office employment, and other employment.

5.1.6 Non-Home Based Work (NHB) Model Estimation

The sample size for the non-home based work purpose was 2,495,400 observations. The best model estimated for HBW trips was used as the starting point, excluding the region variables. The disutility of distance was capped at 30 miles.

Model Estimation Findings:

- The **mode choice logsum** coefficient estimated is approximately 0.9, which is a reasonable result.
- Th**e distance** polynomial included **log of distance**, **distance squared** and **distance cubed**. All exhibited significant coefficients, and a combined distance decay function that decreases with distance.
- The **income** coefficients were not used in this model because this purpose is not stratified by income.
- The **intrazonal** coefficient was positive and significant.
- The **CBD indicator** coefficients for Washington DC and Baltimore were negative and therefore excluded from the final model.
- The **bridge** coefficient was negative and significant.
- The **size term** consists of retail employment, office employment, other employment and households.

5.1.7 Non-Home Based Other (OBO) Model Estimation

The sample size for the non-home based other purpose is 4,054,669 observations. The best model estimated for HBW trips was used as the starting point, excluding the regions variables. The disutility of distance was capped at 30 miles.

Model Estimation Findings:

- The **mode choice logsum** coefficient estimate was greater than 1; therefore it was constrained to 0.8.
- The **distance**, **distance cubed**, and **log of distance** coefficients were negative and significant. The **distance squared** term had a positive and significant coefficient.
- The **income** coefficients were not used in this model because this purpose is not stratified by income.
- The **intrazonal** coefficient was positive and significant.
- The **CBD indicator** coefficients for Washington DC and Baltimore were negative, and for this reason excluded from the final model.
- The **bridgecrossing** coefficient was negative and significant.
- The **size term** consists of retail employment, industry employment, other employment, and households.

5.1.8 Model Calibration

The destination choice model was calibrated to reproduce the trip length frequency distributions from the HTS, regional flows from HTS, and regional flows from the Census Transportation Planning Package (CTPP). Calibration statistics were limited to the model area represented in the household survey, as no data were available for the rest of the model area. The CTPP worker flow comparisons did cover the entire model area. Model calibration consisted of making small incremental adjustments to the estimated coefficients in order to better match observed trip patterns. A key focus was on the segmented distance term to match the short distance portion of the observed trip length frequency curve.

Calibrated model trip length frequency distributions (limited to trips originating in the HTS surveyed region) are compared to HTS survey in the next set of figures. Average trip lengths by purpose comparisons are shown in Table 5-2.

Table 5-2: Observed and estimated average trip distance in miles

Purpose	HTS	Model
HBW	12.6	12.8
HBS	5.2	4.9
HBO	5.9	6.7
NHB	7.4	6.9
OBO	5.3	5.5

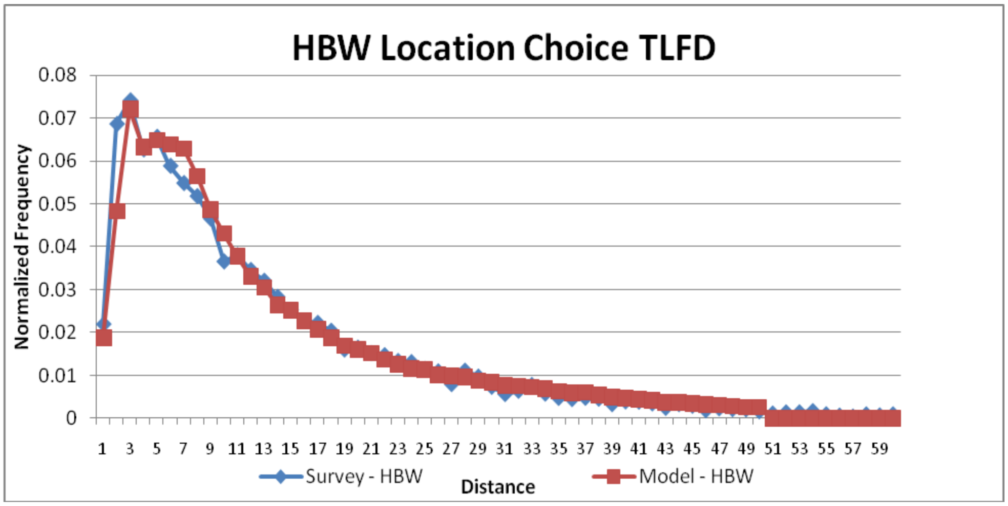
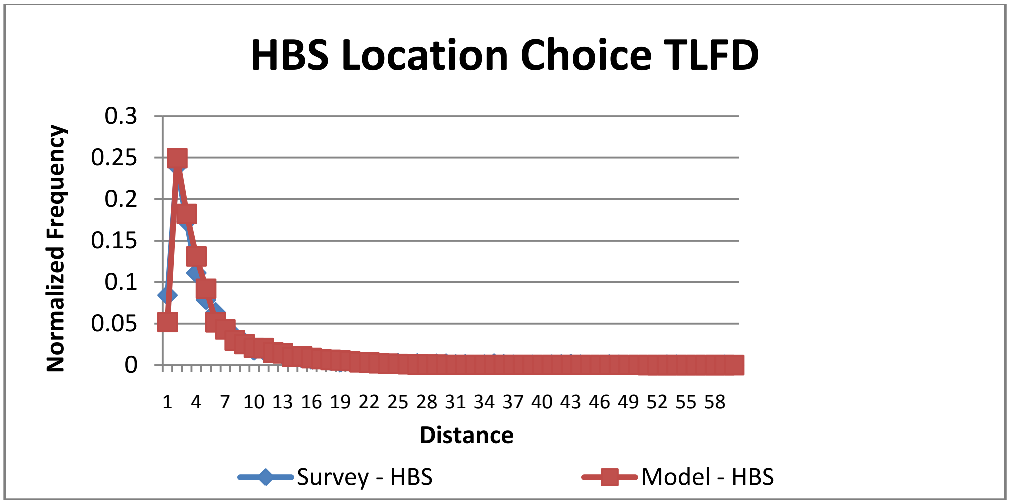
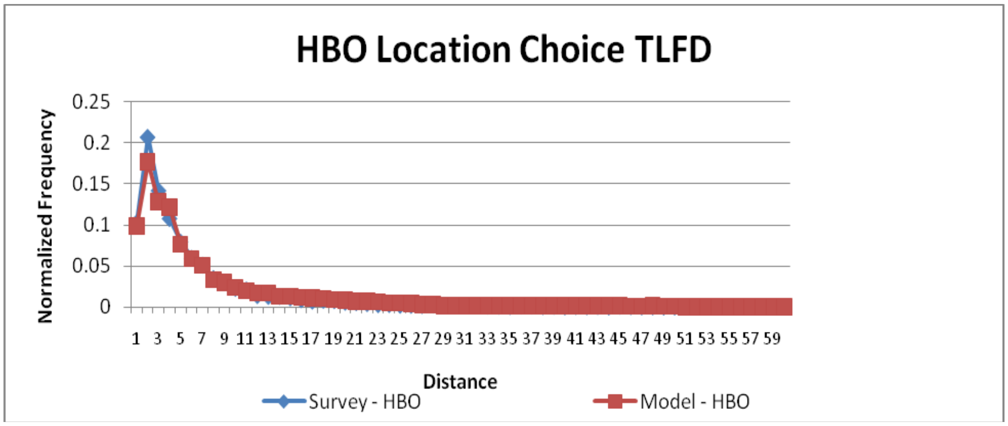
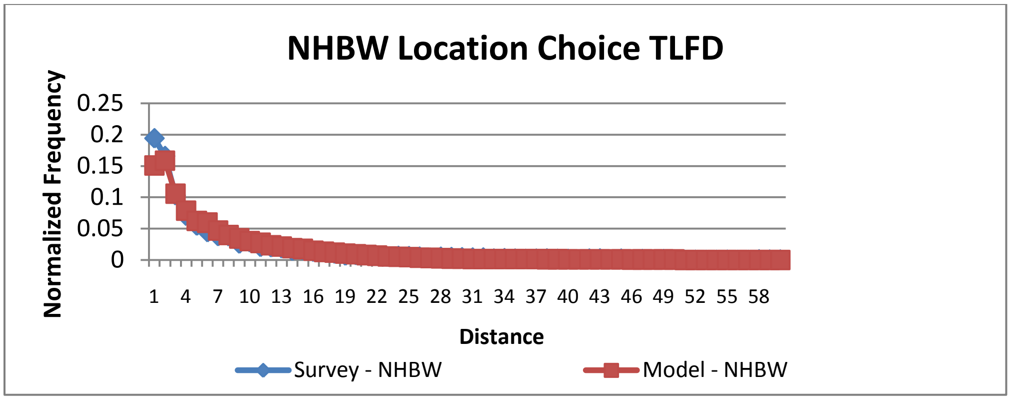
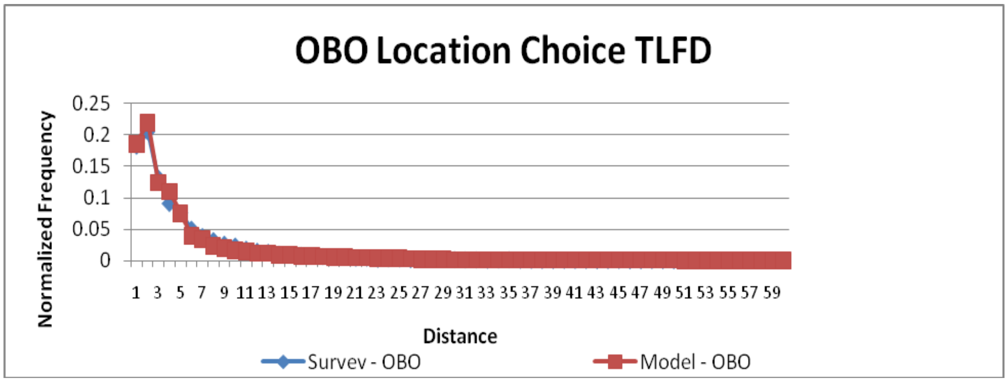


Figure 5-4: Trip length frequency distributions by purpose (HTS region)









The calibrated coefficients implemented in the destination choice model are shown in Table 5-3. The size terms estimated by the destination choice process replaced the size terms calculated with the initial regression analysis.

Table 5-3: Calibrated coefficients for destination choice models

	Trip Purpose				
Explanatory Variable	**HBW**	**HBS**	**HBO**	**NHBW**	**OBO**
Mode choice logsum	0.5769	0.8000	0.8420	0.9078	0.8000
Distance	-0.4383	-0.3986	-0.5788	0.0978	-0.2241
Distance Squared	0.0137	0.0166	0.0261	-0.0032	0.0106
Distance Cubed	-0.0002	-0.0004	-0.0005		-0.0002
Log of Distance	0.7066	-0.9034	-0.4212	-1.5665	-1.0944
Income X Distance interactions					
Income (<30K)					
Income (30-60K)	0.0176	0.0162	0.0345		
Income (60-100K)	0.0470	0.0255	0.0357		
Income (100-150K)	0.0606	0.0263	0.0357		
Income (150K+)	0.0697	0.0263	0.0357		
Intrazonal indicator variable	1.2038		0.6633	0.7228	0.6311
Baltimore CBD indicator					
Bridge Crossing indicator	-0.3013	-1.2928	-0.8054	-0.5280	-0.9768

Explanatory Variable	Trip Purpose HBW	HBS	HBO	NHBW	OBO
Production Region X Distance interactions					
Baltimore CBD (Region 1)	-0.0362				
Washington DC CBD (Region 2)	-0.0882				
Baltimore Semi-Urban (Region 3)	-0.0269				
Wash.DC Semi-Urban (Region 4)	-0.0422				
Baltimore Suburban (Region 5)	-0.0350				
Wash. DC Suburban (Region 6)	-0.0255				
SE Maryland and Halo	-0.0255	-0.0100	-0.0100	-0.0100	-0.0100
SW Maryland and Halo	-0.0350	-0.0100	-0.0100	-0.0100	-0.0100
Size Term (exponentiated)					
Other Employment	1.0000		0.3052	0.4271	0.1470
Retail Employment	1.0134	1.0000	0.1878	1.0000	1.0000
Office Employment	0.2904		0.0446	0.4992	
Industrial Employment	0.3585				0.0874
Households			1.0000	0.2825	0.3243
Distance Cap	25	30	30	30	30
Distance Constants					
0-1mile	0.7729	1.7660	1.4007	0.2417	2.2193
1-2 miles	0.0000	1.9110	0.5347	0.0140	1.1874
2-3miles	-0.1059	1.2765	0.1937	-0.0396	0.6676
3-4 miles	-0.3221	0.8224	0.1937	-0.0396	0.6676
4-5 miles	-0.1424	0.7539	0.1937	-0.0396	0.6676
5-6 miles	-0.1424	0.2023	0.0000	0.0000	0.0000
6-7 miles	-0.1000	0.0721	0.0000	0.0000	0.0000

A gravity formulation, similar to the formulation used in the BMC and MWCOG models, was chosen for HBSC trips in lieu of destination choice. The MSTM HBSC trip distribution model differs from the BMC and MWCOG models in the following ways:

- Friction factor functions incorporate segmented distance terms to facilitate calibration to target trip length frequency distributions over the longer distances encompassed by the statewide model, and
- Gamma function is used to calculate interzonal impedances.

The basic gravity model formulation is:

$$T^k_{ij} = P^k_i * A^k_j * F^k_{ij} / \sum_j (A^k_j * F^k_{ij})$$

Where:
T^k_{ij} = trips for purpose ‘k’ between production SMZ ‘i’ and attraction SMZ ‘j’
P^k_i = productions for trip purpose ‘k’ in SMZ ‘i’
A^k_j = attractions for trip purpose ‘k’ in SMZ ‘j’

F^{k}_{ij} = friction factor for trip purpose ‘k’ between SMZ ‘i’ and ‘j’

Friction factors take the following form:

$$Fkij = \alpha * CT^{\beta} * \exp(\gamma * CT)$$

Where:
CT^{k}_{ij} = Composite Time for purpose ‘k’ between SMZ ‘i’ and ‘j’ defined as follows:

$$CT = \frac{1}{\frac{1}{HT} + \frac{x}{TT}} + \frac{y * TL}{vot}$$

where:
CT = composite time, minutes
HT = highway time, minutes (including terminal time)
TT = total transit time, minutes (best walk-access path)
TL = highway toll, cents
vot = value of time, cents/minute
x, y = coefficients (vary by income and purpose)
□, β and γ = calibrated coefficients

The HBSC gravity model parameters are given in Table 5-4. The β and γ parameters were calibrated for MSTM while the other parameters are consistent with the BMC model. Effectively, since *x* and *y* are zero, the composite time impedance is simply highway time.

Table 5-4: School purpose trip generation gravity model parameters

Parameter	Value	Comment
α	10,000,000	
vot	45.2 cents/minute	Peak period, value per BMC
x	0	Peak period, value per BMC
y	0	Peak period, value per BMC
β	0.1	Calibrated
γ	-0.36	Calibrated
Adj	1.546	Accounts for average trip length difference between skims and travel times reported in the survey

The combination of greater variance in trip rates by area, income market segmentation and inclusion of segmented distance terms in the trip distribution impedance should reduce the need for trip distribution adjustment factors.

The initial phase of model development focused on getting the model implemented and meeting first order calibration targets. Trip distribution calibration efforts focused on meeting trip length frequency and average trip length targets by sub-region. Origin-destination (OD) distribution model adjustments will be implemented in a subsequent phase in concert with refinements to the model structure, which may include adoption of a destination choice structure rather than a gravity model formulation. If the gravity model formulation is retained and OD adjustment factors implemented, they will be implemented in the form of an additional term in generalized impedance rather than multiplicative factors on unadjusted trips as is the convention with K factors.

Calibration targets for the HBSC trip distribution model included average trip length and trip length frequency distribution. The calibration targets were initially developed in the conventional way, using model estimated travel times (skims). It was found that the skim travel times resulted in significantly shorter distances than reported in the survey. Apparently, the skim matrix tended to underestimate congestion, and therefore suggested travel times that were shorter than found in reality. In other words, in the same amount of time, people can travel further in the model than they can in reality.

The survey reports travel time in minutes for each trip. Reported travel times, however, tend to be clustered around (rounded) five-minute intervals, and people tend to underestimate how long it takes them to reach their destination. This is particularly true for discretionary, non-daily travel. To improve the data quality of travel times, it is common in travel demand modeling not to use the reported travel time, but rather reading the travel time from a skim matrix developed for the model. This way, consistent travel times between reported origins and destinations were developed.

To overcome this mismatch, the travel times read from the skim matrixes were scaled to match the average trip length reported in the survey. This procedure allowed using the distribution of travel times according to the skim matrix while reaching the average travel distances as reported in the survey.

Table 5-5 compares average trip length in miles derived from the skim matrix with the average trip distance reported in the survey. Consistently for each purpose, the skim travel times are smaller. A scaling factor has been calculated by dividing the survey distance by the skim distance. For trip distribution, skim travel times are divided by this factor to calculate more reasonable travel times.

Table 5-5: Trip distribution scaling

Purpose	Average Skim	Average Survey	Factor
HBSCHOOL	14.9	23.0	1.546

5.2 Model Validation

To validate the destination choice model, the trip length frequency distribution was calibrated to match average trip length reported in the household travel survey. **Error! Reference source not found.** compares observed trip length with simulated trip length. Trip lengths are slightly overestimated, but overall the patterns of the survey are reflected in the model. This is a calibration result (and not validation) as there is no independent dataset to validate against

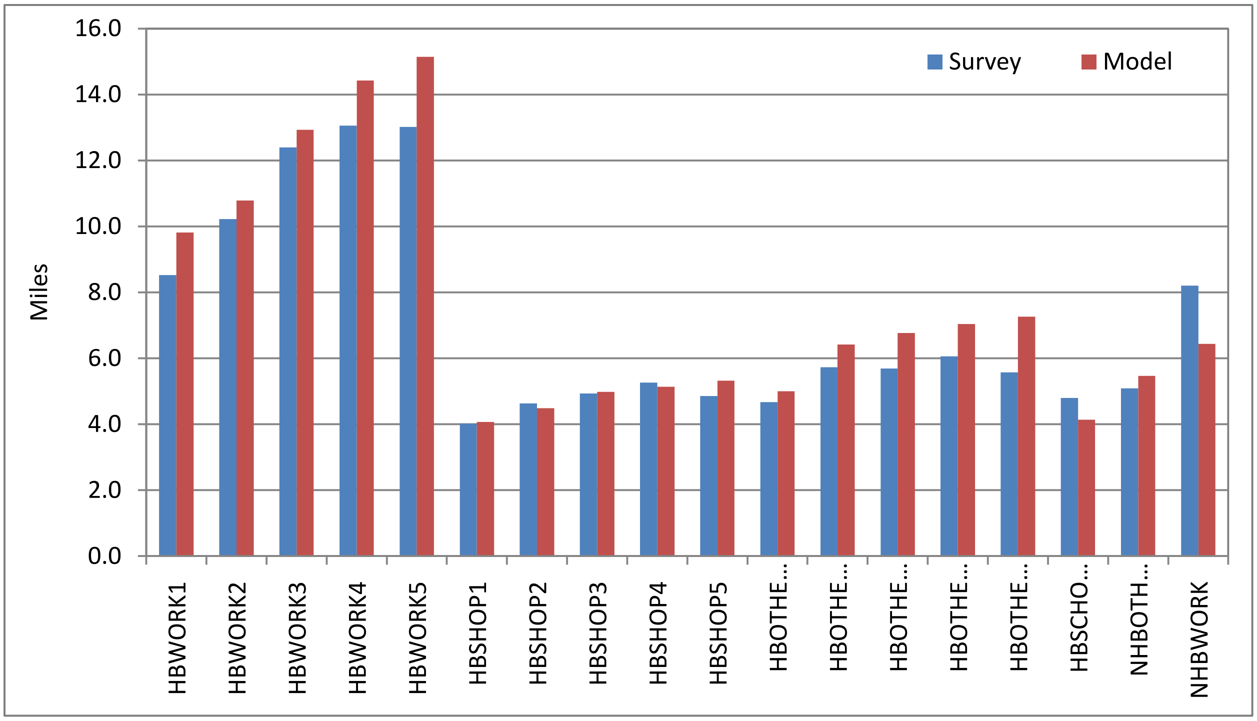


Figure 5-5: Comparison of average trip length in survey and model results for autos

6 Mode Choice Model

6.1 Statewide Layer

Person trip mode choice is an adaptation of the most recent BMC nested logit mode choice model, shown in Figure 6-1[12]. The modes defined in Section 4.2, Consolidated Network Development, were aggregated into these nests. The figure indicates the modes and sub-modes that are incorporated in the model. Rail includes LRT and Metro and the Commuter Rail (CR) includes AMTRAK services as well as MARC commuter rail. All local bus services are included under the Bus and express bus and commuter bus services are included in the ExpBus modes.

```
                          Person
          ________________|_______________
         |                                |
       Auto                           Transit(b)
     ___|___                 ______________|_________________
    |       |               |                                |
  SR(b)    DA              Walk                           PnR(b)
  __|__               ______|______  ______         ______|_____  _______
 |     |             |      |       |      |       |      |      |       |
SR2  SR3+(b)        Bus   Rail(b)  CR(b)  Exp(b)  Bus   Rail(b)  CR(b)  Exp(b)

(b) = carries bias coef
```

Figure 6-1: Structure of MSTM mode choice model

Mode choice is based on generalized utility functions for auto and transit travel. Separate utilities were developed to represent peak and off-peak conditions. Home-based work trips and Non-home based work trips are based on peak period travel characteristics while other purposes are based on off-peak characteristics. Auto utilities for each auto mode include driving time and cost, terminal time and parking costs at the attraction end, and tolls. Transit utilities for each transit mode include walk and drive-access times, initial wait time, in-vehicle time, and transfer time. Bias constants or mode specific constants are included as indicated in Table 6-4 and Table 6-5 below which list all the variables included in the utility expression for each mode and sub-mode.

These variables are described in the BMC Calibration Report as follows. All monetary units were based on year 2000 dollars:

- **In-Vehicle Time** (IVT) (minutes): Run time from the network. This is Single Occupancy Vehicle (SOV) path time for Drive Alone (DA), High Occupancy Vehicle (HOV) path time plus carpool access time for Shared Ride 2 and 3 (SR2 and SR3) (which accounts for additional circulation and pick-up time for carpools). For SR2, access time is defined as the minimum of either 10 minutes or 12% of the in-vehicle time (MIN(0.120*IVT,10)); for SR3, it is the minimum of 15 minutes or 19.9% of the in-vehicle time (MIN(0.199*IVT,15)). Those functions were adopted from the old BMC

[12] This section draws heavily from the BMC Calibration Report: "Travel Forecasting Model Calibration Report," prepared for Maryland Transit Administration by William G Allen Jr., 21 August 2006. Some or all of the modifications made by Parsons Brinckerhoff for Baltimore Region new-starts analyses were incorporated also depending on review of results and experience gained in that work.

model. For Transit, if the run time for each submode does not use that submode, the path is considered invalid and the submode is considered unavailable. Commuter rail run time is factored by 0.75, to reflect the fact that such trips tend to be longer and the riding experience is generally more pleasant than on other types of transit (more seating room, more amenities on-board, etc.).

- **Terminal Time** (minutes): Sum of the times for the production and attraction zones. Computed from a look-up table based on the zonal area types (see section1.4). For SR2, add 1.1 minutes to reflect additional waiting time; for SR3, add 2.5 min.

- **Auto Operating Cost** (cents): Incremental cost of driving (i.e., excludes all fixed costs of vehicle ownership). Computed as distance from the network times: 9.9 cents/mile in year 2000 dollars. About 58% of that cost (5.76 cents/mi) is fuel; the rest (4.14 cents/mi) is maintenance, tires, and oil. The fuel component was calculated using a cost of $1.314/gallon (year 2000 dollars) and an average on-road fuel efficiency of 22.8 mpg. For SR2, divide by 2. For SR3, divide by the average 3+ occupancy by purpose (derived from the Baltimore home interview survey).

- **Auto Tolls** (cents): Toll cost from the network. For SR2, divide by 2. For SR3, divide by the average 3+ occupancy by purpose.

- **Auto Parking Cost** (cents): Computed by the parking cost model for the attraction zone. For SR2, divide by 2. For SR3, divide by the average 3+ occupancy by purpose.

- **Transit Walk Time** (minutes): Sum of transit transfer walk time, from the network, plus computed production zone access to transit time, plus computed attraction zone egress from transit time. Access and egress times are multiplied by adjustment factors to reflect the difficulty or ease of walking.

- **Initial Wait Time** (7.5 min or less, in minutes): Initial wait time is the time spent waiting for the first transit vehicle, from the network. This is the amount of the initial wait time that is equal to or less than 7.5 minutes. Several urban areas have found that the first increment of wait time is more important to mode choice than the second increment. This also helps the modeling of routes with very long headways (e.g., 60+ minutes). TP+, as with most such software packages, computes the wait time as half the headway, but that does not reflect the fact that people tend to schedule their arrivals for long-headway routes, leading to shorter actual wait times than half the headway.

- **Initial Wait Time** (over 7.5 minutes, in minutes): This is the increment of initial wait time that exceeds 7.5 minutes, if any.

- **Transfer Time** (minutes): This is the time spent waiting for the second (and any subsequent) transit vehicles, from the network.

- **Number of Transfers**: In TP+, this is computed from the network as the total number of transit routes boarded, minus one.

- **Transit Fare** (cents): Computed from the network as the sum of the boarding fare and any transfer fares. For drive-access, it also includes the cost of driving to the Park and Ride (PnR) lot, computed as the drive-access distance times: 9.9 cents/mile.

- **Drive-Access Time** (minutes): The time spent driving to a transit PnR lot or station, computed from the network using over-the-road distance and speed.

Table 6-1: Variables included in utility expressions

	Mode								
Variable	**DA/SR**	**Wbus**	**WEBus**	**WRail**	**WCRail**	**Dbus**	**Debus**	**DRail**	**DCRail**
In Vehicle Time	X	X	X	X	X	X	X	X	X
Terminal Time	X								
Auto Operating Cost	X								
Auto Tolls	X								
Auto Parking Cost	X								
Walk Time		X	X	X	X	X	X	X	X
Initial Wait Time (under 7.5 min.)		X	X	X	X	X	X	X	X
Initial Wait Time (over 7.5 min.)		X	X	X	X	X	X	X	X
Transfer Time		X	X	X	X	X	X	X	X
Number of Transfers		X	X	X	X	X	X	X	X
Transit Fare		X	X	X	X	X	X	X	X
Drive Access Time						X	X	X	X

Table 6-2: Nesting coefficients

Nest	Value
Walk Transit Route (Bus, Rail, MARC)	0.30
Drive Transit Route (Bus, Rail, MARC)	0.30
Transit Access (Walk vs. Drive)	0.65
Shared Ride Occupancy (2 vs. 3+)	0.30
Auto Mode (Drive Alone vs. Shared Ride)	0.65

Mode choice coefficients are listed in Table 6-3. Mode specific constants and other bias coefficients, shown in Table 6-4 and Table 6-5, have been calibrated to match the Baltimore and Washington area trips by mode. The income specific bias constants have been added for Transit, Shared Ride, Share Ride3+ and Drive to Transit Nests. Bias constants have been added for express bus, rail and commuter rail modes in both, drive and walk to transit nests. These are meant for each purpose, aggregated by income. The bias constants were calibrated with the 2007 household travel survey, 2007 MTA onboard survey and 2008 WAMTA onboard survey data. The mode choice calibration targets are summarized in Appendix D.

Table 6-3: Mode choice coefficients

Attribute	HBW, NHBW	HBO, HBS, SCH	OBO
In Vehicle Time	-0.025	-0.008	-0.02
Terminal Time	-0.05	-0.02	-0.05
Auto Operating Cost	-0.0042	-0.0018	-0.0044
Auto Parking Cost and Tolls	-0.0084	-0.0036	-0.0088
Walk Time	-0.05	-0.02	-0.05
Initial Wait Time (under 7.5 min.)	-0.05	-0.02	-0.05
Initial Wait Time (over 7.5 min.)	-0.025	-0.01	-0.025
Transfer Time	-0.05	-0.02	-0.05
Number of Transfers	-0.125	-0.06	-0.15
Transit Fare	-0.0042	-0.0018	-0.0044
Drive Access Time	-0.05	-0.02	-0.05

Table 6-4: Mode-specific constants and bias coefficients at 2nd level

Purpose	DA	SR	SR2	SR3	Drive to Transit	Walk to Transit
HBW1	0	0	-0.329	-1.285	-0.856	3.996
HBW2	0	0	-0.351	-1.266	-0.539	2.464
HBW3	0	0	-0.409	-1.586	-1.072	0.771
HBW4	0	0	-0.447	-1.664	-2.503	-1.947
HBW5	0	0	-0.463	-1.695	-3.166	-3.231
HBS1	0	0	-0.094	0.035	-3.127	-1.631
HBS2	0	0	-0.194	0.104	-3.176	-2.417
HBS3	0	0	-0.116	0.09	-4.688	-3.552
HBS4	0	0	-0.043	-0.022	-5.072	-3.585
HBS5	0	0	-0.04	-0.04	-5.428	-3.806
HBO1	0	0	-0.014	0.17	-0.848	0.666
HBO2	0	0	-0.095	0.152	-2.665	-0.616
HBO3	0	0	-0.029	0.19	-3.218	-2.041
HBO4	0	0	0.008	0.197	-4.084	-2.961
HBO5	0	0	-0.001	0.18	-4.188	-3.536
HBSc	0	-0.838	0	-0.132	-0.516	-1.229
NHBW	0	-1.098	0	-0.305	-3.076	-2.419
OBO	0	0.351	0	-0.073	-2.712	-1.784

Table 6-5: Mode-specific constants and bias coefficients at 3rd level

Purpose	Drive to Bus	Walk to Bus	Drive to Express Bus	Walk to Express Bus	Drive to Rail	Walk to Rail	Drive to Commuter Rail	Walk to Commuter Rail
HBW	0	0	-0.437	-5.442	0.378	-0.436	1.107	-3.516
HBS	0	0	0	0	-0.444	1.31	-5.717	0.877
HBO	0	0	0	0	1.398	2.028	3.018	0.272
HBSc	0	0	0	0	-0.126	9.085	41.63	37.091

NHBW	0	0	0	0	-0.33	1.154	2.887	0.792
OBO	0	0	0	0	0.799	2.393	4.36	4.892

Highway and transit networks were developed to be generally consistent with the procedures used in the BMC model although some simplifications were made in recognition of the broader purposes of MSTM and the larger area covered.

GIS techniques were used to define the portion of each zone within walking distance of transit stops and stations and related average walk times. Parking costs by SMZ were calculated as a weighted average of TAZ parking costs from the MPO TAZ data (weighted by employment density). Comparable values were developed for other areas based on employment density.

6.2 Model Validation

The mode split model has been calibrated to resemble the mode split observed in the survey. As no independent data were available, a true validation of mode split was not possible. Instead, a comparison of survey data and model results shows that the mode split model was calibrated to resemble observed travel behavior. Figure 6-2 compares survey and model results for every trip purpose. Given that the statewide model covers a highly heterogeneous study area with parts that have excellent transit service and other parts with almost no transit access, the comparison shows a reasonable picture.

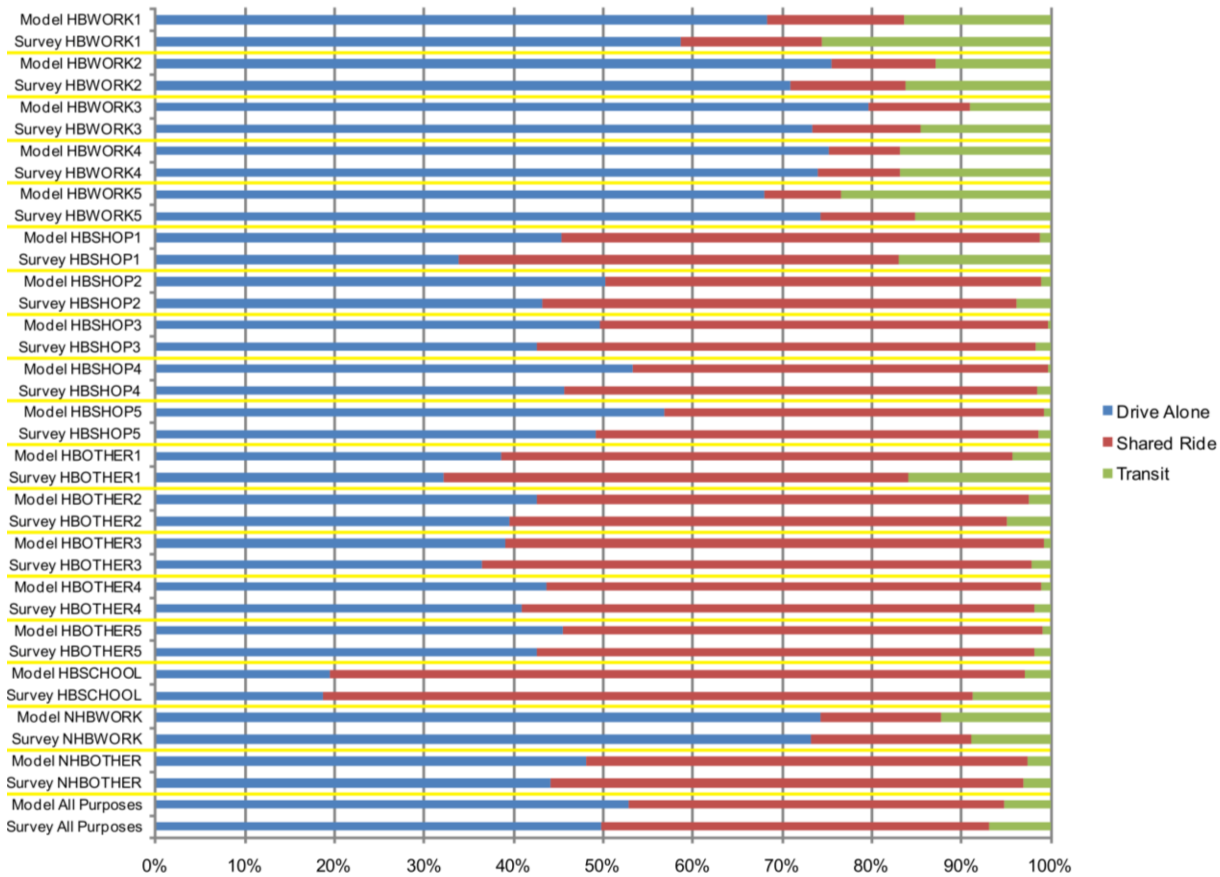


Figure 6-2: Mode split by purpose

7 Regional Person Model

A long-distance model called Nationwide Estimate of Long-Distance Travel (NELDT) has been implemented to cover long-distance travel. The model was presented at the Transportation Research Forum [9], and exchange with international researchers helped to further advance the model design.

This new person long-distance model that is now implemented for MSTM covers all trips traveling a one-way distance of 50 miles or more. In other words, this model handles External-External, External-Internal, Internal-External and Internal-Internal long-distance trips. **Error! Reference source not found.**Figure 7-1 shows the 50 mile range around downtown Baltimore and downtown Washington DC Trips between the two metropolitan areas are within the 50 mile radius, and therefore, covered by the short-distance model. Other trips that exceed the 50 mile range are simulated by NELDT.

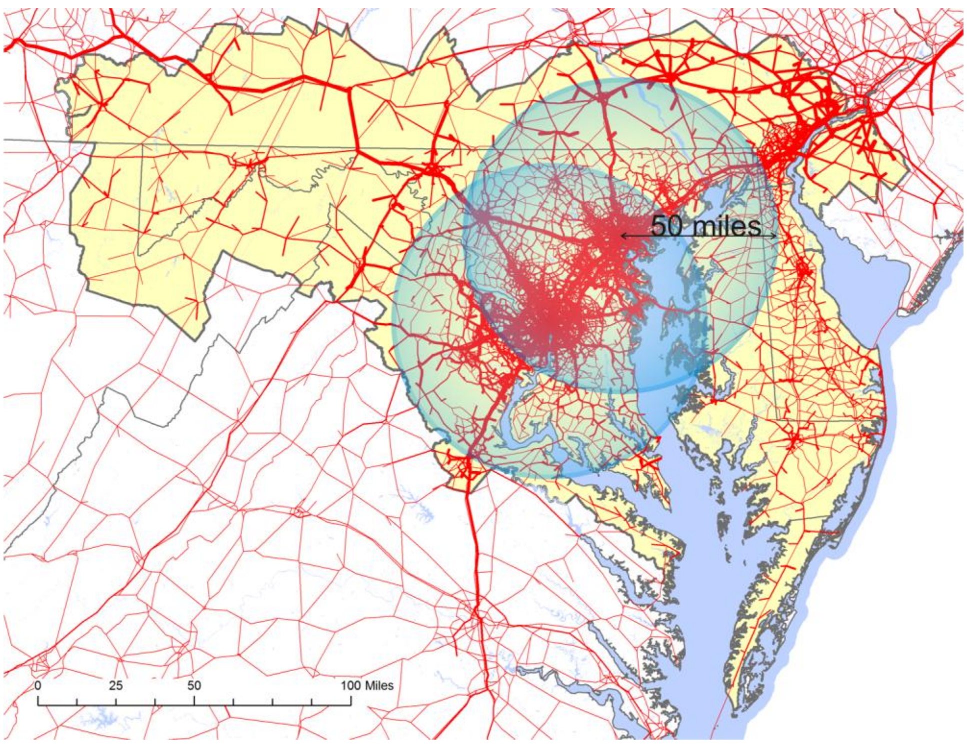


Figure 7-1: MSTM region with 50 miles radius around downtown Baltimore/Washington D.C.

7.1 Data

In 2001/2002, the Federal Highway Administration conducted the National Household Travel Survey (NHTS) [10], which collected data on both daily and long-distance travel within the U.S. [11]. The survey consisted of 69,817 telephone interviews conducted from March 2001 to May 2002. Respondents were asked about their daily travel patterns (short distance) as well as any travel within the past 28 days where the furthest destination was 50 miles or more away from their home (long distance). This data set offers a rich source of information for long distance trips by all modes of transportation within the U.S. A total of 45,165 (raw count) long distance data records are available. In 2010, FHWA published a new NHTS conducted in 2009 [11]. This time, however, interviews focused on daily traffic only, without a special survey for long-distance travel. From this dataset, a total of 28,246 records (raw count) with trip length over 50 miles are available. An analysis of available data records shows that the smaller number of records and the different survey format makes these data unusable for long-distance travel in Indiana. While the NHTS 2002 asked people about their long-distance travel in the last 28 days, the NHTS 2009 asked about trips in a 24h period. As a consequence, long-distance travel is underrepresented in the NHTS 2009.

Table 7-1 summarizes the number of NHTS records for Maryland by destination state. While the number of records is relatively small for travel demand modeling, this area is represented in the NHTS fairly well in comparison to other parts of the country. Particularly neighboring states, which are of most interest to traffic flows to and from Maryland, are fairly well represented.

Table 7-1: NHTS 2002 long-distance records of Maryland residents

Destination	Number of records
MD	202
PA	103
VA	78
DC	43
NY	27
DE	25
WV	20
NJ	19
Abroad	14
NC	7
OH	7
CA	6
FL	6
MA	6
AZ	5
NV	4
SC	4
WA	4
CO	2
GA	2
MO	2
AR	1
HI	1
IA	1

Destination	Number of records
MI	1
NM	1
TN	1
TX	1
Total	593

Air travel data are published by the Bureau of Transportation Statistics based on ticketed passengers [12]. These data provide a ten percent sample of ticketed passengers between all U.S. airports, distinguishing between passengers changing flights and passengers having their final destination at one airport. Data are available by quarter, and to ensure compatibility with the NHTS data, air travel data was retrieved for 3/2001, 4/2001, 1/2002 and 2/2002.

Further data needs are employment and population data at the statewide/regional level, as well as traffic counts for model validation.

7.2 Generate missing NHTS records

For privacy reasons, the NHTS dataset only reports the origin state for trips from states with a population of 2 million or more. Though this does not affect Maryland directly, trips from smaller states such as Delaware or West Virginia are missing in the NHTS. For these smaller states, synthetic data records need to be generated based on travel data of surrounding states for which data are available. There are 15 states and Washington DC for which records need to be synthesized. Figure 7-2 shows the number of data records with a long-distance trip by state. Most states without data records have neighboring states that can be used to synthesize missing data records. Maine records are generated based on the Massachusetts datasets, and Montana records are generated based on Washington and Oregon data.

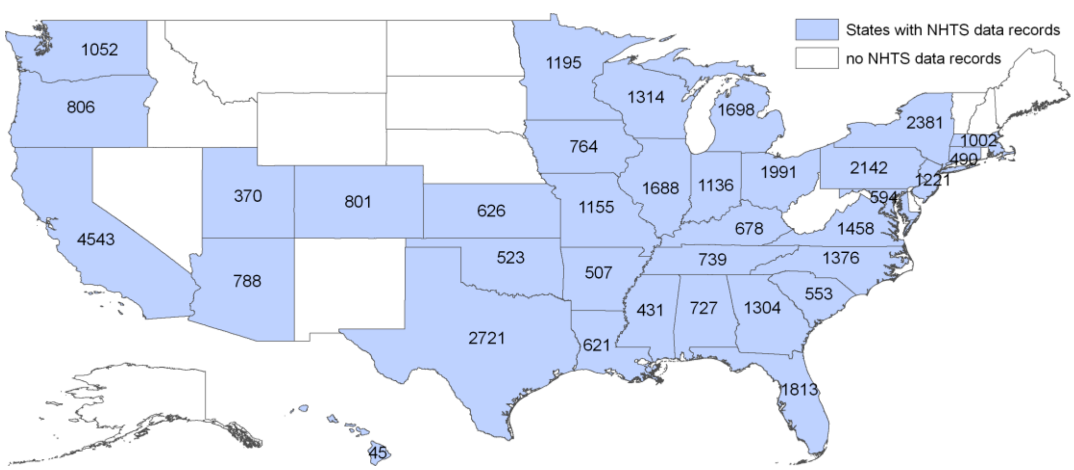


Figure 7-2: Number of NHTS long-distance travel data records by home state

To estimate the number of records that need to be synthesized for the 15 missing states and Washington DC, a multiple regression analysis is done, where population serves as the independent variable. the intercept was forced to be 0 to ensure that if the population of a region is 0, the number of long-distance trips from that region is 0 as well. Table 7-2 summarizes the results of this multiple regression. A reasonable correlation was found for the modes auto, air and bus. The modes train, ship and other are sparsely available across the country and have a small sample sizes, it is little surprising that they show less correlation.

Table 7-2: Revised estimation of NHTS records per state

Auto	Estimate	Std. Error	t value	Pr(>|t|)	**Air**	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)					(Intercept)				
Population	1.23E-04	4.35E-06	28.29	<2e-16***	Population	1.14E-05	3.31E-07	34.47	<2e-16***
Adj. R-squared:	0.9581				Adj. R-squared:	0.9714			
N:	36790				N:	3110			
Bus	Estimate	Std. Error	t value	Pr(>|t|)	**Train**	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)					(Intercept)				
Population	2.89E-06	1.81E-07	15.96	<2e-16***	Population	1.53E-06	3.35E-07	4.56	6.34E-05***
Adj. R-squared:	0.8788				Adj. R-squared:	0.3612			
N:	833				N:	370			
Ship	Estimate	Std. Error	t value	Pr(>|t|)	**Other**	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)					(Intercept)				
Population	1.37E-07	2.82E-08	4.864	0.0000258***	Population	1.69E-07	7.24E-08	2.336	0.0255*
Adj. R-squared:	0.393				Adj. R-squared:	0.113			
N:	36				N:	70			
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1									

These factors were used to estimate the number of trip records for states that were excluded from the NHTS survey as their population was below 2 million. A corresponding number of trip records were synthesized for these states, as shown in Table 7-3. Only auto, air and bus trips are analyzed subsequently as the modes train and ship are only available in selected areas and cannot be estimated with a general regression analysis.

Table 7-3: NHTS records synthesized for each state and Washington D.C.

State	Auto	Air	Bus
Alaska	79	7	2
Delaware	99	9	2
District of Columbia	70	7	2
Idaho	165	15	4
Maine	159	15	4
Montana	112	10	3
Nebraska	213	20	5
Nevada	267	25	6
New Hampshire	157	15	4
New Mexico	228	21	5
North Dakota	78	7	2
Rhode Island	132	12	3
South Dakota	94	9	2
Vermont	76	7	2
West Virginia	222	21	5
Wyoming	61	6	1

For each state listed in Table 7-3, up to four neighboring states were chosen. From these neighboring states, NHTS records were selected randomly to synthesize records for each state of Table 7-3. The destination of each synthesized record is set to ensure that the share of intrastate trips is the same as the average share of intrastate trips in neighboring states. This way, the characteristics of the travelers of neighboring states is copied, while the average trip length of neighboring states is approximately achieved. Table 7-4 shows the synthesizing of auto long-distance travel records for New Mexico as an example. First, the number of intra-state, to-neighboring-states and other-destination travel records are summarized for the four neighboring states AZ, CO, OK and TX, resulting in an average of 84% of travelers who stay in the same state, 10% who visit neighboring states and 6% who travel further away. A corresponding number of records are chosen for New Mexico from the four neighboring states. After selecting a record, the origin is replaced with New Mexico, and the destination is replaced with NM for intra-state trips, with AZ, CO, OK or TX for trips to neighboring states, and for trips to other destinations the destination is given by the selected record. The same procedure is applied to all 15 states and Washington DC, for which NHTS records are not published.

Table 7-4: Process to synthesize auto long-distance travel records for New Mexico

State	Intra-state	Neighboring state	Other destination
AZ	543	82	42
CO	605	55	50
OK	340	128	20
TX	2,110	159	130
Sum	3,598	424	242
Share	0.844	0.099	0.057
Total # of records NM	228 records for auto trips		
Records NM by mode	192	23	13
Choose destination from	NM	AZ, CO, OK, TX	given by sampled states

Alaska is particularly difficult as it has no neighboring US states, and – given its size – it has a very unique long-distance travel pattern. As an interim solution, Washington State was chosen as a “neighboring” state to Alaska. Though distances are big in Alaska, the absolute number of long-distance travelers is very small, and they rarely reach the Maryland region.

Because the NHTS is a national survey that interviewed long-distance travelers in their home state, no international visitors are included in the NHTS data set. International travelers need to be synthesized based on air travel data and land-border crossings from Canada and Mexico. Their characteristics are assumed to be comparable to American long distance travelers.

7.3 Nationwide number of long-distance travelers

As the NHTS data set is a sample of long distance travel, not all long distance trips of the entire population are included. Even though the NHTS data set includes weights for every data record, simply expanding the records based on these weights is not recommended [13]. Long-distance travel is an event that is too rare to expand from single records. If, for instance, a person reported two trips in a 28-day period, expanding this trip to

2 trips / 28 days x 365 days = 26 trips per year

it cannot be carried out with statistical confidence. This person may have done far fewer trips greater than 50 miles in this year. Because long distance trips are relatively rare, a simple expansion produces statistically insignificant results. Instead, the total number of air travelers provided by BTS air travel data is used to expand the NHTS nationwide.

Table 7-5: Expanded number of long-distance travelers in the U.S.

	Auto	Air	Bus	Train	Ship	Other
NHTS Records	36,790	3,110	833	370	36	70
Synthesized records	5,687	233	102			
Total number of records	42,477	3,343	935	370	36	70
BTS air statistics		84,640,725				
Expansion factor	25318.793					
Number of yearly travelers	1,075,466,370	84,640,725	23,673,071	9,367,953	911,477	1,772,316
Number of daily travelers	2,946,483	231,892	64,858	25,666	2,497	4,856

Table 7-5 shows the expanded number of long-distance travelers on an average day in the U.S. after synthesizing NHTS records for missing states, 3,343 air travel records are available. This only includes clean records that have all required data attributes. Given the number of air passengers according to BTS database, an expansion factor of 25,318.793 was calculated, which led to a yearly number of travelers for all modes.

The assumption behind this expansion is that the NHTS is a representative sample across all modes. If the share of auto and air records in the NHTS represents the mode split in reality, air travel data may be used to expand the NHTS data. Next, the yearly number was converted into daily travelers by dividing by 365. In urban travel demand models, it is common to use a smaller number than 365 to convert yearly in daily traffic volumes, as it is assumed that weekday traffic carries more trips than weekend traffic. For long-distance travel, however, weekends carry at least a similar number of trips as weekdays, particularly for personal trips. For lack of better information -the NHTS records do not report the weekday- yearly data was divided by 365 to derive travel on an average day.

It should be noted that Table 7-5 shows how many long-distance trip are started on a given day. Each record, however, describes a journey including both the outbound trip and the return trip. In the expansion process, NHTS records are duplicated until the number of air trips matches the observed total of 231,892 trips.

7.4 Direction of Travel

The NHTS data records describe tours, including outbound trip, possibly staying overnight at the destination, and return trip. For each long-distance traveler, the number of nights stayed away from home is provided by NHTS. As an average day shall be simulated, both the outbound and the inbound trip need to be represented. If someone is staying away from home for 0 nights, it is assumed that this person has the outbound trip and the return trip on the same day, thus the trip of this person is added to the trip table twice, from home state to destination state and from destination state to home state. Travelers that stay one night are assigned with half a trip from their home state to the destination state and another half trip from the destination state to the home state. For a two-night trip, one third of an outbound trip and one third of a return trip is added for the simulation of an average day, and so on.

$$trips_{state_a,state_b} = longtrips_{state_a,state_b} \cdot \frac{1}{nights+1} \qquad (1)$$

where $trips_{state_a,state_b}$ is the number of average daily trips from $state_a$ to $state_b$

$longtrips_{state_a,state_b}$ is the number of all trips from NHTS origin to NHTS destination

nights is the number of nights away from home

In addition, the number of trips is influenced by the distance traveled, at least for auto trips. Someone traveling from San Francisco to Chicago has to drive approximately a day and half. Even if there were several drivers allowing the vehicle to travel without overnight stays, traffic would be overestimated if the entire trip from San Francisco to Chicago was assigned to the network as traveled on the one day simulated. The assumption was made that the average traveler would drive for up to 750 miles per day, and then rest for an overnight stay. Trips below 750 miles are not adjusted, but trips longer than this threshold are reduced proportionally to the distance traveled.

$$trips_{state_a,state_b} = longtrips_{state_a,state_b} \cdot \frac{\sigma}{\max\left(\sigma, dist_{state_a,state_b}\right)} \qquad (2)$$

where $trips_{state_a,state_b}$ is the number of average daily trips from $state_a$ to $state_b$

$longtrips_{state_a,state_b}$ is the number of all trips from NHTS origin to NHTS destination

σ is a threshold the average traveler is assumed to be able to travel per day, for auto travel it is set to 750 miles

$dist_{state_a,state_b}$ is the travel distance from $state_a$ to $state_b$

This way, long-distance trips of more than 750 miles are scaled down to account for the fact that it is impossible to drive from coast to coast in a single day. A trip from San Francisco to Chicago (2,133 miles) would be assigned as 0.35 trips.

Finally, long-distance travel journeys need to be converted into trips. A journey from *i* to *j* is converted into an outbound trip from *i* to *j* and a return trip from *j* to *i*, assuming that each trip was a one-destination, one-purpose, one-mode trip.

7.5 Disaggregation

The NHTS reports trip origins and destinations by state. The simulation of travel demand in Maryland requires a geography much smaller than states, at least in the Chesapeake Bay region. To make these long distance trips usable for MSTM, trip origins and destinations are disaggregated to the Statewide and Regional level. This disaggregation is done based on population and employment. Zones with more population and employment are expected to generate and to attract more long-distance trips than less populated zones. Furthermore, the larger the distance between two zones, the smaller is the attraction between them. This reasoning follows common gravity theory. The following equation is applied to disaggregate trips between states to trips between counties and zones:

$$tripsDisagg_{taz_i,taz_j} = trips_{state_a,state_b} \cdot \frac{weight_{taz_i,taz_j}}{\sum_{taz_k \in State_a} \left(\sum_{taz_l \in State_b} weight_{taz_k,taz_l} \right)} \qquad (3)$$

where taz_i (= SMZ or RMZ in MSTM) is located in $state_a$
taz_j (= SMZ or RMZ in MSTM) is located in $state_b$
taz_k are all zones located in $state_a$
taz_l are all zones located in $state_b$

The weights for disaggregation are calculated differently for personal and business trips.

Business trips:

$$weight_{taz_i,taz_j} = (\alpha_1 \cdot pop_i + \alpha_2 \cdot oEmp_i + \alpha_3 \cdot tEmp_i) \cdot (\alpha_4 \cdot pop_j + \alpha_5 \cdot oEmp_j + \alpha_6 \cdot tEmp_j) \cdot \exp(\gamma \cdot d_{i,j}) \qquad (4)$$

Personal trips:

$$weight_{taz_i,taz_j} = (\beta_1 \cdot pop_i + \beta_2 \cdot tEmp_i) \cdot (\beta_3 \cdot pop_j + \beta_4 \cdot rEmp_j + \beta_5 \cdot tEmp_j) \cdot \exp(\gamma \cdot d_{i,j})$$

where pop_i is population in zone *i*
$tEmp_i$ is total employment in zone *i*
$rEmp_i$ is retail employment in zone *i*
$oEmp_i$ is office employment in zone *i*

$d_{i,j}$ is the travel distance from county *i* to county *j*

Alpha and beta are parameters to weight the impact on trip production and trip attraction of different population and employment types.

Table 7-6 shows the parameters used to weight production and attraction factors. With the exception of β_3, which was based on NHTS data, all values were asserted and should be subject to careful reevaluation if additional data become available.

Table 7-6: Parameters for long-distance trip production and attraction

Parameter	Value	Reasoning
α_1	0.5	A business trip starting in the morning is likely to start from the home location
α_2	0.4	A business trip starting later in the day is likely to start from the work location, which commonly is office employment
α_3	0.1	A few trips are generated by total employment, which accounts for other employment types, but purposefully accounts for office employment for a second time
α_4	0.1	A very few business trips are attracted by households (such as sales call)
α_5	0.2	Several long-distance trips are attracted by office employment
α_6	0.7	Most business trips are attracted tototal employment, accounting for hotels, office employment and other employment.
β_1	0.9	Almost all personal trips start at home
β_2	0.1	A few personal long-distance trips start at their work location
β_3	0.5	Population is a major attractor of personal trips (value based on NHTS share of personal trips that visit friends or relatives)
β_4	0.4	A few personal trips visit general employment (such as hotels)
β_5	0.1	Many personal trips visit retail employment

The parameter γ was calibrated to resemble the average long-distance trip length of 136 miles as reported in the NHTS dataset. The parameter γ has been set differently for each origin state to reflect different travel behavior patterns across the county.

The result of this module is a trip table with daily trips between all SMZ and all RMZ zones. This trip table may be split into time-of-day periods and be assigned to the highway network for long-distance auto travel.

8 Freight Model

8.1 Statewide Layer

The statewide level truck trip model is an adaption of the BMC and MWCOG truck and commercial vehicles models.

Two truck types, Medium Truck and Heavy Truck, and commercial vehicles are distinguished. Trip generation is based on employment by category and total households. BMC truck generation rates are shown in Table 8-2. Comparative generation rates for other areas are given in Table 8-2, showing that BMC truck trip generation rates are comparable to rates applied in other regions. Trips ends are calculated for the statewide level model area.

Table 8-1: BMC commercial vehicle generation rates

Generation Variable	Commercial Vehicle Generation Rates		
	Light (4-Tire)	Medium Truck	Heavy Truck
Employment:			
Industrial	0.454	0.125	0.179
Retail	0.501	0.124	0.127
Office	0.454	0.034	0.026
Households	0.146	0.048	0.061

Table 8-2: Comparative commercial vehicle generation rates

Model	Households	Employment Agriculture	Manufacture	Wholesale	Retail	Service	Other
QRFM	0.251	1.110	0.938	0.938	0.888	0.437	0.663
Phoenix	0.154	0.763	0.641	0.763	0.591	0.309	0.763
Columbus	0.134	0.506	0.506	0.506	0.437	0.233	0.506
Atlanta	0.140	0.482	0.482	0.482	0.643	0.232	0.232
Huston	0.020	0.300	0.480	0.300	0.360	0.300	0.300
Seattle	0.093	0.410	0.347	0.347	0.328	0.162	0.245
Vancouver	0.019	0.096	0.069	0.071	0.143	0.043	0.229

8.2 Regional Layer

Truck trip distribution is based on a gravity model formulation using truck generalized cost incorporating truck travel times, travel cost and tolls. The current implementation uses truck travel time in the off-peak time period. The initial truck distribution parameters were borrowed from the BMC Truck Model and the BMC Commercial Vehicles Model. As the gamma parameter was set to 0 in the BMC model, the gravity formulation technically becomes an exponential function (because exp(0) = 1).

$$F_{i,j} = \alpha \cdot t_{i,j}{}^{\beta} \cdot \exp(\gamma \cdot t_{i,j})$$

Where
$F_{i,j}$ Friction factor from zone i to j
$T_{i,j}$ Off peak travel time from i to j
α, β, γ Parameters defined below

Table 8-3: Friction factors for the statewide truck model

Original BMC Parameters

Parameter	CommercialVehicles	MediumHeavyTrucks	HeavyHeavyTrucks
Alpha	1,202,604.28	1,202,604.28	3,269,017.37
Beta	-3.75	-5.8	-2.9
Gamma	0	0	0

Adjusted MSTM Parameters

Parameter	CommercialVehicles	MediumHeavyTrucks	HeavyHeavyTrucks
Alpha	1,202,604.28	1,202,604.28	3,269,017.37
Beta	*-8.75*	*-6.8*	*-3.9*
Gamma	0	0	0

8.3 Freight-Economy Reconciliation

This section describes the reconciliation of the economic data with the FAF. Inforum[13] has assembled a database of historical and projected freight shipments published in the 2002 Freight Analysis Framework (FAF), which is produced by the U.S. Department of Transportation. The FAF "estimates commodity flows and related freight transportation activity among states, regions, and major international gateways." This database covers the periods 2002, 2010, 2015, 2020, 2025, 2030, and 2035. Shipments are measured in thousands of tons; shipments in millions of dollars also are available but are not included in this work. The data are published in four sets: domestic freight; US-Canada and US-Mexico land freight; international sea freight; and international air freight. Detail is available for 138 regions, including 114 domestic regions, 17 domestic ports, and 7 international regions. Detail also includes 43 commodities and 7 transportation modes. For each commodity and each mode, nonzero values are published for shipments from region to region. For international shipments, either the origin or destination may be

[13] Inforum, an economic forecasting and research group at the University of Maryland that has been in operation since 1967, employs interindustry-macroeconomic general equilibrium models to examine past employment trends and to forecast future employment across sectors of the economy. Their primary model, LIFT (Long-term Interindustry Forecasting Tool), uses a bottom-up approach to make such predictions, meaning that it uses component data within each of its defined industries to estimate future employment rather than starting with top-level macroeconomic indicators. In this regard, the model is well-suited to address the questions posed in this report, which focus on commodity shipments. The LIFT model aggregates the North American Industry Classification System (NAICS) industries into 97 industries that span the economy. Inforum maintains a second US model, Iliad, that offers detail on 360 commodities formed from NAICS data.

a foreign region. For these international exchanges, a US port is listed; ports may be one of the 17 designated ports, or the "port" may be one of the 114 domestic regions.

After assembling the published FAF data, the data were aggregated in three parts, preserving detail on 131 regions and all 43 commodities. The three parts are total domestic-domestic shipments, exports, and imports. Because the focus of this study is the trucking mode, a second corresponding set of databases were constructed from FAF Truck and International Air records. For each commodity, there are region-region tables of total shipments and truck shipments for domestic trade, exports, and imports.

For each commodity, the regional detail was aggregated to calculate total shipments, shipments by truck, total consumption, and total consumption of goods shipped by truck. Shipments were defined as domestic-domestic trade plus exports. Consumption was defined as domestic-domestic trade plus imports.

The FAF database is compiled from information published in Bureau of Transportation's Commodity Flow Survey (CFS); Surface Transportation Board's Carload Waybill Sample; U.S. Army Corps of Engineers (USACE) waterborne commerce data; Bureau of Transportation Statistics' Transborder Surface Freight database; and the Air Freight Movements database from BTS. Each of the 43 commodities employed in the FAF is defined according to the Standard Classification of Transported Goods (SCTG).[14]

These classification codes were compared to the commodity detail employed in the Inforum Lift and Iliad inter-industry macroeconomic models, where the industry detail are derived from data published according to the North American Industrial Classification System (formerly the Standard Industrial Classification system). Industry production data employed by Inforum models primarily are derived from BEA's Gross Output by Industry. Gross output represents the market value of an industry's production of goods and services. Data are compiled at the Bureau of Economic Analysis (BEA) using publications from U.S. Department of Agriculture (USDA), U.S. Geological Survey (USGS), Department of Energy (DOE), Census, Bureau of Labor Statistics (BLS), and BEA.[15] In addition to Gross Output, other sources of industry production information utilized by Inforum's models include BEA's Input Output tables and Foreign Trade data from Census. For each commodity defined in the models, the models offer real output, exports, and imports. For each SCTG commodity employed in the FAF, a match was found in the Inforum models, where the match sometimes was the sum of several narrowly defined commodities. This information is used as the basis of model-derived indexes for each of the SCTG commodities for output, exports, and imports.

For each FAF commodity, for domestic shipments, exports, and imports, we calculate from the FAF projections a forecast of the share of truck shipments relative to total (all transportation modes) shipments. These projected shares are employed to adjust our indexes for domestic supply, exports, and imports. These adjustments yield indexes for truck shipments of domestically produced and consumed products, truck shipments of exported goods, and truck shipments of imported goods, where the shipments are measured in constant dollars. Next, these constant-

[14] Information on SCTG was found at http://www.statcan.ca/english/Subjects/Standards/sctg/sctg-class.htm#19.
[15] More information may be found at http://www.bea.gov/scb/account_articles/national/0600gpi/tablek1.htm.

dollar truck shipment levels are scaled to the corresponding 2002 FAF levels, for domestic shipments, exports, and imports for each commodity. This yields model-based history and forecasts of tons of each commodity shipped by truck. The model-based indexes are consistent with the FAF 2002 survey data.

The FAF projections of shipments by truck were updated by scaling the sum of the regional detail to corresponding model-derived totals. For each FAF commodity, the sum (across domestic regions) of domestic shipments was scaled to the model-derived total. This was done both for domestic shipments and domestic consumption. In similar fashion, the sum of FAF exports and imports were scaled to the model-derived totals. Finally, total truck shipments were calculated by adding the detail. Total shipments are the sum of domestic shipments plus exports. Total consumption is the sum of domestic consumption plus imports.[16]

Total shipments and total consumption projections are provided for each commodity and each region. 2002 levels are consistent with FAF levels. Data for 2005 to 2030, in five-year intervals, are provided according to the methodology described above, where the sums of the original FAF figures are controlled to model-derived totals. Estimates for 2000 are constructed by using 2002 FAF regional distributions and trucking shares and where the total shipments are controlled to the model-derived index levels for 2000.

A series of 43 worksheets contain information on each FAF commodity. Total truck shipments and total consumption of truck freight, calculated from the FAF database, are provided, together with corresponding model-derived aggregate indexes. FAF figures are provided for 2002 and 2010-2030 in 5-year increments. Model-derived updates are provided for 2000, 2002, and 2005-2030 in 5-year increments. For each commodity, shipments and consumption figures also are provided for each domestic region and port, where the regional detail is consistent with the model-derived totals.

The methodology described here depends on several assumptions that warrant additional investigation. A crucial assumption is that growth of constant-dollar indexes for output, real exports, and real imports correspond to growth of shipments by weight. This assumption may fail if the economic data are adjusted for quality change or if the nature of the commodity changes over time.

The updated projections and historical estimates seem to offer improvements over the FAF projections. In particular, the effects of the recent recession are clear, though the recession effects are still more clear in the annual economic data. In general, the long-run shipments estimates do not differ dramatically from the FAF projections but arguably are more plausible. Further, the production and consumption totals by zone are classified into the 130x130 matrix format by the internal proportion fitting (IPF) method. INFORUM provides the Production and Consumption

[16] Note that the current work is done slightly differently. The FAF-based detail for commodity shipments by truck are scaled to the model-derived estimates for total shipments, where total shipments are the sum of domestic supply and exports. Similarly, FAF-based detail for receipts are scaled to the sum of model-derived figures for domestic receipts plus imports. This change from the original procedure minimizes problems with the initial results. These problems arose where the FAF forecasts of imports and exports differ substantially from the Inforum forecasts. In the current work, we assume that the foreign shares of commodity shipments implied in the FAF forecasts will hold.

by FAF zone as control totals (marginals), and FAF provides the starting pattern of flows connecting the FAF zones (seed). The IPF process modifies the flows between zones until it matches the INFORUM FAF zones totals. The result OD flows is the commodity flow forecast used as the starting demand in the regional truck model.

8.4 Update truck model data

The most important input data for the truck model is the Freight Analysis Dataset (FAF), published by the Federal Highway Administration (FHWA). When the truck model was developed initially, the most recent version available was FAF2. In Spring 2010, FWHA released the next update of this dataset, called FAF3. Comparisons between FAF2 and FAF3 showed that the differences are substantial, and FHWA recommends not to use FAF2 anymore. Furthermore, the MSTM methodology to convert FAF data into truck trips was updated significantly. For clarity reasons, the complete revised methodology is documented below, rather than attempting to explain piece-meal-wise every change.

The changes only affect the long-distance model (modeling trips greater than 50 miles). The short-distance model was recalibrated slightly to adjust for changes in the long-distance model. This calibration step is documented below, otherwise the short-distance truck model remains unchanged.

8.4.1 Data

The third generation of the FAF data, called FAF^3, was released in summer 2010 and contains flows between 123 domestic FAF regions and 8 international FAF regions. The MSTM truck model is using the third release of FAF^3, also called FAF3.3. Figure 8-1 shows Maryland in Yellow and the Size of the Zones Provided by FAF.

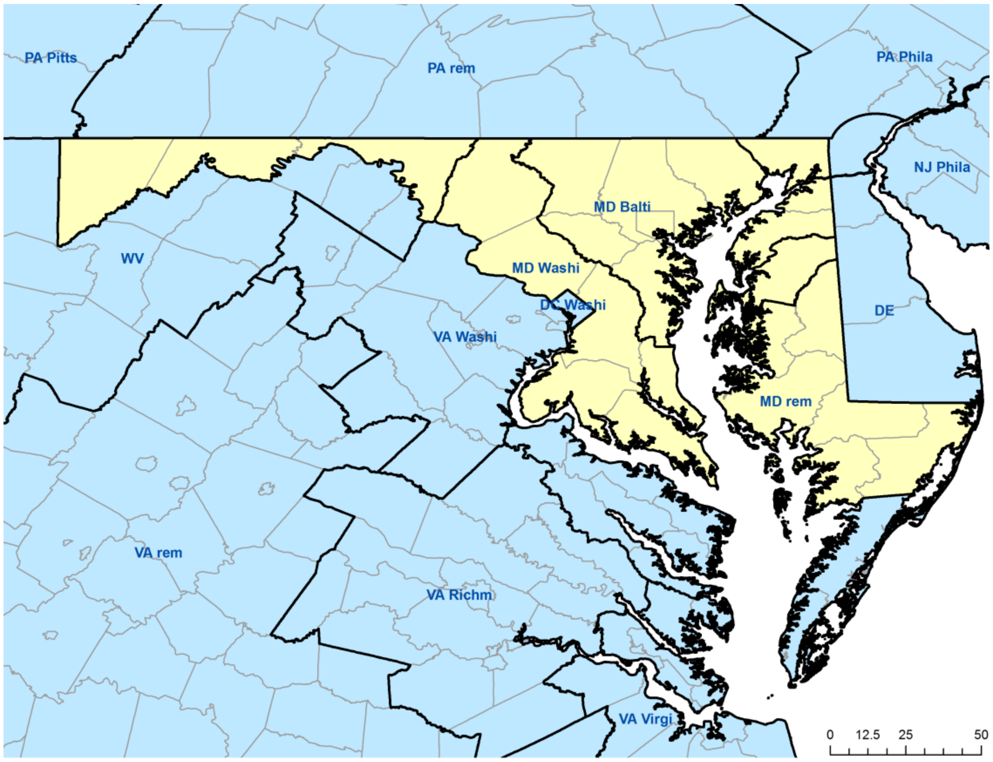


Figure 8-1: FAF zones in Maryland

FAF[3] data provide commodity flows in tons and dollars by

- FAF zones (123 domestic + 8 international zones)
- Mode (7 types)
- Standard Classification of Transported Goods (SCTG) commodity (43 types)
- Port of entry/exit for international flows (i.e. border crossing, seaport or airport)

The base year is 2007, and freight flow forecasts are provided for the years 2015 to 2040 in five-year increments. At this point, the FAF base year 2007, which is coincident with the current MSTM base year, and the forecast for 2030 are used.

The FAF data contain different modes and mode combinations. For the ILLIANA project, the mode Truck is used. Further data required for the truck model include the Vehicle Inventory and Use Survey (VIUS) that was done for the last time in 2002. The U.S. Census Bureau publishes the data with survey records of trucks and their usage[17]. County employment by 10 employment types were collected from the Bureau of Labor Statistics[18], and county-level employment for agriculture was collected from the U.S. Department of Agriculture[19]. Input/Output coefficients used for flow disaggregation were provided by the Bureau of Economic Analysis[20]. Finally, MSTM population and employment data are used for truck disaggregation, and truck counts are necessary to validate the model.

[17] http://www.census.gov/svsd/www/vius/products.html
[18] ftp://ftp.bls.gov/pub/special.requests/cew/2010/county_high_level/
[19] http://www.nass.usda.gov/Statistics_by_Subject/index.php
[20] http://www.bea.gov/industry/io_benchmark.htm

8.4.2 Truck model design

The resolution of the FAF data with 123 zones within the U.S. is too coarse to analyze freight flows in Maryland. Hence, a method has been developed to disaggregate freight flows from FAF zones to counties and further to MSTM zones. An overview of the truck model design is shown in Figure 8-2. First, the FAF[3] data are disaggregated to counties across the entire U.S. using employment by eleven employment types in each county. Within the MSTM region, detailed employment categories are used to further disaggregate to SMZ. Finally, commodity flows in tons are converted into truck trips using average payload factors.

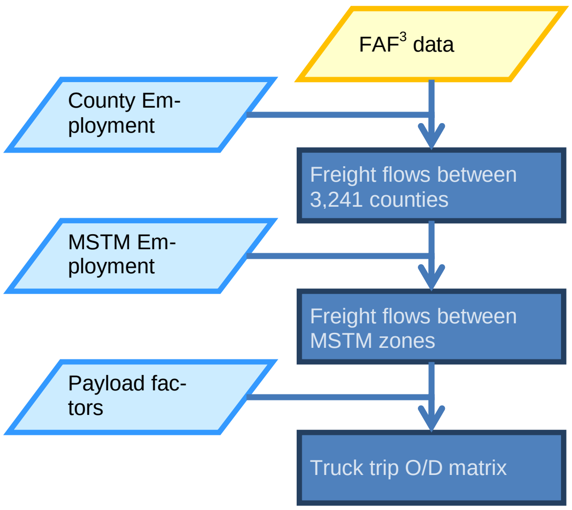


Figure 8-2: Model design of the regional truck model

Output of this module is a truck trip table between all MSTM zones for two truck types, single-unit trucks and multi-unit trucks.

8.4.3 Commodity flow disaggregation

Freight flows are given by FAF zones. For some states, such as New Mexico, Mississippi or Idaho, a single FAF region covers the entire state. Flows from and to these large states would appear as if everything was produced and consumed in one location in the state's center (or the polygon's centroid). To achieve a finer spatial resolution, truck trips are disaggregated from flows between FAF zones to flows between counties based on employment distributions (Figure 8-3). Subsequently, trips are further disaggregated to SMZ in the MSTM model area.

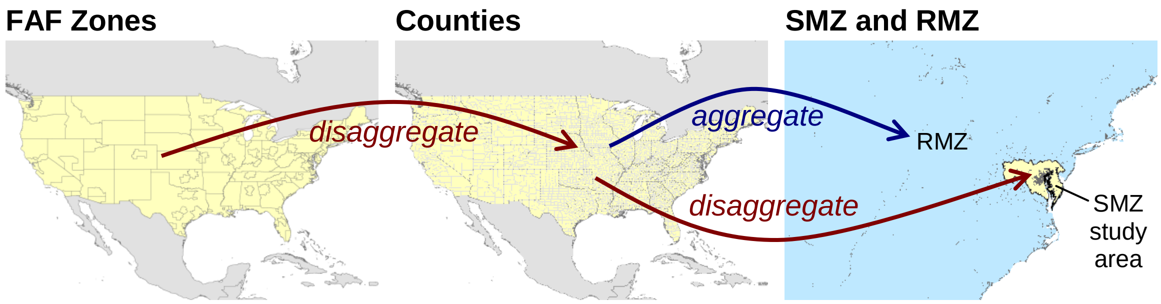


Figure 8-3: Disaggregation of freight flows

In the first disaggregation step from FAF zones to counties employment by county in eleven categories is used:

- Agriculture
- Construction Natural Resources and Mining
- Manufacturing
- Trade Transportation and Utilities
- Information
- Financial Activities
- Professional and Business Services
- Education and Health Services
- Leisure and Hospitality
- Other Services
- Government

County-level employment for agriculture was collected from the U.S. Department of Agriculture[21]. For all other employment categories, data were retrieved from the Bureau of Labor Statistics[22]. These employment types serve to ensure that certain commodities are only produced or consumed by the appropriate employment types. For example, SCTG25 (logs and other wood in the rough) is produced in those zones that have forestry employment (the model uses agricultural employment as a proxy for forestry); this commodity is shipped to those zones that have employment in industries consuming this commodity, particularly manufacturing and construction. At the more detailed level of MSTM zones, four employment categories are available:

- Retail
- Office
- Other
- Total

The following equation shows the calculation to disaggregate from FAF zones to counties. A flow of commodity *c* from FAF zone *a* to FAF zone *b* is split into flows from county *i* (which is located in FAF zone *a*) to county *j* (which is located in FAF zone *b*) by:

[21] http://www.nass.usda.gov/Statistics_by_Subject/index.php
[22] ftp://ftp.bls.gov/pub/special.requests/cew/2010/county_high_level/

$$flow_{i,j,com} = flow_{FAF_a,FAF_b} \cdot \frac{weight_{i,com} \cdot weight_{j,com}}{\sum_{M \in FAF_a} \sum_{N \in FAF_b} weight_{m,com} \cdot weight_{n,com}} \tag{6}$$

where $flow_{i,j,com}$ = flow of commodity *com* from county *i* to county *j*
$county_i$ = located in FAF_a
$county_j$ = located in FAF_b
$county_m$ = all counties located in FAF_a
$county_n$ = all counties located in FAF_b

To disaggregate flows from FAF zones to counties, employment in the above-shown eleven categories and make/use coefficients are used. The make/use coefficients were derived from input/output coefficients provided by the Bureau of Economic Analysis[23]. These weights are commodity-specific. They are calculated by:

Production

$$weight_{i,com} = \sum_{ind} \left(empl_{i,ind} \cdot mc_{ind,com} \right) \tag{7}$$

Consumption

$$weight_{j,com} = \sum_{ind} \left(empl_{j,ind} \cdot uc_{ind,com} \right) \tag{8}$$

where $emp_{i,ind}$ = the employment in zone *i* in industry *ind*
$mc_{ind,com}$ = make coefficient describing how many goods of commodity *com* are produced by industry *ind*
$uc_{ind,com}$ = use coefficient describing how many goods of commodity *com* are consumed by industry *ind*

Table 8-4 shows the make coefficients applied. Many cells in this table are set to 0, as most commodities are produced by a few industries only. No value was available for commodities SCTG09 (tobacco products) and SCTG15 (coal). They were assumed to be produced by agricultural employment and mining, respectively. As only the relative importance of each industry for a single commodity is required, it is irrelevant to which value the entry for these two commodities is set, as long as the industry that produces this commodity is set to a value greater than 0 and all other industries are set to 0.

Table 8-4: Make coefficients by industry and commodity

Commodity	Agriculture	Construction	Health	Leisure	Manufacturing	Mining	Retail	Wholesale	Service
SCTG01	811.6238	0	0	0	0	0	0	0	0
SCTG02	198.234	0	0	0	0	0	0	0	0
SCTG03	3669.689	0	0	0	0	324.679	0	0	0
SCTG04	159.456	0	0	0	114.4688	0	0	0	0

[23] http://www.bea.gov/iTable/index_industry.cfm

Commodity	Agriculture	Construction	Health	Leisure	Manufacturing	Mining	Retail	Wholesale	Service
SCTG05	0	0	0	0	786.7564	220.2534	0	0	0
SCTG06	0	0	0	0	1289.469	0	0	0	0
SCTG07	205.8607	0	0	0	6551.506	0	0	0	0
SCTG08	0	0	0	0	1150.509	0	0	0	0
SCTG09	1	0	0	0	0	0	0	0	0
SCTG10	0	0	0	0	4.254867	211.2682	0	0	0
SCTG11	0	0	0	0	0.643628	25.07928	0	0	0
SCTG12	0	0	0	0	3.647224	142.1159	0	0	0
SCTG13	0	0	0	0	3.740241	95.63332	0	0	0
SCTG14	0	0	0	0	0	42.32755	0	0	0
SCTG15	0	0	0	0	0	1	0	0	0
SCTG16	0	0	0	0	0	138.1041	0	0	0
SCTG17	0	0	0	0	46.14806	12.86544	0	0	0
SCTG18	0	0	0	0	46.14806	12.86544	0	0	0
SCTG19	0	0	0	0	222.981	156.6388	0	0	0
SCTG20	0	0	0	0	1133.067	7.601936	0	0	0
SCTG21	0	0	0	0	393.104	0	0	0	0
SCTG22	0	0	0	0	267.6962	0	0	0	0
SCTG23	0	0	0	0	1082.518	0	0	0	0
SCTG24	0	0	0	0	1839.762	0	0	0	0
SCTG25	93.52182	5031.908	0	0	0	0	0	0	0
SCTG26	0	0	0	0	7578.98	0	0	0	0
SCTG27	0	0	0	0	392.5042	0	0	0	0
SCTG28	0	0	0	0	3254.577	0	0	0	0
SCTG29	0	0	0	0	621.0631	0	0	0	561.9978
SCTG30	0	0	0	0	747.4527	0	0	0	0
SCTG31	0	0	0	0	1439.455	9.26281	0	0	0
SCTG32	0	0	0	0	3039.151	0	0	0	0
SCTG33	0	0	0	0	4198.737	0	0	0	0
SCTG34	0	0.067042	0	0	3546.295	0	0	0	0
SCTG35	0	0	0	0	12377.87	0	0	0	0
SCTG36	0	0	0	0	6003.092	0	0	0	0
SCTG37	0	0	0	0	1785.718	0	0	0	0
SCTG38	0	0	0	0	3133.745	0	0	0	0
SCTG39	0	0	0	0	711.9008	0	0	0	0
SCTG40	0	0	0	0	1088.497	0	0	0	0
SCTG41	0	0	0	1.052104	29.10704	0	0	0	8.608894
SCTG43	0.06671	0.041744	0	1.37E-05	0.84238	0.041744	0	0	0.007408
SCTG99	0.06671	0.041744	0	1.37E-05	0.84238	0.041744	0	0	0.007408

Table 8-5 shows this reference in the opposite direction, indicating which industry consumes which commodities.

Table 8-5: Use coefficients by industry and commodity

Commodity	Agriculture	Construction	Health	Leisure	Manufacturing	Mining	Retail	Wholesale	Service
SCTG01	166.435	8.623	1.006	0.576	11.188	8.623	26.532	26.532	87.325
SCTG02	2.810	7.737	0.583	0.110	8.045	7.737	6.805	6.805	28.851
SCTG03	107.551	182.070	8.192	3.078	105.791	182.070	127.262	127.262	291.450
SCTG04	6.897	4.603	0.353	0.796	17.855	4.603	12.377	12.377	38.949
SCTG05	190.286	8.577	9.624	3.631	60.307	8.577	43.047	43.047	74.914
SCTG06	27.336	3.295	0.003	6.097	57.220	3.295	103.089	103.089	181.644
SCTG07	854.169	16.416	0.240	17.500	727.346	16.416	406.972	406.972	574.950
SCTG08	44.799	1.365	0.018	1.568	104.258	1.365	80.459	80.459	113.579
SCTG09	0	0	0	0	1	0	0	0	0
SCTG10	0.324	0.432	0	0.216	1.807	0.432	9.840	9.840	20.447
SCTG11	0.052	0.034	0	0.025	0.367	0.034	1.138	1.138	2.850
SCTG12	0.292	0.193	0	0.142	2.082	0.193	6.446	6.446	16.150
SCTG13	0.210	0.119	0	0.100	1.519	0.119	5.224	5.224	11.377
SCTG14	0.089	0.271	0	0.006	0.770	0.271	1.391	1.391	1.881
SCTG15	0	0	0	0	1	0	0	0	0
SCTG16	0	14.709	0.001	0.021	5.266	14.709	4.810	4.810	40.067
SCTG17	0	4.504	0.001	0.062	0.214	4.504	0.587	0.587	0.684
SCTG18	0	4.504	0.001	0.062	0.214	4.504	0.587	0.587	0.684
SCTG19	0	19.706	0.002	0.292	10.691	19.706	9.784	9.784	47.663
SCTG20	5.555	6.648	0.003	2.795	124.747	6.648	69.714	69.714	98.951
SCTG21	0.007	0.927	0.003	0.446	54.918	0.927	21.135	21.135	85.901
SCTG22	0	1.962	0	0.427	23.736	1.962	34.287	34.287	21.988
SCTG23	0	2.086	0.004	2.092	130.089	2.086	43.369	43.369	139.217
SCTG24	0	5.313	0.012	10.806	170.388	5.313	71.067	71.067	166.788
SCTG25	1.192	439.025	0.773	0.534	14.600	439.025	84.419	84.419	116.618
SCTG26	4.259	682.990	0.021	44.158	1013.975	682.990	364.036	364.036	492.067
SCTG27	0	13.153	0	0.753	24.780	13.153	14.936	14.936	18.074
SCTG28	0	130.718	0.022	12.418	262.769	130.718	273.317	273.317	271.229
SCTG29	0	3.585	0.421	18.980	63.615	3.585	74.467	74.467	354.167
SCTG30	1.170	1.011	0.001	4.451	44.320	1.011	41.063	41.063	103.563
SCTG31	0	9.376	0.005	8.515	79.061	9.376	117.192	117.192	138.139
SCTG32	0	25.823	0.009	7.868	107.547	25.823	231.599	231.599	225.025
SCTG33	0	13.984	0.020	20.462	189.055	13.984	170.017	170.017	414.986
SCTG34	0	6.001	0.019	16.051	206.897	6.001	139.227	139.227	329.660
SCTG35	0	26.945	0.128	24.231	1573.704	26.945	602.492	602.492	1576.753
SCTG36	0	9.136	0.003	4.341	487.881	9.136	316.719	316.719	294.676
SCTG37	0	1.969	0.012	5.082	149.155	1.969	61.745	61.745	159.730
SCTG38	0	4.902	0.036	19.310	353.619	4.902	111.608	111.608	418.334
SCTG39	0	1.783	0.006	5.501	103.988	1.783	36.846	36.846	84.256
SCTG40	0.547	1.445	0.007	6.542	64.723	1.445	42.580	42.580	122.633
SCTG41	0	0	0	0	0	0	0	0	1
SCTG43	0.054	0.064	0.001	0.010	0.244	0.064	0.144	0.144	0.275
SCTG99	0.054	0.064	0.001	0.010	0.244	0.064	0.144	0.144	0.275

The subsequent disaggregation from counties to zones within the MSTM study area follows the same methodology as the disaggregation from FAF zones to counties. As fewer employment categories are available at the MSTM SMZ level, make/use coefficients of Table 8-4 and Table 8-5 were aggregated from eleven to four employment categories. Equations 5, 6 and 7 were used accordingly for the disaggregation from counties to SMZ.

The disaggregated commodity flows in short tons need to be transformed into truck trips. Depending on the commodity, a different amount of goods fits on a single truck. FAF provides payload factors that were applied to convert flows from tons into trucks. FAF distinguishes five truck types, which were aggregated to the two truck types used in this model (single-unit [FAF category 1] and multi-unit trucks [FAF categories 2 to 5]). First, the share of truck types by distance class is calculated based on Table 8-6.

Table 8-6: Share of truck type by distance class

Minimum Range (miles)	Maximum Range (miles)	Single Unit	Truck Trailer	Combination Semitrailer	Combination Double	Combination Triple
0	50	0.793201	0.070139	0.130465	0.006179	0.0000167
51	100	0.577445	0.058172	0.344653	0.019608	0
101	200	0.313468	0.045762	0.565269	0.074434	0.000452
201	500	0.142467	0.027288	0.751628	0.075218	0.002031
501	10000	0.06466	0.0149	0.879727	0.034143	0.004225

For every truck type, tons are converted into tons separately. As an example, Table 8-7 shows payload factors for single-unit truck[24]. These payload factors describe how many single-unit trucks are used to carry one ton of this commodity on the average. Multiplying these values with the tons traveling provides number of trucks needed to carry this flow. The nine body types (auto, livestock, bulk, flatbed, tank, day van, reefer, logging and other) are not used further but aggregated to a single truck type, in this case single-unit trucks.

Table 8-7: Payload factors for single unit trucks by commodity

Commodity	Auto	Livestock	Bulk	Flatbed	Tank	Day Van	Reefer	Logging	Other
1	0	0	0.0066	0.04922	0.00111	0.00419	0.00173	0	0
2	0	0	0.02675	0.0086	0.00103	0.00032	0.00003	0	0.00003
3	0	0	0.01069	0.01981	0.00102	0.00996	0.00942	0	0.00147
4	0	0	0.01463	0.02657	0.00562	0.00334	0.00137	0	0.00034
5	0	0	0.00004	0.00089	0	0.03835	0.04837	0	0.00033
6	0	0	0	0.00025	0	0.15767	0.00216	0	0.00011
7	0	0	0.00001	0.00032	0.00073	0.02096	0.02048	0	0.02192
8	0	0	0	0.00002	0	0.02133	0.00286	0	0.02956
9	0	0	0	0	0	0.06785	0.04242	0	0.01498

[24] Payload factors for all FAF truck types can be found at http://faf.ornl.gov/fafweb/Data/Freight_Traffic_Analysis/faf_fta.pdf, pages A-1 to A-5

10	0	0	0.01399	0.01865	0.00029	0.00115	0	0	0.00185
11	0	0	0.02362	0.00638	0	0.00107	0	0	0.00058
12	0	0	0.02337	0.00292	0	0	0	0.00002	0.00034
13	0	0	0.02393	0.00255	0.00119	0.0008	0.00002	0	0.00048
14	0	0	0.01773	0.01261	0	0	0	0	0
15	0	0	0.01973	0.00307	0	0	0	0	0.001
16	0	0	0.00685	0.02455	0.01041	0.00086	0	0	0.01333
17	0	0	0	0.00186	0.02298	0.02755	0	0	0.00225
18	0	0	0.00026	0.00328	0.03386	0.00038	0	0	0.00261
19	0	0	0.00116	0.01074	0.0466	0.00273	0	0	0.00122
20	0	0	0.00171	0.02421	0.0146	0.01697	0	0	0.00266
21	0	0	0	0	0	0.10537	0.0122	0	0
22	0	0	0.01074	0.00974	0.01882	0.00302	0	0	0.00063
23	0	0	0.00145	0.01277	0.00987	0.03153	0	0	0.00539
24	0	0	0.00109	0.04904	0.00199	0.04913	0.00147	0	0.00863
25	0	0	0.0177	0.0167	0	0.00013	0	0.00831	0.00291
26	0	0	0.01437	0.03091	0.00002	0.01721	0	0.00017	0.00205
27	0	0	0	0.00142	0	0.07422	0	0	0
28	0	0	0.00262	0.00222	0	0.06609	0.00109	0	0.00223
29	0	0	0	0.00909	0	0.0857	0	0	0.00038
30	0	0	0.00154	0.0146	0	0.09299	0.00181	0	0.00251
31	0	0	0.00404	0.00588	0.00034	0.00436	0	0	0.01456
32	0	0	0.00076	0.06023	0	0.01594	0	0	0.01038
33	0	0	0.004	0.03186	0.00005	0.02246	0	0.00005	0.02908
34	0	0	0.00271	0.03187	0	0.03959	0	0.00002	0.00814
35	0	0	0.00033	0.01488	0	0.08017	0.00164	0	0.01258
36	0	0	0.00041	0.0073	0	0.00756	0	0	0.0548
37	0	0	0.00649	0.0228	0	0.00782	0	0	0.0141
38	0	0	0.00064	0.04872	0	0.11375	0	0	0.0006
39	0	0	0.00007	0.00432	0	0.11805	0.00166	0	0.00382
40	0	0	0.00027	0.01702	0.00117	0.07196	0.00051	0	0.01452
41	0	0	0.01372	0.00869	0.00221	0.00069	0.00011	0	0.01908
42	0	0	0.00215	0.01208	0.02291	0.00117	0	0	0.00181
43	0	0	0	0.00415	0	0.09378	0	0	0

Furthermore, an average empty-truck rate of 19.36 percent of all truck miles traveled (estimated based on U.S. Census Bureau data[25]) was assumed. As FAF provides commodity flows, empty trucks need to be added. Furthermore, the empty truck model takes into account that commodity flow data may be imbalanced. For example, to produce one ton of crude steel, 1.4 tons of iron

[25] http://www.census.gov/svsd/www/sas48-5.pdf

ore, 0.8 tons of coal, 0.15 tons of limestone and 0.12 tons of recycled steel are commonly used[26], i.e. commodity flows into and out of such a plant are highly imbalanced. While it is reasonable to assume that commodity flows are imbalanced, trucks are assumed to always be balanced, i.e. the same number of trucks is assumed to enter and leave every zone in the long run. Figure 8-4 shows a simplified example of flows between three zones. Blue arrows show truck flows based on commodity flows that are imbalanced, and red arrows show necessary empty truck trips to balance the number of trucks entering and leaving every zone.

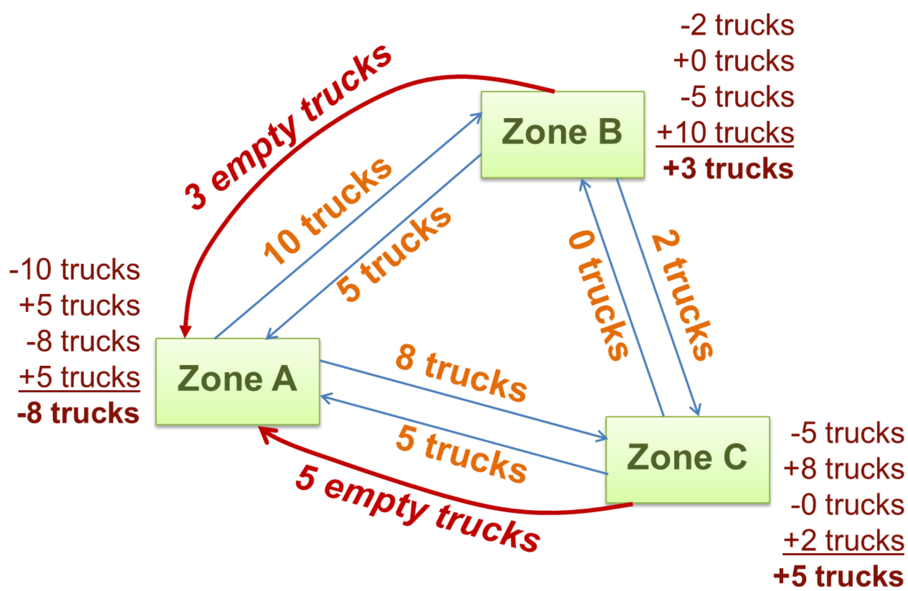


Figure 8-4: Example of imbalanced commodity flows (blue) and required empty trucks (red)

The concept of the empty truck model is shown in Figure 8-5. All zones that have a positive balance of trucks (i.e. more trucks are entering than leaving the zone based on commodity flows) need to generate empty trucks, and their number of excess trucks are put into the empty truck trip matrix as row totals (purple cells in Figure 8-5). Zones with a negative balance (i.e. more trucks are leaving than entering the zone based on commodity flows) need to attract empty truck trips, and their balance is put (as a positive number) as column totals into the empty truck trip matrix (yellow cells in Figure 8-5).

[26] http://worldsteel.org/dms/internetDocumentList/fact-sheets/Fact-sheet_Raw-materials2011/document/Fact%20sheet_Raw%20materials2011.pdf

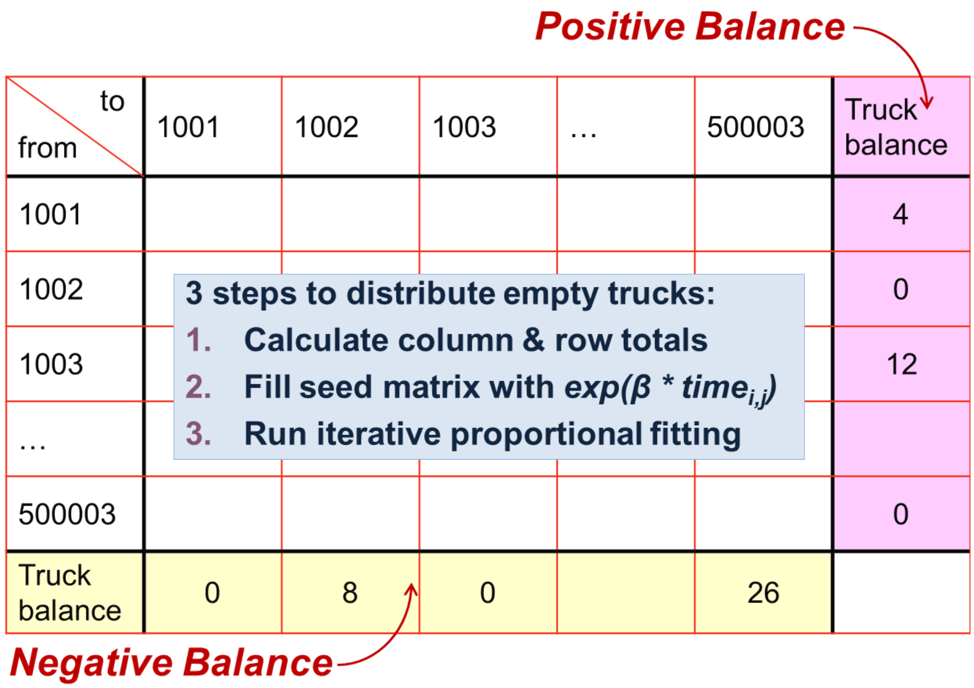


Figure 8-5: Matrix of empty truck trips

The cells within the empty truck trip matrix are filled with an impedance value calculated by a gravity model. It is assumed that empty trucks attempt to pick up another shipment in a zone close by, thus the travel time is used to calculate the impedance:

$$friction_{i,j} = \exp\left(\beta \cdot \mathrm{d}_{\mathrm{i,j}}\right) \qquad (11)$$

where $friction_{i,j}$ is the friction for empty truck trips from zone i to zone j
β is the friction parameter, currently set to -0.1
$d_{i,j}$ is the distance from zone i to zone j

A matrix balancing process is used to distribute empty truck trips across the empty truck trip matrix. Empty trucks are balanced separately for single-unit and multi-unit trucks. These empty trucks are added to the truck trip table of loaded trucks. The first and the second row in Table 8-8 show the number of trucks generated based on commodity flows and the number of trucks generated to balance flows into and out of every zone. The number of empty truck trips necessary to balance truck trips by zone is significantly lower than the 19.36 percent empty trucks according the U.S. census bureau. Thus, another 17.2 percent of empty trucks needs to be added to account for the larger number of observed empty truck trips. These additional empty truck trips are added globally, i.e. all truck trips are scaled up to match the observed empty truck trip rate.

Table 8-8: Number of long-distance trucks generated nationwide

Purpose	SUT	MUT	Share

Trucks based on commodity flows (FAF3)	348,940	1,146,330	80.6%
Trucks returning empty for balancing	9,512	31,103	2.2%
Additional empty truck trips (Census data)	74,295	244,073	17.2%
Total trucks trips	432,747	1,421,506	100.0%

This is an interesting finding by itself. If all trucking companies were perfectly organized and cooperated on the distribution of shipments between trucks that are available close by, only 2.3 percent empty truck trips would be necessary. But because there is competition between trucking companies and because of imperfect information about available shipments, a much higher empty truck rate is observed in reality. Granted, this is a simplified empty-truck model, and the 2.3 percent empty-truck rate may not be achievable for two reasons: First, the model works with fractional numbers, i.e. the model may send 0.5 trucks from zone a to zone b, which is acceptable as the model simulates an average day but not possible in reality. Secondly, only two truck types are distinguished. It might be considered in future phases of this project to explicitly handle truck types, such as flatbed, livestock or reefer trucks.

Finally, yearly trucks need to be converted into daily trucks to represent an average weekday. As there are slightly more trucks traveling on weekdays than on weekends, a weekday conversion factor needs to be added.

$$trucks_{daily} = \frac{trucks_{yearly}}{365.25} \cdot \frac{AAWDT}{AADT} \qquad (12)$$

where: $trucks_{daily}$ is the number of daily truck trips
$trucks_{yearly}$ is the number of yearly truck trips
AAWDT is the average annual weekday truck count
AADT is the average annual daily truck count

Based on ATR (Automatic Traffic Recorder) truck count data the ratio *AAWDT/AADT* was estimated to be 1.02159, meaning that the average weekday has just 2 percent more long-distance truck traffic than the average weekend day. The resulting truck trip table with two truck types, single-unit and multi-unit trucks, is added to the multi-class assignment.

8.5 Model Validation

The truck model was originally developed by Bill Allen for BMC and MWCOG. It made heavy use of geographically specific k-factors, which were all removed in the MSTM application. As a rigorous validation of the BMC or MWCOG truck model was never published, it is unknown how well the model performed when all k-factors were included.

For commercial vehicles and trucks, no survey data were available. Instead, data reported in the BMC and MWCOG reports were used to estimate the reasonability of the MSTM model output. The bright red bars show the model output of MSTM in phase 1, and the dark red show recalibrated the model output of phase 2 (Figure 8-6). Green bar show target data reported in the MWCOG truck model report, and salmon and blue colored bars show the model output of the BMC and MWCOG models. It should be cautioned to consider the reported trip length of the

BMC and MWCOG models as target data, as no observed data exists. Overall, the longer trip lengths may be due to the larger study area of MSTM. No further changes were made to the commercial vehicle and truck models in phase 3.

The current truck model is based on the BMC truck model, which mostly uses parameters of the FHWA Quick Response Freight Manual (QRFM). Those parameters were developed from a truck survey for Phoenix in 1992. These parameters are not only outdated but also originate from an urban form that is very different to Maryland. For future model updates, it would be desirable to conduct a truck survey to improve these modules by using local and recent data.

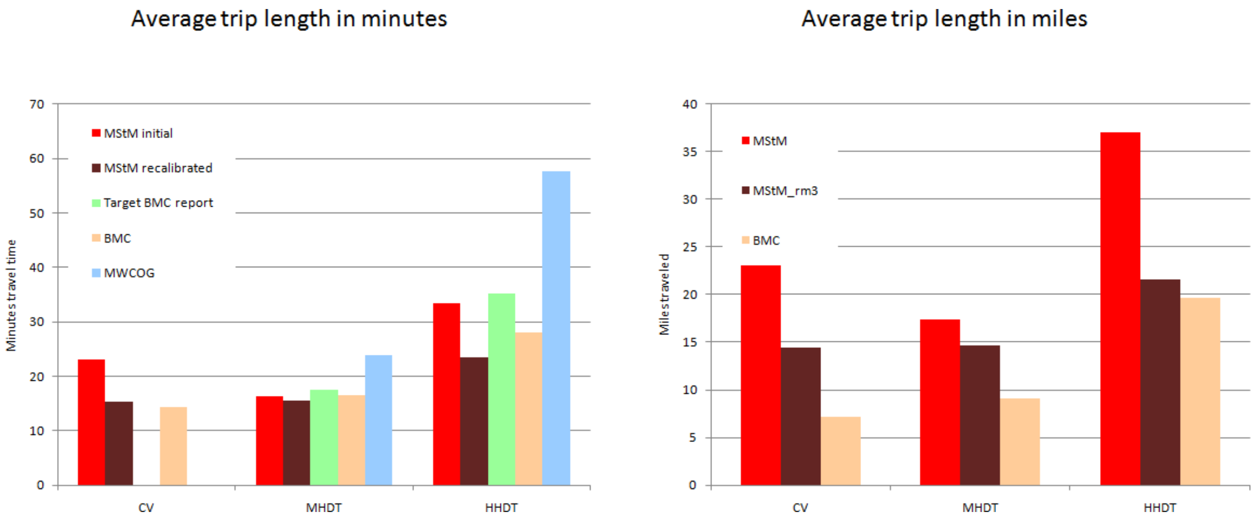


Figure 8-6: Comparison of average trip length

Figure 8-7 shows the percent root mean square error for five different volume classes. It is common that the simulation of trucks does not perform as well as the simulation of autos. There is too much heterogeneity in truck travel behavior, and a large number of trips are not A-to-B and B-to-A trips but rather tours from A-to-B-to-C-to…to-Z, which are particularly difficult to model in trip-based approaches. Furthermore, there is no truck survey that was used to estimate truck trip rates. The rates applied are borrowed from the BMC truck model, which in turn copied and slightly modified these rates from the Phoenix truck survey from1992. The person travel demand model, in contrast, uses a local survey for the BMC/MWCOG region from 2007, and thus, provides local recent data to calculate trip rates.

In light of these general difficulties in truck modeling, the MSTM truck model performs reasonably well. While the midrange from 500 to 5,000 observed truck trips results in a %RMSE of just over 100%, the highest volume range (>=5,000 observed trucks) with 337 truck counts achieves a fairly good %RMSE (by truck modeling standards) of 52%. It is expected that future phases could improve the truck model quite a bit a conducting a local truck survey and by splitting the four employment types currently used in MSTM into a larger number of types (such as ten employment types).

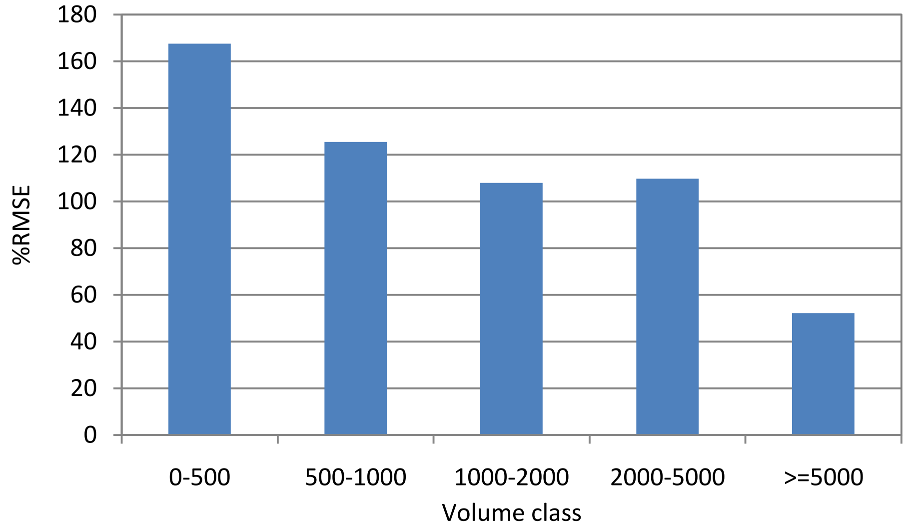


Figure 8-7: Truck percent root mean square error by volume class

9 Trip Assignment

9.1 Model Integration and Time-of-Day Processing

Temporal allocation of the person, commercial and truck vehicle trips was accomplished by applying factors to the respective daily trip matrices to derive peak (AM and PM) and off-peak (MD and NT) trip matrices for network assignment. The process was taken from the BMC models. Factors for person trips are derived from household survey data on a production-to-attraction (PA) basis for home-based travel for application to person trip matrices in PA format. These factors produce directional flow matrices replicating observed average peaking characteristics. Factors for non-home-based person trips are derived on an OD basis and applied to the corresponding OD trip matrices. Vehicle trips are assigned by time of day period. Separate assignments were done for the AM and PM peak periods and for the rest of the day combined. Transit trips were assigned on a daily basis with work trip assignment based on peak service characteristics and assignment of all other trip based on off-peak service characteristics. BMC factors for auto person trips and the drive access component of transit drive-access trips are given in Table 9-1. They sum to 100% by purpose for the P-A and A-P directions individually.

Table 9-1: Person trip time of day factors

Purpose	PA_AM	AP_AM	PA_MD	AP_MD	PA_PM	AP_PM	PA_NT	AP_NT
HBW1	55.27%	3.61%	18.96%	27.45%	5.57%	45.00%	20.20%	23.95%
HBW2	60.72%	2.30%	14.26%	20.22%	4.44%	53.03%	20.57%	24.45%
HBW3	63.56%	1.34%	11.57%	19.98%	3.32%	60.17%	21.54%	18.51%
HBW4	68.04%	1.50%	9.45%	18.62%	2.42%	61.94%	20.09%	17.94%
HBW5	71.47%	0.69%	9.10%	15.98%	1.91%	64.32%	17.52%	19.01%
HBS1	18.44%	3.27%	50.53%	43.71%	19.04%	29.45%	11.99%	23.58%
HBS2	17.31%	2.80%	42.50%	38.25%	21.43%	28.27%	18.76%	30.68%
HBS3	16.04%	2.53%	39.67%	37.77%	26.57%	27.63%	17.72%	32.07%
HBS4	15.55%	2.00%	36.14%	33.34%	26.83%	28.48%	21.48%	36.18%
HBS5	17.91%	2.23%	32.72%	33.73%	24.68%	26.43%	24.69%	37.61%
HBO1	38.17%	9.31%	38.69%	39.86%	13.02%	28.33%	10.12%	22.50%
HBO2	32.41%	8.72%	35.66%	32.05%	17.06%	27.42%	14.87%	31.81%
HBO3	31.51%	10.08%	33.74%	31.98%	20.40%	27.24%	14.34%	30.70%
HBO4	31.49%	9.15%	30.86%	27.91%	22.04%	30.56%	15.61%	32.38%
HBO5	31.69%	9.72%	28.98%	27.47%	22.71%	31.08%	16.62%	31.73%
HBSc	89.92%	0.21%	4.11%	62.86%	2.79%	29.16%	3.19%	7.77%
NHBW	4.62%	29.34%	50.44%	58.38%	38.88%	5.89%	6.07%	6.39%
OBO	7.46%	9.08%	57.40%	55.57%	21.16%	22.55%	13.97%	12.80%

Time of Day (TOD) factors for regional and statewide trucks are shown in Table 9-2. These are derived from TOD factors reported for the BMC commercial and truck model.

Table 9-2: Regional and statewide truck time of day factors

Assignment Period (P->A Only)	Com. Veh.	MHDT	HHDT	Regional Trucks	Regional Autos
AM 6:30-9:30	16.982	16.982	16.982	20	Defined explicitly by the NELDT model
Midday 9:30a-3:30p	42.845	42.845	42.845	50	
PM 3:30-6:30	15.426	15.426	15.426	20	
Night 6:30p-6:30a	24.747	24.747	24.747	10	
Total	100.00%	100.00%	100.00%	100.00%	100.00%

9.2 Highway Assignment (Autos and Trucks)

Bridge crossings are a particular challenge to calibrate. On the one hand, bridges are a bottlenecks for many trips, and on the other hand research in travel demand shows that rivers form a mental barrier. To the model, a bridge crossing simply represents a link on the network as any other road, and a trip across the river is as likely in the model as a trip on the same side of the river. In reality, however, bridge crossings tend to form a mental barrier. Many trips tend to have their origin and destination on the same side of the bridge, as a river forms a natural border that tends to limit travel across. This is particular true for the Potomac River, as for large parts this river also forms the border between Maryland and Virginia. To account for this psychological barrier, the destination choice model included a factor that impacted travel from one river zone to another. No further adjustment or factoring has been applied.

Figure 9-1 shows which bridges were analyzed. These bridges were chosen as count data were available and as they serve major traffic flows in the region.

Figure 9-1: Bridge crossings analyzed in MSTM

In Figure 9-2, green bars show the count data, and the colored bars show simulated volumes of different vehicle classes. The Woodrow Wilson Bridge has less traffic in the simulation than suggested by count data, while the American Legion Bridge has more traffic than observed. It is possible that too many trips are taking the western part of the beltway for driving around Washington, while some of them should be using the eastern part of the beltway. Given the high levels of congestion in the Washington DC area and an almost identical travel time when using the eastern or the western part of the beltway for many origin-destination pairs, this deviation appears to be acceptable.

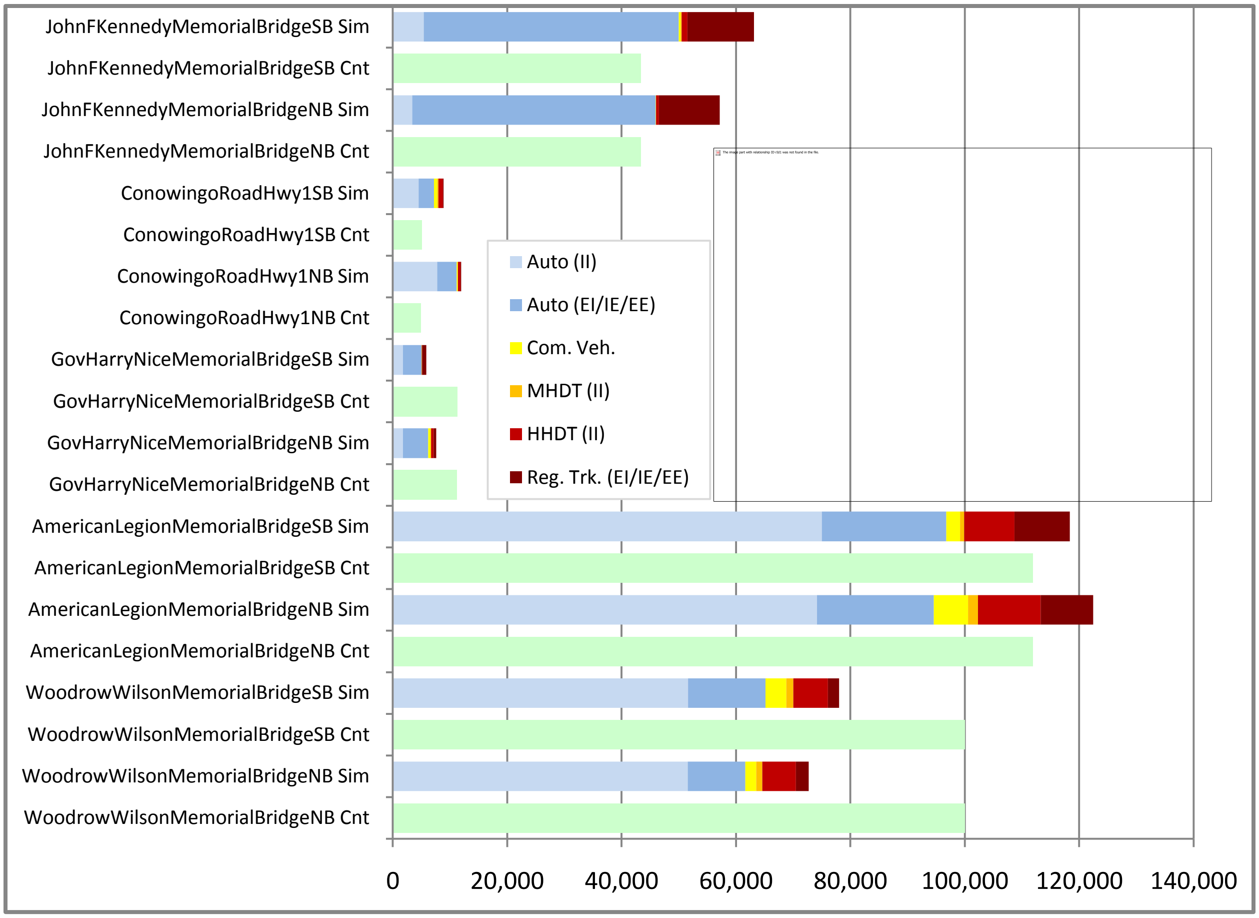


Figure 9-2: Validation of traffic volumes on selected bridge crossings

Figure 9-3 compares the MSTM model results with results from other statewide models for which detailed validation data were available to the authors. Percent Root Mean Square Error (Percent RMSE) of different volume ranges was used as the validation criteria.

The plot shows the Maryland model results in blue. There are two models, Ohio and Oregon, for which a lot of count data were available, and therefore, a very detailed analysis was feasible. In general, these two models have performed better than the MSTM model, which is mainly due to two reasons. For one, these two models were developed over more than a decade, and thus had more iterations to evolve than MSTM, which was developed over the course of approximately two years. Secondly, the geographies of Ohio and Oregon are easier to model than Maryland. Ohio and Oregon have a limited number of metropolitan areas, and density declines rapidly at the border of the study area. Much of Maryland, on the other hand, is covered by a huge Mega-Region that extends all the way from Boston, MA to Richmond, VA. Therefore, a statewide model for Maryland has to deal with a lot of through traffic, and there are a lot of local trips crossing the northern and southern border of the MSTM study area.

Task 91 in Figure 9-3 is a mix of several statewide models across the U.S. for which these validation data were available. Some of these models have performed better, while others performed worse. Overall, the validation of MSTM is within the range of many other statewide models.

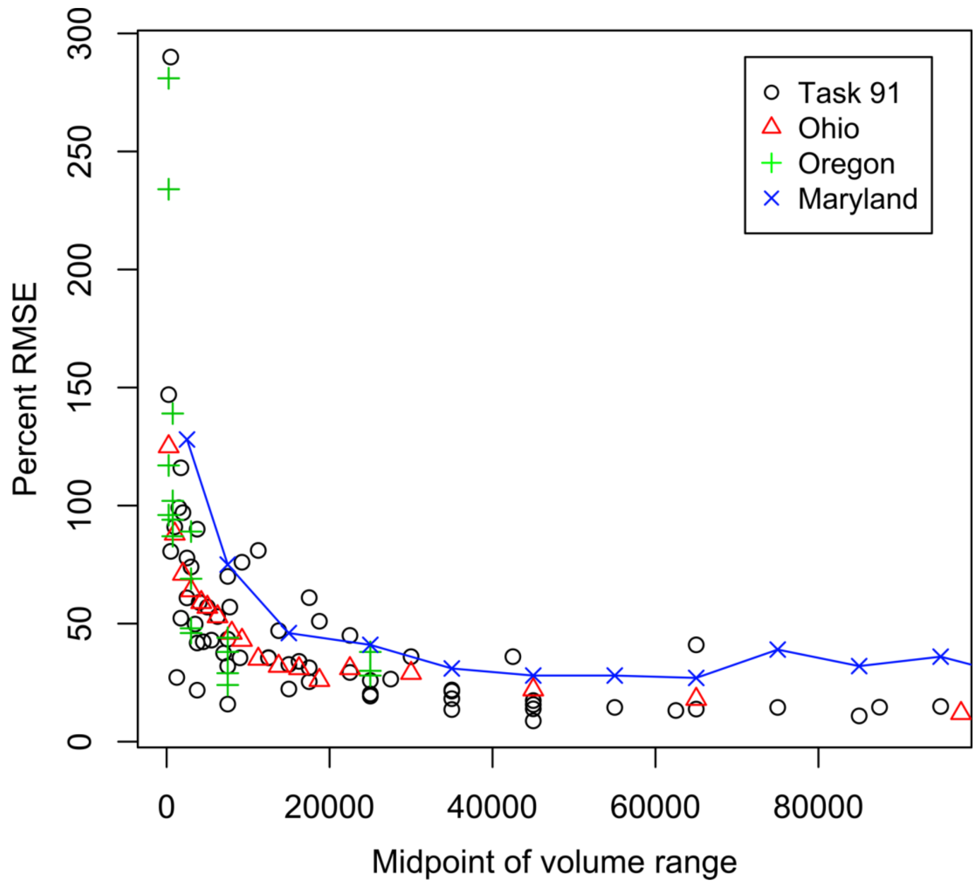


Figure 9-3: Comparison of MSTM with other statewide models

10 Implementation of a feedback loop

A crucial input for the model is travel time on the network. Initially, congested travel times were assumed based on free-flow speed, link length, area type and facility. Congested travel times were an exogenous input that did not change with congestion. To overcome this shortcoming, a feedback loop was implemented that uses travel times calculated by the assignment and feeds them back into trip generation. The procedure is visualized in Figure 10-1.

Transit skimming and transit assignment are not included in the feedback loop, as these two processes do not affect highway travel times, nor do transit travel times change with congestion. As these two transit modules are computationally relatively intensive, excluding them from the feedback accelerates a model run.

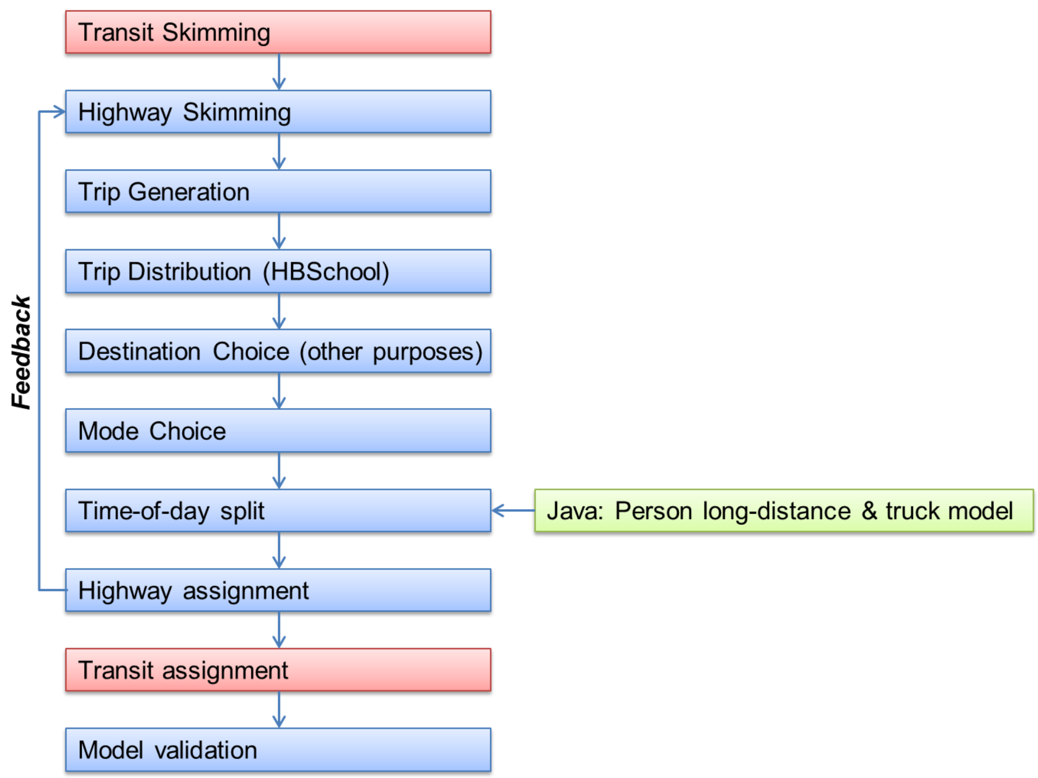


Figure 10-1: Feedback loop design

The initial skim values are calculated using free-flow travel time. All subsequent modules use these skim matrices. After the assignment has been completed, skim matrices are recalculated using the travel times generated in the assignment. To avoid oscillating model results, the new highway skims are not used directly but rather averaged with the previous skim values. By using the average between the previous skim values and the recalculated skim values, changes happen more gradually and the model is able to converge more quickly.

Figure 10-2 shows the convergence of the feedback loop by iteration. The x-axis shows the iteration, and the y-axis shows the percent root mean square error (%RMSE) between the skim values of two subsequent loop iterations. If the %RMSE is 0, the skim values did not change when using the speed of the latest two assignments. A non-zero %RMSE indicates that the resulting speed of the assignment has been different from the speed used to calculate the skim values. The blue line shows the %RMSE of a model setup in which the speed of the latest assignment is used to calculate the skim values directly. The %RMSE is reduced continuously over the first six iterations, and then starts oscillating around 12%. As a test, 75 iterations were run and the %RMSE did not improve much over 12%; therefore, the graphic only shows the first 16 iterations. The red line in Figure 10-2 shows the convergence of a feedback implementation, in which the revised skim matrices are averaged with the skim matrix from the previous iteration. The model does not oscillate and reaches convergence after a couple of iteration. The latter version has been implemented in the MSTM model.

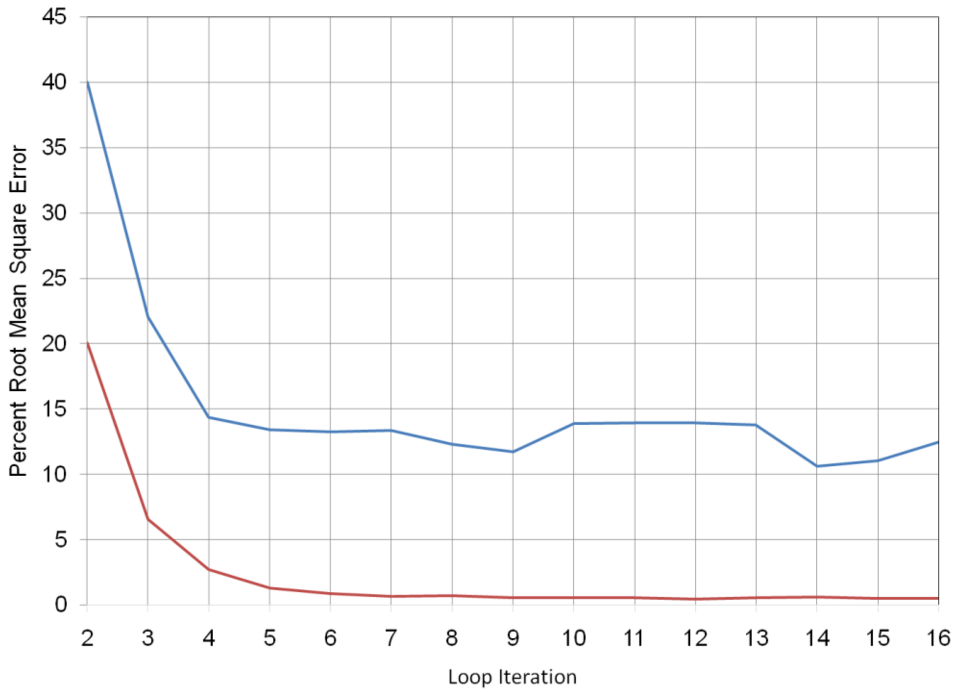


Figure 10-2: Feedback loop conversion with averaging (red) and without averaging (blue)

Given the insignificant changes after six iterations, the total number of iterations has been set to be six. Figure 10-2 suggests that six iterations is a good compromise between model convergence and run time. The current run time of the model is about 11 hours on a 12-core machine with 24GB of memory.

11 Calibration

The revisions implemented in phase IV of MSTM required recalibrating the model. The three elements that have changed, non-motorized share of trips, truck trips and feedback loop, have altered travel demand and congestion significantly enough that the model required fine-tuning. In particular, three elements where adjusted. First, the share of unreported trips for non-home-based trips was increased, secondly, the time-of-day choice was recalibrated, and lastly, the scaling factor on medium trucks and heavy trucks was fine-tuned.

11.1 Trip Rates

Acknowledging that household travel surveys tend to underreport trip making, the team concluded based on a literature review in phase III of this project that the trip generation rates are likely to underrepresent actual trip making. For lack of better data, the team assumed that people reported work trips correctly, but underreported all other trips. Trip rates of all trip purposes except home-based work (HBW) were scaled up by 20%.

Conventional wisdom further suggests that non-home based trips are even further underreported in household travel surveys. It is not uncommon that the trip from work to lunch and back to work is omitted when responding to a travel demand survey. Therefore, it was hypothesized in phase IV of MSTM that non-home based trips need to be scaled up by a factor of 1.4 instead of 1.2. Table 11-1 shows the scaling factors currently applied for every trip purpose.

Table 11-1: Trip rate scaling factors by trip purpose

HBW	HBS	HBO	HBSchool	NHBW	NHBO
1.0	1.2	1.2	1.2	1.4	1.4

Though values in Table 11-1 are partly based on literature review completed in phase III and partly based on conventional wisdom, no "true" scaling factor could be developed. While the guesstimate is based on a commonly accepted shortcoming of household travel surveys and improved the model results, it would be desirable to develop factors that are driven by observed travel behavior. If trip making was tracked passively either by GPS data or cell phone data, true trip rates could be calculated that include all relevant trips. It is recommended to explore this option in future phases of the MSTM project.

11.2 Time-of-Day Choice

The model distinguishes four time periods: AM Peak, Midday, PM Peak and Night. The time-of-day choice for person travel is provided by the survey. These parameters were fine-tuned to better match the observed time-of-day shares.

For long-distance truck travel, no data were available. In lack of better assumptions, long-distance trucks were spread out evenly across 24 hours in previous phases of MSTM. This appeared to be oversimplifying truck travel. In this phase, the time-of-day split for long-distance trucks was revised based on the assumption that more trucks travel during the day time than at night. The new time-of-day split assumes that 20% of all long-distance trucks travel in the AM Peak, 50% during Midday, 20% during PM Peak and another 10% during the night.

11.3 Truck Trips

An important improvement of phase VI was to replace truck trips generated by FAF^2 with truck trips generated by FAF^3 (compare section 8.4). A comparison of FAF^2 and FAF^3 trips revealed that commodity flows reported in these two datasets have changed dramatically. According to FHWA, who releases these data, flows have improved significantly in FAF^3. The validation shown in section 0 confirms this notion by showing an improved match between counts and truck volumes.

However, implementing FAF^3 for long-distance truck travel required adjustments for short-distance trucks. Short-distance trucks are based on the Quick Response Freight Manual (QRFM)[27] published by FHWA. QRFM is based on a truck survey conducted in Phoenix in 1992, and as such it was expected that the model has to be fine-tuned when being implemented for the state of Maryland. Scaling factors were implemented to adjust short-distance trucks. After implementing FAF^3, both medium and heavy short-distance trucks were scaled down with a factor of 0.8 across the SMZ study area. Lacking a local truck survey, no further refinements were made.

[27] Beagan, D., Fischer, M., Kuppam, A. (2007) Quick Response Freight Manual II. FHWA: Washington, D.C.

12 Validation

In model validation, model results are compared to independent observed data, i.e. data that have not been used in model development. If the model results resemble independent observed data, the model is assumed to reasonably represent real-world travel behavior. This section is organized in two parts. First, travel demand after trip generation, destination choice, mode choice and time-of-day choice are compared to the survey. This is not a validation in the traditional sense, as the survey was used in model development. However, comparing survey results with model results ensures that the model was specified correctly and represents observed travel demand reasonably well. The second section shows the validation of assignment results. The assignment is compared with traffic counts and HPMS VMT estimates, which are considered to be independent observed data, and thus count as true model validation.

12.1 Modeled Travel Demand and Survey Summaries

The number of trips was followed through different modules to ensure that no trips get lost in the process of the model. Figure 12-1 shows the number of trips by purpose across different modules. A significant drop in number of trips happens in mode choice, when person trips are converted into vehicle trips. Otherwise, the number of trips is stable across different modules confirming that all trips are carried through. Given the insufficient number of sample records, which resulted in large expansion factors that had to be capped, the survey could not be used to calculate a total of observed number of trips.

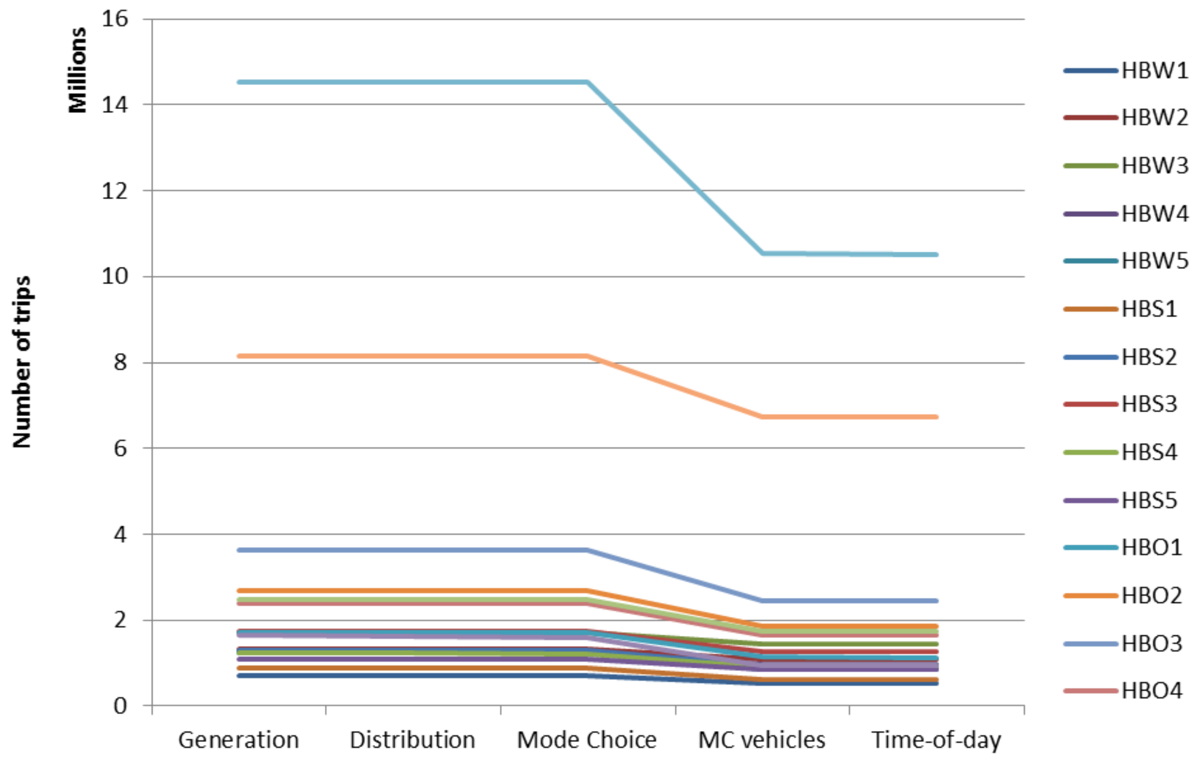


Figure 12-1: Number of trips generated by purpose

The modeled average trip length was compared to the observed average trip length. Figure 12-2 shows a bar chart comparing the two. While single trip purposes deviate to some degree from the survey, the overall pattern given by the survey is replicated by the model. Higher income groups tend to make longer trips, and the longest trip lengths are found for home-based work trips.

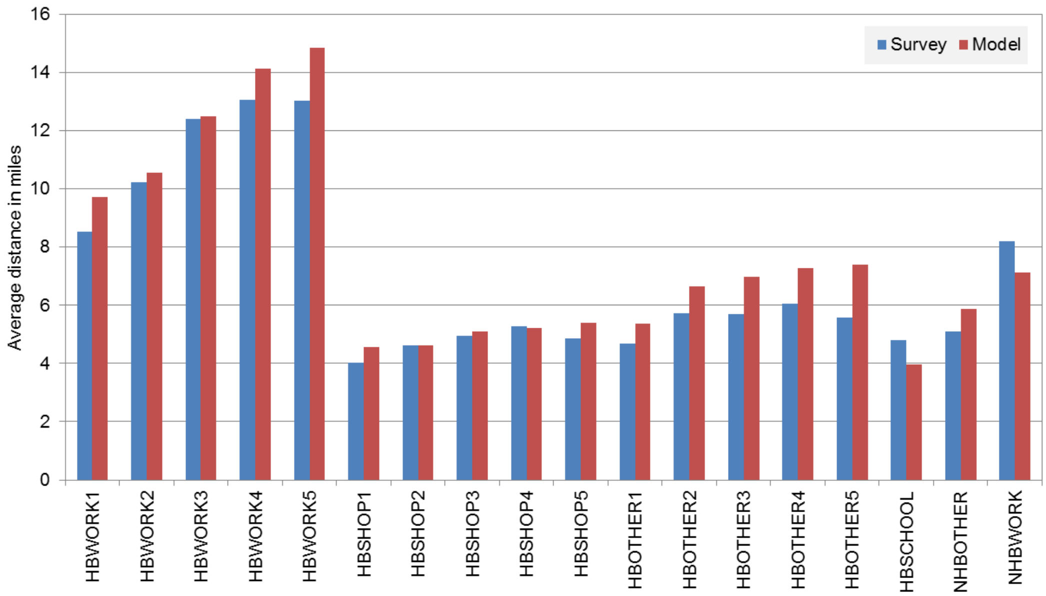


Figure 12-2: Average trip length observed in the survey and modeled by MSTM

To validate the modal split, the mode shares of drive-alone auto, shared-ride auto and transit trips were compared with the observed mode split from the survey. Figure 12-3 shows this mode split in a bar graph. Two rows need to be seen together to compare observed with modeled mode split. Overall, the model represents the mode split reasonably well. Considering that the mode choice model has to deal with mode choice across the entire SMZ study area with very different transit options and user market, the match between the model and the observed mode split is considered to be reasonable.

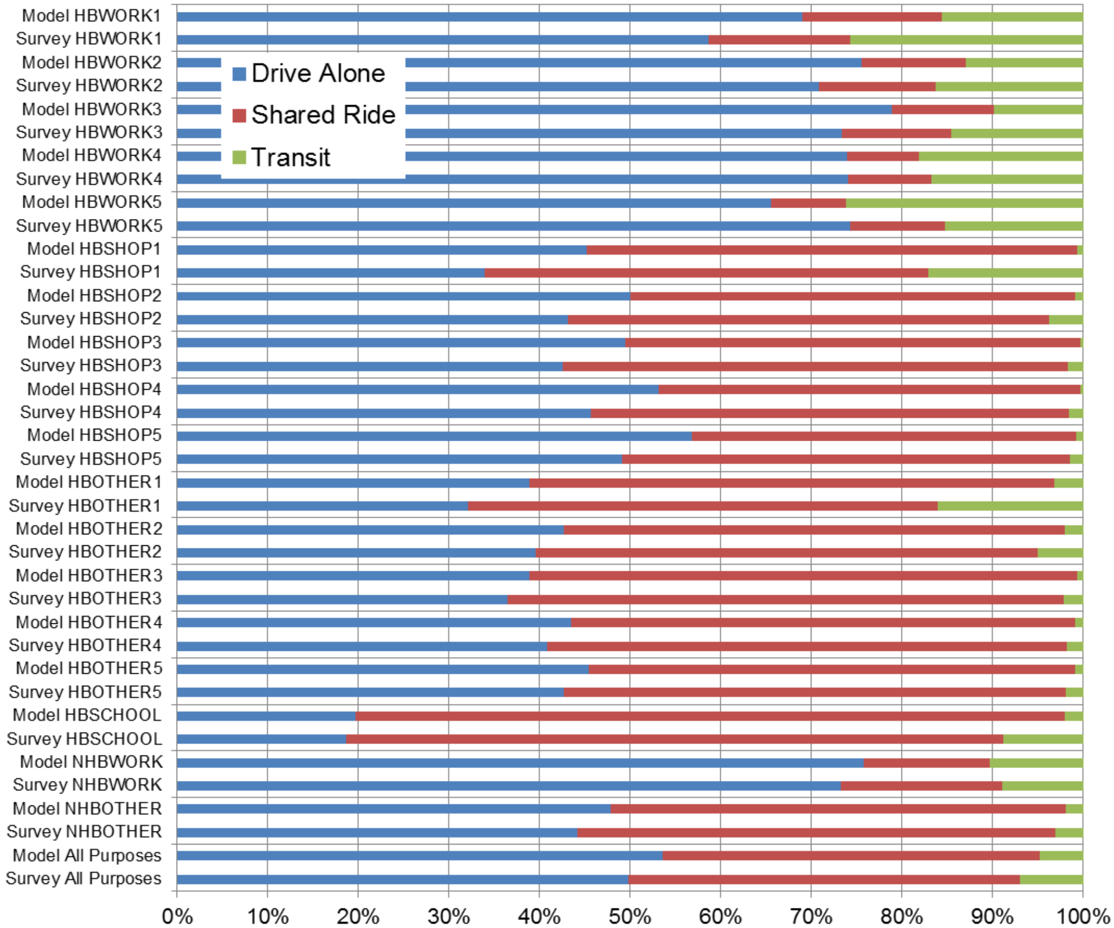


Figure 12-3: Validation of mode split

Finally, the time-of-day split between the model and survey was analyzed (Figure 12-4). For all purposes, the time-of-day split is deviating by less than two percent, and only school trips deviate by more than one percent. Given the uncertainties in every survey, these results suggest that the model is matching observed time-of-day split as well as possible.

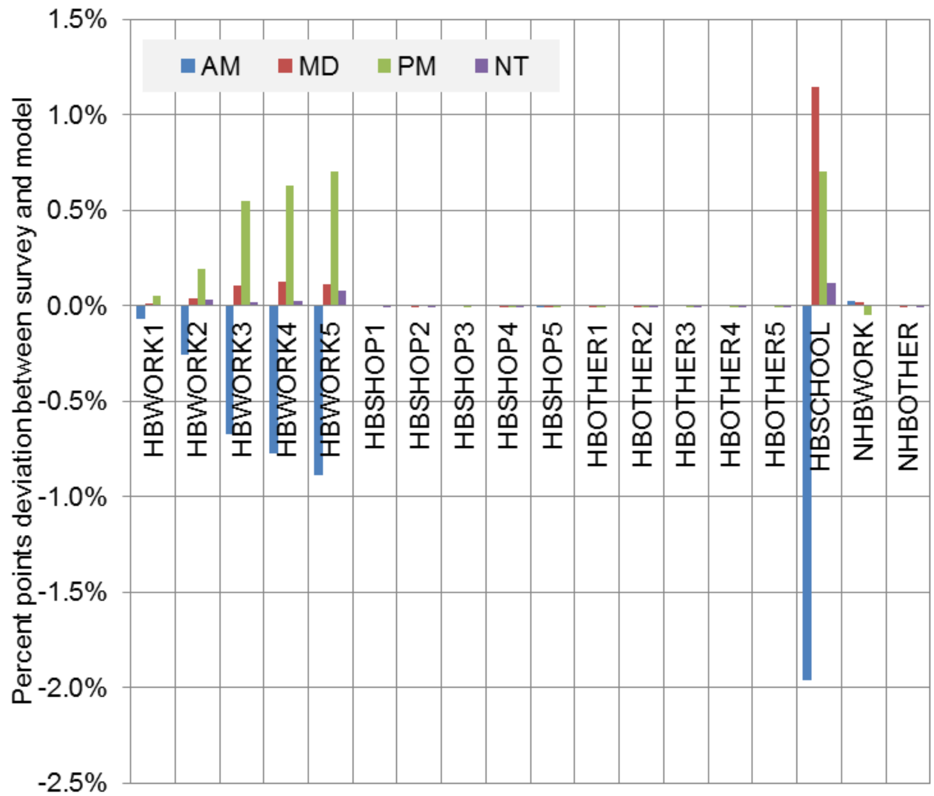


Figure 12-4: Difference in time-of-day choice between survey and model results

The comparison of the survey and the model results for trip length, mode split and time-of-day split suggest that the model replicates reasonably well the observed travel behavior.

12.2 Assignment Validation

True validation can only be performed with observed data that have not been used in the model development. Two datasets were available for this purpose: traffic counts and HPMS estimates of vehicle miles traveled by county.

Figure 12-5 compares the simulated volumes with count data in Maryland. Points were not expected to line up on the diagonal, as count data commonly have a 20% standard deviation from the average volume. Furthermore, the network and zone system of a statewide model are simplified, which reduces the ability to match count data. Nevertheless, the general pattern is represented fairly well. Across all count locations, a Root Mean Square Error (RMSE) of 3,763 is achieved, or a Percent Root Mean Square Error (%RMSE) of 25%. This is reasonable for a statewide model.

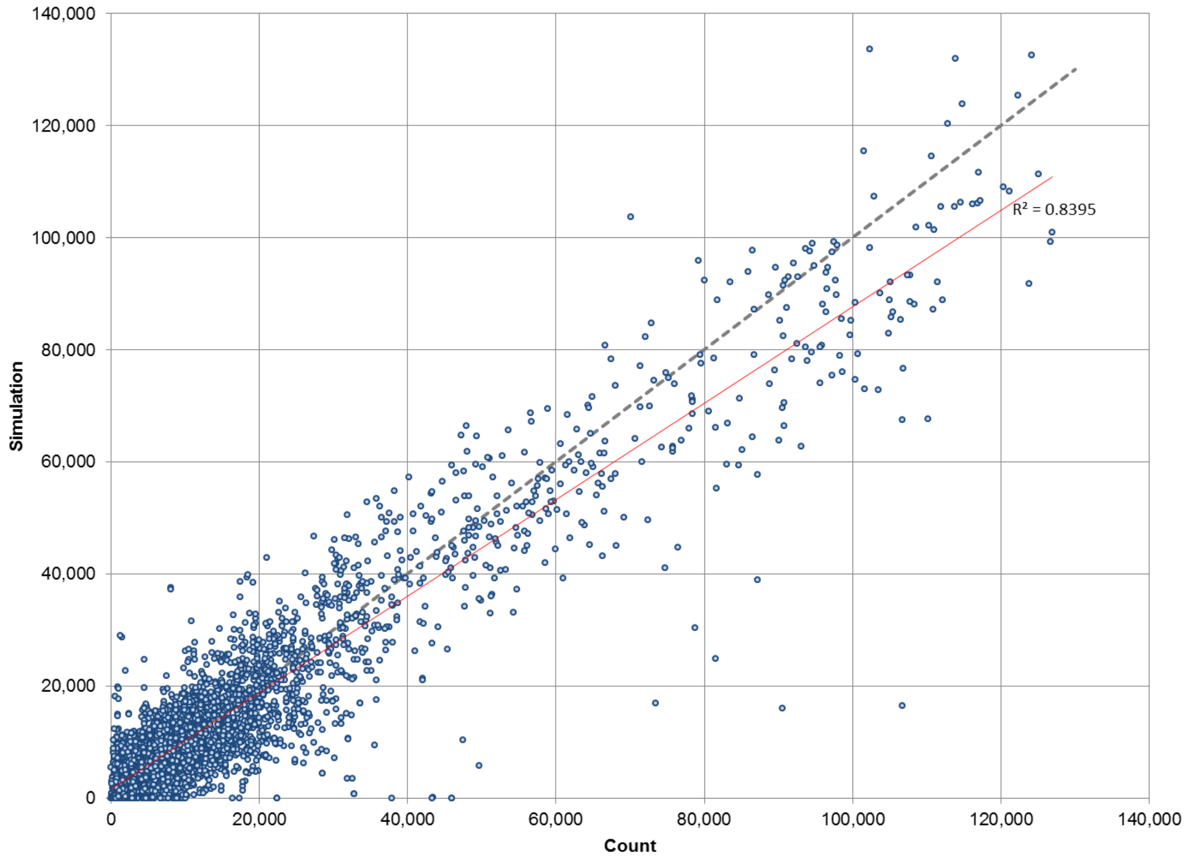


Figure 12-5: Comparison of counts with model volumes, all vehicles

To provide a spatial image of which links are over- and underestimated, Figure 12-6 shows graphically where the model deviates from the observed count data. While a model is not expected to replicate every count location, it is worth understanding which parts of the network are captured well by the model and where there is room for improvements. Green links are links where the model and the count data agree on the traffic volume within a range of -5,000 to +5,000 vehicles. Red links are those where the model volumes are larger than the count volumes, and blue links show where the model underestimates volumes. Most of the beltway around Baltimore and Washington is underestimated, while I-95 northeast of Baltimore tends to be overestimated. Otherwise, there is very little pattern to be recognized.

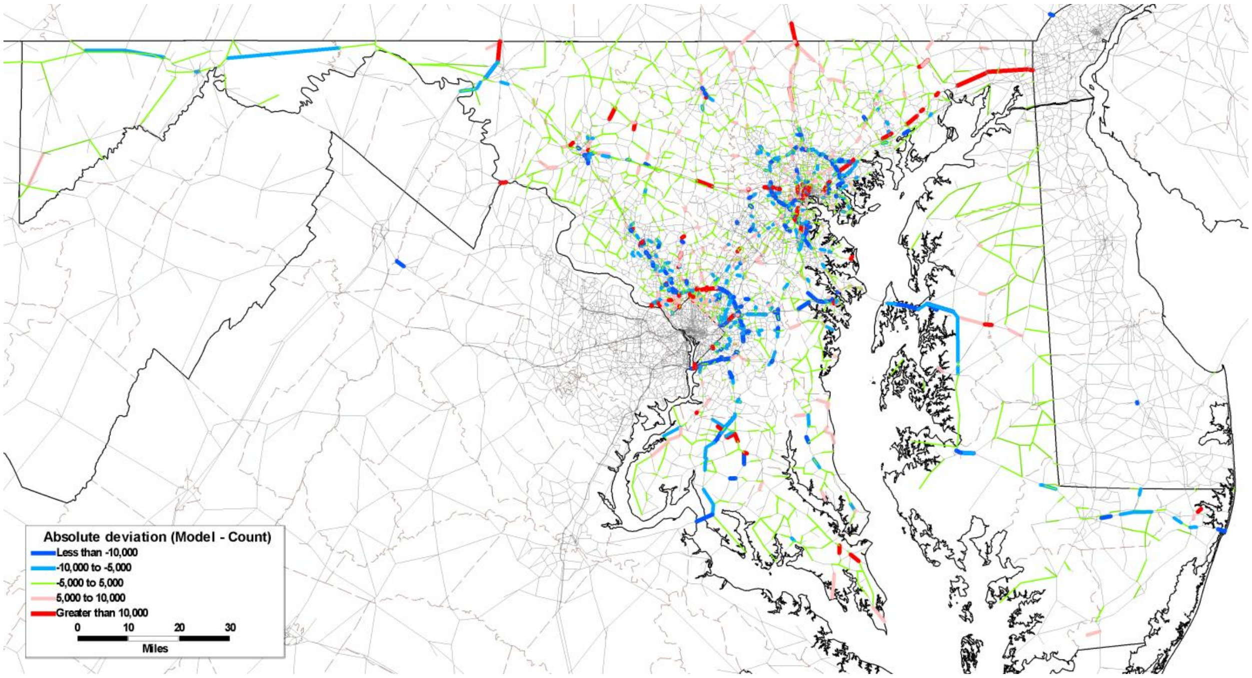


Figure 12-6: Geographic distribution of links over- and underestimated

It is common in model evaluation to not only look at single count locations but also at screenlines. Screenlines combine a series of count locations across major corridors, such as across parallel routes between Baltimore and Washington D.C. While single count locations may blur the picture of the overall model performance with too many points, screenlines help understanding whether the model is able to replicate major traffic flows between different regions. Figure 12-7 shows the validation across 61 screenlines that have been defined for MSTM. Every dot in this scatter diagram represents one screenline, which is an aggregation of several counts. The color indicates how many links on a given screenline actually have count data. Green dots show screenlines for which at least ¾ of all links have count data. Yellow dots are screenlines on which 50% to 75% of its links have count data, and red dots show screenlines with less than 50% of its links filled with counts. The green screenlines are considered to be reliable, while yellow and red screenlines are less informative given the higher uncertainty due to missing counts. Green screenlines show a close resemblance of model volumes and count data, and most of the yellow and red screenlines match count data quite well, too.

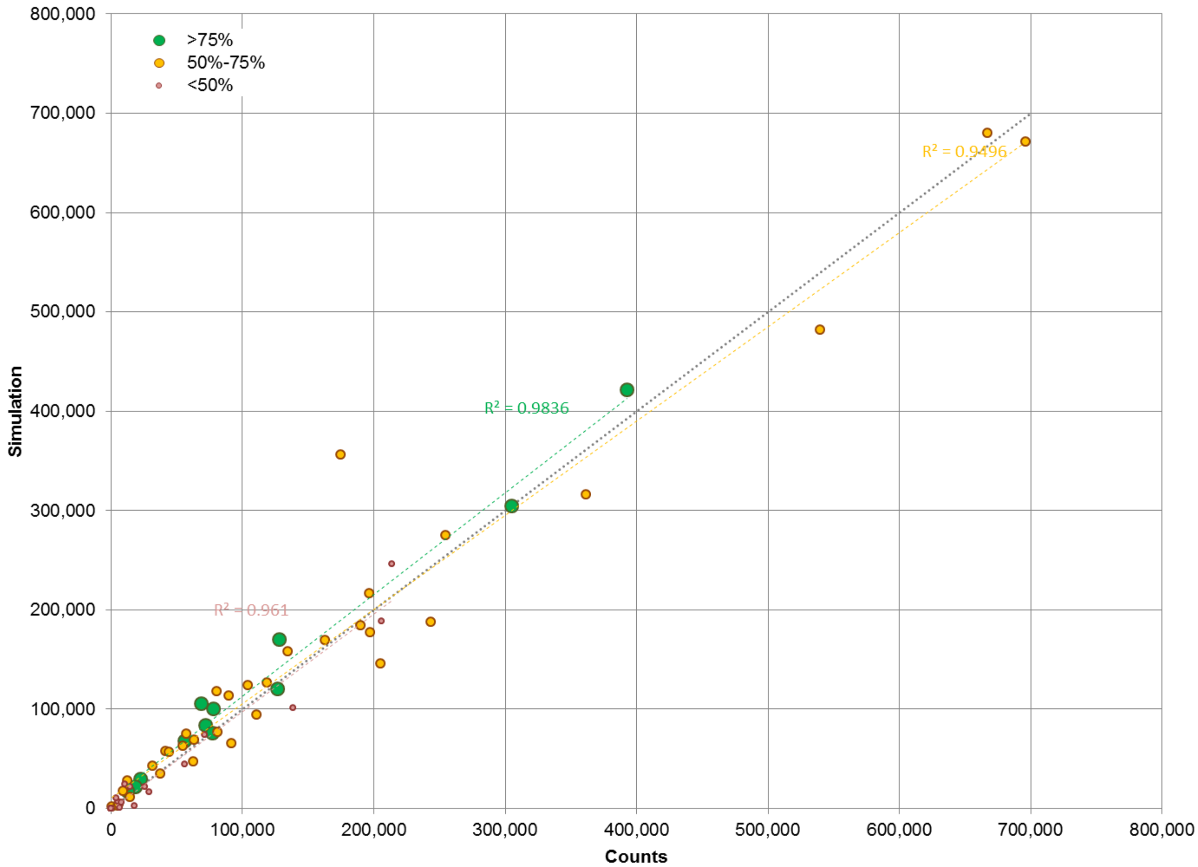


Figure 12-7: Validation by screenlines

It is common in travel demand modeling that autos perform better than trucks. This is mostly due to two reasons: First, person travel has been studied much more than truck travel, and therefore, the knowledge of travel behavior of autos it larger by an order of magnitude. Secondly, truck travel is generally assumed to be much more heterogeneous than auto travel. Trucks serve many different industries with different requirements, carry a large variety of goods, and there are many more truck types than auto types that could be relevant for travel behavior. Figure 12-8 shows that MSTM makes no exception here, truck travel matches count data less well than auto travel. However, in comparison to other truck models, the match is comparatively satisfying. A RMSE of 1,301 or a %RMSE of 77% was achieved. This is significantly better than the RMSE of 2,284 and the %RMSE of 135% that was achieved for trucks at the end of phase III of MSTM.

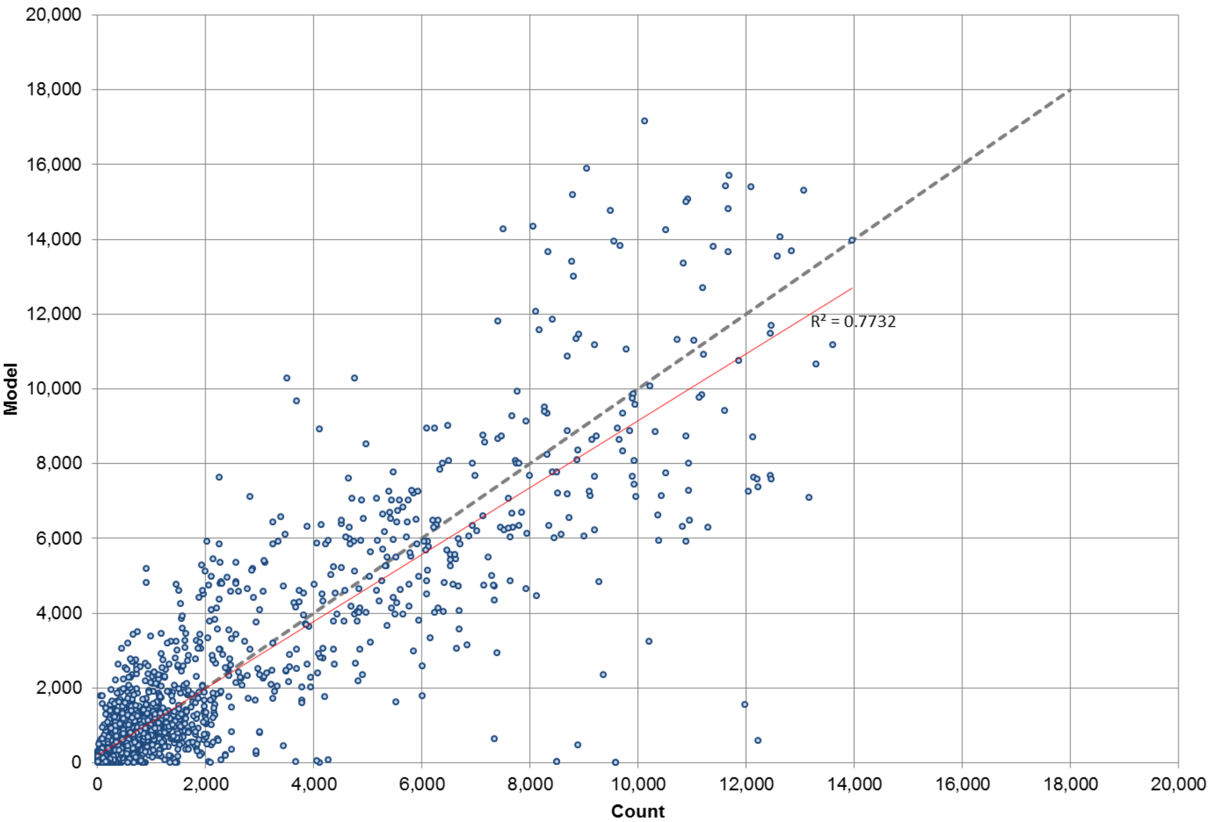
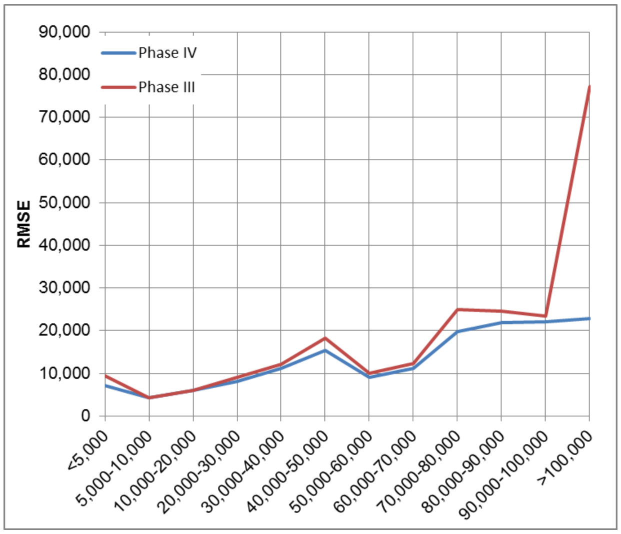
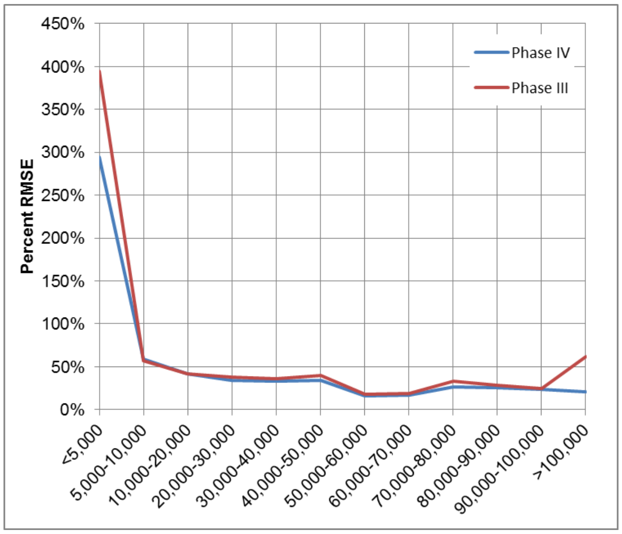


Figure 12-8: Comparison of counts with model volumes, trucks only

Figure 12-9 validates the model by volume class, and shows the improvement of the model at the end of this phase (Phase IV) in comparison with the model results at the end of Phase III. The left plot shows the absolute error (RMSE), and the right side shows the relative error (%RMSE). The curves show the expected shapes with the relative error being smaller for higher volume facilities. With the exception of the second volume class (5,000 – 10,000 vehicles), where the model validates slightly worse, validation across all volume groups has improved with the revisions implemented in Phase IV.



Root Mean Square Error



Percent Root Mean Square Error

Figure 12-9: Validation by volume class

The Highway Performance Monitoring System (HPMS) collects vehicle miles traveled (VMT) by county across the U.S. For this reason, VMT was estimated using count data. Though this is only an estimate with a significant amount of uncertainty, the simulated VMT of the model was compared to the HPMS VMT. As the HPMS VMT estimate includes all roads, including minor residential roads, and the MSTM network is a simplified network that only includes major roads, the HMPS VMT numbers had to be adjusted. Based on estimated VMT by facilities type and based on which share of each facility type by county is represented in the MSTM network, the official HPMS numbers were adjusted to reflect only roadways that are included in the MSTM network.

Figure 12-10 compares estimated VMT with modeled VMT, ordered by estimated VMT. While the overall pattern is replicated, some significant differences can be found for a few counties. Most importantly, Prince Georges's County is underestimated by about 16%. Part of this deviation is likely a function of the statewide mode choice model that has been implemented to capture mode split in many, very different regions across the state. While MSTM models a transit share of 6.5 percent, the Red Line model has a transit share of 5.1% and the Purple Line model has a transit share of 5.5% for this county. It is possible that MSTM overestimates transit in this county, and therefore, does not send a large enough number of vehicle trips on the network to generate VMT. The next phase of MSTM is going to significantly improve the mode choice model. It is expected that improvements to the mode choice model will help to replicate HPMS VMT estimates by county more closely than today.

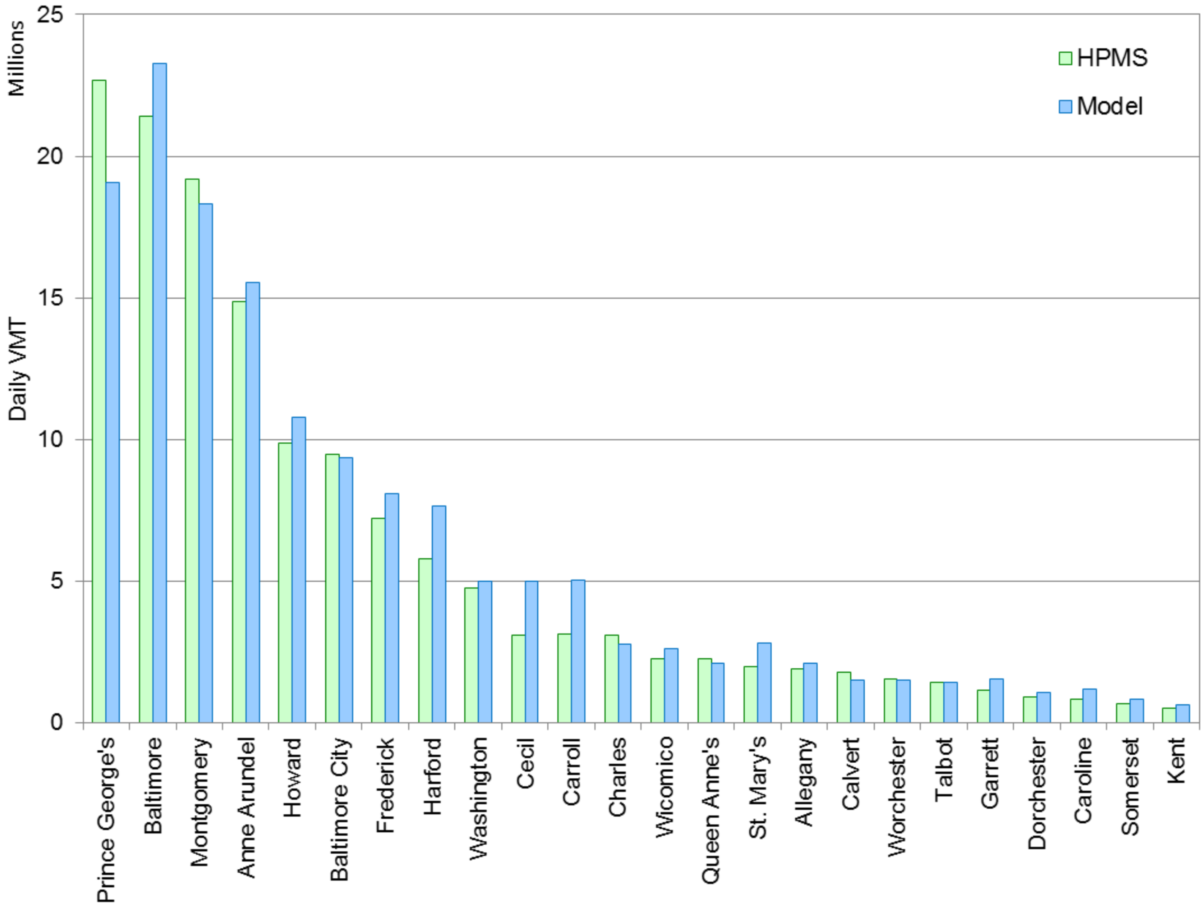


Figure 12-10: Comparison of HPMS and MSTM VMT by county

Across the entire state of Maryland, the model generates 5.3 percent more VMT than the HPMS numbers suggest. Given that the HPMS is an estimate and not a truly observed number, this relatively small deviation is considered to be insignificant.

Figure 12-11 shows a comparison between HPMS VMT estimates and MSTM VMT results in a map. The numbers show the absolute deviation between the two, and the colors represent the deviation normalized by population (which removes the size effect). Baltimore City and its surrounding counties closely replicate the HPMS VMT estimates. The worst deviation can be found in Prince George's County, which is underestimated by 16 percent or 3.6 Million VMT. It is anticipated that the revision of the mode choice model will help improving VMT comparisons, particularly in Prince George's County.

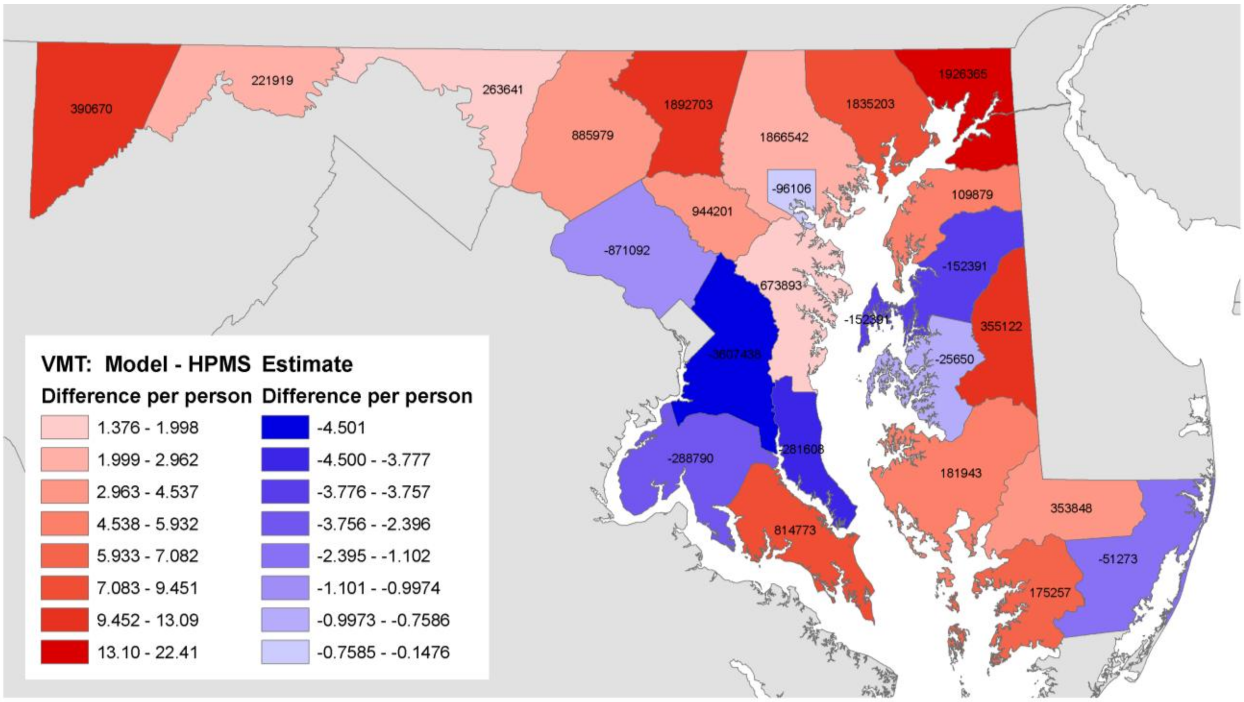


Figure 12-11: Deviation between HPMS VMT estimates and modeled VMT by county

13 Model Application Overview

The MSTM model is implemented in CUBE. The statewide models use CUBE scripts, while the Regional models and Statewide truck model are implemented in Java and called from CUBE.

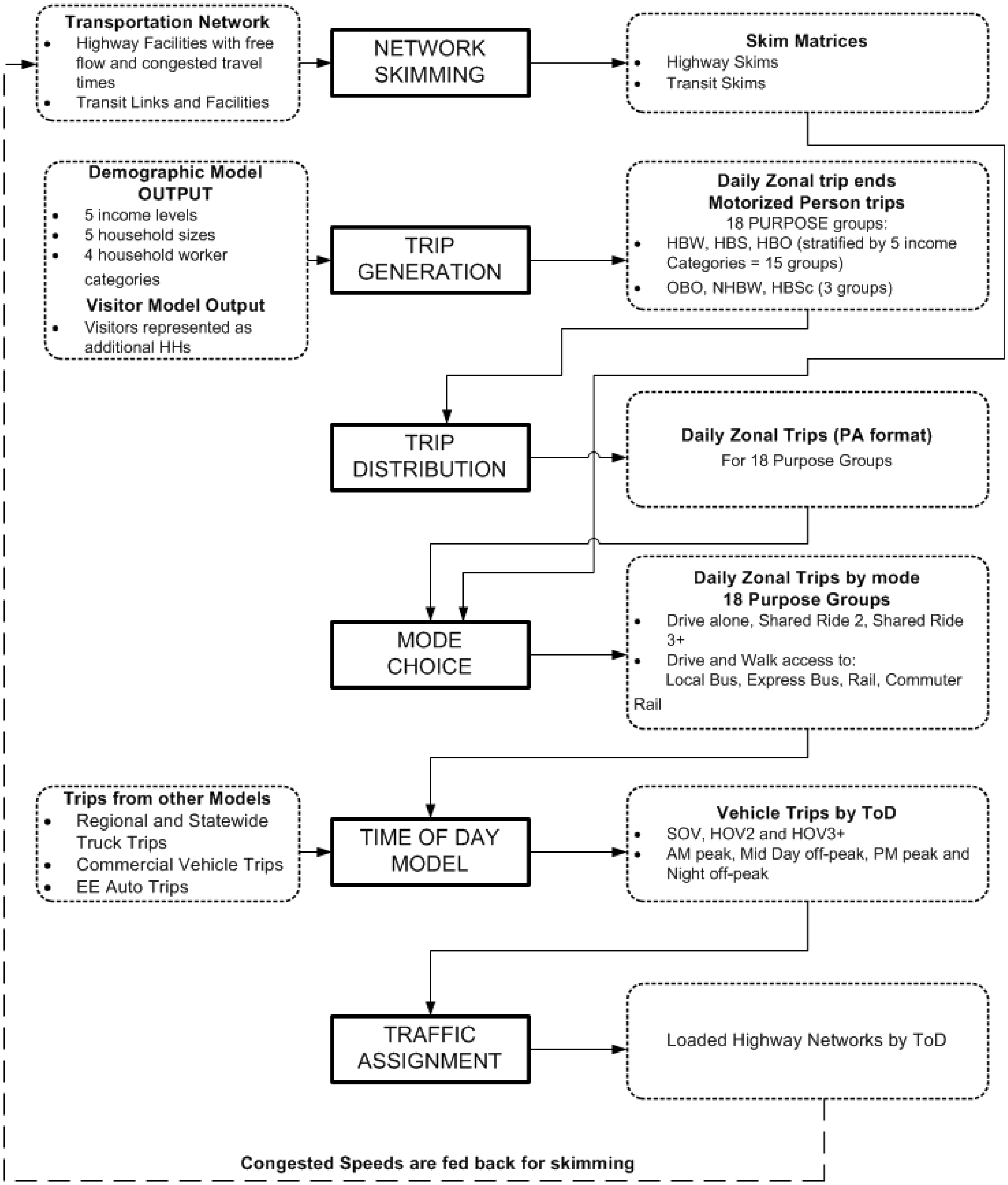


Figure 13-1: MSTM module flowchart

The MSTM travel model is implemented in cube. Some modules, however, require some substantial data preprocessing and are therefore implemented in java. The java model is built highly modular, allowing to easily plugging in and out different model components. The java MSTM package executes the regional and the statewide model as well as the regional auto model (

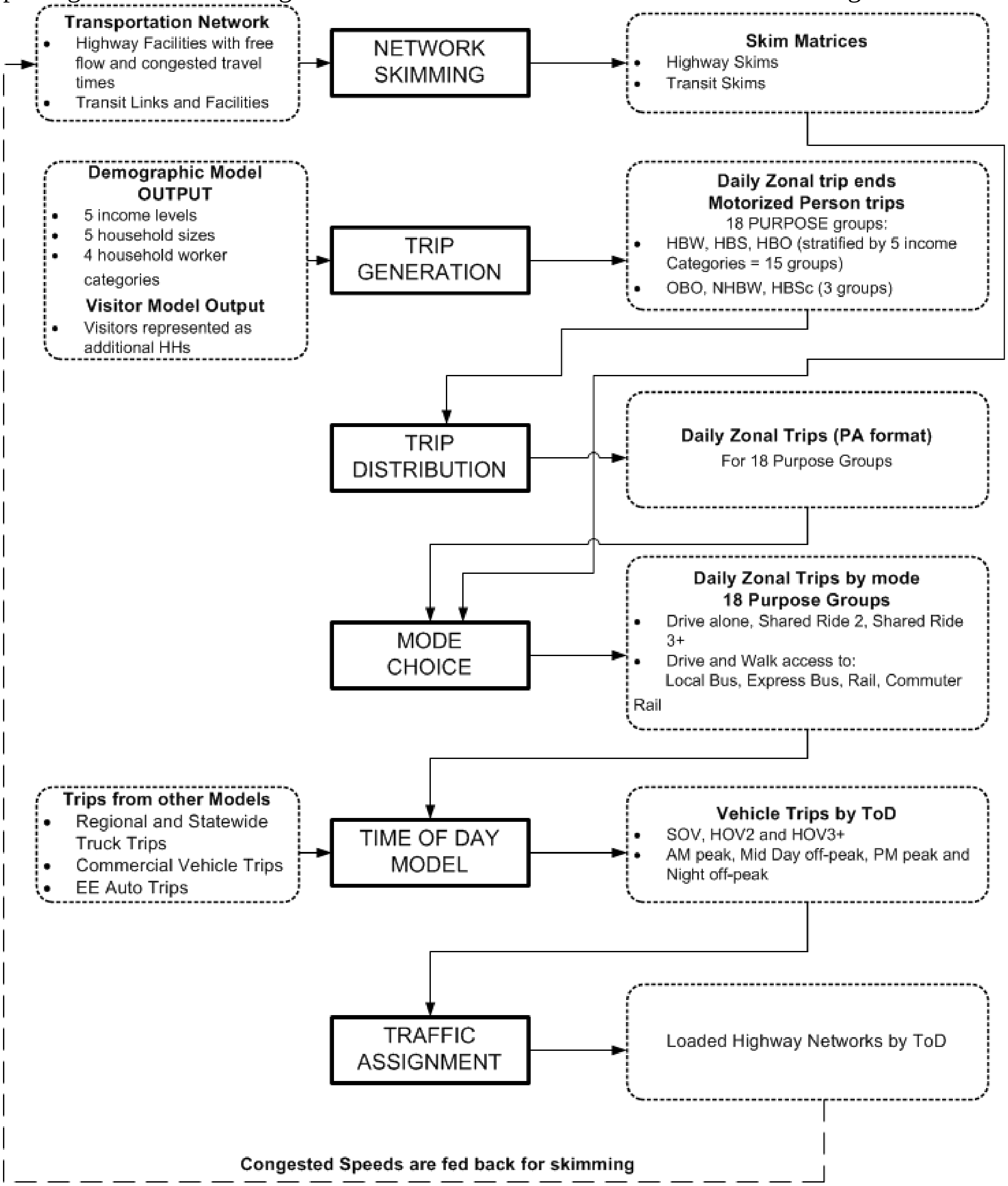


Figure 13-1).

14 User's Guide

This section describes MSTM model components from a user's perspective. The input and output files from various steps of the model are also discussed. The next section gives a summary table of the files for reference in later sections, and to give the reader an idea about where files reside.

14.1 Running the model

To facilitate running the model, an "MSTM Desktop Reference" was developed as a companion document. The Desktop Reference is designed to be less technical and provide a more user-friendly experience while also being more convenient and accessible.

This section highlights the regional which is written in java and called upon in the model through DOS batch commands. The **Regional Model** folder consists of the input files for the java processes used in the Regional freight and long distance person model.

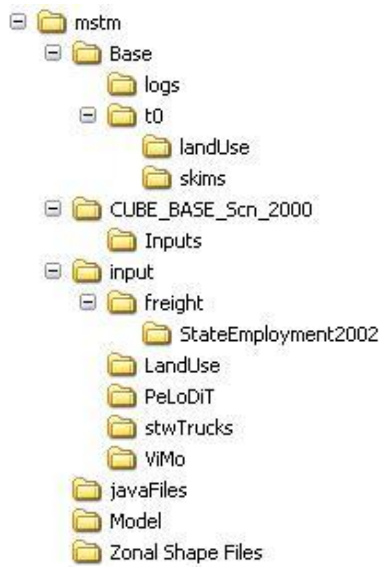


Figure 14-1: MSTM folder structure

Some of the folders consist of input files, summarized in Table 14-1 and Table 14-2. Various output files that are produced from a full model run are summarized in Table 14-3.

Table 14-1: Summary of input files in model folder

File Name	Description
Areatype.dat	Area-type information of each SMZ: 1 for least activity-density and 9 for most
ModeChoiceCoeff.dat	Mode Choice Coefficients and Bias factors
parameter.dat	General parameters for all the steps of model
Destinationparameter.dat	Destination Choice parameters
ParkCost.dat	Parking costs for all SMZs in cents
WBusOP.fac, WBusPK.fac, WCROP.fac, WCRPK.fac, WExpOP.fac, WExpPK.fac, WRai-	Walk to transit skim factors for skimming and assignment

File Name	Description
IOP.fac, WRailPK.fac, WTrnOP.fac, WTrnPK.fac	
DBusOP.fac, DBusPK.fac, DCROP.fac, DCRPK.fac, DExpOP.fac, DExpPK.fac, DRailOP.fac, DRailPK.fac	Drive to transit skim factors for skimming and assignment
WTrnOP.fac, WTrnPK.fac, DTrnOP.fac, DTrnPK.fac	Walk and Drive to all transit modes skimming factors
SYSFILE.PTS	Transit System Information file, contains mode numbers, names and operator info
SMZ_WalkShare.csv	Walk percentages of all the SMZs
MSTM_Ps_2000.csv,MSTM_As_2000.csv	Production & Attraction rates by purpose and SMZ.

Table 14-2: Input files of the regional model folder

File Name
countData\bmcScreenlines.csv countData\extStationsMDwithCountData.csv countData\extStationsSMZ.csv countData\maryland_05272010.csv countData\maryland_06082010.csv countData\mwcogScreenlines.csv countData\outerScreenlines.csv countData\riverScreenlines.csv countData\selectedLinks.csv countData\transitScreenline.csv
freight\commodityConsBy21Ind.csv freight\commodityConsBy4Ind.csv freight\commodityProdBy21Ind.csv freight\commodityProdBy4Ind.csv freight\countyIDs.csv freight\payloadByCommodity.csv freight\RegionList.csv freight\ROWRegionList.csv freight\SCTGtoSTCCconversion.csv freight\shortestPathFAF.csv freight\stateList.csv freight\stcgToStccReference.csv
LandUse\RMZpopulation.csv LandUse\smzCountyEmployment.csv LandUse\smzPopEmp.csv
neldt\countyPopulation.csv neldt\frictionFactorByState.csv neldt\rowList.csv neldt\stateList.csv neldt\statesToSynthesizeNHTS.csv
nhts2002\LDTPUBshort.csv nhts2002\LDTPUBshort.txt skims\frictionFactorByState.csv stwTrucks\stwTruckParameters.csv

The Regional model can be adjusted by the mstm.properties file. This file lists all input and output files. If the model is run with different input data, the file names of input files can be changed. Also, this file has an option to turn on a switch to write out additional validation files. An optional list of visitors, long-distance travelers, regional truck trips and statewide truck trips may be turned on by changing the corresponding entry from "false" to "true." These additional files are not written in the default settings, since they are not required in the model data flow and are only used in supplemental model analysis.

Table 14-3: Output files of the java model

Module	File Name	Description
Trucks	Base/t0/regionalAndStatewideTruckFlows.csv	Regional and statewide truck flows (SMZ/RMZ)
	*Base/t0/statewideTrucks.csv	*Statewide truck flows (SMZ)
	*Base/t0/freightDailyTruckFlowsList.csv	*Regional truck flows (SMZ/RMZ)
Person travel	Base/t0/regionalAutoTravelers.csv	Visitors and long-distance car travelers
	Base/t0/regionalPublicTransitTravelers.csv	Visitors and long-distance public transp. travelers
	Base/t0/regionalPopulationChangeThroughTravelers.csv	Change of population in SMZ due to visitors and long-distance travelers
	*Base/t0/LongDistanceTravelers.csv	*Long-distance travelers (from SMZ to RMZ)
	*Base/t0/Visitors.csv	*Visitors (from RMZ to SMZ)

* Files are optional output files that can be turned on in the mstm.properties file for further analysis

The folder MSTM also contains a batch file called: RunMSTM.bat. The MSTM model can be run by double clicking this file. This file can be edited to run specific step of the model.
Contents of this file are described below:

Step 0: Sets the run path for the model:

CODE
```
setrunpath=%CD%
sThe model-path is "%var%"
cd"%runpath%"\
```

Step 1: Runs the Highway Skims process.

CODE
```
ECHO Running Highway Skims...
cdCUBE_BASE_Scn_2000
runtpp "HighwaySkim.s"
if ERRORLEVEL 2 goto done
```

Step 2: Starts CUBE cluster nodes and runs Transit Skims step. Closes the nodes, once the process is finished.

CODE
```
ECHO Running Transit Skims...
runtpp "TransitSkims.s"
if ERRORLEVEL 2 goto done
```

Step 3: Runs the Iterative Proportional Fitting step.

CODE

```
ECHO Running IPF...
runtpp " IPF.s"
if ERRORLEVEL 2 goto done
```

Step 4: Runs the Trip Generation step.

CODE

```
ECHO Running Trip Generation...
runtpp " TripGeneration.s"
if ERRORLEVEL 2 goto done
```

Step 5: Runs the Trip Distribution step.

CODE

```
ECHO Running Trip Distribution...
runtpp "TripDistribution.s"
if ERRORLEVEL 2 goto done
```

Step 6: Runs the java based model that produces statewide and regional trucks trips as well as visitors and long-distance travelers.

CODE

```
ECHO Running Truck and Regional Model...
cd "%runpath%"
call runMSTMfromConsole.bat
```

Step 7: Runs Destination Choice Model.

CODE

```
ECHO Running DestinationChoice...
cd "%runpath%"\CUBE_BASE_Scn_2000\
runtpp "DCModel.s"
if ERRORLEVEL 2 goto done
```

Step 8: Starts cube cluster nodes, runs mode choice step and closes the nodes.

CODE

```
ECHO Running Mode Choice...
cd "%runpath%"\CUBE_BASE_Scn_2000\
clusterMSTM1-19 START EXIT
runtpp "ModeChoice(MSTM).s"
clusterMSTM1-19 CLOSE EXIT
if ERRORLEVEL 2 goto done
```

Step 9: Runs the time of day model.

CODE

```
ECHO Running Time of Day...
runtpp " TOD.s"
if ERRORLEVEL 2 goto done
```

Step 10: Starts cube cluster nodes, runs Highway Assignment step and closes the nodes.

CODE

```
ECHO Running Highway Assignment...
cd "%runpath%"\CUBE_BASE_Scn_2000\
cluster HwyAssign 1-8 START EXIT
runtpp "HwyAssign.s"
cluster HwyAssign 1-8 CLOSE EXIT
if ERRORLEVEL 2 goto done
```

Step 11: Runs VMT validation scrpts..

CODE

```
ECHO Running VMT Validations...
runtpp "VMT_VHT_ByCountyOnly.s"
if ERRORLEVEL 2 goto done
runtpp "ComVeh_Truck_TLFD.s"
if ERRORLEVEL 2 goto done
```

Step 12: Starts cube cluster nodes, runs Transit Assignment step and closes the nodes.

CODE

```
ECHO Running Transit Assignment...
runtpp "TransitAssign.s"
if ERRORLEVEL 2 goto done
```

Step 13: Model Date/Time and other Outputs.

CODE

```
echo FINISHED RUN %DATE% %TIME%
:done
PAUSE
```

The Setup shown above assumes that there are eight processors or cores in the machine. If available cores differ, then some cluster settings need to be changed in the scripts. This will be described later. If the user wishes not to use a cluster setup, then all the lines that begin with word "`cluster`" should be prefixed with "`rem `", which stands for remark in DOS. Each step is discussed in more detail in the following sections.

14.2 Step 1: Highway Skims

This step skims the network for distance, travel time and tolls. These are computed for single occupancy vehicles (SOV), vehicles with two occupants (SR2 or HOV2) and vehicles with three or more occupants (SR3+ or HOV3+). Additionally, truck paths are skimmed. The shortest path cost parameter is the sum of travel time, the time value of tolls imposed and a quarter of total distance. Intrazonal travel times and distances are assumed to be 60% of the average of nearest three zones. Terminal times, assumed to be a function of area-type of a zone, are also added to the skims for both, origin and destination zone. Skimming is done for peak as well as off-peak periods using suitable attributes from the network.

INPUTS: Highway network (MSTM.net), parameters file (parameter.dat).

OUTPUTS: Peak and Off-peak skim matrices (HwyOP.skm, HwyPK.skm).

14.3 Step 2: Transit Skims

14.3.1 Pre-Transit Network Processing

Prior to skimming the network for transit parameters, Non-transit legs are added to the transportation network. A Non-transit leg is a representation of a bundle of walk and drive links that can be combined to form a path. This path is represented as a leg. It does not exist like a link on the network, but is a representation of the sum of distance, time and other parameters of the underlying network links. There are four kinds of non-transit legs: walk-access, walk-egress, auto-access, and walk-transfer. Code snippets that produce these legs are:

CODE
```
GENERATE, ; Zonal WALK Access and Egress Legs
NTLEGMODE=13, COST=LI.DISTANCE, MAXCOST=8*1.0,
FROMNODE=1-1832, TONODE=1833-120000, DIRECTION=3,
INCLUDELINK=(LI.SWFT=4-6,10-13,21-26), MAXNTLEGS=8*99,
ONEWAY=FALSE, EXTRACTCOST=(60*(LI.DISTANCE/2.5))
```

CODE
```
GENERATE, ; WALK TRANSFER LEGS (From all nodes to all nodes!)
NTLEGMODE=12, COST=LI.DISTANCE, MAXCOST=8*0.5,
FROMNODE=1833-120000, TONODE=1833-120000, DIRECTION=3,
INCLUDELINK=(LI.SWFT=4-6,10-13,21-26), MAXNTLEGS=8*99,
ONEWAY=FALSE, EXTRACTCOST= (60*(LI.DISTANCE/2.5))
```

Zonal auto access is discussed next.

14.3.2 Auto Access Link Development

There are park and ride nodes built in the network, along with the auto access links and walk egress links, for those PnR nodes to the highway system. Links connecting the PnR lots to the Rail or Bus routes are also present. Some additional commands like the following are used to build the Non-Transit Legs in CUBE transit skims process:

CODE
```
GENERATE, ; ZONAL AUTO ACCESS LEGS (BMC PNR Stations)
NTLEGMODE=11, COST=LI.DISTANCE, MAXCOST=8*10,
FROMNODE=1-1832, DIRECTION=1,
INCLUDELINK=(LI.SWFT=1-13), MAXNTLEGS=8*2,
EXTRACTCOST=60*((LI.DISTANCE/LI.FFSPD)+(LI.TOLLOP/1400)),
ACCESSLINK=
3002-4002,,,       ; Explanation:
3003-4003,,,       ; 3002 - 4002           ,      ,           ,
3005-4005,,,       ; PNR - STATION SERVED , COST , DISTANCE ,
3014-4014,,,       ; (WHERE THESE WILL BE ADDED TO THE CORROSPONDING
VALUES OBTAINED FROM THE ROUTE FROM fromNODES TO THE PNRs)
```

14.3.3 Transit Skims

This process is fairly complex compared to the highway skimming described above. Travel time for transit modes is first calculated as a function of the facility type (i.e. freeways, arterials, ramps, etc., see Table 2-6). The formula changes according to the facility type. Skimming is done separately for Peak and Off-peak periods, as well as for walk and drive access to transit. Four modes are skimmed: Bus, Express Bus, Rail and Commuter Rail. In addition to these, walk to all transit and drive to all transit modes are also skimmed. The shortest path parameter is the transit time calculated above. Prior to skimming, the network is augmented with drive access links and walk access links to facilitate the access, egress and transfers.

A variety of quantities are skimmed, these include: initial wait times, transfer wait times, total walk time, auto times, auto distances (meant for auto access, will be zeros for walk access), number of transfers, total bus time (including local, express and premium MTA buses), rail time (light, commuter and metro rail included), actual times on all transit modes, shortest journey times, local bus times, express bus times (would be zero when only local bus is allowed, etc.), metro and light rail times, commuter rail times, transfer and boarding penalties, times and distance of Amtrak and Greyhound modes.

INPUTS: Highway network (MSTM.net), parameters file (parameter.dat), system file (SYSFILE.PTS), factor files (*.fac files for each mode), route files (*.lin files)

OUTPUTS: Skim matrices: **α**Bus**XX**.skm, **α**CRail**XX**.skm, **α**ExpBus**XX**.skm, **α**Rail**XX**.skm where **α**can be W for walk access or D for drive access**, XX** can be PK or OP; Route files (.RTE), report files (.txt files). Route files and report files have similar naming convention as the skim matrices.

14.3.4 Transit Fare Development

Transit fare matrices are developed using the fare matrices from the BMC and MWCOG models for their respective areas. The fare matrices and person trips from these model areas are first aggregated to the SMZ level. Fares are then computed by weighing them using person trips. These are then combined to a single matrix by filling in the zeros in BMC matrix with values from MWCOG matrix. Hence, if a fare value exists for any SMZ in the BMC region as well as MWCOG region, then MWCOG's value is ignored. The inputs to this section are matrices that are internal to the model.

14.4 Step 3: Iterative Proportional Fitting

This step creates households at the SMZ level by the size and income (HH_By_SIZ_INC.csv) and by the workers and income (HH_By_WRKS_INC.csv) groups.

INPUTS: Census 2000 household distribution by size and income groups (Cen2000Seed_HH_By_SIZ_INC.csv), census 2000 household distribution by worker and income groups (Cen2000Seed_HH_By_WRK_INC.csv), scenario year households by SMZ (Target_Size_Wrk_Inc.csv) and aggregate Scenario year households by income, size and worker groups (Target_HH_Size_Wrks.dat).

OUTPUTS: Households by size and income (HH_By_SIZ_INC.csv) and households by workers and income (HH_By_WRKS_INC.csv).

14.5 Step 4: Trip Generation

This step generates the person trips based on the productions and attractions. Production and attraction rates are multiplied with the corresponding zonal socio-economic data. For most purposes, productions and attractions are balanced towards productions. Only for the two purposes NHBW and NHBO, productions are balanced towards attractions.

INPUTS: parameters file (parameter.dat), Socioeconomic data (Activities.csv), Purpose rates and coefficients (XX_rates.txt, XX is the purpose group name), Attraction shares by purpose (HBWAttrShares.csv), SMZ to region definition (ZonesToRegions.csv), Household workers and income categories (HH_By_WRKS_INC.csv), Household size and income categories (HH_By_SIZ_INC.csv) and motorized shares (MotorizedShares.csv).

OUTPUTS: Trip Zonal productions (MSTM_Ps.csv) zonal attractions (MSTM_As.csv).

14.6 Step 5: Trip Distribution

This step distributes the trips across zonal matrix using the interzonal impedances. Composite travel time is used as the impedance which is described in Section 5.2. Once composite times are calculated, friction factors are generated assuming exponential distribution using respective gamma and beta values for each purpose. There are six trip purposes, and the home based purposes are further categorized by income into five groups. Hence, there are a total of eighteen groups for which distribution can be performed, but currently applied only to Home based school purpose. A standard gravity model is applied iteratively to balance the productions and attractions for each zone.

INPUTS: parameters file (parameter.dat), Highway skims (HwyPK.skm, HwyOP.skm), Walk to transit skims (Wtrn*.skm), Zonal productions (MSTM_Ps.csv) zonal attractions (MSTM_As.csv).

OUTPUTS: Trip matrices (**XX**.trp, **XX** is the purpose group name).

14.7 Step 6: Regional Model

The Java Module is called from the CUBE Program and processes four models: the Statewide Truck Model, the Regional Truck Model, the Visitor Model and the Person Long Distance Model. The Statewide Model simulates truck trips with an origin and a destination within the SMZ area. The Regional Truck Model simulates truck trips having origin or destination or both end of the trip outside the SMZ area. The Visitor Model creates trips of visitors that live outside the SMZ area and visit the SMZ area. The opposite direction of long-distance travel, residents who live in the SMZ area and travel to a destination outside this area is modeled by the Person Long Distance Model.

14.8 Step 7: Destination Choice

The destination choice model is applied to distribute trips between production and attraction zone pairs. The zonal productions are distributed based on utility between the interchanges. This model is for all the trip purposes except for Home based school trip purpose. For more details on the terms see Section 5.2 on destination choice model.

INPUTS: parameters file (parameter.dat), destination.parameters(destchoiceParameters.dat) trip productions (**MSTM_Ps.csv)**, households by income (**HHbyIncome.csv)**, zonal activities (**Activities.csv)**, peak and off-peak highway and transit skims (**Hwy<TM>.skm**and**<ACCESS><MODE><TM>.skm**where <TM> = time period and <ACCESS> = walk or drive access mode to transit and <MODE> = transit mode (ex:WBusPk.skm), fare matrix (FareByModes.mtx), parking cost information (ParkCost.dat),area-type file (AreaType.dat), Mode choice coefficients (ModeChoiceCoeff.dat)

OUTPUTS: Trip matrices (DEST_**XX**.trp,**XX** is the purpose group name).

14.9 Step 8: Mode Choice

This step distributes trips for each interchange among eleven alternative modes for each of the 18 purpose groups. Zones are segmented into groups of people who have access to transit via walking or driving or both, and those who do not have transit access. Utilities are then calculated for all the 11 modes available using the appropriate skims and other costs like fares, parking costs, vehicle operating costs, tolls, etc. A logit-based mode choice is then run assuming the (dis)utilities of modes not available to a particular market segment are very low, so that no trips are assigned to those sub-modes. Trips for all the sub-modes are aggregated among all the market segments to yield total trips by sub-modes for each interchange.

INPUTS: parameters file (parameter.dat), trip matrices (DEST_**XX**.trp, **XX** is the purpose group name), fare matrix (FareByModes.mtx), area-type file (AreaType.dat), parking cost information (ParkCost.dat), Zonal walk-shares (SMZ_WalkShare.csv), average occupancy and terminal time information (embedded in the script), Mode choice coefficients (ModeChoiceCoeff.dat)

OUTPUTS: Trip matrices (MC_**XX**.trp, **XX** is the purpose group name)

14.10 Step 9: Time of Day

Trips by mode are distributed across the four periods (see Table 14-4). Percent factors are used to distribute production zone to attraction zone trips and vice versa. These are then averaged to get total OD trips for any interchange. Similar calculations are done for regional & statewide trucks and commercial vehicles, which the java-based model produces, and the exogenous external-external auto trip table. Only highway trips (SOV, HOV2 and HOV3+) are factored by time of day.

Table 14-4: Time of day periods

Time Period	Time Range

AM Peak Period	6:30 am to 9:30 am
MD Off-Peak Period	9:30 am to 3:30 pm
PM Peak Period	3:30 pm to 6:30 pm
NT Off-Peak Period	6:30 pm to 6:30 am (of next day)

INPUTS: parameters file (parameter.dat), Trip matrices (MC_**XX**.trp, **XX** is the purpose group name).

OUTPUTS: Highway Trips (Veh_**XX_YY**.trp, **XX** is purpose group and **YY** is ToD period).

14.11 Step 10: Highway Assignment

This process is similar to highway skimming (see Section 12.2). Each shortest path is loaded with all the trips for an OD interchange. In the next iteration, this volume is averaged out with current and older shortest path distances. This process of balancing the trips continues to obtain equilibrium assignment of the highway network. Different values of time are assumed for each income group and all five income groups are assigned as separate user classes. Commercial vehicles, Regional autos, Regional Trucks, and medium and heavy Statewide trucks are also assigned as separate user classes for analysis purposes. All the non-home based purposes are added to Income group 3 for assignment purposes.

In regular model runs, the standard assignment script (HwyAssign.s) is used. For some applications, a full distinction of all user classes is not necessary, and collapsing certain user classes helps to improve runtime. The alternative assignment script (HwyAssignUnc.s) collapses user classes INC1VEH, INC2VEH,INC3VEH,INC4VEH,INC5VEH and COMVEH into one user group. This way, fewer vehicles classes are assigned to the network and the runtime improves. The scripts HwyAssign.s and HwyAssignUnc.s can be used interchangeably and do not affect the performance of other parts of the model.

INPUTS: parameters file (parameter.dat), network file (inputs\MSTM.net), matrix file with commercial vehicles, internal trucks, regional trucks and regional autos (Veh_Regional_@YY@.tmp, **YY** is ToD period), internal auto trip matrices (Veh_**XX_YY**.trp, **XX** is purpose group and **YY** is ToD period).

OUTPUTS: Assigned networks for four ToD periods and for daily traffic, which is a summation of the four ToD periods (MSTMHwyAsgn_@YY@.tmp, **YY** is ToD period).

14.12 Step 11: Validation

VMT by county and by statewide functional type are summarized after the highway assignment is completed.
INPUTS: Daily assigned network (MSTM_Veh_Dly.net)
OUTPUTS: Assigned network attributes table (ValidationLinks.dbf), VMT by county (VMT_VHT_byCty.csv), VMT by functional type (VMT_BySWFT.csv).

14.13 Step 12: Transit Assignment

This process is similar to transit skimming (see Section 11.5). 'All or Nothing' assignment of trips is done for the shortest paths between any interchange.

INPUTS: Mode choice trip tables (MC_XX.trp, XX is the purpose).

OUTPUTS: Peak and Off-peak assigned networks (MSTMAsgn_XXYY.net, XX is the sub mode and YY is the ToD period), loaded legs without route (loadedlegs_asgn_XXYY,XX is the sub mode and YY is the ToD period), loaded legs with route (loaded-legs_withroute_asgn_XXYY,XX is the sub mode and YY is the ToD period).

14.14 Step 13: Model Date/Time and other Outputs

This process prints the Date/Time when the model finishes running. Additionally, a batch file is setup to enable standard output processing of VMT and assignment results. This process produces the outputs, as specified in the batch file C:/…./*.bat.

15 References

1. Donnelly, Rick and Costinett, Patrick (2008) *A High-Level Specification of the First Generation Maryland Statewide Travel Model*, SHA MSTM Task 3 deliverable, 15 March 2008.
2. Donnelly, Rick, Costinett, Patrick, Moeckel, Rolf, Weidner, Tara (2008) *Maryland Statewide Transportation Model (MSTM) Design and Specification*, Draft Report v3.0, SHA MSTM Task 8 deliverable, 3September 2008.
3. Parsons Brinckerhoff (2008) *MSTM Transportation Models Review*, Draft Report v1.0, SHA MSTM Task 2 deliverable, 23 April 2008.
4. Parsons Brinckerhoff (May 2009) *MSTM Users Guide*, Draft Report v1.0.
5. Parsons Brinckerhoff (2008) *MSTM Model Zones Development*, Draft Report v1.0, SHA MSTM Task 5 deliverable, 26 March 2008.
6. Parsons Brinckerhoff (2008) *MSTM Model Networks Development*, Draft Report v1.0, SHA MSTM Task 4 deliverable, forthcoming.
7. NCSG (2008). Socioeconomic Forecasting Methodology, Scenario Project, Tommy Hammer Inc., Prepared for National Center for Smart Growth.
8. BMC, MWCOG, 2007-2008 Highway Travel Survey, draft results obtained via Maryland State Highway Authority, http://www.baltometro.org/transportation-planning/household-travel-survey.
9. Moeckel, Donnelly (2011) Nationwide Estimate of Long-Distance Travel (NELDT). Generating External Trips for Local Travel Demand Models. In Proceedings of the Annual Meeting of the Transportation Research Forum 2011 in Long Beach, CA. 10-12 March 2011.
10. Baik, H., A. A. Trani, et al. (2008). "Forecasting Model for Air Taxi, Commercial Airline, and Automobile Demand in the United States." Transportation Research Record: Journal of the Transportation Research Board **2052**: 9-20.
11. Dargay, J. and S. Clark (2010). The determinants of long distance travel: An analysis based on the British national travel survey. World Conference on Transport Research. Lisbon, Portugal**:** 21.
12. Haupt, T., T. Friderich, et al. (2004). Validate: A New Method To Generate Nationwide Traffic Data. Online resource www.ptvamerica.com/support/library. Accessed 20 October 2010.
13. Gur, Y. J., S. Bekhor, et al. (2009). Use of massive cell phone data to obtain inter-city person trip tables for nationwide transportation planning in Israel. Annual Meeting of the Transportation Research Board. Washington, D.C.**:** 14.
14. FHWA (2004) 2001 National Household Travel Survey. User's Guide**.** Online resource http://nhts.ornl.gov/2001/usersguide/usersguide.zip. Accessed 12 September 2009.
15. FHWA (2010) "National Travel Household Survey." Online resource http://nhts.ornl.gov/index.shtml. Accessed 25 July 2010.
16. BTS (2009) "Origin and Destination Survey: DB1BCoupon." Online resource: http://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=289&DB_Short_Name=Origin%20and%20Destination%20Survey. Accessed 12 September 2009.
17. FHWA (2004) "2001 National Household Travel Survey. User's Guide." Online resource http://nhts.ornl.gov/2001/usersguide/usersguide.zip. Accessed 12 September 2009.

18. Federal Highway Administration (2008) *Freight analysis framework (FAF)*. Internet Resource: http://ops.fhwa.dot.gov/freight/freight_analysis/faf/index.htm (accessed 30 June 2008)
19. U.S. Census Bureau (2008) *Truck Transportation (NAICS 484) —Estimated Number of Truck Miles Traveled by Employer Firms: 1998 Through 2003.* Internet resource: http://www.census.gov/svsd/www/sas48-5.pdf (accessed on 22 Dec. 2008).
20. U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau (2004) *2002 Economic Census. Vehicle Inventory and Use Survey*. Geographic Area Series. Washington.
21. Battelle (2002) *Freight Analysis Framework Highway Capacity Analysis. Methodology Report to Office of Freight Management and Operations*, U.S. Department of Transportation. Columbus, Ohio.
22. BMC, MWCOG, 2007-2008 Highway Travel Survey, draft results obtained via Maryland State Highway Authority, http://www.baltometro.org/transportation-planning/household-travel-survey

16 Abbreviations

The following abbreviations are used throughout the report.

BMC: BaltimoreMetropolitan Council
MWCOG: Metropolitan Washington Council of Governments
VDOT: Virginia Department of Transportation
PennDOT: Pennsylvania Department of Transportation
DELDOT: Delaware Department of Transportation
MPO: Metropolitan Planning Organization
BEA: Bureau of Economic Analysis
QCEW: Quarterly Census Employment and Wages
CTPP: Census Transportation Planning Package
MSTM: Maryland Statewide Transportation Model
HH: Household
Pop: Population
Emp: Employment
JL: Jurisdictional Level
SMZ: Statewide Modeling Zone

17 Appendix A: Methodology for Cleaning QCEW Data

17.1 Methodology for Cleaning Qrtrly Census Employment/Wage (QCEW) Data

To develop employment data for the areas of Maryland not covered by BMC or MWCOG QCEW data was used. The QCEW dataset was created by the Maryland Department of Labor, Licensing and Regulation (DLLR) to comply with federal unemployment insurance regulations. The dataset is generally not made available to the public due to confidentiality rules; however, the National Center for Smart Growth (NCSG) was able to obtain the data under a strict confidentiality agreement. To preserve confidentiality the NCSG agreed to display information only at the SMZ level. Each record in the QCEW database corresponds to an individual workplace. The data are collected quarterly and provide monthly summaries of employment by workplace. This section describes the characteristics of the raw dataset obtained from DLLR including a discussion of (1) the time period of the data, (2) how master account records were treated, and (3) how workplaces with zero employment were treated.

17.1.1 Date of Dataset

NCSG used QCEW data from the second quarter of 2007, the most recent quarter available. QCEW provides employment by month. To create a composite value for the quarter, the employment values for each of the three months in the quarter were averaged for each workplace. These average quarterly employment figures were used for the remainder of the analysis.
It should be noted that revisions to this dataset were received in March of 2010 but were not incorporated because the analysis had already been completed. An investigation of the revisions showed only minor changes: the total number of workplaces remained the same and average quarterly employment was revised downwards only 0.2%.

Also, implicit in our methodology is the assumption that employment in the second quarter is typically representative of employment in other quarters. To verify this assumption, a comparison was done between statewide annual average employment and quarterly employment from 2002 through 2008 using data available from DLLR. The result showed that, on average, second quarter employment was 100.16% of average annual employment over the time period. Of all four average quarterly employment figures, the second quarter figures were closest to average annual employment.

17.1.2 Treatment of Master Account Records

The records in the QCEW dataset are by workplace but many firms (businesses) operate at more than one location in Maryland. For firms with multiple locations, the QCEW database contains a redundant record, called a "master account record", that shows total statewide employment for that firm. The database also splits out the total employment for each of the firm's locations in the state. Thus, keeping the master account record in place for the analysis would result in double-counting of employment for firms that have multiple establishments. To prevent this, the master account records were removed from the database. Table 17-1 summarizes single and multiple establishments in the raw QCEW dataset. Records containing the multi-code "2" (i.e. master account) were removed to prevent double counting: that left 167,587 records (169,713 – 2,126) encompassing 2,563,505 statewide employees for further analysis.

Table 17-1: Multi establishment employment indicator

Multi Code	Description	Count
1	Single establishment unit	143,320
2	Multi-unit master record	2,126
3	Subunit establishment level record for a multi-unit employer	23,916
4	Multi-establishment employer reporting as a single unit due to unavailability	323
5	A subunit record that actually represents a combination of establishments	18
6	A known multi-establishment employer reporting as a single unit	10
Total		**169,713**

17.1.3 Treatment of Records with Zero Employment

Many workplaces in the database had zero employment recorded. This raised a red flag and was investigated. DLLR confirmed that these zeros were "legitimate" and would occur when:

A new firm has been registered and DLLR notified of this but the firm has not yet filed its first annual tax return. DLLR receives information about employment through the tax filing.

A firm has gone out of business

A firm was relocated or changed its name and the old workplace record was not deleted

A workplace is seasonally operated and is in the off-season in quarter two.

Given that these records are considered legitimate, they were left in the dataset and no effort was made to develop employment totals for them.

The QCEW dataset used is not a complete count of workers in Maryland by workplace location for two reasons: (1) some employees are not required to pay unemployment insurance and (2) physical location addresses are not available for all workplace locations. This section will describe the reasons for these omissions.

17.1.4 Employment Not Counted in QCEW Data

The QCEW database does not include all employees working in the state: employees that are not required to pay unemployment insurance are not in the dataset. The largest omissions of this type include military service members and the self employed. Omissions with more minor impacts include:

"Railroad workers

Agricultural labor where cash wages are less than $20,000 or fewer than 10 workers are employed during the current or preceding quarter

Domestic service unless during any quarter of the current or preceding calendar year the employer pays cash wages of at least $1,000 to individuals performing the employment

Crew members and officers of vessels having a capacity of 10 tons or less

State and local government elected officials

Religious organization workers except where employment is elected to be covered as provided for in the law

Insurance and real estate agents that receive payment solely by commission[28]"

The dataset includes most other civilian state, local, and federal government workers although some federal civilian employees are omitted for national security reasons.

17.1.5 Physical Location Addresses Not Available for all Workplaces

DLLR does not have the physical location address for every workplace in the dataset; in some cases, only the tax address is provided. The tax address refers to the location that processes an establishment's payroll (many of which are located outside of Maryland), not necessarily to the actual location where the employees listed work. Given geo-referencing issues, workplaces where only a tax address was available or no address information was available were not included in our analysis. Table 17-2 provides a summary of the number of records that contain tax addresses or no addresses as opposed to physical location addresses. Note that these tables do not include master account records. Altogether, the lack of pertinent address information results in the removal of approximately 4% of all employees in the raw QCEW data. The adjustments described later in this section are designed to compensate for these omissions.

Table 17-2: QCEW address data

	Address Availability			**Totals**
	Physical Location Address	*Tax Address*	*No Address*	
# of Workplaces	144,198	23,337	52	167,587
# of Employees	2,450,529	109,344	3,632	2,563,505

NOTE: Multi-unit firm master records not included

17.1.6 Geo-referencing the QCEW Data

Employment records were tied to locations on the ground (geo-referenced) using latitude and longitude values in the dataset when they were available, and doing new geocoding when they were not. This section describes (1) how the latitude and longitude values included in the raw dataset were used, (2) how the geocoding was conducted, and (3) the overall results and caveats of the geo-referencing step.

[28] Source: Appendix A of the DLLR 2006 Report: http://www.dllr.state.md.us/lmi/emppay/emplpayrpt2006.pdf

17.1.7 Points Geo-referenced Using Latitude and Longitude Values

The geo-referencing effort was helped greatly by the fact that DLLR had already geocoded most of the dataset as part of its work with the Bureau of Labor Statistics (BLS). These workplace points, containing 96% of the retained QCEW employment, were already assigned latitude and longitude values in the QCEW database and could be easily plotted to designate workplace location points.

One complication with using the workplace points derived from the provided latitude and longitude values is that not all of them represent precise workplace locations. If BLS could not locate the points at their proper address (due to data issues in the QCEW dataset or in the street layer referencing), they were assigned to street intersections, centroids of nine-digit zipcode areas (zip +4), or other less precise geographies. Approximately 17% of the retained workers in the QCEW dataset were located in this manner. Table 17-3 provides a breakdown of geo-referencing information.

Table 17-3: Summary of geo-referencing information

	Lat. & Long. Provided		**No Lat. & Long. Provided**	**Totals**
	Exact Address Location	*Non-Address Location*		
# of Workplaces	119,159	17,898	7,141	144,198
# of Employees	1,941,217	420,158	89,154	2,450,529

Unfortunately, some of these more crudely estimated workplace locations might fall into the wrong zoning district and therefore could distort the employment densities calculated in our employment analysis. To address this problem, employment locations that were geocoded by zip code centroids, a subset of the non-address locations shown in Table A-3, were removed from the analysis. We retained workplaces that were geocoded to the proper street and block, but not the correct side of the street. The amount of employees dropped due to imprecise workplace locations amounted to about 0.5% of the retained employment within the QCEW dataset. As described later in this section, we made adjustments to the dataset in order to account for the dropping of the poorly geo-referenced workplace locations.

17.1.8 Points Assigned Through Geocoding

As Table 17-3 shows, 7,141 of the remaining workplace records were missing the latitude and longitude data provided by DLLR. This might have happened, for example, due to new employment establishments being incorporated into the dataset subsequent to the last round of BLS geocoding. To incorporate this employment information, we geocoded those records missing coordinates using the physical location address provided and Environmental Systems Research Institute (ESRI) Street Map USA geocoding service. Where this geocoding service could not locate a record, the task was performed manually using Google Earth and other sources. Despite these efforts, there were some workplaces (representing approximately 400 employees) that could not be georeferenced using the address information provided and were dropped from our analysis.

17.1.9 Results and Caveats

The final result of all the geocoding was a statewide GIS layer of points that included the approximate locations of 135,261 workplaces encompassing 2,243,486 workers.

One important caveat involves cases where the georeferenced points do not align with the county zoning layers used to compute employment densities. A preliminary inspection of this issue indicated that it may be most problematic in counties where both the latitude and longitude values and the geocoded points were assigned by BLS using ESRI's Street Map USA shapefile[29]. This street centerline layer does not align well with the underlying zoning layers and some employment points may be assigned to the incorrect zoning polygon thereby distorting the employment density estimates. Even with the above caveat, the data is more than adequate to support the model.

17.1.10 Adjustment Technique to Compensate for Omitted Employment

After applying all the filtering described previously in this document, we arrived at a count of 2,243,486 employees (Table 17-4) in the state of Maryland in the second quarter of 2007 that could be tied to a specific location with tolerable accuracy (i.e., georeferenced). However, a total of 320,019 employees appearing in the raw QCEW dataset had to be dropped for the reasons discussed above. In addition, an unknown amount of employees were never counted by DLLR as part of the QCEW data collection effort due to the fact that not all employees must pay unemployment insurance.

An adjustment technique was created by comparing the retained QCEW quarter two employment totals with county-level average annual employment totals from the Bureau of Economic Analysis (BEA). The BEA employment totals include military service members, the self-employed, and the other workers not counted in the QCEW dataset. To compensate for the shortfall in the QCEW counts, we used the BEA estimate of total employment for each county (which includes all of the omitted employment) as a control total, and then adjusted each QCEW workplace record upwards (in a few cases downwards) to match the BEA data at the county level (by two-digit NAICS code). The following sections provide more detail.

Table 17-4: Comparison of 2007 employment totals from various data sources

	NCSG QCEW*	BEA	BLS	MSTM**
Total Statewide Employment	2,243,486	3,437,502	2,547,350	2,774,238

*NCSG QCEW data draws on the QCEW data as outlined in Table 17-2 but provides only the total employment that NCSG was able to georeference

**MSTM data refers to the SHA approved SMZ totals used for the MD Stawide Transportation Model (MSTM). These data make use of the BMC, MWCOG, CTPP, and BEA wage and salary data sources. See the text for a more tho-

[29] Counties where ESRI's shapefile appears to have been used, and for which employment estimates might have a greater chance of being off, are Allegany, Baltimore (City), Baltimore (County), Caroline, Carroll, Cecil, Charles, Dorchester, Harford, Prince George's, Queen Anne's, and Worcester.

rough description of how these totals were derived.

17.1.11 Adjustment by NAICS Code

We made our adjustments at the most disaggregate scale possible (the county level), given available data. We also differentiated the adjustments by industrial classification, using two-digit North American Industry Classification System (NAICS) codes, which were available from BEA at the county level and also included in the QCEW dataset. Thus, it was possible to give each industry in each county a separate adjustment factor. For example, all workplaces in Baltimore County that were coded as NAICS code 44, retail trade, received an upward adjustment (i.e. were multiplied) by 1.199 to equate with the Baltimore County BEA control total for code 44. In Howard County, a separate factor of 1.168 was computed and applied to each workplace with NAICS code 44. This process was repeated for each county and each industry.

After the adjustment factors had been computed, some outliers (i.e. very high adjustment factors) were noted. These primarily involved military employment and a few NAICS categories in a couple of counties. Special efforts were made to address these outliers as described below.

17.1.12 Special Military Adjustments

Military employment is not included in the QCEW but is included in the BEA data as Public Administration employment. Because military employment is high in Maryland (Fort Meade, Fort Detrick, Patuxent Naval Air Station etc.), the adjustment factors for the Public Administration NAICS code, which includes military employment, are unusually high compared to other industries. Rather than adjusting all Public Administration employment sites in a county to include military employees, we manually allocated military employees to bases in seven counties: Anne Arundel, Charles, Frederick, Harford, Montgomery, Prince George's and St, Marys.

Military bases were extracted from ESRI base data. A centroid (i.e. center point) was created for each base using the feature to point tool. The bases allocated to were Fort Meade in Anne Arundel, the US Naval Surface Warfare site in Charles County, Fort Detrick in Frederick County, Aberdeen Proving Grounds in Harford County, the US Naval Surface Weapons Facility in Montgomery County, Andrews Air Force Base in Prince George's County, and Patuxent Naval Air Station in St. Mary's County.

Although we noted that all counties in Maryland have some military employment (due to National Guard installations), these numbers are quite low for most counties. For counties without major U.S. military bases, military employment was not considered separately and Public Administration job sites were adjusted by the total number of government employees. BEA data includes a sub-category of employment called "Military." For counties with major military installations, the Public Administration adjustment factors were determined by subtracting the military employment from total Public Administration employment.

17.1.13 Other Special Adjustments

In two counties, factors in some sectors exceeded 10, meaning that the BEA data by NAICS code was ten times higher than the QCEW data Because the issue only affected two employment categories (NAICS codes) in two specific counties, we treated these issues individually:

- **Queen Anne's County**: Prior to additional adjustment, the Public Administration (NAICS 92) factor was around 45. This was due to the omission of the centroid for Queen Anne's County Government with 575 employees, which was georeferenced to the zip code centroid level.
- **St. Mary's County**: Prior to additional adjustment, Management (NAICS 55) was approximately 45. This occurred because four of the five management employment locations were georeferenced to the zip code centroid level. Upon inspecting the physical location of these centroids, we realized that multiplying one management location by a factor of 45 would be less accurate than including employment locations georeferenced to zip code centroids. Management employment locations were distributed throughout the county, not concentrated in a single area. To address this issue, we merged the Management employment locations georeferenced to the zip code centroid level with all other employment locations in the data set. By including these locations, the factor dropped to around 1.3.

Note that establishments with NAICS code 99 (unclassified) in the QCEW were not adjusted because BEA data does not include NAICS code 99. However, this impacts only approximately 650 employees across the state.

An implicit assumption of this type of adjustment is that the employment not counted at all by QCEW reporting, and not capable of being tied accurately to the ground even if counted by QCEW, is (1) properly accounted for by BEA at the county level, so BEA estimates can be used as control totals, and (2) is more or less uniformly undercounted by county and by NAICS code, so applying adjustment factors to individual workplace records is a reasonable way to simulate the distribution of the employment for which we have no precise location.

17.1.14 Results and Caveats

The final output of this effort is a GIS point layer of individual workplace locations. Each workplace point is associated with an adjusted average 2007 quarter-two employment estimate. Total statewide adjusted employment is estimated to be 3,434,267 (1,190,781 employees more than with the unadjusted data). This total does not precisely match the BEA total due to rounding.

The large amount of the adjustment, representing approximately 35% of total employment, was a cause for concern and prompted further review. The review revealed that a substantial portion of the adjusted employees (59%) resulted from self employment that is not counted in QCEW employment but is counted by BEA. Further caveats are in order regarding the adjustments. First, the necessary adjustment of quarter two employment figures using *annual* average employment from BEA is likely to introduce some error due to the different timeframes involved. Second, at a point level, the estimated employment at any give workplace location is not an accurate measure of true employment. This iscaused by our adjusting the (accurate) QCEW data for that site to account for all the employment QCEW either not counted by QCEW or not georeferenced.

Thus within a county, all employment in a given industry which can not be georeferenced has been reassigned to sites where employment of the same type is known to be located. This adjusted data provides a much better estimate of total employment than the unadjusted QCEW.

18 Appendix B: Jurisdictional Level (JL)[30] Totals to SMZ

BMC
BEA JL Retail * BMC TAZ Retail Σ SMZ / BMC JL Retail
BEA JL Office * BMC **2000 Employment Disaggregation Procedure**
TAZ Office Σ SMZ / BMC JL Office
BEA JL Industrial * BMC TAZ Industrial Σ SMZ / BMC JL Industrial
BEA JL Other * BMC TAZ Other Σ SMZ / BMC JL Other

MWCOG-within Maryland – Adjusted MWCOG total employment distributed down to the SMZ level using MWCOG SMZ total employment first. Then SMZ total employment will be distributed by employment category using CTPP.

SMZ Total Emp. = MWCOG TAZ Total Emp Σ SMZ / MWCOG JL Total Emp (Adjusted)

SMZ Total Emp * CTPP Retail (2000) Σ SMZ / CTPP Total Emp (2000) Σ SMZ
SMZ Total Emp * CTPP Office (2000) Σ SMZ / CTPP Total Emp (2000) Σ SMZ
SMZ Total Emp * CTPP Industrial (2000) Σ SMZ / CTPP Total Emp (2000) Σ SMZ
SMZ Total Emp * CTPP Other (2000) Σ SMZ CTPP Total Emp (2000) Σ SMZ

MWCOG-outside Maryland – Adjusted MWCOG total employment distributed down to the SMZ level using MWCOG SMZ total employment first. Then SMZ total employment will be distributed by employment category using CTPP.

SMZ Total Emp. = MWCOG TAZ Total Emp Σ SMZ / MWCOG JL Total Emp (Adjusted)

SMZ Total Emp * CTPP Retail (2000) Σ SMZ / CTPP Total Emp (2000) Σ SMZ
SMZ Total Emp * CTPP Office (2000) Σ SMZ / CTPP Total Emp (2000) Σ SMZ
SMZ Total Emp * CTPP Industrial (2000) Σ SMZ / CTPP Total Emp (2000) Σ SMZ
SMZ Total Emp * CTPP Other (2000) Σ SMZ CTPP Total Emp (2000) Σ SMZ

Non-MPO Region Maryland
BEA JL Retail * QCEW Retail Σ SMZ / QCEW JL Retail
BEA JL Office * QCEW Office Σ SMZ / QCEW JL Office
BEA JL Industrial * QCEW Industrial Σ SMZ / QCEW JL Industrial
BEA JL Other * QCEW Other Σ SMZ / QCEW JL Other

New Jersey and Remainder West Virginia
BEA JL Retail * CTPP Retail Σ SMZ / CTPP JL Retail
BEA JL Office * CTPP Office Σ SMZ / CTPP JL Office
BEA JL Industrial * CTPP Industrial Σ SMZ / CTPP JL Industrial
BEA JL Other * CTPP Other Σ SMZ / CTPP JL Other

[30] BEA employment category and totals are equivalent to Tommy Hammers estimations in the year 2000.

Delaware(DelDOT) – DelDOT has several categories that were collapsed to 4.
Retail = DelDOT: Business
Office = DelDOT: Information + Finance + Hospital + Health + Service + Public Adm
Industrial = DelDOT: Manufacturing
Other = DelDOT: Natural Resources + Construction + Utilities

BEA JL Retail * DelDOT TAZ Retail Σ SMZ / DelDOT JL Retail
BEA JL Office * DelDOT TAZ Office Σ SMZ / DelDOT JL Office
BEA JL Industrial * DelDOT TAZ Industrial Σ SMZ / DelDOT JL Industrial
BEA JL Other * DelDOT TAZ Other Σ SMZ / DelDOT JL Other

Pennsylvania and Virginia(P&VDOT) – P&VDOT do not separate industrial from other.
Retail = P&VDOT: Retail
Office = P&VDOT: Service
Industrial = P&VDOT: Other * CTPP Industrial / (CTPP Industrial + CTPP Other)
Other = P&VDOT: Other * CTPP Other / (CTPP Industrial + CTPP Other)

BEA JL Retail * P&VDOT TAZ Retail Σ SMZ / P&VDOT JL Retail
BEA JL Office * P&VDOT TAZ Office Σ SMZ / P&VDOT JL Office
BEA JL Industrial * P&VDOT TAZ Industrial Σ SMZ / P&VDOT JL Industrial
BEA JL Other * P&VDOT TAZ Other Σ SMZ / P&VDOT JL Other

19 Appendix C: 2030 Employment Disaggregation Procedure

(Jurisdictional Level Totals to SMZ)

BMC
Σ BMC TAZ Retail (2030)
Σ BMC TAZ Office (2030)
Σ BMC TAZ Industrial (2030)
Σ BMC TAZ Other (2030)

MWCOG-within Maryland
Step 1: (Intermediate Step-I)

SMZ Retail 2030-I = (JL Retail 2030 / JL Retail 2000) * MWCOG Retail (2000)
SMZ Office 2030-I = (JL Office 2030 / JL Office 2000) * MWCOG Retail (2000)
SMZ Industrial 2030-I = (JL Industrial 2030 / JL Industrial 2000) * MWCOG Retail (2000)
SMZ Other 2030-I = (JL Other 2030 / JL Other 2000) * MWCOG Retail (2000)

Revised JL Total Emp 2030 = (Retail + Office + Industrial + Other)Σ SMZ

Step 2:

JL Total Emp 2030 / Revised JL Total Emp 2030 = Factor 2030

SMZ Retail 2030 = SMZ Retail 2030-I* Factor 2030
SMZ Office 2030 = SMZ Office 2030-I* Factor 2030
SMZ Industrial 2030 = SMZ Industrial 2030-I* Factor 2030
SMZ Other 2030 = SMZ Other 2030-I* Factor 2030

MWCOG-outside Maryland
Step 1: (Intermediate Step-I)

SMZ Retail 2030-I = (JL Retail 2030 / JL Retail 2000) * MWCOG Retail (2000)
SMZ Office 2030-I = (JL Office 2030 / JL Office 2000) * MWCOG Retail (2000)
SMZ Industrial 2030-I = (JL Industrial 2030 / JL Industrial 2000) * MWCOG Retail (2000)
SMZ Other 2030-I = (JL Other 2030 / JL Other 2000) * MWCOG Retail (2000)

Revised JL Total Emp 2030 = (Retail + Office + Industrial + Other)Σ SMZ

Step 2:

JL Total Emp 2030 / Revised JL Total Emp 2030 = Factor 2030

SMZ Retail 2030 = SMZ Retail 2030-I* Factor 2030
SMZ Office 2030 = SMZ Office 2030-I* Factor 2030
SMZ Industrial 2030 = SMZ Industrial 2030-I* Factor 2030

SMZ Other 2030 = SMZ Other 2030-I* Factor 2030

Remainder of Maryland
Hammer JL Retail (2030) * QCEW Retail (2007) Σ SMZ / QCEW JL Retail (2007)
Hammer JL Office (2030) * QCEW Office Σ SMZ (2007)/ QCEW JL Office (2007)
Hammer JL Industrial (2030)* QCEW Industrial (2007) Σ SMZ / QCEW JL Industrial (2007)
Hammer JL Other (2030)* QCEW Other Σ SMZ (2007) / QCEW JL Other (2007)

New Jersey and Remainder West Virginia
Hammer JL Retail (2030) * CTPP Retail (2000) Σ SMZ / CTPP JL Retail (2000)
Hammer JL Office (2030) * CTPP Office (2000) Σ SMZ / CTPP JL Office (2000)
Hammer JL Industrial (2030) * CTPP Industrial (2000) Σ SMZ / CTPP JL Industrial (2000)
Hammer JL Other (2030) * CTPP Other (2000) Σ SMZ / CTPP JL Other (2000)

Delaware (DelDOT) – DelDOT has several categories that were collapsed to 4.
Retail = DelDOT: Business
Office = DelDOT: Information + Finance + Hospital + Health + Service + Public Adm
Industrial = DelDOT: Manufacturing
Other = DelDOT: Natural Resources + Construction + Utilities

Hammer JL Retail (2030)* DelDOT TAZ Retail Σ SMZ / DelDOT JL Retail
Hammer JL Office (2030) * DelDOT TAZ Office Σ SMZ / DelDOT JL Office
Hammer JL Industrial (2030) * DelDOT TAZ Industrial Σ SMZ / DelDOT JL Industrial
Hammer JL Other (2030) * DelDOT TAZ Other Σ SMZ / DelDOT JL Other

Pennsylvania and Virginia(P&VDOT) – P&VDOT do not separate industrial from other.
Retail = P&VDOT: Retail
Office = P&VDOT: Service
Industrial = P&VDOT: Other * CTPP Industrial / (CTPP Industrial + CTPP Other)
Other = P&VDOT: Other * CTPP Other / (CTPP Industrial + CTPP Other)

Hammer JL Retail (2030) * P&VDOT TAZ Retail Σ SMZ / P&VDOT JL Retail
Hammer JL Office (2030)* P&VDOT TAZ Office Σ SMZ / P&VDOT JL Office
Hammer JL Industrial (2030) * P&VDOT TAZ Industrial Σ SMZ / P&VDOT JL Industrial
Hammer JL Other (2030)* P&VDOT TAZ Other Σ SMZ / P&VDOT JL Other

20 Appendix E: HTS Survey Overview

In 2007-2008, the Baltimore Metropolitan Council (BMC) on behalf of the Baltimore Regional Transportation Board, teamed with the Transportation Planning Board at the Metropolitan Washington Council of Governments (MWCOG) to conduct a household travel survey in both the Baltimore and Washington regions (HTS Survey) [18]. Data for the survey was collected from randomly selected households in the Baltimore and Washington DC region. Each household completed a travel diary that documented the activities of all household members on an assigned day. Demographic information was also collected. The surveys have been stored in a database, which contains records for approximately 4,500 households, 10,000 persons, 49,000 trips, and 6,000 vehicles.

The HTS data consist of four separate files – a household, person, trip and vehicle file. The data fields are in Table 20-1 through Table 20-4. The four survey files can be linked based on the common 'sampn' field. Processing of the survey for MSTM assumed several regions. Figure 20-1 identifies the aggregation to urban, suburban and rural regions used in the trip generation process. Figure 20-2 identifies the aggregation used in the destination choice model. Each region was assigned based on the FIPS code of the home location of the household record corresponding to the trip.

Table 20-1: HTS household records

Variable Name	Description
sampn	Sample Number
tpb_mod	TPB Modeled Area
bmc_mod	BMC Modeled Area
msa	MSA
home_fips2	Residence Jurisdiction
home_tract	Residence Census Tract
home_tpb_taz	Residence TPB Transportation Analysis Zone
home_bmc_taz	Residence BMC Transportation Analysis Zone
housing_type	Housing Type
o_housing_type	Other, Housing Type
tenure	Housing Tenure
o_tenure	Other, Housing Tenure
hhsiz	Household Size
rc_hhsiz	Household Size - Recoded
hhstu	Number of Students in HH
hhlic	Number of Licensed Drivers in HH
hhwrk	Number of Workers in HH
hhdis	Person with Disability in HH
hhveh	Number of HH Vehicles Available
rc_hhveh	Number of Vehicles - Recoded
bikes	Number of HH Bicycles Available
incom	Household Income
imhousing	Housing Type - Imputation Flag

Variable Name	Description
imtenure	Housing Tenure - Imputation Flag
impedis	Household Disability - Imputation Flag
imbikes	Household Bicycle - Imputation Flag
imincom	Household Income - Imputation Flag
stratum	Stratum Number
home_cluster_id	Activtivity Cluster ID Number

Table 20-2: HTS person records

Variable Name	Description
sampn	Sample Number
personid	Personid Number
age	Age in Years
ageg	Age Group
gend	Gender
race	Race/Hispanic Ethnicity
relate	Relationship to Reference Person
lic	Have Drivers License?
pedis	Personal Disability that limits Mobility?
wkstat	Work Status
emply	Currently Employed?
jobs	Number of Current Jobs
etype	Type of Employment/Classification
hours	Number of Hours Worked Last Week
reason	Reason Did Not Work Last Week
wloc	Work Location
work_jur	Place of Work
gtowk	Usual Means of Transportation to Work Last Week
start01	Typical Work Start Time for Primary Job
end01	Typical Work End Time for Primary Job
fixd1	Job Work Schedule Flexibility for Primary Job
wkdy1	Work Days for Primary Job
start01_w2	Typical Work Start Time for 2nd Job
end01_w2	Typical Work End Time for 2nd Job
fixd2	Job Work Schedule Flexibility for 2nd Job
wkdy2	Work Days for 2nd Job
start01_w3	Typical Work Start Time for 3rd Job
end01_w3	Typical Work End Time for 3rd Job
fixd3	Job Work Schedule Flexibility for 3rd Job
wkdy3	Work Days for 3rd Job
start01_w4	Typical Work Start Time for 4th Job
end01_w4	Typical Work End Time for 4th Job
fixd4	Job Work Schedule Flexibility for 4th Job
wkdy4	Work Days for 4th Job
start01_w5	Typical Work Start Time for 5th Job

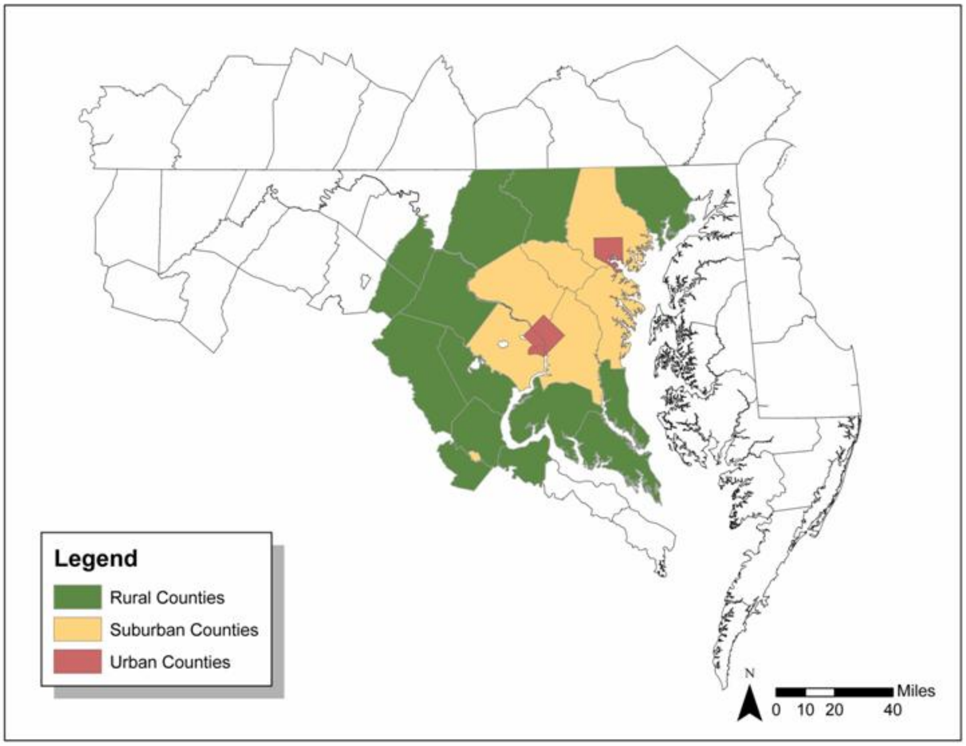
Variable Name	Description
end01_w5	Typical Work End Time for 5th Job
fixd5	Job Work Schedule Flexibility for 5th Job
wkdy5	Work Days for 5th Job
eltlc	Eligible to Telecommute
datlc	Days Telecommuted Last Week
tb01	Employer Provides Free Parking
tb02	Employer and Employee Share Parking Cost
tb03	Employer Provides Preferential Parking for Carpools/Vanpools
tb04	Employer Provides Subsidies for Carpool/Vanpools
tb05	Employer Provides Subsidies for Transit/Vanpooling
tb06	Guaranteed Ride Home Available to Employee
tb07	Employer Provides Bike/Pedestrian Facilities or Services
tb08	Employer Provides Information on Commute Options
tb09	Employer Does Not Offer Transportation Benefits
secbf	Secure Bicycle Facility at Work Location
btrvl	Number of Weekdays Used Bicycle Last Week
buser	Type of Bikeway Mostly Used Last Week
stud	Attend School?
schol	Current Grade Level
sloc	School Location
sbypk	Secure Bicycle Location at School
smode	Usual Means to School Last Week
sdays	Days Attended School Last Week
volun	Volunteer on a Regular Basis
vloc	Volunteer Location
vdays	Volunteer Days Per Week
ffactor	Final Weighting Factor
impage	Age - Imputation Flag
impageg	Age Group - Imputation Flag
impgend	Gender - Imputation Flag
imprace	Race/Hispanic Ethnicity - Imputation Flag
implic	Driver License - Imputation Flag
impwkstat	Work Status - Imputation Flag
imppedis	Personal Disability Status - Imputation Flag

Table 20-3: HTS trip records

Variable Name	Description
sampn	Sample Number
personid	Personid Number
rtripid	Linked Trip ID
opurp	Origin Trip Purpose
oact1	Origin Activity
ofips	Origin Fips Code
otaz_tpb	Origin TPB TAZ Number
otaz_bmc	Origin BMC TAZ Number
dpurp	Destination Trip Purpose
dact1	Destination Activity
dfips	Destination Fips Code
dtaz_tpb	Destination TPB TAZ Number
dtaz_bmc	Destination BMC TAZ Number
begt	Begin Trip Time
endt	End Trip Time
pmode	Primary Travel Mode
mode	Detailed Travel Mode
accmode	Transit Access Mode
egrmode	Transit Egress Mode
vehid	Vehicle ID Number
oocc	Origin Vehicle Occupancy
docc	DestinationVehicle Occupancy
tt	Reported Travel Time
dist	Estimated Trip Distance
ffactor	Final Trip Weighting Factor

Table 20-4: HTS vehicle records

Variable Name	Description
sampn	Sample Number
vhtno	Household Vehicle Number
body	Vehicle Body Type
o_body	Vehicle Body Type, Other
fuel	Vehicle Fuel Type
o_fuel	Vehicle Fuel Type, Other
year	Vehicle Model Year
make	Vehicle Make
o_make	Vehicle Make, Other
model	Vehicle Model
ffactor	Final Vehicle Weighting Factor



Region	Activity Density Criteria	
	Lower	Upper
Urban	10	High
Suburban	1.5	9.9
Rural	0.01	1.49

$$Activity\ Density = \frac{(Households + Retail\ Emp. + Total\ Emp.)}{Area}$$

Figure 20-1: Map of HTS regions used in MSTM trip generation

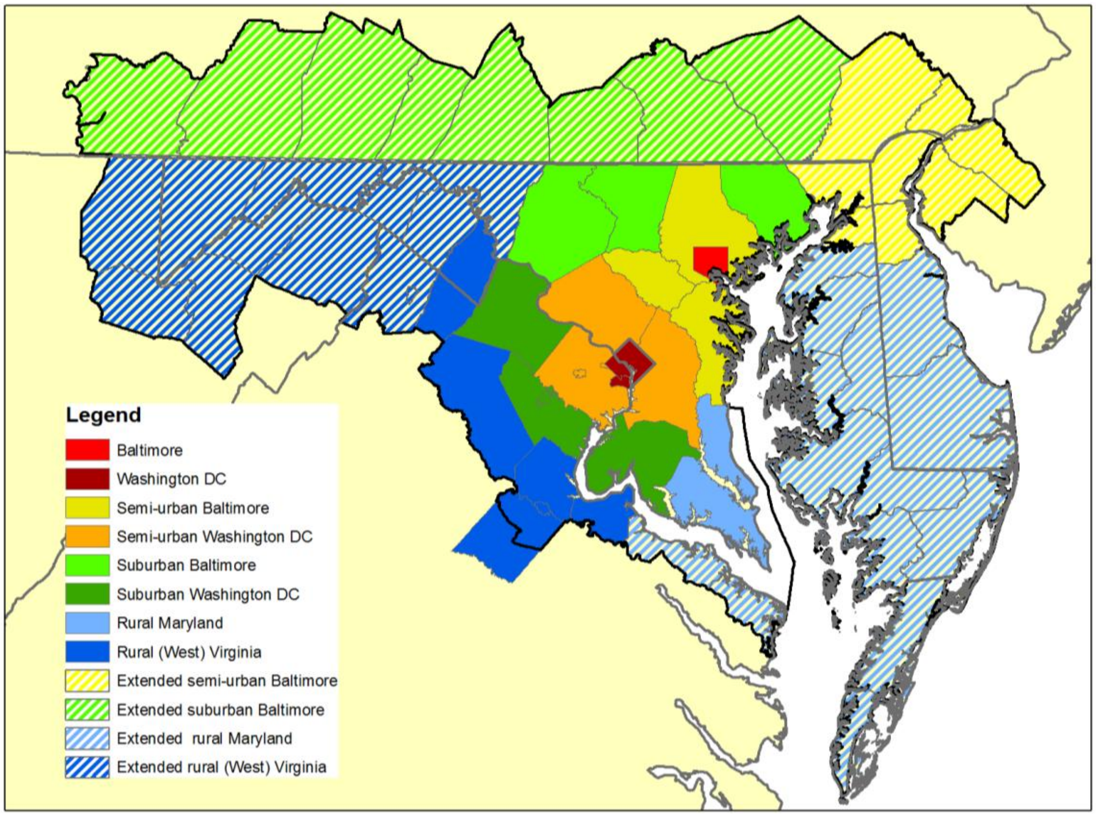


Figure 20-2: HTS data processing used in MSTM destination choice

Table 20-5: List of counties within the SMZ study area and the corresponding applied region

CountyName	Region	CountyName	Region	CountyName	Region
Accomack County, VA	Rural, 1	Fayette County, PA	Rural,5	Northampton County, VA	Rural,7
Adams County, PA	Rural,2	Franklin County, PA	Rural,5	Northumberland Co, VA	Rural,7
Alexandria, VA	Urban,2	Frederick County, MD	Rural,5	Preston County, WV	Rural,8
Allegany County, MD	Rural,2	Frederick County, VA	Rural,5	Prince George's Co, MD	Suburban,8
Anne Arundel C, MD	Suburban,3	Fredericksburg Co, VA	Rural,5	Prince William County, VA	Rural,8
Arlington County, VA	Urban,3	Fulton County, PA	Rural,5	Queen Ann's County, MD	Rural,8
Baltimore City, MD	Urban,3	Garrett County, MD	Rural,5	Salem County, NJ	Rural,8
Baltimore County, MD	Suburban,3	Gloucester County, NJ	Rural,6	Somerset County, MD	Rural,8
Bedford County, PA	Rural,3	Grant County, WV	Rural.6	Somerset County, PA	Rural,8
Berkeley County, WV	Rural,3,	Hampshire County, WV	Rural,6	Spotsylvania County, VA	Rural.8
Calvert County, MD	Rural,3	Harford County, MD	Rural,7	St. Mary's County, MD	Rural,8
Caroline County, MD	Rural,3	Howard County, MD	Suburban.7	Stafford County, VA	Rural,8
Carroll County, MD	Rural,3	Jefferson County	Rural,7	Sussex County, DE	Rural,8
Cecil County, MD	Rural,3	Kent County, DE	Rural,7	Talbot County, MD	Rural,8

Charles County, MD	Rural,3	Kent County, MD	Rural,7	Tucker County, WV	Rural,8
Chester County, PA	Rural4,	King George, VA	Rural,7	Washington County, MD	Rural,8
Clarke County, VA	Suburban,4	Lancaster County, PA	Rural,7	Westmoreland County, VA	Rural,8
Delaware County, PA	Rural,4	Loudoun County, VA	Rural,7	Wicomico County, MD	Rural,8
District of Columbia	Suburban,5	Mineral County, WV	Rural,7	Winchester County, VA	Rural,8
Dorchester County, MD	Urban,5	Montgomery Co, MD	Suburban,7	Worchester County, MD	Rural,8
Fairfax County, VA	Rural,5	Morgan County, WV	Rural,7	York County, PA	Rural,8
Fauquier County, VA	Suburban,5	New Castle County, DE	Suburban,7		

21 Appendix F: Recalculation of HTS Expansion Factors

Most commonly, a household travel survey (HTS) is not a full survey but a sample of travelers whose travel behavior shall be analyzed. If the sample was perfectly representative, meaning all segments of the population were surveyed by the same share as they are appear in the population as a whole, the survey could be used without any adjustments. In practice, however, certain parts of the population are oversampled, why other part of the population are underrepresented. It is common for household travel surveys to under-sample young households and oversample older households and retirees, as the latter tend to be more at home, and therefore, are easier to reach to respond to a survey. Very low income households as well as very high income households tend to show less willingness in participating in surveys. Particularly rare household types, such as a five-person household with no car, are difficult to sample by the same rate as they appear in the population.

To make a survey representative of the population, expansion factors are assigned to every survey record. Survey records of household types that were under-sampled receive a higher expansion factor than survey records that were oversampled. Summing up all expansion factors by household type leads to the same relative distribution of household types as found in reality.

The BMC/MWCOG HTS provides expansion factors that were used in phase II of the MSTM project. A closer review of these expansion factors revealed incompatibility with the MSTM socio-economic data. Using the provided expansion factors led to an overrepresentation of mid-income households and an underrepresentation of low- and high-income households. It is not uncommon to recalculate expansion factors for every purpose at hand. With a different household segmentation in different analyses, expansion factors become skewed. The only option to well-represent the target population (in this case the MSTM socio-economic data) is to recalculate expansion factors that help replicating the population of interest.

As an expansion factor describes how many households in reality a survey record represents, the factor is simply calculated by dividing the number of records by the number of households.

$$f_h = \frac{p_h}{r_h}$$

where f_h = Expansion factor for household type h
p_h = Number of households of household type h in population
r_h = Number of records in survey that interviewed household type h

Finally, the expansion factor f_h is assigned to each survey record that interviewed household type h. Household types that had been oversampled get a smaller expansion factor, while household types that were under-sampled receive a larger expansion factor.

A review of calculated expansion factors showed that some calculated factors turned out to be undesirably large. This was also true for the expansion factors originally provided by the BMC/MWCOG HTS, where the largest expansion factors were above 1,000. In other words, single survey records were supposed to represent the travel behavior of over 1,000 households. This happens in cases were too few survey records are supposed to represent a large number of households. Statistically, large expansion factors are problematic. In essence, the large expansion factors expand the travel behavior of a very few survey records to a larger part of the population. If the few surveyed records of this household type had an unusual day while surveyed or if the surveyed households for some reason had an atypical travel behavior, using large expansion factors would replicate this non-representative travel behavior in the analysis. To avoid using statistically insignificant expansions, the expansion factor in this task was limited to 500. In other words, each record may never represent more than 500 households in reality. Limiting the expansion factor increases the confidence in the analyses travel behavior, at the expense of slightly under-representing very rare household types.

Commonly, one single expansion factor is calculated for each record. In the MSTM model, however, households are segmented by two different classifications. Households by number of workers and income class are used for all work trips, and households by household size and income class are used for all non-work trips. To improve the linkage between the survey data and the model segmentation, two separate sets of expansion factors were calculated, one matching households by workers and income and the other one matching households by size and income. As calculating two expansion factors is an advanced procedure, a more traditional single expansion factor was calculated in addition. This allows future user to the model to go back to a single expansion factor if that shall be desired. At this point, only the work and non-work expansion factors are used. Table 21-1 summarizes the available expansion factors for each survey record.

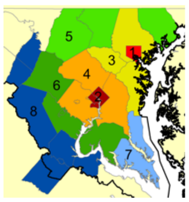
Table 21-1: Available expansion factors

Set	Description	Attribute name	Number of household types	Usage
1	Original	ffactor	unknown	Currently not used
2	By workers	expFW	20 (0 to 3+ workers, 1 to 5 income)	Used for work trips
3	By household size	expFnW	25 (1 to 5+ hh size, 1 to 5 income)	Used for non-work trips
4	By workers and size	expFboth	100 (0 to 3+ workers, 1 to 5+ hh size, 1 to 5 income)	Currently not used

Figure 21-1 summarizes newly calculated expansion factors by number of workers (columns), income (colors) and region (rows). Each field shows the expansion factor and in parentheses the number of surveyed records as well as the number of households in the MSTM model data.

	HHwrks 0	HHwrks 1	HHwrks 2	HHwrks 3+
HTS Region 1	207.5 (283sr/ 58711hh) 246.2 (97sr/ 23886hh) 194.1 (30sr/ 5823hh) 104.8 (12sr/ 1257hh) 366.5 (2sr/ 733hh)	167.7 (155sr/ 25989hh) 134.0 (234sr/ 31359hh) 255.9 (102sr/ 26105hh) 271.5 (38sr/ 10318hh) 500.0 (13sr/ 6868hh)	402.0 (21sr/ 8443hh) 223.1 (64sr/ 14279hh) 155.1 (100sr/ 15510hh) 81.8 (93sr/ 7612hh) 118.8 (50sr/ 5940hh)	500.0 (1sr/ 2008hh) 369.4 (9sr/ 3325hh) 500.0 (5sr/ 3578hh) 406.8 (4sr/ 1627hh) 229.8 (4sr/ 919hh)
HTS Region 2	208.3 (218sr/ 45415hh) 120.5 (204sr/ 24585hh) 104.0 (123sr/ 12796hh) 107.1 (67sr/ 7178hh) 362.7 (28sr/ 10155hh)	257.5 (121sr/ 31161hh) 120.2 (365sr/ 43867hh) 101.5 (460sr/ 46701hh) 124.2 (281sr/ 34891hh) 281.3 (136sr/ 38251hh)	441.4 (21sr/ 9270hh) 316.0 (56sr/ 17695hh) 175.4 (139sr/ 24384hh) 90.4 (255sr/ 23054hh) 113.5 (276sr/ 31325hh)	500.0 (3sr/ 2089hh) 433.2 (9sr/ 3899hh) 500.0 (6sr/ 5352hh) 234.7 (21sr/ 4929hh) 222.3 (25sr/ 5558hh)
HTS Region 3	210.1 (131sr/ 27529hh) 193.5 (160sr/ 30959hh) 221.5 (76sr/ 16837hh) 230.4 (33sr/ 7602hh) 500.0 (10sr/ 5709hh)	499.9 (49sr/ 24493hh) 178.2 (213sr/ 37964hh) 194.5 (211sr/ 41045hh) 231.4 (119sr/ 27534hh) 500.0 (42sr/ 23693hh)	500.0 (11sr/ 12679hh) 379.1 (70sr/ 26536hh) 284.5 (133sr/ 37840hh) 161.2 (200sr/ 32250hh) 348.6 (98sr/ 34162hh)	(2665hh) 500.0 (8sr/ 5556hh) 469.2 (17sr/ 7977hh) 172.6 (39sr/ 6732hh) 345.7 (17sr/ 5877hh)
HTS Region 4	131.5 (233sr/ 30634hh) 106.7 (375sr/ 40005hh) 165.7 (209sr/ 34638hh) 170.0 (144sr/ 24473hh) 414.4 (63sr/ 26110hh)	500.0 (93sr/ 50285hh) 159.4 (500sr/ 79713hh) 163.1 (587sr/ 95729hh) 158.4 (496sr/ 78547hh) 457.0 (190sr/ 86837hh)	500.0 (34sr/ 25574hh) 356.8 (153sr/ 54593hh) 260.7 (325sr/ 84739hh) 153.8 (576sr/ 88614hh) 276.3 (435sr/ 120179hh)	500.0 (6sr/ 6891hh) 500.0 (11sr/ 14623hh) 500.0 (40sr/ 22773hh) 311.5 (75sr/ 23363hh) 459.0 (57sr/ 26165hh)
HTS Region 5	113.5 (128sr/ 14527hh) 153.4 (122sr/ 18710hh) 161.2 (79sr/ 12733hh) 184.6 (37sr/ 6832hh) 475.9 (8sr/ 3807hh)	274.2 (56sr/ 15353hh) 183.1 (131sr/ 23985hh) 188.1 (177sr/ 33286hh) 183.4 (130sr/ 23839hh) 415.2 (43sr/ 17855hh)	500.0 (8sr/ 9813hh) 271.2 (76sr/ 20614hh) 216.7 (170sr/ 36846hh) 168.5 (195sr/ 32851hh) 335.3 (84sr/ 28162hh)	500.0 (1sr/ 2400hh) 423.2 (12sr/ 5078hh) 269.7 (34sr/ 9169hh) 232.0 (36sr/ 8352hh) 364.4 (17sr/ 6194hh)
HTS Region 6	137.2 (43sr/ 5899hh) 162.3 (56sr/ 9088hh) 112.4 (59sr/ 6629hh) 307.7 (15sr/ 4616hh) 500.0 (5sr/ 3198hh)	432.7 (21sr/ 9087hh) 188.5 (102sr/ 19227hh) 174.2 (142sr/ 24736hh) 194.1 (111sr/ 21542hh) 500.0 (34sr/ 22964hh)	500.0 (6sr/ 5805hh) 313.7 (52sr/ 16310hh) 246.4 (109sr/ 26855hh) 157.5 (182sr/ 28656hh) 288.0 (124sr/ 35708hh)	(1458hh) 500.0 (5sr/ 4166hh) 357.6 (19sr/ 6794hh) 282.4 (25sr/ 7060hh) 500.0 (13sr/ 7214hh)
HTS Region 7	137.9 (58sr/ 8000hh) 128.4 (66sr/ 8474hh) 148.4 (46sr/ 6826hh) 179.5 (17sr/ 3051hh) 500.0 (1sr/ 1309hh)	300.7 (26sr/ 7817hh) 148.0 (77sr/ 11393hh) 175.6 (93sr/ 16334hh) 264.9 (43sr/ 11392hh) 500.0 (10sr/ 7551hh)	500.0 (7sr/ 4954hh) 437.2 (22sr/ 9619hh) 214.6 (84sr/ 18027hh) 155.4 (100sr/ 15544hh) 500.0 (21sr/ 12267hh)	(1073hh) 302.4 (7sr/ 2117hh) 293.6 (14sr/ 4110hh) 172.3 (21sr/ 3619hh) 183.3 (14sr/ 2566hh)
HTS Region 8	261.2 (36sr/ 9405hh) 169.0 (44sr/ 7435hh) 184.0 (22sr/ 4048hh) 202.5 (11sr/ 2228hh) 213.0 (4sr/ 852hh)	300.7 (29sr/ 8720hh) 174.0 (66sr/ 11481hh) 225.7 (61sr/ 13766hh) 163.9 (55sr/ 9012hh) 339.8 (18sr/ 6116hh)	500.0 (7sr/ 5393hh) 314.9 (31sr/ 9761hh) 231.0 (66sr/ 15243hh) 146.9 (84sr/ 12337hh) 291.2 (33sr/ 9609hh)	(1186hh) 275.0 (8sr/ 2200hh) 390.7 (9sr/ 3516hh) 240.0 (12sr/ 2880hh) 386.8 (5sr/ 1934hh)
All Regions	177.1 (1130sr/ 200120hh) 145.1 (1124sr/ 163142hh) 155.8 (644sr/ 100330hh) 170.3 (336sr/ 57237hh) 428.7 (121sr/ 51873hh)	314.4 (550sr/ 172905hh) 153.4 (1688sr/ 258989hh) 162.4 (1833sr/ 297702hh) 170.5 (1273sr/ 217075hh) 432.4 (486sr/ 210135hh)	500.0 (115sr/ 81931hh) 323.3 (524sr/ 169407hh) 230.4 (1126sr/ 259444hh) 143.0 (1685sr/ 240918hh) 247.4 (1121sr/ 277352hh)	500.0 (11sr/ 19770hh) 500.0 (69sr/ 40964hh) 439.4 (144sr/ 63269hh) 251.3 (233sr/ 58562hh) 371.2 (152sr/ 56427hh)

Region Definition



Income 1
Income 2
Income 3
Income 4
Income 5

179.0 Expansion factor
Data in model, but < 1 survey records
Data in survey but not in model
X Data neither in survey nor model
sr Number of survey records
hh Number of households in model data

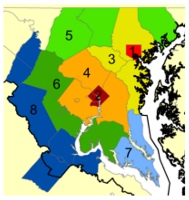
Figure 21-1: Expansion factors by number of workers, income and region

There are four cases in which no survey records were available, which are marked by a red dot. Several expansion factors had to be capped at 500. The summary shows that there are a couple of cases where only few survey records were available, particularly for households with three or more workers.

Figure 21-2 provides the same overview for households by household size (columns), income (color) and regions (row). Though survey records were available in each category, a small number of records particularly for household size 5+ required to cap expansion factors at 500.

	HHsize 1	HHsize 2	HHsize 3	HHsize 4	HHsize 5+
HTS Region 1	182.1 (278sr/ 50618hh) 126.1 (195sr/ 24585hh) 125.3 (77sr/ 9651hh) 83.4 (24sr/ 2001hh) 192.1 (7sr/ 1345hh)	204.0 (103sr/ 21011hh) 185.3 (126sr/ 23347hh) 162.8 (111sr/ 18067hh) 107.5 (79sr/ 8496hh) 104.0 (44sr/ 4577hh)	227.2 (40sr/ 9087hh) 283.0 (37sr/ 10472hh) 433.4 (24sr/ 10402hh) 175.6 (28sr/ 4917hh) 500.0 (7sr/ 4289hh)	293.5 (21sr/ 6164hh) 238.5 (30sr/ 7156hh) 470.1 (15sr/ 7052hh) 301.6 (11sr/ 3318hh) 247.9 (10sr/ 2479hh)	470.3 (18sr/ 8466hh) 450.3 (16sr/ 7205hh) 500.0 (10sr/ 5521hh) 417.2 (5sr/ 2086hh) 500.0 (1sr/ 1732hh)
HTS Region 2	216.4 (221sr/ 47828hh) 93.3 (457sr/ 42621hh) 84.5 (419sr/ 35412hh) 112.0 (220sr/ 24636hh) 230.1 (83sr/ 19099hh)	219.1 (93sr/ 20375hh) 199.7 (121sr/ 24163hh) 115.7 (224sr/ 25913hh) 87.8 (271sr/ 23794hh) 167.1 (234sr/ 39098hh)	252.4 (31sr/ 7825hh) 327.6 (29sr/ 9500hh) 225.2 (54sr/ 12159hh) 138.6 (72sr/ 9978hh) 196.0 (72sr/ 14111hh)	436.8 (12sr/ 5241hh) 303.5 (21sr/ 6374hh) 314.7 (26sr/ 8181hh) 158.1 (43sr/ 6797hh) 144.1 (59sr/ 8503hh)	500.0 (6sr/ 7209hh) 500.0 (6sr/ 6565hh) 500.0 (5sr/ 6580hh) 241.4 (18sr/ 4345hh) 352.8 (17sr/ 5998hh)
HTS Region 3	294.4 (109sr/ 32093hh) 202.4 (203sr/ 41092hh) 184.2 (132sr/ 24311hh) 178.9 (45sr/ 8051hh) 500.0 (6sr/ 3062hh)	144.3 (57sr/ 8224hh) 141.0 (159sr/ 22419hh) 206.3 (161sr/ 33215hh) 240.5 (150sr/ 36082hh) 446.4 (75sr/ 33483hh)	500.0 (15sr/ 10322hh) 284.2 (54sr/ 15346hh) 246.9 (80sr/ 19755hh) 162.3 (82sr/ 13310hh) 407.7 (35sr/ 14270hh)	500.0 (7sr/ 8254hh) 500.0 (19sr/ 12588hh) 359.9 (47sr/ 16917hh) 159.5 (76sr/ 12123hh) 325.2 (39sr/ 12681hh)	500.0 (3sr/ 7465hh) 500.0 (16sr/ 8396hh) 500.0 (17sr/ 9117hh) 144.1 (38sr/ 5475hh) 500.0 (12sr/ 6966hh)
HTS Region 4	192.1 (212sr/ 40722hh) 126.0 (533sr/ 67154hh) 152.8 (424sr/ 64781hh) 164.2 (243sr/ 39907hh) 326.8 (58sr/ 18953hh)	149.3 (95sr/ 14181hh) 115.1 (314sr/ 36156hh) 151.6 (388sr/ 58827hh) 147.3 (547sr/ 80596hh) 334.5 (361sr/ 120751hh)	500.0 (29sr/ 19433hh) 327.8 (94sr/ 30813hh) 254.4 (173sr/ 44018hh) 171.8 (222sr/ 38149hh) 368.5 (137sr/ 50484hh)	500.0 (22sr/ 16921hh) 430.5 (63sr/ 27122hh) 338.4 (116sr/ 39256hh) 188.5 (190sr/ 35812hh) 316.4 (139sr/ 43986hh)	500.0 (8sr/ 20890hh) 500.0 (35sr/ 25098hh) 479.4 (60sr/ 28764hh) 238.0 (89sr/ 21180hh) 500.0 (50sr/ 29593hh)
HTS Region 5	138.5 (127sr/ 17590hh) 184.0 (130sr/ 23917hh) 195.5 (84sr/ 16421hh) 221.0 (32sr/ 7071hh) 178.0 (13sr/ 2314hh)	83.2 (47sr/ 3912hh) 92.2 (142sr/ 13098hh) 163.1 (194sr/ 31645hh) 193.4 (173sr/ 33459hh) 319.2 (78sr/ 24895hh)	475.6 (15sr/ 7134hh) 338.5 (35sr/ 11846hh) 215.2 (83sr/ 17859hh) 171.9 (77sr/ 13239hh) 406.3 (31sr/ 12596hh)	500.0 (1sr/ 6477hh) 496.5 (22sr/ 10924hh) 241.4 (69sr/ 16656hh) 168.2 (76sr/ 12785hh) 500.0 (17sr/ 10738hh)	500.0 (3sr/ 6215hh) 500.0 (12sr/ 7749hh) 314.3 (30sr/ 9429hh) 149.6 (40sr/ 5983hh) 461.5 (13sr/ 5999hh)
HTS Region 6	181.3 (40sr/ 7251hh) 158.8 (93sr/ 14773hh) 186.6 (73sr/ 13619hh) 208.4 (34sr/ 7085hh) 494.8 (6sr/ 2969hh)	84.7 (22sr/ 1864hh) 116.9 (77sr/ 9005hh) 127.2 (133sr/ 16915hh) 221.8 (111sr/ 24615hh) 314.6 (87sr/ 27374hh)	500.0 (4sr/ 4203hh) 377.7 (23sr/ 8688hh) 284.0 (45sr/ 12778hh) 160.2 (74sr/ 11856hh) 500.0 (26sr/ 15851hh)	500.0 (2sr/ 4007hh) 500.0 (12sr/ 8384hh) 250.7 (50sr/ 12536hh) 169.4 (71sr/ 12025hh) 356.6 (40sr/ 14263hh)	500.0 (2sr/ 4661hh) 500.0 (10sr/ 7352hh) 308.4 (28sr/ 8634hh) 156.6 (43sr/ 6733hh) 500.0 (17sr/ 9385hh)
HTS Region 7	179.4 (50sr/ 8971hh) 177.2 (57sr/ 10098hh) 167.4 (49sr/ 8204hh) 247.5 (10sr/ 2475hh) 500.0 (1sr/ 751hh)	78.3 (26sr/ 2036hh) 96.3 (76sr/ 7317hh) 140.1 (97sr/ 13591hh) 243.6 (67sr/ 16323hh) 500.0 (12sr/ 8962hh)	500.0 (7sr/ 3637hh) 323.4 (16sr/ 5175hh) 303.1 (30sr/ 9094hh) 144.1 (42sr/ 6054hh) 342.1 (17sr/ 5816hh)	500.0 (6sr/ 3473hh) 387.8 (13sr/ 5042hh) 243.7 (37sr/ 9018hh) 134.7 (46sr/ 6198hh) 438.0 (12sr/ 5256hh)	500.0 (2sr/ 3387hh) 371.4 (10sr/ 3714hh) 219.2 (24sr/ 5262hh) 185.9 (16sr/ 2975hh) 500.0 (4sr/ 3002hh)
HTS Region 8	392.9 (29sr/ 11394hh) 187.1 (48sr/ 8980hh) 289.2 (19sr/ 5494hh) 175.2 (11sr/ 1927hh) 171.7 (3sr/ 515hh)	94.4 (31sr/ 2927hh) 160.0 (54sr/ 8639hh) 170.8 (70sr/ 11954hh) 194.6 (63sr/ 12260hh) 197.5 (38sr/ 7505hh)	500.0 (5sr/ 3578hh) 294.3 (17sr/ 5003hh) 228.3 (33sr/ 7534hh) 153.3 (33sr/ 5058hh) 376.2 (12sr/ 4514hh)	500.0 (3sr/ 3194hh) 255.6 (18sr/ 4600hh) 295.3 (24sr/ 7088hh) 136.8 (36sr/ 4923hh) 500.0 (3sr/ 3739hh)	500.0 (4sr/ 3201hh) 293.8 (12sr/ 3526hh) 372.4 (12sr/ 4469hh) 137.9 (19sr/ 2621hh) 500.0 (4sr/ 2381hh)
All Regions	203.1 (1066sr/ 216467hh) 135.9 (1716sr/ 233220hh) 139.3 (1277sr/ 177893hh) 150.5 (619sr/ 93153hh) 276.9 (177sr/ 49008hh)	157.2 (474sr/ 74530hh) 134.8 (1069sr/ 144144hh) 152.5 (1378sr/ 210127hh) 161.3 (1461sr/ 235625hh) 287.0 (929sr/ 266645hh)	446.7 (146sr/ 65219hh) 317.5 (305sr/ 96843hh) 255.9 (522sr/ 133599hh) 162.8 (630sr/ 102561hh) 361.8 (337sr/ 121931hh)	500.0 (74sr/ 53731hh) 415.1 (198sr/ 82190hh) 303.9 (384sr/ 116704hh) 171.2 (549sr/ 93981hh) 318.6 (319sr/ 101645hh)	500.0 (46sr/ 61494hh) 500.0 (117sr/ 69605hh) 418.2 (186sr/ 77776hh) 191.8 (268sr/ 51398hh) 500.0 (118sr/ 65056hh)

Region Definition



- Income 1
- Income 2
- Income 3
- Income 4
- Income 5

179.0 Expansion factor
Data in model, but < 1 survey records
Data in survey but not in model
X Data neither in survey nor model
sr Number of survey records
hh Number of households in model data

Figure 21-2: Expansion factors by household size, income and region

It is common practice to base expansion factor on at least 30 survey records. Using fewer than 30 records bears the risk of extrapolating unusual travel patterns. If at least 30 records are used, averaging across all records helps extracting a representative travel behavior.

In MSTM phase II, eight HTS regions were differentiated in trip generation and mode split. While the use of regions in mode split was meant to be a placeholder, the use of regions in trip generation becomes doubtful when looking at the survey records availability by region in Figure 21-1 and Figure 21-2. Given the small number of records by region in many categories, it was decided that all regions need to be collapsed into one when estimating trip rates in MSTM phase III. This way, the number of survey records is large enough to ensure robust and statistically significant trip rates across all household categories. Using one region only, all categories have significantly more than 30 survey records except one: Household type 3+ workers income 1 has 11 records only. While this is unfortunate, this single exception appears to be acceptable given the large reliability across all other categories.

Figure 21-3 shows the expanded number of MSTM households in the area that is covered by the survey. Bars show the number of households by household type, defined here by number of workers (0, 1, 2 or 3+) and income (1, 2, 3, 4 or 5). The blue bars show the original expansion factors that were provided by the survey data. As that original expansion was not geared towards the MSTM household types, it is not unexpected that those bars do not match up nicely with the grey bars, which show the MSTM household data for the same area. The red bars show the number of expanded households when using the newly calculated expansion factors. In most cases, the red and the grey bars line up nicely. There are a few cases where the two do not match, for example 2w_inc1 and 3+w_inc1. Even though the newly calculated expansion factors are doing better than the original expansion factors, the target population is not quite reached. This is because expansion factors were capped at 500 to avoid over-fitting the expansion. It is fairly rare that a household has 2 or 3+ workers, yet belongs to the lowest income group. The survey does not represent such rare households very well, and thus the expansion does not this household type very well. Given the comparatively small number of households in that category, the deviation is assumed to be acceptable.

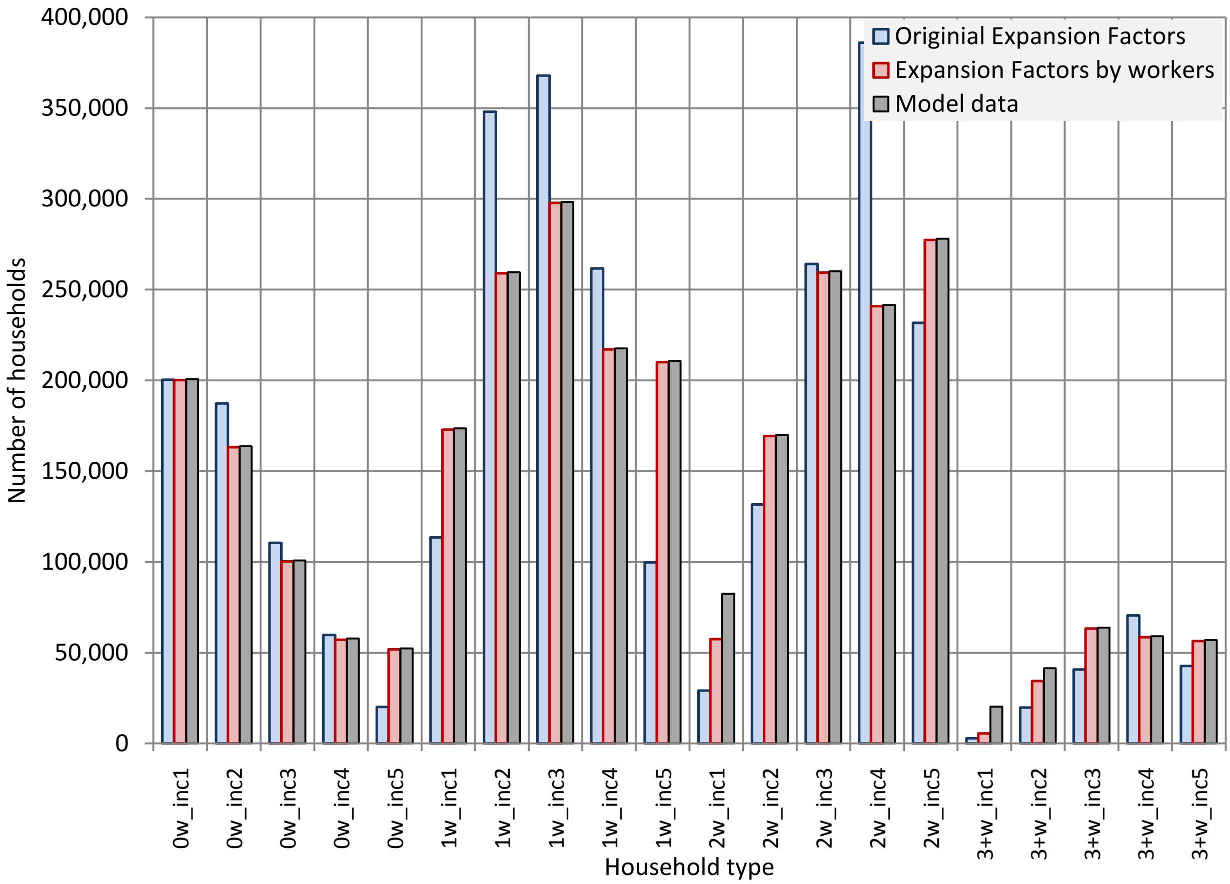


Figure 21-3: Expanded number of households by workers

Figure 21-4 shows the same comparison for households by household size (1, 2, 3, 4 or 5+) and income (1, 2, 3, 4 or 5). Again, most household types are closely matched by the new expansion factors. Exceptions are size4_inc1 and size5+_inc1. Again, these are rare household types that are not well captured by the household travel survey. However, given the comparatively small number of households in these categories, the error introduced is fairly minor. If the cap of 500 for expansion factors was removed, the number of households would be matched precisely.

However, this precision would be bought by accepting expanding the HTS based on a very small number of records, which is likely to overemphasize outliers. Therefore, the small errors shown in Figure 21-3 and Figure 21-4 are assumed to be more acceptable than over-specifying the model.

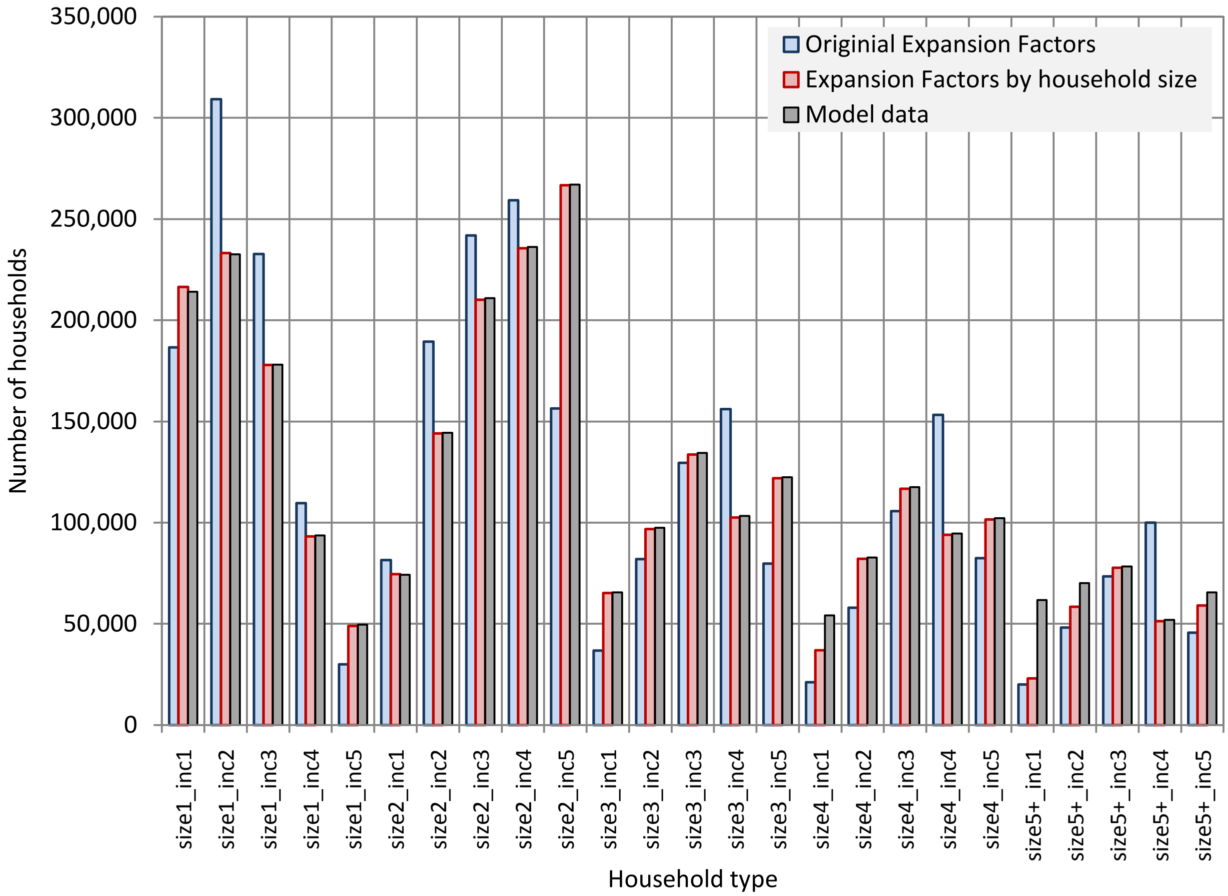


Figure 21-4: Expanded number of households by size

Next, the data has been summarized by income category to show in Figure 21-5. The light blue bars show the deviation of the original expansion factors provided by the HTS from the MSTM model data. The brown bars show the deviation reached with the new expansion factors. Most categories match very well. Income group 1 is underrepresented by 9 percent, which is more than desired yet three-times better than the original expansion factors.

Finally, Figure 21-6 shows the impact of the new expansion factors on the number of trips generated within the HTS area. As no target data are available for the number of trips generated, only the trips based on the original expansion factors are compared with the number of trips based on the recalculated expansion factors. While the total number of trips is only 1 percent larger with the new expansion factors, quite some shifts may be observed across different purposes. These new expansion factors are expected to better connect the survey data with the household data in the MSTM model.

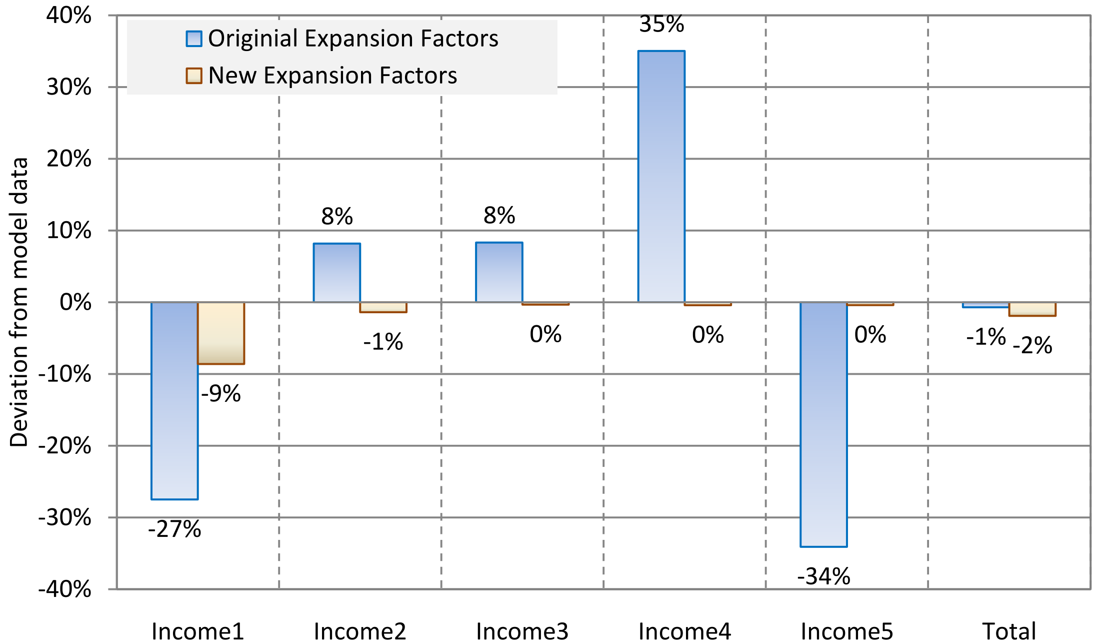


Figure 21-5: Comparison of expanded number of households by income

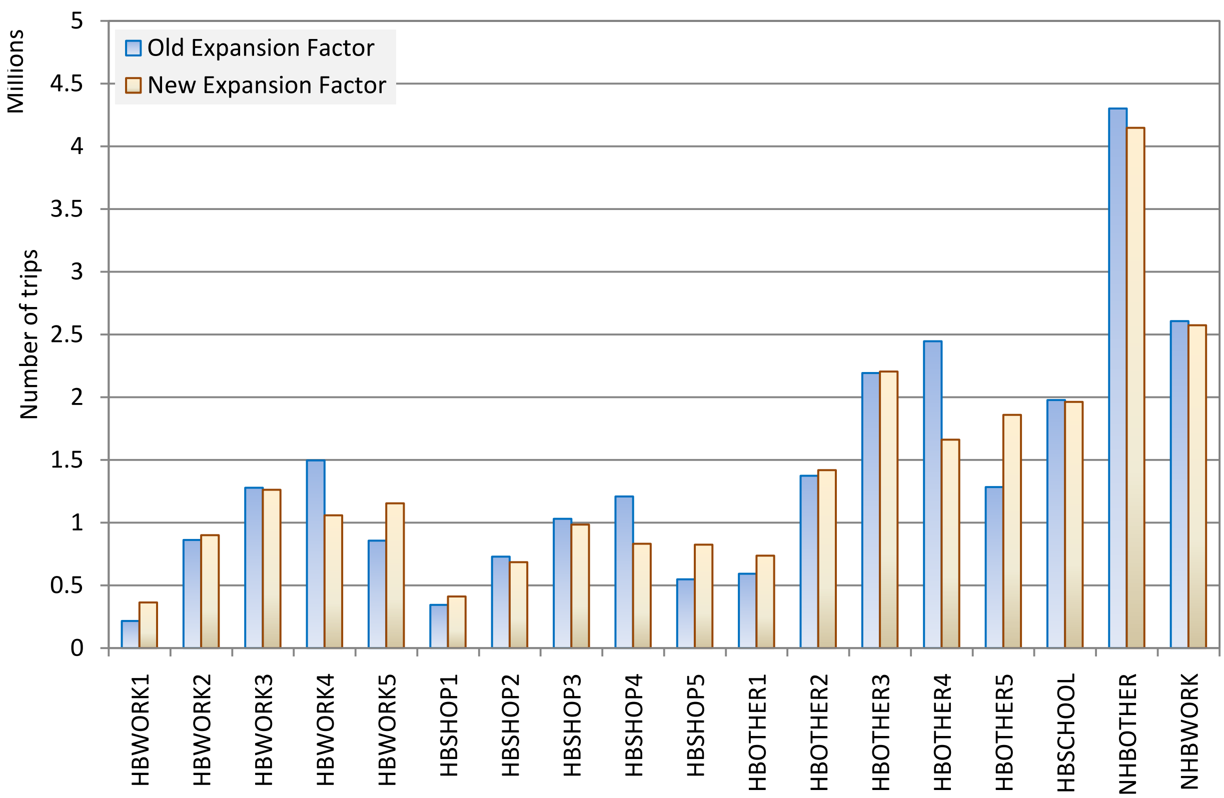


Figure 21-6: Number of expanded trips by purpose

22 Appendix G: MSTM Productions & Attractions Parameters

The HTS was processed to develop updated parameters to use in the MSTM Trip Generation components of the MSTM statewide model. The parameters were developed specific to the regions identified in Figure 20-2. This will allow MSTM to vary these parameters by region, where sufficient survey data is available. Records in the individual data sets were also grouped into income, worker, and household size categories, as shown in the tables below.

Table 22-1: Income categories

Income Range	Category
$0 < $29,000	1
$30,000 - $59,999	2
$60,000 - $999,999	3
$100,000 - $149,999	4
$150,000+	5

Table 22-2: Worker categories

Workers	Category
0 Workers	1
1 Worker	2
2 Workers	3
3+ Workers	4

Table 22-3: Household size categories

Household Size	Category
1 Person	1
2 Person	2
3 Person	3
4 Person	4
5 Person	5

22.1 Productions

The MSTM Trip Productions parameter was developed using the survey's household and trip data files. The work-related trips were categorized into a combination group of income, number of workers per household, and region. The non-work related trips were categorized by income, household size, and region only, since number of workers is not relevant for those trips. In a few cases where there were very few survey records in a particular grouping, it was combined with another group that was somewhat similar (i.e., same income and workers, different region) and the same rate was applied across the combination of groupings.

The survey data expansion factors were used to get the total number of trips. The script then classified the survey records by the grouped variables and regions (see Table 21-1, Table 22-2 and Table 22-3). The trip purpose was determined by mapping the origin purpose, destination purpose, and income class to the generalized model purposes. Once the data was classified, the production rate by the grouping was calculated. Input and output files to the Productions HTS processing script are listed in Table 22-4.

Table 22-4: Production HTS processing input and output files

File name	Description
Inputs	
tpb_bmc_hts07_hf.csv	HTS household data
tpb_bmc_hts07_tf_w_smz.csv	HTS trip data
RegionalDefinition.csv	Mapping of County to Region
Purposes.csv	Mapping of O-D-Purpose to general Purpose category
Outputs	
FactoredWorkRelatedObservations.csv	Summary for work related trips
FactoredNonWorkRelatedObservations.csv	Summary for non work-related trips
FactoredallHHwkrs.csv	Summary of workers by households
FactoredallHHsiz.csv	Summary by household size
FactoredWorkRelatedRates.csv	Work-related production rates
FactoredNonWorkRelatedRates.csv	Non work-related production rates

The outputs were compared to the BMC and MWCOG rates from the survey, as well as actual income data and the rates calculated for the first generation MSTM model. These comparisons showed a clear relationship between the calculated rates for the survey area and the survey results. Figure 22-1 shows a comparison of the BMC model rates, the MSTM Gen 1 rates, and the new MSTM rates. The income classes used are slightly different for the three models, but all follow the same general pattern. Trips rates are lower in lower income categories, but level out once income reaches about the 50K per year.

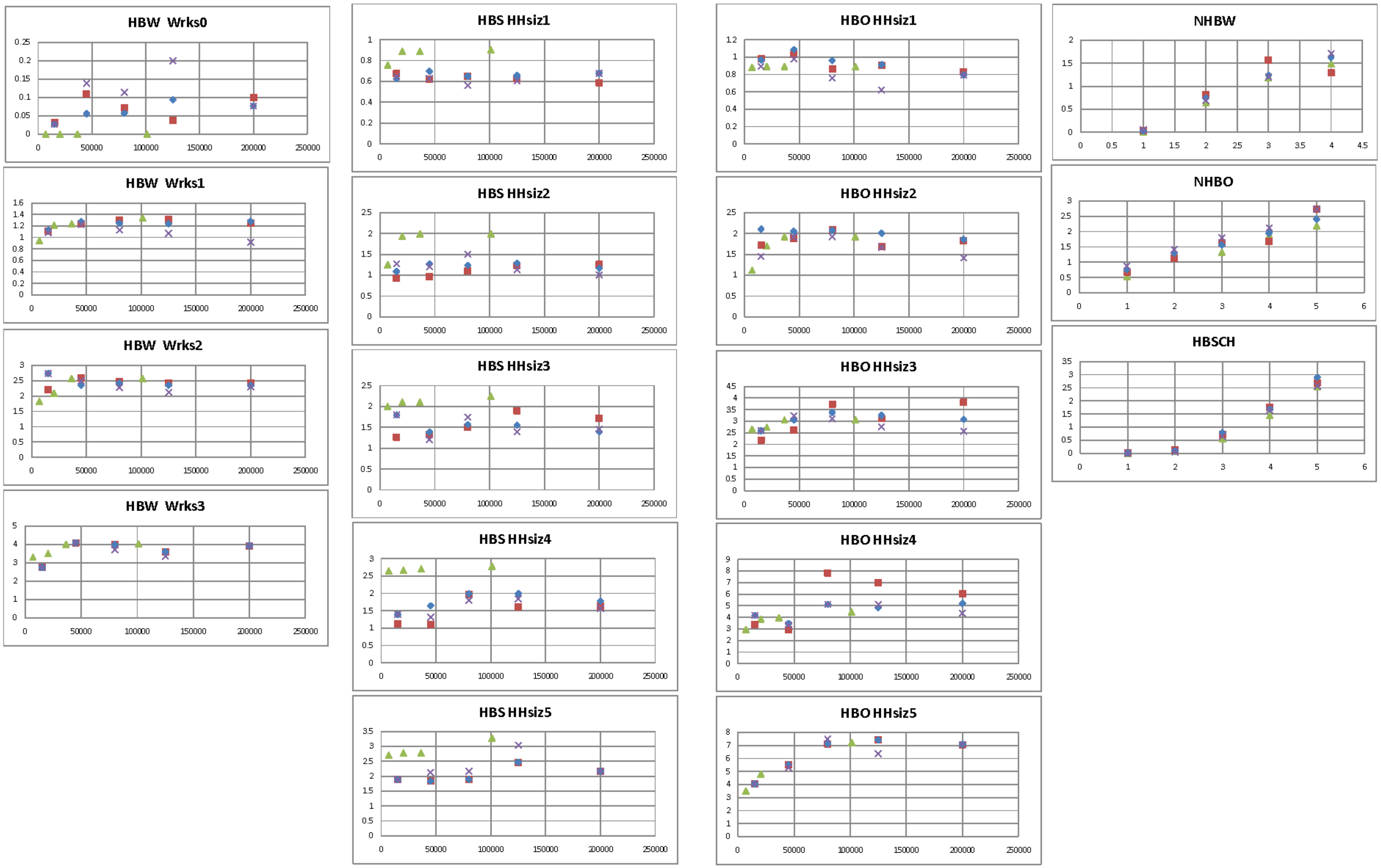


Figure 22-1: Trip production rates

The resulting MSTM input production rates by purpose and region that were updated through this processing of the HTS data are contained in the files<purpose>_rates.txt.

22.2 Attractions

Trip attractions were processed using regression analysis that estimates how many trips are attracted by the number of households and employment. The HTS trip file was the only survey file used for the attractions. The trip purposes were appended based on the income, origin purpose and destination purpose.

Table 22-5: Attraction HTS processing input and output files

File name	Description
Inputs	
tpb_bmc_hts07_tf_w_smz.csv	HTS trip data file, appended County FIPS code and Purpose
Outputs	
tripAttrRatesCounty.pdf	Regression plots of expanded trips vs demographic variables, coefficient parameters and other summary statistics

The surveyed origin-destination trip format was converted into productions-attractions format so that the attraction end of the trips could be isolated. Then the summarized survey expansion factor was regressed against the number of households and employees by county. The dependent variable, the number of households or employment (by employment classification), were applied specifically to each purpose based on its unique characteristics. For example, the total number of Retail employment was used to regress the Home-Based Shop purpose. Table 22-6 shows the purposes and the independent variables. The coefficient for households or employment was calculated by region and by purpose.

Table 22-6: Trip purpose and independent variables

Purpose		Independent variable
HBW	Home Based Work	Total Employment
HBS	Home Based Shop	Retail Employment
HBO	Home Based Other	Households, Other Employment
HBSchool	Home Based School	School Employment
NHBWork	Non Home Based Work	Office Employment, Other Employment
NHBOther	Non Home Based Other	Households

The R-squared values of the estimated attraction rates versus observed survey rates and the associated scatterplots show a clear relationship between the estimated and observed values. Figure 22-2 and Figure 22-3 show the scatterplot for all purposes purpose (where each dot represents a county). The coefficients calculated by the regression are used for the attraction rates SMZ-regionwide.

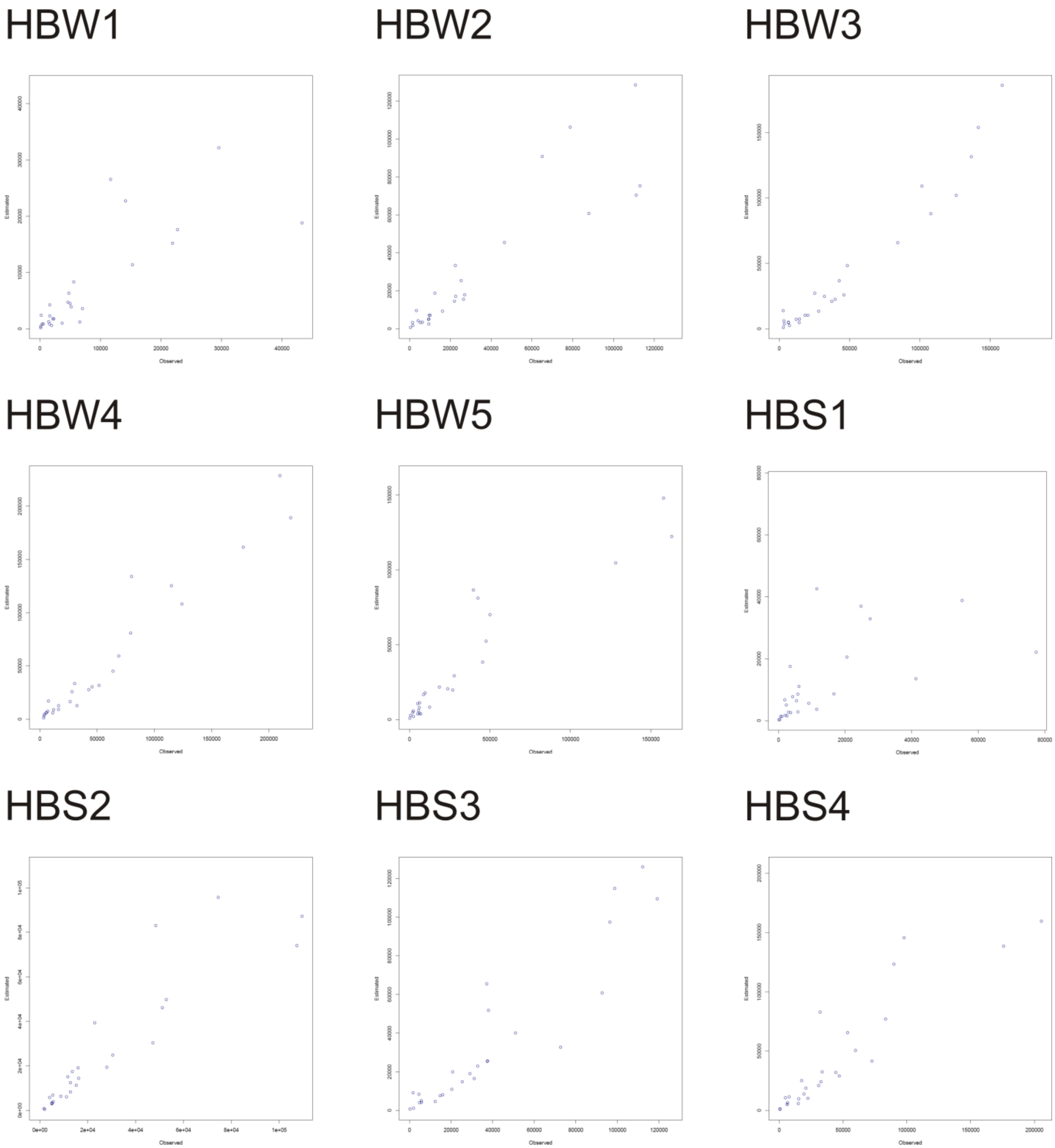


Figure 22-2: Trip attractions by purpose, part 1

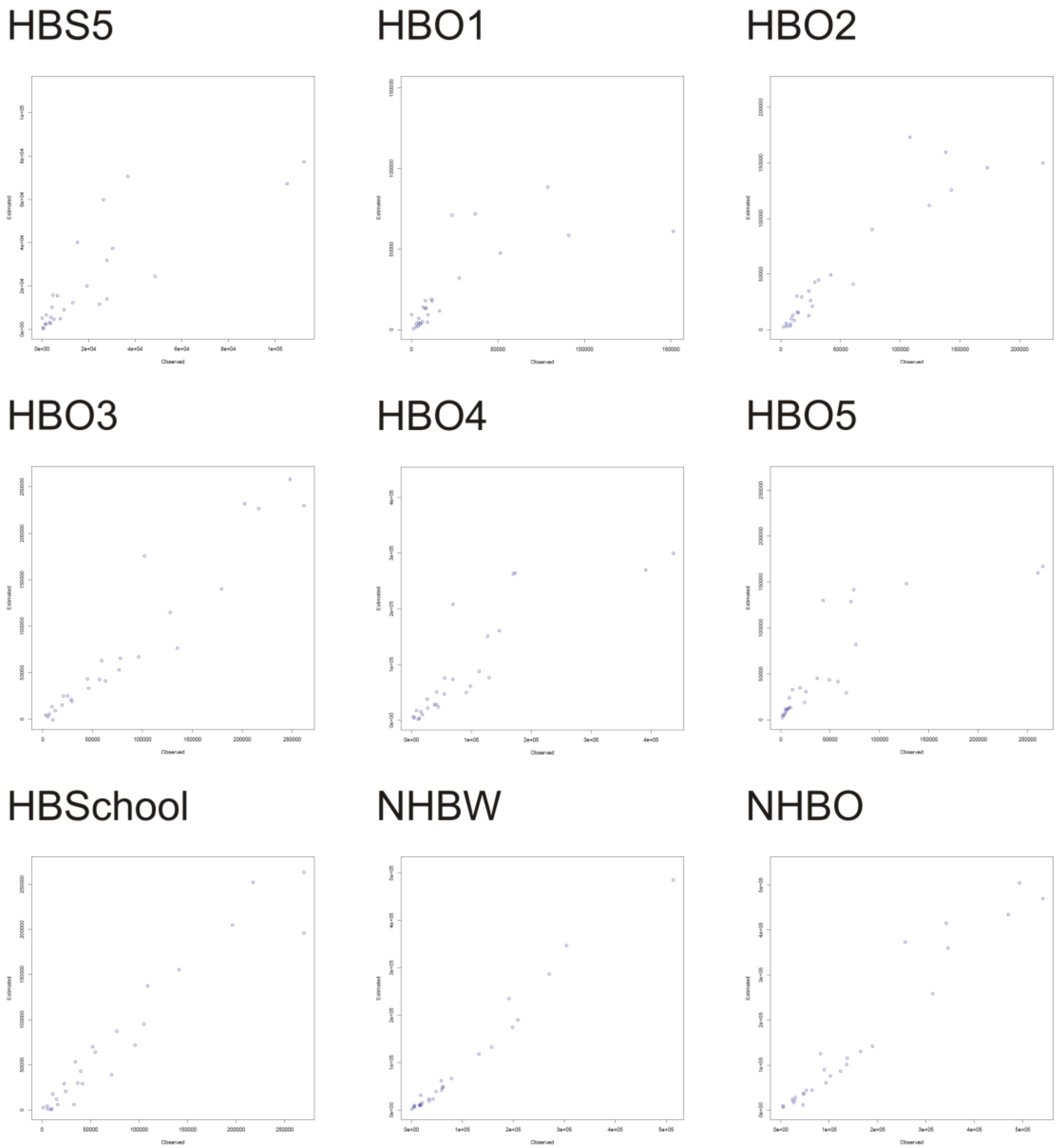


Figure 22-3: Trip attractions by purpose, part 2

The resulting MSTM input attraction rates by purpose and region that were updated through this processing of the HTS data are contained in the files<purpose>_rates.txt.

23 Appendix H: MSTM Destination Choice Calibration Targets

The HTS data was the primary source to develop observed trip length distributions by trip purpose, region, and income category for use in the MSTM trip distribution model calibration.

23.1 Home Based School Trip Distribution Targets

Since the gravity model was replaced by the destination choice model for all purposes except Home-Based School, this HTS processed target data now only applies to the HBSC. The HTS survey trip file was used to create the input data file for this script. The trip purpose was appended based on the income, origin purpose and destination purpose, and the skims data was appended to get trip lengths in minutes. Histograms were created for each region and purpose.

The trip length frequency distribution data was used to generate parameters, based on the shape of the line that was fit to each of the curves. The parameter values are included in Section 6.2.

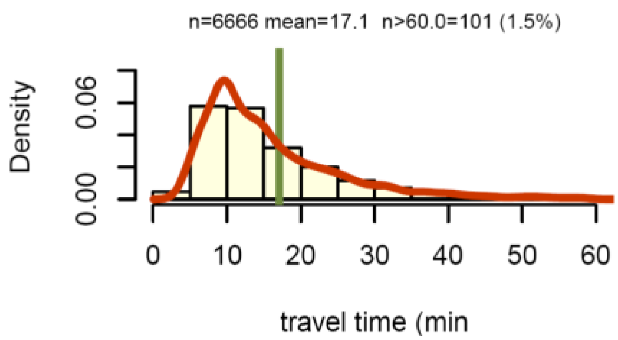


Figure 23-1: Trip length frequency distribution, home-based school purpose

The MSTM trip distribution parameters to be updated by this HTS data during calibration are listed at the start of the TripDistribution.s CUBE script file.

23.2 Destination Choice Model Targets

Calibration targets included region-to-region worker flows from the 2005-2009 American Community Survey Census Transportation Planning Package (CTPP), and HTS-based trip flows by region (Table 23-2 through Table 23-7). The less frequent region-to-region trip flows are based on very small sample sizes, and therefore are not considered accurate point estimates of the real region flows.

Table 23-1: MDHTS observed distance by purpose

Purpose	Average Trip Distance (miles)
HBW	12.6
HBS	5.2
HBO	5.9
NHBW	7.4

OBO	5.3

Table 23-2: Observed region-to-region worker flows (CTPP)

	Destination								
Origin	1	2	3	4	5	6	7	8	Total
1	154,465	6,970	81,115	5,500	2,950	240	30		**251,270**
2	2,350	423,655	5,200	147,870	430	6,210	610	1,147	**587,472**
3	146,540	66,500	483,880	78,255	14,075	2,300	1,510	606	**793,666**
4	8,065	777,810	31,460	901,830	5,805	36,570	1,790	4,350	**1,767,680**
5	23,115	10,340	59,835	34,500	158,305	2,643	89	1,208	**290,035**
6	385	89,970	1,660	114,270	1,045	119,320	3,093	5,164	**334,907**
7	240	14,430	2,260	14,020	85	4,845	52,435	366	**88,681**
8	308	25,373	1,452	31,734	6,132	27,030	498	168,924	**261,451**
Total	**335,468**	**1,415,048**	**666,862**	**1,327,979**	**188,827**	**199,158**	**60,055**	**181,765**	**4,375,162**

Table 23-3: HBW observed region-to-region trip flows (HTS)

	Destination								
Origin	1	2	3	4	5	6	7	8	**Row Totals**
1	235,575	8,931	119,165	13,173	20,443	871	171		**398,329**
2	8,344	638,283	35,721	353,855	5,006	40,916	4,511	12,220	**1,098,855**
3	130,197	39,171	456,590	70,577	51,563	2,197	3,422	804	**754,522**
4	13,092	383,002	63,821	876,617	25,367	92,959	5,166	13,479	**1,473,504**
5	25,053	7,107	55,940	27,986	267,853	1,630		3,225	**388,795**
6	850	46,831	2,808	102,959	1,334	170,251	4,417	18,843	**348,293**
7	171	5,413	3,946	4,951		4,154	54,103	171	**72,907**
8		11,236	1,035	16,966	2,919	21,858	171	110,753	**164,937**
Total	**413,282**	**1,139,975**	**739,026**	**1,467,085**	**374,484**	**334,836**	**71,959**	**159,496**	**4,700,143**

Table 23-4: HBS observed region-to-region trip flows (HTS)

	Destination								
Origin	1	2	3	4	5	6	7	8	**Row Totals**
1	177,208	521	52,713	2,081	2,107	271			**234,899**
2		469,739	2,539	78,751	1,446	3,097	443		**556,015**
3	63,565	3,571	591,814	22,540	20,575	526	3,274	152	**706,018**
4	1,293	100,980	23,960	1,095,224	4,303	37,911	4,349	1,709	**1,269,731**
5	2,263	655	16,088	3,174	323,964	629		710	**347,482**
6		3,566	470	38,338	641	321,820	3,735	7,937	**376,506**
7		319	2,221	4,175	135	3,624	49,444		**59,919**
8		287		958	745	3,703		143,492	**149,186**
Total	**244,329**	**579,638**	**689,805**	**1,245,241**	**353,916**	**371,580**	**61,246**	**154,001**	**3,699,756**

Table 23-5: HBO observed region-to-region trip flows (HTS)

Origin	Destination 1	2	3	4	5	6	7	8	Row Totals
1	444,982	3,531	113,236	5,706	9,578			500	**577,532**
2	2,738	1,006,440	11,375	240,018	2,060	15,458	1,242	2,080	**1,281,411**
3	114,485	8,090	1,125,312	43,352	43,505	2,044	3,225	287	**1,340,302**
4	6,907	224,974	46,635	2,322,641	9,368	55,844	3,810	4,753	**2,674,933**
5	12,955	1,239	43,921	14,241	673,555	328		315	**746,554**
6	161	15,151	771	58,030	328	593,606	3,761	10,556	**682,365**
7		1,655	3,171	3,321		3,703	133,316	304	**145,470**
8	574	1,333	440	4,631	495	11,088	630	289,824	**309,014**
Total	**582,802**	**1,262,414**	**1,344,860**	**2,691,941**	**738,889**	**682,071**	**145,983**	**308,619**	**7,757,580**

Table 23-6: NHB observed region-to-region trip flows (HTS)

Origin	Destination 1	2	3	4	5	6	7	8	Row Totals
1	137,735	1,908	62,466	2,613	5,035	247			**210,004**
2	1,525	669,270	8,858	123,075	171	13,337	1,352	4,727	**822,315**
3	42,672	6,192	282,168	18,730	16,170	1,018	994		**367,945**
4	3,964	101,757	24,951	571,940	6,525	27,633	963	4,949	**742,681**
5	3,515		13,884	6,476	136,153	162		571	**160,762**
6		10,116	162	26,617		104,387	1,030	2,902	**145,213**
7		1,297	1,103	2,046		1,304	35,273		**41,023**
8		6,399		2,759	450	3,328		46,304	**59,240**
Total	**189,412**	**796,938**	**393,591**	**754,257**	**164,503**	**151,417**	**39,612**	**59,452**	**2,549,183**

Table 23-7: OBO observed region-to-region trip flows (HTS)

Origin	Destination 1	2	3	4	5	6	7	8	Row Totals
1	193,136	513	64,469	918	4,313			574	**263,923**
2	622	510,847	1,778	114,132	1,557	4,468	363	885	**634,654**
3	54,978	3,902	720,740	31,124	24,891	152	2,149	192	**838,128**
4	1,397	99,229	23,970	1,157,485	8,004	36,157	2,881	1,813	**1,330,935**
5	3,429	304	19,721	4,510	374,153		163	783	**403,064**
6	287	3,214	331	25,932	749	333,385	2,897	7,857	**374,651**
7		731	2,383	2,057	163	3,169	75,911		**84,415**
8		1,037	192	2,130	723	3,615	314	161,180	**169,191**
Total	**253,849**	**619,777**	**833,584**	**1,338,288**	**414,554**	**380,947**	**84,678**	**173,284**	**4,098,961**

24 Appendix I: Destination Choice Sampling Correction Factors

Notation:

$i \in C$	=	unique alternatives from the full set
$i \in D \subset C$	=	unique alternatives from the sample
$q(i)$	=	selection probability (probability to be drawn)
n_i	=	selection frequency in the sample
N	=	sample size
V_i	=	utility of a choice alternative
$P(i)$	=	choice probability

Note that the selection frequencies in the sample over unique alternatives are totaled to the sample size:
$\sum_{i \in D} n_i = N$.

However, the number of unique alternatives in the sample D can be any number between 1 and N inclusive.

The choice probability with sampling correction factors can be calculated by the following formula:

$$P(i) = \frac{\exp\left[V_i + \ln\left(\frac{n_i}{N \times q(i)}\right)\right]}{\sum_{j \in D} \exp\left[V_j + \ln\left(\frac{n_j}{N \times q(j)}\right)\right]} = \frac{\left(\frac{n_i}{N \times q(i)}\right) \times \exp(V_i)}{\sum_{j \in D}\left(\frac{n_i}{N \times q(j)}\right) \times \exp(V_j)} . \qquad (1)$$

Since N is a fixed number it can be cancelled out and the formula (1) can be equivalently re-written in a simpler form:

$$P(i) = \frac{\exp\left[V_i + \ln\left(\frac{n_i}{q(i)}\right)\right]}{\sum_{j \in D} \exp\left[V_j + \ln\left(\frac{n_j}{q(j)}\right)\right]} = \frac{\left(\frac{n_i}{q(i)}\right) \times \exp(V_i)}{\sum_{j \in D}\left(\frac{n_i}{q(j)}\right) \times \exp(V_j)} . \qquad (2)$$

Formula (1) assumes a utility correction factor of $\ln\left(\frac{n_i}{N \times q(i)}\right)$, while formula (2) assumes a correction factor of $\ln\left(\frac{n_i}{q(i)}\right)$. Since both formulas yield the same probabilities, the simpler correction factor from the formula (2) is normally applied in the choice context.

25 Appendix J: MSTM Mode Choice Targets

The MSTM mode choice calibration targets were developed from the 2007 HTS, 2007 MTA on-board survey and 2008 WAMTA onboard survey data. The MSTM mode choice parameters to be updated by this HTS data during calibration are contained in file modeChoiceCoeff.dat. This included mode shares by purpose and income. For the HTS processing script uses the HTS household and trip data files. The trip purposes were appended based on the income, origin purpose and destination purpose. The modes classifications are shown in Table 25-2.

Table 25-1: Mode choice HST processing input and output files

File name	Description
Inputs	
tpb_bmc_hts07_hf.csv	HTS household data
tpb_bmc_hts07_tf_w_smz.csv	HTS trip data
Inputs	
ModeShares.csv	Un-expanded mode share

Table 25-2: Mode classification

Mode	Mode Choice Classification
Transit	Transit
Auto Driver	Auto D
Auto Passenger	Auto P
Walk	Non-Motorized
Bike	Non-Motorized
Other	N/A

The mode share was calculated by mode choice classification, and purpose category. Mode share is a ratio of mode share, calculated by dividing the total number of trips by each mode by the total number of trips in that category.

To enhance the mode choice targets, the HTS survey were augmented with the mode choice calibration targets used in the calibration of the BMC and MWCOG MPO models, consisting of the 2007 MTA and 2007 WMATA onboard surveys respectively. The same calibration targets from these two surveys were aggregated to develop the MSTM transit targets. However, the income group definitions were not same in the BMC, MWCOG and MSTM models for HBW, HBO and HBS purposes. Due to this difference, the transit targets by income used in the BMC and MWCOG model calibration were redistributed to the MSTM income groups based on the total households in the each income group. For the remaining MSTM trip purposes (HBSCH, NHBW and OBO), the transit targets from the BMC and MWCOG models were added straightway.

Once the MSTM total transit and drive to transit trips were developed, the modes specific targets from the BMC and MWCOG calibration targets were aggregated and then adjusted to match to the MSTM transit total and drive to transit trips. The drive alone (DA), share ride 2 (SR2) and share ride 3+ (SR3+) targets were developed from the 2007 household travel survey. Table 25-3 shows the MSTM mode choice calibration targets.

Table 25-3: MSTM transit targets

MWCOG	ACS 2008 Households			2007 WAMTA On Board Survey					
				HBW		HBS		HBO	
	INC Group	Total HHs	Distri-bution	Transit	Drive2 TRN	Transit	Drive2 TRN	Transit	Drive2 TRN
INC1	< 36 K	361,845	20%	171,407	15,184	13,985	289	68,859	6,501
INC2	36K - 65K	341,777	19%	120,951	32,616	5,394	540	17,891	4,148
INC3	65K - 99K	447,703	25%	146,217	53,967	3,543	862	15,851	4,948
INC4	> 99K	647,697	36%	241,961	120,156	3,493	35	18,854	11,829
Total		1,799,022	100%	680,536	221,923	26,415	1,726	121,455	27,425

BMC	ACS 2008 Households			2007 MTA On Board Survey					
				HBW		HBS		HBO	
	INC Def	Total HHs	Distri-bution	Transit	Drive2 TRN	Transit	Drive2 TRN	Transit	Drive2 TRN
INC1	< 15K	173,069	9%	12,632	1,013	2,402	105	14,625	1,469
INC2	15K-30K	205,380	11%	34,077	2,022	3,667	210	15,042	1,007
INC3	30K - 50K	309,918	16%	32,526	7,142	1,192	248	6,591	887
INC4	> 50K	1,267,542	65%	38,829	24,201	1,001	13	6,677	2,661
Total		1,955,909	100%	118,065	34,378	8,262	575	42,935	6,024

MSTM	ACS 2008 Households			2007 MTA + WAMTA On Board Survey					
				HBW		HBS		HBO	
	INC Def	Total HHs	Distri-bution	Transit	Drive2 TRN	Transit	Drive2 TRN	Transit	Drive2 TRN
INC1	< 20K	421382	11%	101,710	16,776	9,227	384	56,174	5,643
INC2	20K-40K	537317	14%	143,152	25,916	11,103	1,013	49,433	4,328
INC3	40K - 60K	585141	16%	129,036	31,155	4,043	791	22,367	3,011
INC4	60K- 100K	918665	24%	166,790	50,628	3,237	35	19,043	7,591
INC5	> 100 K	1292426	34%	257,913	131,826	7,065	78	17,373	12,877
Total		3754931	100%	798,601	256,301	34,677	2,300	164,390	33,449

All In-come Purposes	2007 MTA + WAMTA On Board Survey					
	SCH		NHBW		OBO	
	Transit	Drive2 TRN	Transit	Drive2 TRN	Transit	Drive2 TRN
MWCOG	22,451	1,464	145,350	45,128	107,189	33,522
BMC	7,773	507	11,499	1,843	8,480	1,369
MSTM	30,224	1,971	156,849	46,971	115,669	34,891

Table 25-4: Mode choice calibration targets

Purpose/Mode (NEW)	Drive Alone	Shared Ride 2	Shared Ride 3+	Bus (W)	Express Bus (W)	Rail (W)	Commuter Rail (W)	Bus (D)	Express Bus (D)	Rail (D)	Commuter Rail (D)
HBOTHER1	313,087	437,335	68,903	26,925	-	17,365	454	1,570	-	9,188	672
HBOTHER2	541,279	650,989	108,195	23,695	-	15,281	399	1,381	-	8,086	591
HBOTHER3	738,732	1,051,151	195,760	10,721	-	6,914	181	625	-	3,659	268
HBOTHER4	619,558	745,822	122,951	9,128	-	5,887	154	532	-	3,115	228
HBOTHER5	738,712	809,082	152,554	8,327	-	5,370	140	485	-	2,842	208
HBSCHOOL	216,608	838,577	5,233	23,173	-	5,080	-	1,304	-	668	-
HBSHOP1	186,427	234,876	34,575	5,184	-	3,344	87	84	-	492	36
HBSHOP2	287,201	299,780	53,977	6,238	-	4,023	105	101	-	592	43
HBSHOP3	388,815	434,657	74,595	2,272	-	1,465	38	37	-	216	16
HBSHOP4	352,417	355,295	53,088	1,819	-	1,173	31	29	-	173	13
HBSHOP5	376,824	336,022	43,043	3,970	-	2,560	67	64	-	377	28
HBWORK1	272,487	61,884	10,767	29,191	75	38,978	823	2,390	545	25,892	3,815
HBWORK2	595,244	95,924	12,176	41,085	105	54,860	1,159	3,364	767	36,442	5,369
HBWORK3	868,536	126,321	17,036	37,033	95	49,450	1,045	3,032	692	32,849	4,840
HBWORK4	698,660	76,101	10,665	47,869	123	63,919	1,350	3,920	894	42,460	6,256
HBWORK5	688,675	85,695	11,572	74,021	190	98,840	2,088	6,061	1,383	65,657	9,673
NHBOTHER	1,477,524	1,571,543	193,976	32,430	-	48,348	-	1,392	-	33,499	-
NHBWORK	913,678	188,919	33,697	44,112	-	65,765	-	1,874	-	45,097	-
TOTAL				427,193	588	488,624	8,121	28,246	4,281	311,304	32,053

26 Appendix K: MSTM Time of Day Parameters

Processing of the HTS data was done to develop time of day distributions by trip purposefor use in the MSTM Temporal Allocation component of the MSTM model. The input files for the development of these parameters were the HTS household and trip data files. The trip purposes were appended based on the income, origin purpose and destination purpose. The trip purposes were appended in Production-Attraction format.

Table 26-1: Time of day HTS processing inputs and output files

File name	Description
Input	
tpb_bmc_hts07_hf.csv	HTS household data
tpb_bmc_hts07_tf_w_smz.csv	HTS trip data
Purposes_PA_AP	Trip purposes, in Productions-Attractions format
Output	
ToD.csv	Time of day output

Based on the beginning time of each trip, the record was assigned to one of four time periods based on **Error! Reference source not found.**Table 26-2.

Table 26-2: Time periods

Abbreviation	Time period
AM	6:30 am - 9:30 am
MD	9:31 am - 15:30 pm
PM	15:31 pm - 18:30 pm
NT	18:31 pm - 6:29 am (next day)

The number of surveyed trips was then summarized by time period as well as by direction (Production to Attraction, or Attraction to Production). The percentage of trips by time period and direction, by purpose, defines the input parameters used in the MSTM Time of Day model, as shown in the Appendix.