New York Best Practice Model 2012 Update

Summary Report

final

report

prepared for

New York Metropolitan Transportation Council

prepared by

Cambridge Systematics, Inc.

with

Arizona State University
University of Texas, Austin
Gallop Corporation
EA Harper Consulting
CDM Smith
Watchung Transportation
Florida International University
David Rubin

July 26, 2021

draft report

New York Best Practice Model 2012 Update

Summary Report

prepared for

New York Metropolitan Transportation Council

prepared by

Cambridge Systematics, Inc.
101 Station Landing, Suite 410
Medford, MA 02155

with

Arizona State University
University of Texas, Austin
Gallop Corporation
EA Harper Consulting
CDM Smith
Watchung Transportation
Florida International University
David Rubin

date

July 26, 2021

Table of Contents

1.0 Introduction ... 1-1

1.1 Project Objectives ... 1-1

1.2 Products and Additional References ... 1-2

2.0 Model Structure ... 2-1

2.1 Activity-Based Demand Components ... 2-1

2.1.1 Mode Choice Model Alternatives ... 2-7

2.2 Non-Activity-Based Model Components ... 2-7

2.2.1 Air Passenger Models ... 2-7

2.2.2 External Travel ... 2-8

2.2.3 Truck/Commercial Vehicles ... 2-8

2.2.4 Visitor Model ... 2-8

2.3 Trip Assignment ... 2-9

2.3.1 Highway Assignment ... 2-9

2.3.2 Transit Assignment ... 2-10

3.0 Data Development ... 3-1

3.1 Zone Systems ... 3-1

3.2 Socioeconomic Data ... 3-3

3.3 Networks ... 3-4

3.3.1 Highway Network ... 3-4

3.3.2 Transit Network ... 3-5

3.4 Data for Model Estimation and Validation ... 3-5

3.4.1 Survey Data ... 3-5

NYMTC Regional Household Travel Survey ... 3-5

NYMTC Regional Establishment Survey ... 3-6

Transit Rider Surveys ... 3-6

3.4.2 Traffic and Transit Counts ... 3-6

4.0 Model Estimation ... 4-1

5.0 Model Implementation ... 5-1

5.1 TransCAD Interface and Components ... 5-1

5.1.1 TransCAD GISDK Code Updates ... 5-1

5.1.2 User Interface and Reporting ... 5-3

5.1.3 Reporting ... 5-3

5.2 PopGen ... 5-3

5.3 CEMSELTS/CEMDAP Code 5-4

6.0 Model Validation 6-1

6.1 Validation of Activity-Based Demand Components 6-1

6.1.1 CEMSELTS Components 6-2

6.1.2 CEMDAP Components 6-5

GA Series 6-5

WSCH Series 6-7

NWSCH Series 6-10

JASCH Series 6-11

CSCH Series 6-12

6.2 Highway and Transit Assignment Validation 6-13

6.2.1 Highway Assignment 6-13

VMT Checks 6-13

Link Volume Checks 6-14

Screenlines 6-18

6.2.2 Transit Assignment 6-20

Linked Transit Trips 6-21

Station Groups 6-21

Hub-Bound Summary 6-24

6.3 Sensitivity Testing 6-25

6.3.1 Toll change test 6-26

6.3.2 Parking cost change test 6-26

6.3.3 Major development test 6-27

6.3.4 AV “package” test 6-27

7.0 Lessons Learned 7-1

8.0 References 8-0

List of Tables

Table 2-1. CEMSELTS Components 2-3
Table 2-2. CEMDAP Components – GA Series 2-4
Table 2-3. CEMDAP Components – WSCH Series 2-5
Table 2-4. CEMDAP Components – NWSCH Series 2-6
Table 2-5. CEMDAP Components – JASCH Series 2-6
Table 2-6. CEMDAP Components – CSCH Series 2-6
Table 3-1. TAZ Numbering 3-2
Table 6-1. Average Modeled and Observed Home-School Distances (miles) 6-4
Table 6-2. Average Modeled and Observed Home-Work Distances (miles) 6-4
Table 6-3. Modeled Subregion Level Home-Work Flows Compared to ACS 6-5
Table 6-4. Modeled and Observed Percentages of Child School Start and End Times in Peak Periods 6-7
Table 6-5. Modeled and Observed Percentages of Work Start and End Times in Peak Periods 6-7
Table 6-6. Modeled and Observed Regional Child School Mode Shares 6-7
Table 6-7. Modeled and Observed Regional Worker Commute Mode Shares 6-9
Table 6-8. Average Modeled and Observed Home-Work Distances (miles) 6-9
Table 6-9. Modeled and Observed Regional Trip Mode Shares on Work Commute 6-10
Table 6-10. Average Modeled and Observed Non-Worker Trip Distances (miles) 6-11
Table 6-11. Modeled and Observed Regional Non-Worker Trip Mode Shares 6-11
Table 6-12. Modeled and Observed Regional Joint Trip Mode Shares 6-12
Table 6-13. Modeled and Observed Regional Child Trip Mode Shares 6-13
Table 6-14. Modeled and Observed Daily VMT by Facility Type 6-14
Table 6-15. Modeled and Observed Daily VMT by Subregion 6-14
Table 6-16. %RMSE by Facility Type 6-15
Table 6-17. %RMSE by Volume Group 6-15
Table 6-18. Modeled and Observed VMT on Major Routes 6-16
Table 6-19. Modeled and Observed Volumes on Major Crossings 6-17
Table 6-20. Modeled and Observed Volumes on Screenlines 6-19
Table 6-21. Aggregate Screenline Summary 6-19
Table 6-22. Mode Shares (trips) 6-20
Table 6-23. Linked Transit Trip Summary by Time Period and Subregion 6-22
Table 6-24. Station Group Transit Assignment Summary 6-24
Table 6-25. Hub-Bound Transit Summary 6-25

List of Figures

Figure 2-1. 2012 NYBPM Model Structure 2-2

1.0 Introduction

This report documents a project undertaken by the New York Metropolitan Transportation Council (NYMTC) to develop a new activity-based travel demand model for the greater New York region. This model, named the New York Best Practice Model 2012 Update (2012 NYBPM), replaces the previous version of the NYBPM, from 2010. The new model was developed and validated using a base year of 2012. This model was developed for NYMTC by a team led by Cambridge Systematics, Inc. (CS). Other team members included University of Texas, Austin, Arizona State University, CDM Smith, Inc., Watchung Transportation, Inc., Gallop Corporation, Florida International University, EA Harper Consulting, and David Rubin.

1.1 Project Objectives

Metropolitan Planning Organizations such as NYMTC are required to establish a performance based approach to transportation decision making. This includes establishment of performance targets that address the region's performance measures and tracking progress toward the targets. These requirements apply to the regional transportation plan and any scenarios that are analyzed. Also included is any required analysis of air quality conformity for the various plans and scenarios.

The 2012 NYBPM was developed to be capable of providing the necessary model-related information to determine whether performance targets are being met and the effects of the policies and investments involved with any planning scenarios on the achievement of these targets. With these requirements in mind, the 2012 NYBPM has the following capabilities:

- Production of measures of travel demand at aggregate and disaggregate levels. These include roadway traffic volumes and speeds/delays and transit boardings at stations and routes, as well as aggregate measures such as vehicle miles traveled.

- Sensitivity to the demographics of the traveling population and how they are changing over time.

- Sensitivity to changes in transportation level of service, due to both changes in regional land use and development over time (including changes in congestion levels) and changes in transportation service resulting from transportation policy changes and transportation investments.

While no travel demand model can perfectly replicate the complexity of human behavior that affects the way that people travel, the activity-based approach adopted by NYMTC provides a robust method, based on the state-of-the-art in travel demand modeling, for approximating this behavior and the diversity of perceptions, constraints, and preferences that cause people to react differently to changes in conditions. The approach used for the 2012 NYBPM—including PopGen, CEMSELTS, and CEMDAP (see Section 2.1)—is an advanced approach that attempts to capture the effects of the interrelationships among household members that can affect travel behavior and the constraints in time and space that are faced by travelers. Some of the relevant features of this approach include the following:

- It accommodates intra-household decisions of activity-travel choices among all individuals (children and adults) in a household in a compact and computationally efficient manner.

- It incorporates spatial-temporal dependencies and constraints in activity-travel patterns between and within individuals of a household.

- It adopts a true activity-based approach by focusing explicitly on activity episode generation and their characteristics (including chaining into tours—chains of trips starting and ending at home or work—and travel characteristics associated with the activity episodes).
- The approach allows enhanced sensitivity to land use, built environment and development patterns, and multi-modal (and inter-modal) transportation policies and demographic changes in the population, by using an agent-based micro-simulation platform that is designed conceptually to accommodate any level of spatial resolution and incorporate time-varying levels of accessibility.
- The approach enables a holistic assessment of the effects of land-use, built environment, and transportation policies on entire activity-travel patterns through time availability considerations, spatial-temporal dependencies, and inter-individual constraints and interactions.
- It facilitates environmental justice analyses by having the ability to examine the effects of policies on any defined segment of the population by type of activity, by spatial unit of interest, by travel mode, and for any time-of-day.
- The approach considers land use using a multitude of variables that encompass the “four Ds” (density, diversity, design, and destinations) as explanatory variables of its behavioral equations from the long term (location/relocation) to the very short term (stop location), offering increased behavioral realism and behavioral sensitivity to the combined impact of land use and level of service improvements.

The structure of the 2012 NYBPM is described in detail in Chapter 2.0.

1.2 Products and Additional References

The main products of the 2012 NYBPM project include the following:

- The operational, validated model.
- The executable software to run the model, including source code, to run the activity-based demand modeling components.
- A model interface that uses NYMTC’s chosen modeling software, TransCAD, to execute the model and provide key reports of model results.
- Scripts to run various model components, such as network skimming and highway and transit assignment, using TransCAD’s GISDK programming platform.
- Socioeconomic data files at the transportation analysis zone (TAZ) level for the base year of 2012 and selected forecast years.
- Synthetic populations representing the model region’s population for the base year of 2012 and selected forecast years.
- An integrated transportation network representing the major roadways in the region’s highway system and the transit service operated in the region by public agencies and private operators.

- Compiled data for model validation, including traffic counts on screenlines and other key roadways and various measures of transit ridership.

- Data files for model estimation, using information from the network and from data sources including NYMTC's Regional Household Travel Survey (RHTS), Regional Establishment Survey (RES), National Household Travel Survey (NHTS), and the Public Use Microdata Sample (PUMS) from the U.S. Census Bureau's American Community Survey (ACS).

- A variety of detailed reports documenting specific work items, including:

    - Model design report (Cambridge Systematics, Inc. et al., 2017)

    - Model implementation plan (Cambridge Systematics, Inc., 2017)

    - Model validation plan (Cambridge Systematics, Inc. and EA Harper Consulting, 2017)

    - Network development report (Cambridge Systematics, Inc. et al., 2018a)

    - Model estimation report (Cambridge Systematics, Inc. et al., 2018b)

    - Visitor model memo (Cambridge Systematics, 2018)

    - Model validation report (Cambridge Systematics, Inc. et al., 2020)

- Various data files with model results, including TransCAD loaded networks with assignment results and a database containing activity-based demand model results.

References for the reports listed above are provided in Chapter 8.0.

2.0 Model Structure

The model design is documented by Cambridge Systematics, Inc. et al (2017). The structure of the NYBPM is illustrated in Figure 2-1.

2.1 Activity-Based Demand Components

The overall activity-based model system is defined by the integration of three key components:

- **PopGen**, the synthetic population generator;
- **CEMSELTS**, the socioeconomic modeling system; and
- **CEMDAP**, the activity-based modeling engine.

PopGen is an open source synthetic population generator developed by Arizona State University. PopGen Version 2.0 was used to generate synthetic populations for the NYMTC model region. The synthetic populations serve as input to subsequent microsimulation model components embedded within CEMSELTS and CEMDAP. PopGen is written in the open source programming language Python and can be seamlessly integrated with the activity-based microsimulation model system. The operational controls for the PopGen software have been integrated within the TourCast platform.

CEMSELTS (Comprehensive Econometric Microsimulator of Socioeconomics, Land use and Transportation Systems) is the component used to produce additional socioeconomic and demographic attributes for each person in the synthetic population with a view to develop a rich set of input data for the activity-based microsimulation model system. All of the variables that can be simulated by CEMSELTS are stripped away from the synthetic population generated by PopGen and replaced with simulated values from CEMSELTS. The resulting richer set of inputs is then fed to CEMDAP, to simulate complete daily activity-travel patterns for the population of the model region. The CEMSELTS components are shown in Table 2-1.

CEMDAP (Comprehensive Econometric Microsimulator for Daily Activity-travel Patterns) is a microsimulation implementation of a continuous-time activity-travel modeling system. It takes as input the disaggregate agent level socio-demographics, land use patterns, and transportation system level-of-service characteristics, and model parameters for the model region, to provide as outputs the detailed individual level daily activity-travel patterns for all the individuals in the study area. “Agents” in this case refer to the individuals who live in the model region, who are performing the activities that result in the travel being modeled. The CEMDAP components are further subdivided into segments based on travel type:

- Generation-allocation (GA)
- Worker (WSCH)
- Non-worker (NWSCH)
- Child (CSCH)
- Joint (JASCH)

Descriptions of the CEMDAP components in each of these five segments are provided in Table 2-2 through Table 2-6.

Figure 2-1. 2012 NYBPM Model Structure

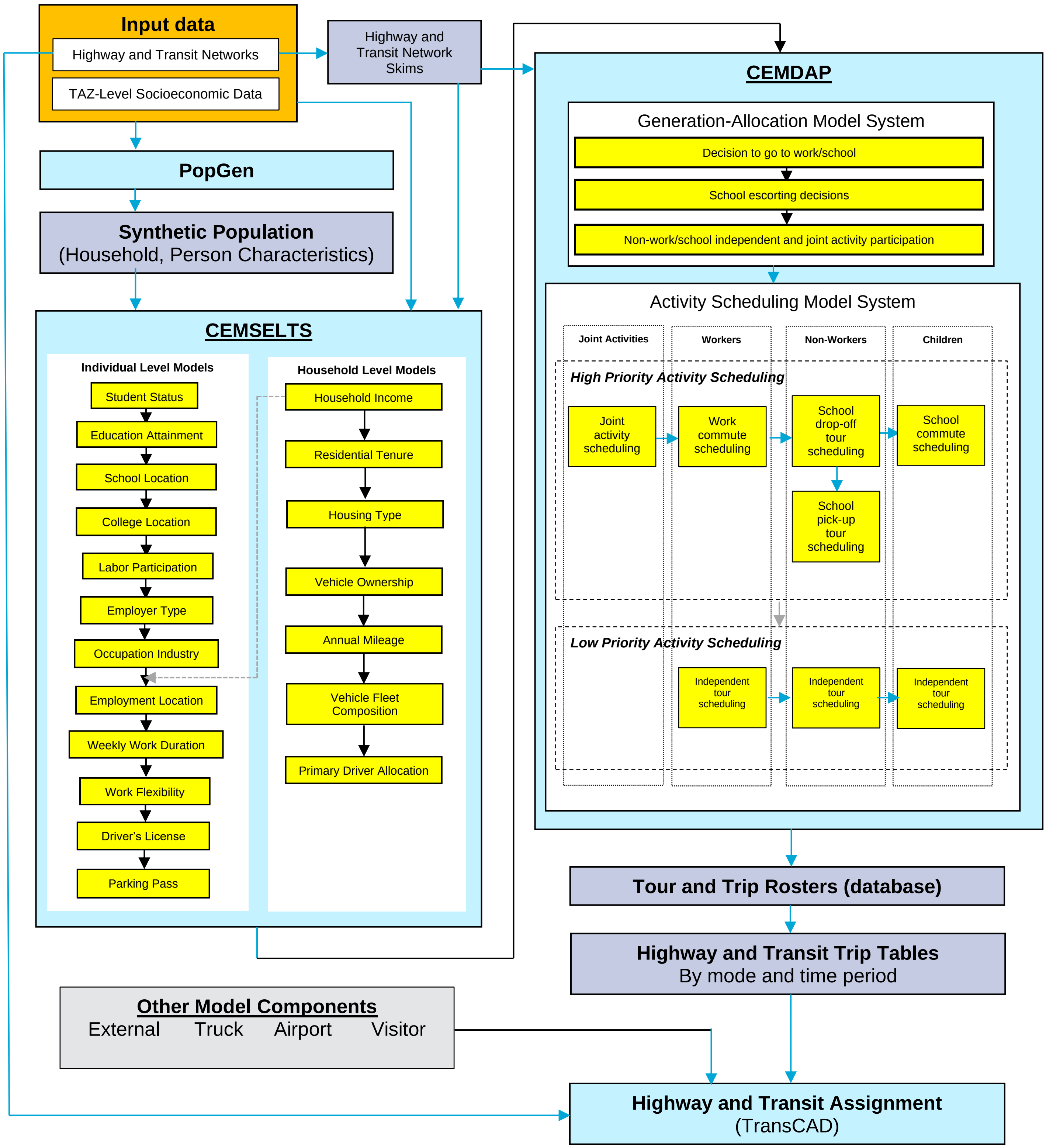

Table 2-1. CEMSELTS Components

Component	Description	Model Unit	Model Type	Data Source
Student status[1]	Student status - Grade level/college status for each person based on age	Person	Lookup tables	RHTS/PUMS
Education attainment	Less than high school/high school/some college/college graduate/any grad school	Person	MNL (5 alts)	RHTS/PUMS
School location	School location - TAZ for each K-12 student	Student	MNL (TAZ alts)	RHTS
College location	College location - TAZ for each college student	Student	MNL (TAZ alts)	RHTS
Labor force participation	Labor force participation - binary choice	Person	Binary logit	RHTS
Employer type	Employer type	Worker	MNL (5 alts)	RHTS
Occupation industry	Occupation industry	Worker	MNL (6 alts)	RHTS
Household income	Household income level	Household	ORL (8 alts)	RHTS
Residential tenure	Residential tenure - own/rent	Household	Binary logit	PUMS
Housing type	Housing unit type	Household	MNL (3/4 alts)	RHTS/PUMS
Employment location	Work location - Regular workplace TAZ for each worker	Worker	MNL (TAZ alts)	RHTS
Weekly work duration	Work duration - <35 hours, 35-45 hours, or >45 hours per week	Worker	MNL (3 alts)	RHTS
Work flexibility	Work flexibility - none, low, medium, and high	Worker	ORL (4 alts)	RHTS
Driver's license	Person holding of driver's license	Person	Binary logit	RHTS
Parking pass	Worker holding of parking pass	Worker	Binary logit	RHTS
Vehicle ownership[2]	Number of vehicles owned by the household	Household	MNL (5 alts)	RHTS
Annual mileage	Household mileage (annual)	Household	Log-linear regression	NHTS
Vehicle fleet composition	Vehicle fleet - number of household vehicles by type/vintage category	Household	MDCEV	NHTS
Primary driver allocation	Primary driver - which person in the household is the primary driver of each vehicle	Household	MNL (2-8 alts)	NHTS

Model structure abbreviations: MNL – multinomial logit, ORL - ordered response logit, MDCEV – multiple discrete-continuous extreme value.

Data source abbreviations: RHTS – NYMTC Regional Household Travel Survey, PUMS – Public Use Microdata Sample from the U.S. Census Bureau's American Community Survey (ACS), NHTS – National Household Travel Survey.

Notes:

1. Lookup table obtained directly from RHTS/PUMS – no validation required
2. New component added after model design plan was completed

Table 2-2. CEMDAP Components – GA Series

Code	Component	What's Modeled	Unit	Model Type
GA1	Child's decision to go to school	Yes/no	Tour	Binary logit
GA2	Child's school start time	Continuous	Person	Hazard-duration
GA3	Child's school end time	Continuous	Person	Hazard-duration
GA4	Adult's decision to go to work	Yes/no	Person	Binary logit
GA5	Adult's work start and end times	32 periods	Tour	Multinomial logit
GA6	Adult's decision to go to school	Yes/no	Person	Binary logit
GA7	Adult's school start time	Continuous	Person	Log-linear regression
GA8	Adult's school end time	Continuous	Person	Log-linear regression
GA9	Child's travel mode to school	Modes	Trip	Multinomial logit
GA10	Child's travel mode from school	Modes	Trip	Multinomial logit
GA11	Allocation of drop off episode to parent	Mother/father	Household	Binary logit
GA12	Allocation of pick up episode to parent	Mother/father	Household	Binary logit
GA13	Determination of households with non-zero out-of-home duration	Out-of-home activities: yes/no	Household	Binary logit
GA14	Determination of total OH time of a household	% time in-home/% out-of-home/% travel	Household	Fractional split
GA15	Independent and joint activity participation for households of size ≤ 5	Activity purpose/ # of participants	Household	MDCEV
GA16	Independent activity participation for households of size > 5	Activity purpose/ # of participants	Household	MDCEV
GA17	Decision of adult to undertake other serve-passenger activities	Yes/no	Person	Binary logit

Table 2-3. CEMDAP Components – WSCH Series

Code	Component	What's Modeled	Unit	Model Type
WSCH1	Worker commute mode	Modes	Tour	Nested logit
WSCH2	Number of before-work tours	0, 1, or 2+ tours	Person	Multinomial logit
WSCH3	Number of work-based tours	0, 1, or 2+ tours	Person	Multinomial logit
WSCH4	Number of after-work tours	0, 1, or 2+ tours	Person	Multinomial logit
WSCH5	Before-work tour mode	Modes	Tour	Multinomial logit
WSCH6	Work-based tour mode	Modes	Tour	Multinomial logit
WSCH7	After-work tour mode	Modes	Tour	Multinomial logit
WSCH8a	Worker number of stops on commute tour	0, 1, or 2 stops	Tour	Ordered probit
WSCH8b	Worker number of stops on before work/after work/at-work tour	1, 2, 3, 4, or 5 stops	Tour	Ordered probit
WSCH9	Worker home or work stay duration before tour	Minutes	Tour	Log-linear regression
WSCH10	Worker activity type at stop	Activity purpose	Trip	Multinomial logit
WSCH11	Worker activity duration at stop	Minutes	Trip	Log-linear regression
WSCH12	Worker travel distance to a stop	Miles	Trip	Log-linear regression
WSCH13	Worker location of a stop	Restricted set of 50 TAZs	Trip	Multinomial logit
WSCH14	Worker Commute Trip Mode Choice	Modes	Trip	Multinomial logit

Table 2-4. CEMDAP Components – NWSCH Series

Code	Component	What's Modeled	Unit	Model Type
NWSCH1	Non-worker number of independent tours	1, 2, 3, or 4 tours	Person	Ordered probit
NWSCH2	Non-worker decision to undertake independent tour before pick-up/joint discretionary tour	Performs tour: yes/no	Tour	Binary logit
NWSCH3	Non-worker decision to undertake an independent tour after pick-up/joint discretionary tour	Performs tour: yes/no	Tour	Binary logit
NWSCH5	Non-worker number of stops in a tour	1, 2, 3, 4, 5, or 6 stops	Tour	Ordered probit
NWSCH6	Non-worker number of stops following pick-up/drop-off	0, 1, 2, or 3 stops	Tour	Ordered probit
NWSCH7	Non-worker home stay duration before tour	Minutes	Tour	Log-linear regression
NWSCH8	Non-worker activity type at stop	Activity purpose	Trip	Multinomial logit
NWSCH9	Non-worker activity duration at stop	Minutes	Trip	Log-linear regression
NWSCH10	Non-worker travel distance to a stop	Miles	Trip	Log-linear regression
NWSCH11	Non-worker stop location	Restricted set of 50 TAZs	Trip	Multinomial logit
NWSCH4	Non-worker trip mode	Modes	Trip	Nested logit

Table 2-5. CEMDAP Components – JASCH Series

Code	Component	What's Modeled	Unit	Model Type
JASCH2	Joint activity start time	Minutes from 3:00 a.m.	Trip	Log-linear regression
JASCH3	Joint activity distance to stop	Miles	Trip	Log-linear regression
JASCH4	Joint Activity location	Restricted set of 50 TAZs	Trip	Multinomial logit
JASCH6	Joint discretionary trip mode choice	Modes	Trip	Nested logit
JASCH5	Vehicle Used for Joint Home-Based Tour	Household vehicles (up to 7)	Tour	Multinomial logit

Table 2-6. CEMDAP Components – CSCH Series

Code	Component	What's Modeled	Unit	Model Type
CSCH4	Child departure time from home for independent discretionary tour	Minutes from 3:00 a.m.	Trip	Log-linear regression
CSCH5	Child activity duration at independent discretionary stop	Minutes	Trip	Log-linear regression
CSCH6	Child travel distance to independent discretionary stop	Miles	Trip	Log-linear regression
CSCH7	Child location of independent discretionary stop	Restricted set of 50 TAZs	Trip	Multinomial logit
CSCH3	Child mode for independent discretionary trip	Modes (see list)	Trip	Nested logit

2.1.1 Mode Choice Model Alternatives

NYMTC made the decision to define the following set of alternatives to be used in mode choice:

- Auto SOV
- Auto HOV – 2 occupants
- Auto HOV – 3+ occupants
- Taxi
- Commuter rail/bus – auto access (includes commuter rail, zone based ferries such as NY Waterways, and commuter buses)
- Commuter rail/bus – walk access (includes commuter rail, zone based ferries, and commuter buses)
- Other rail – auto access (includes subway/el, PATH, LRT, and flat fare ferries such as the Staten Island ferry)
- Other rail – walk access (includes subway/el, PATH, LRT, and flat fare ferries)
- Local bus – auto access
- Local bus – walk access
- Walk
- Bicycle

Not all modes are available in every mode choice model. For example, auto SOV is unavailable for the child and joint mode choice models. Alternatives that were never or rarely chosen in the observed survey data sets were excluded from some models; for example, transit auto access modes are not available for work-based subtour mode choice.

It is important to note that differences among the types of transit services within each defined mode are considered in transit path building. As discussed in Section 2.3.2, a multipath transit assignment process is used, and therefore, for each zonal origin-destination pair, multiple paths are chosen during assignment.

2.2 Non-Activity-Based Model Components

The other demand components of the overall NYBPM consist of the non-activity based components shown at the lower left of Figure 2-1 as “Other Model Components.” These include the external travel, special generators, commercial vehicle travel, and visitor model components. The first three of these components were taken directly from the 2010 NYBPM, and their structures were not updated as part of the 2012 model development (although the 2012 model inputs (e.g., socioeconomic data) were used in running them as part of the 2012 NYBPM). The visitor model is a new component developed for the 2012 NYBPM.

2.2.1 Air Passenger Models

Since air passengers traveling to and from airports use the same transportation infrastructure that other travelers use, their ground travel must be considered as part of the demand on infrastructure and are introduced into the NYBPM as “special generator” trips by mode. Air passengers who use highway modes are added to the TransCAD trip tables for highway assignment; however, air passengers using transit to access the airport are assigned as a separate mode, making it possible to account for differing values of time and different route and mode choices. The 2012 NYBPM adopted the methodology in place from the 2010 model.

2.2.2 External Travel

External travel in the 2010 NYBPM was derived from cordon traffic counts. A gravity model was used to distribute trips within the model region, and growth factors were used to project external demand for future scenarios. These trips were added as separate cores to the trip tables by time period. In more recent updates, an effort was made to integrate the external model into the core modeling process, by accounting for jobs held by workers who resided outside the region and vice-versa. The Census Transportation Planning Products (CTPP) home to work survey and American Community Survey data were used to develop a seed matrix for County to County auto and total trips, which were adjusted using a Fratar process to produce future year matrices. Long distance non-work trips were derived from the NHTS. This methodology was maintained as part of this model update, using updated traffic counts to reflect 2012 conditions.

2.2.3 Truck/Commercial Vehicles

The NYBPM truck model underwent significant improvements during the 2010 update. The current methodology includes an explicit representation of external traffic at true origins and destinations rather than locking in external trips at external stations. FAF3 data were used for long distance trips and an enhanced Quick Response Freight Manual (QRFM) method was used to model shorter trips. A gravity model was used to distribute the trips. The model uses a form of generalized cost impedance, different from the generalized cost used in highway assignment. The structure of the truck/commercial vehicle component was not updated for the 2012 NYBPM.

2.2.4 Visitor Model

A new visitor model was developed as part of the 2012 NYBPM. The visitor model estimates the average weekday travel made within the model region by people who do not live in the region. This travel component is not covered by the RHTS, and since the travelers' residences are outside the region, the concept of home based tours used in the activity-based model is not relevant.

The visitor model is a person trip based component applied separately in TransCAD and is incorporated into the 2012 NYBPM interface. Transit person and auto vehicle trip tables are the outputs of the visitor model, and these tables are combined with the outputs of other model components for transit and highway assignment.

The data source for estimating the visitor model was the hotel component of the Regional Establishment Survey (RES). This is an intercept survey of visitors staying at hotels in the region that obtained information about all trips made by hotel guests within the past day, including trip purpose, time of day, origin/destination, and mode. The estimated models are applied to reflect the entire universe of visitors to the region, segmented by traveler type (e.g., business vs. leisure).

The visitor model consists of the following components:

- **Trip generation** – The number of trips generated per visitor is estimated. Trips are segmented by purpose (e.g., business, meal, other, non-hotel based). Trip production rates reflect average trip rates by purpose per traveler in each segment.

- **Time of day** – The generated daily trips are split into trips by time period using percentages derived from the data from the RES hotel component by area type. Factors vary by area type.

- **Destination choice** – Multinomial logit destination choice models were estimated for each traveler segment and trip purpose. The models estimate the probability of choosing each destination zone for each origin zone. Similar to (but simpler than) the destination choice models used for the resident activity-based models, the utility variables include measures of impedance (e.g., mode choice logsum and distance) and size variables that include zone level measures of activity (e.g., employment by type).

- **Mode choice** – Mode choice is performed using multinomial logit models. Similar to (but simpler than) the mode choice models used in the resident activity-based models, these models estimate for each trip the probabilities of choosing the auto, rail, taxi, and walk modes. The models are segmented by traveler type and area type.

2.3 Trip Assignment

While the outputs of the activity-based model components are individual trip rosters, the trip rosters are combined into trip tables for use in aggregate highway and transit assignment processes. Walk and bicycle trips are not assigned in the 2012 NYBPM.

2.3.1 Highway Assignment

The NYBPM uses the TransCAD General User Equilibrium multi-class highway assignment procedure. Highway assignment is performed for four time periods:

- A.M. Peak (6:00 -10:00 AM)
- Mid-Day (10:00 AM – 3:00 PM)
- P.M. Peak (3:00 – 7:00 PM)
- Night (7:00 PM – 6:00 AM)

Multiple internal iterations of equilibrium highway assignment are performed, within global iterations (feedback loops) of the entire model, with an averaging of trip tables and link volumes in the intermediate iterations to promote convergence. The final iteration, which generates the scenario forecasts, is done without any averaging of demand or link volumes, so that a fully consistent set of model measures from each stage of the model is available to report impacts of the scenario.

The NYBPM loads seven highway mode vehicle trip tables:

- Single Occupant Vehicle (SOV)
- High Occupant Vehicle – 2 person (HOV2)
- High Occupant Vehicle – 3 or more person (HOV3+)
- Taxi
- Heavy Truck
- Medium Truck
- Other Commercial

The equilibrium highway assignment procedure is applied in an iterative fashion, where travel times are updated after each iteration to reflect congestion occurring on the network. These updates to travel time are based on a volume-delay function for each link. The volume-delay functions are modified versions of BPR functions with parameters varying by facility types. The free-flow times are based initially on the network data provided for each link and then updated in each iteration to represent the travel time resulting from the assigned traffic volumes from the last iteration. The method currently used is the minimization of generalized cost.

Turn penalties are included in the highway assignment to prohibit certain turn movements or penalize movements. These are included in the model by identifying specific turn movements by their node numbers, and then coding the penalty function that will apply to these turn movements.

2.3.2 Transit Assignment

Transit assignment is performed for the a.m. peak period. The mode choice component of the 2012 NYBPM has the following transit submodes:

- Commuter rail/bus – auto access
- Commuter rail/bus – walk access
- Other rail (except commuter) – auto access
- Other rail (except commuter) – walk access
- Local bus – auto access
- Local bus – walk access

(The Staten Island Ferry is treated as part of the subway mode, and other ferry services are treated as part of the commuter rail mode.) This is because the Staten Island Ferry costs nothing to ride and is often used as part of multimodal trips that use other modes such as subway. Other ferry services have fares that are zone-based, similar to the way that commuter rail fares are collected.

The 2012 NYBPM uses the Pathfinder algorithm, which reflects crowding and transit line capacities. The multi-class capability of Pathfinder accounts for the fact that bus and subway lines are used by commuter rail riders in addition to other riders, and thus limited line capacities reflect both classes of riders. The assignment procedure is iterative and considers the volume/capacity ratio in the final assignment solution, with the objective of minimizing a travelers' generalized cost. Both route level capacities and link level capacities are used.

3.0 Data Development

3.1 Zone Systems

Like all travel models, the NYBPM uses a system of transportation analysis zones (TAZ) to represent the locations of homes and workplaces, activities, and trip ends. The model region is divided into 5,418 internal TAZs and 111 TAZs representing external stations, where network roadways crossed the regional boundary. A summary of the TAZ numbering by geographic location is provided in Table 3-1.

In New Jersey and Connecticut, the TAZ system was revised from the 2010 NYBPM to ensure consistency with those used in other models whose regions overlap with that of the NYBPM, such as the North Jersey Transportation Planning Authority (NJTPA) model. This resulted in a greater number of TAZs in these states compared to the 2010 NYBPM.

For some model components, TAZ-level data is disaggregated to micro-analysis zones (MAZ). This is done mainly to better represent model features where a finer level of spatial data is needed due to relatively short trip lengths, such as transit walk access and egress and non-motorized travel.

Table 3-1. TAZ Numbering

State	County			District		Sub-Region	TAZs Range		# of TAZs
	#	FIPS	Name						
NY	1	36061	New York	1	CBD: Lower	CBD	1	14	14
				2	CBD: Valley		15	107	93
				3	CBD: Midtown		108	165	58
				4	Other Manhattan	Upper Manhattan	166	335	170
	2	36081	Queens	5	Queens	Other NYC	336	1004	669
	3	36005	Bronx	6	Bronx		1005	1343	339
	4	36047	Kings	7	Kings		1344	2103	760
	5	36085	Richmond	8	Richmond		2104	2212	109
	6	36059	Nassau	9	Nassau	Long Island	2213	2491	279
	7	36103	Suffolk	10	Suffolk		2492	2813	322
	8	36119	Westchester	11	Westchester	Mid-Hudson	2814	3036	223
	9	36087	Rockland	12	Rockland		3037	3101	65
	10	36079	Putnam	13	Putnam		3102	3120	19
	11	36071	Orange	14	Orange		3121	3200	80
	12	36027	Dutchess	15	Dutchess		3201	3279	79
NJ	13	34003	Bergen	16	Bergen	New Jersey NJTPA Core Area	3280	3467	188
	14	34031	Passaic	17	Passaic		3468	3568	101
	15	34017	Hudson	18	Hudson		3569	3754	186
	16	34013	Essex	19	Essex		3755	3983	229
	17	34039	Union	20	Union		3984	4098	115
	18	34027	Morris	21	Morris	New Jersey NJTPA Other	4099	4199	101
	19	34035	Somerset	22	Somerset		4200	4280	81
	20	34023	Middlesex	23	Middlesex		4281	4489	209
	21	34025	Monmouth	24	Monmouth		4490	4642	153
	22	34029	Ocean	25	Ocean		4643	4778	136
	23	34019	Hunterdon	26	Hunterdon		4779	4810	32
	24	34041	Warren	27	Warren		4811	4837	27
	25	34037	Sussex	28	Sussex		4838	4881	44
	26	34021	Mercer	29	Mercer	New Jersey (DVRPC)	4882	5005	124
CT	27	09001	Fairfield	30	Fairfield	Connecticut	5006	5215	210
	28	09009	New Haven	31	New Haven		5216	5404	189
Special Generator Zones							5405	5418	14
Total TAZs									5418

Note: Unused zone numbers – 5419 - 6000. External Zone numbers – 6001 - 6111

3.2 Socioeconomic Data

Socioeconomic data are used in numerous ways in the 2012 NYBPM, including the following:

- In the development of the synthetic populations using PopGen, as control totals
- To compute size variables for location choice models, to represent the amount of activity at potential destinations
- In the computation of accessibility variables used in some model components

NYMTC produced totals for a variety of socioeconomic data items at the TAZ level, for the base year of 2012 and for five-year increments from 2015 to 2050. Variables include the following:

- Persons/households
  - Total population
  - Total households
  - Population in households
  - Group quarters population
  - Persons by age group and gender
  - Persons by employment status
  - Labor force
- Employment by type
  - Agriculture
  - Mining
  - Utilities
  - Construction
  - Manufacturing
  - Wholesale Trade
  - Retail Trade
  - Transportation & Warehousing
  - Information
  - Finance & Insurance
  - Real Estate, Rental & Leasing
  - Professional, Scientific & Technical
  - Management of Companies & Enterprises
  - Administrative, Support, Waste Management
  - Educational Services
  - Health Care & Social Assistance
  - Arts, Entertainment & Recreation
  - Accommodation & Food Services
  - Other Services
  - Government

3.3 Networks

The CS team updated the 2010 NYBPM highway and transit networks to the 2012 base year and implemented various corrections and improvements. The network development process is documented in a technical memorandum (Cambridge Systematics, Inc. et al., 2018a).

An important feature of the 2012 NYBPM is the integration of the highway and transit networks. The exercise of integrating the highway and transit networks was made possible by the availability of the General Transit Feed Specification (GTFS) files for the major transit operators. Highly accurate shapes (or lines) describing the bus network were conflated to the highway link layer and used to associate bus routes with highway links. Bus frequencies and speeds were also imported.

The integrated network link layer consists of three types of links:

- Highway network links representing the regular roadway segments, used for autos, trucks, taxis, and buses;
- Fixed guideway transit links exclusively used by transit modes like subway, rail, ferry, and trams; and
- Transit station connection links connecting the fixed guideway stations and road network links for station/stop access and egress, or for transferring between transit stations/stops.

3.3.1 Highway Network

In preparation for updating the highway network, a complete review of all network attributes and the GISDK scripting that makes use of the available highway data fields was performed. During the development of the integrated network, fields relevant to the transit model were added, including walk time and transit in vehicle travel time (IVTT). The changes necessary to update the highway network were provided by NYMTC, along with their consultants and member agencies. In addition to the basic 2010-to-2012 changes, coding for tolling and truck policies was updated. Updates included:

- Toll amounts were updated to 2012 dollars, for the facilities operated by the Port Authority of New York and New Jersey (PANYNJ), Metropolitan Transportation Authority (MTA), New York State Bridge Authority (NYSBA), New York State Thruway Authority (NYSTA), and New Jersey Turnpike Authority.
- Analysis of the changes in the 2010 NYBPM truck routes and restrictions was performed. A variety of data sources were reviewed in detail and cross checked with the network. The revised truck route designations and restrictions were reviewed for consistency using thematic mapping.
- NYMTC performed a review of the National Highway System (NHS) and provided a list of NHS links to be reviewed for potential inclusion in the 2012 NYBPM highway network.
- Changes were made to the highway link layer to improve the model function and utility, or to correct other known errors. These changes included centroid connector edits to accommodate changes in the TAZ definitions, updated ramp types based on NYMTC recommendations, improvements and corrections to the various highways, and updated locations and operations of HOV facilities.

3.3.2 Transit Network

As noted above, the highway and transit link layers were integrated by importing GTFS route locations and schedules into the 2012 NYBPM. The integration required the realignment of the GTFS shapes to the highway link layer (conflation) and the conversion of the GTFS route data into TransCAD format (route system development). The details of these processes can be found in the network development technical memorandum (Cambridge Systematics, Inc. et al., 2018a).

Additional items that were updated as part of the 2012 transit network development included the following:

- Updating fares to represent those charged by the various transit operators in 2012, in 2012 dollars. Since the operators use a variety of fixed, variable, and zone-based fares, this included both direct coding and modeling of fares.
- Updating the transit walk network, including walk access/egress, transfer, and connector links. This included prohibiting walking along higher classified roadways and ramps, and adding walk links where needed.
- Automatically generating transit auto access links by updating a script from the 2010 NYBPM, to be consistent with the 2012 highway and transit network.

3.4 Data for Model Estimation and Validation

Besides the socioeconomic and network data described earlier in this chapter, a variety of other data sources were used in the development of the 2012 NYBPM, for model estimation, validation, and application. These sources are summarized below.

3.4.1 Survey Data

The following survey data sets were used in the 2012 NYBPM development process.

NYMTC Regional Household Travel Survey

The NYMTC Regional Household Travel Survey (RHTS) was conducted by NYMTC and NJTPA in 2010-2011. Like most household travel surveys, the survey collected information specific to each household, including information related to each person living in the household and each vehicle owned by the household. In addition, each household was assigned a travel day, and household members were asked to record all travel and the characteristics of that travel for a 24-hour period. In total, nearly 19,000 households completed the travel diary information and made nearly 144,000 (linked) trips during their designated travel days. For a linked trip, the “From” place represents a trip Origin and the “To” place represents a trip destination. The linked trips may or may not have stops. Households were surveyed from each of the 28 counties within the NYBPM model region within New York, New Jersey, and Connecticut.

THE RHTS data set was the primary source for the estimation most of the CEMSELTS components (see Table 2-1) and all of the CEMDAP components. It served as the primary source for the validation of the CEMSELTS and CEMDAP components as it provided the most comprehensive data set that represented the travel behavior of the region’s residents.

NYMTC Regional Establishment Survey

The NYMTC Regional Household Travel Survey (RES) was conducted by NYMTC in 2014-2015. Like the RHTS, it covered the entire NYBPM model region. RES data was used as a supplemental source in the estimation and validation of some of the CEMDAP components, and the hotel sample was the primary data source used in the estimation of the visitor model components.

Transit Rider Surveys

These surveys—which were the most recent available at the time—were valuable in validating the transit related model components such as mode choice and transit assignment. Major operators/systems that have provided surveys include:

- Long Island Railroad (2012-2014)
- Metro-North Railroad (East of Hudson, 2007)
- PATH (Port Authority of New York and New Jersey) (2012)
- New York City Transit (NYCT) subway/bus (2008)
- Newark City Subway (2008)
- Hudson-Bergen Light Rail (2008)
- NY Waterways ferries (2013)
- Nassau Inter-County Express (NICE) (2013)
- New Jersey Transit (NJ Transit) Bus (2013)

Data from 2013 are also available in a merged survey for a group of 11 bus private companies.

3.4.2 Traffic and Transit Counts

A variety of available traffic count data was used for highway assignment validation. Screenline count locations were identified where hourly count data were available. Other count data, including classification counts, from New York City, New York State, river crossings, and major New Jersey screenlines from the NJTPA travel demand model, were incorporated into a database for highway assignment validation. The database development is documented in the Task 5A report (Cambridge Systematics, Inc. and EA Harper Consulting, 2019).

Transit boarding counts were used for transit assignment validation. These were not uniformly available for all transit services due to the wide variety across operators and services of station entry, fare collection, and transferring types. The database development is documented in the Task 5B report (Cambridge Systematics, Inc. and EA Harper Consulting, 2018).

4.0 Model Estimation

The 2012 NYBPM model estimation was the process of obtaining statistical estimates for the parameters of the mathematical functions that are part of the activity-based demand model components in CEMSELTS and CEMSELTS.

Model estimation consisted of using statistical processes to estimate the most likely values of the parameters of the mathematical formulations used in the model to relate the model outputs to the input data, i.e., the activity-based demand model components in CEMSELTS and CEMSELTS. Different types of these formulations are used, depending on the specific component; these are specified in Table 2-1 through Table 2-6. The estimation procedure used observed data from several sources, which are described in detail in Section 3.4.1, including:

- Regional Household Travel Survey (RHTS)
- Regional Establishment Survey (RES)
- The 2010-2012 3-year Public Use Microdata Sample (PUMS), part of the American Community Survey (ACS)
- The 2009 National Household Travel Survey (NHTS)

The model estimation process and results are described in detail in the model estimation report (Cambridge Systematics, Inc. et al., 2018b). This report provides a summary of the mathematical formulations used for the various components and the following information about estimation of every CEMSELTS and CEMDAP component:

- Model estimation data sources
- Assumptions in model estimation
- Model type (e.g., multinomial logit, log-linear regression)
- Model structure and alternative definitions for choice models
- Variable definitions and descriptive statistics
- Estimation results and analysis

5.0 Model Implementation

The 2012 NYBPM is implemented using a combination of proprietary, open source, and newly written software and code. These include the following:

- TransCAD, proprietary travel modeling software from Caliper Corporation
- TransCAD GISDK scripts from the 2010 NYBPM, adapted to reflect the specifics of the 2012 NYBPM
- Newly written (by CS) TransCAD GISDK scripts
- PopGen version 2 open source software (programmed in Python)
- Existing CEMSELTS and CEMDAP C# application code, adapted by CS and UTA to reflect the specifics of the 2012 NYBPM.

There are two main documents that provide details about the model implementation process. The 2012 NYBPM User Guide (Cambridge Systematics, Inc. and EA Harper Consulting, 2021) provides specific information to model users on the model structure, hardware requirements, software structure, interface, and data files. A separate memorandum (Cambridge Systematics, Inc., 2021) documents the work done to develop the model implementation, including the TransCAD interface development, the adaptation of the CEMSELTS/CEMDAP code, the PostgreSQL database, and reporting functions.

5.1 TransCAD Interface and Components

NYMTC decided that continuing to use the TransCAD software platform, from Caliper Corporation, and its flowchart Interface as used in the previous version of the NYBPM, was the preferred approach for running the 2012 NYBPM. This approach streamlined model application and analysis. The platform was updated to use TransCAD version 8, including Caliper's most current flowchart interface and scenario management system, for the 2012 NYBPM.

CS developed the model implementation using TransCAD as the model interface and to perform many of the model functions, including network skimming and assignment, application of non-activity-based model components, and reporting. The TransCAD interface calls the applications for the activity-based NYBPM components that run using separate code, including the synthetic population generation using PopGen and the activity-based demand application (CEMSELTS and CEMDAP). Scenario and file management is achieved through a scenario management system integrated into the user interface.

5.1.1 TransCAD GISDK Code Updates

While the TransCAD GISDK code from the 2010 NYBPM was used as the starting point for the implementation of the 2012 NYBPM, CS undertook a thorough review of the GISDK code and made a number of revisions. These revisions included the following:

- <u>Removal of many of the hardcoded parameters and filenames</u> included throughout the 2010 NYBPM codebase. In many cases, these hardcoded parameters were undocumented, and the effects on model results were unknown, especially given the changes to the model structure. As appropriate, parameters and filenames were moved to the scenario management screens, increasing transparency and streamlining future model updates.

- Removal of numerous legacy calibration adjustments, many of which were undocumented, that had been implemented for the 2010 NYBPM (and possibly earlier versions). Since the activity-based components were fully re-designed for the 2012 NYBPM, the model was completely revalidated, and some new calibration adjustments were made. These included the following:
  - Representing one-way tolls for river crossings as two-way tolls for better interaction with location choice models; this is known as "balanced tolling."
  - Adjusting network lookup tables (e.g., speeds) to improve traffic assignment validation by facility type and area type.
  - Adjusting value of time parameters.
  - Refinements to the transit and highway pathfinding settings.
- Correction of bugs in the model process. The code review identified numerous bugs that were corrected.
- Removal of legacy utility functions that were no longer used by NYMTC.
- Updates to utility functions to improve useability and to function properly with the current network formats.
- Updates to truck and external trip tables, expanding and disaggregating 2010 trip tables based on new TAZ layers, changes in base year demographics, and forecast year demographics for years beyond 2012.

The structures for most of the non-activity-based demand components, including truck trips, airport trips and external travel, were not revised as part of the 2012 NYBPM development. A new non-activity-based component, the visitor model, was included for the first time in the 2012 NYBPM (see Section 2.2.4).

The truck and external models were not revised, but the trip tables from the 2010 NYBPM were updated for 2012 and forecast years using a factoring process. The external trip component of the model is based on 2012 traffic counts at all external stations along with disaggregated external trip tables from the 2010 model. Disaggregated 2010 trip tables are used as a seed and expanded using an iterative proportional fitting (IPF) procedure to match traffic volumes at all external stations. In forecast year models, updated external volumes are input to the process. The model scripts expand the external seed matrices for consistency with forecasted external volumes.

Truck trip tables for the 2012 model are based on disaggregated truck trip tables from the 2010 model. Truck trip tables from the 2010 model were factored based on results of initial assignment of 2010 trucks. Resulting truck VMT on links with counts was compared to count VMT, resulting in a global adjustment factor. The adjustment factor was applied uniformly for the entire truck trip table. These adjusted truck trip tables are input directly to the model.

Forecast year truck trip tables were developed by inflating the 2012 truck tables using TAZ level socioeconomic data. The inflation process used the IPF procedure to increase trip activity-based on employment growth rates in each TAZ. The resulting forecast year truck tables have been provided with the travel model dataset and are input directly into the model.

5.1.2 User Interface and Reporting

The TransCAD flowchart interface provides a streamlined approach to running the travel model and managing multiple scenarios.

CS implemented the 2012 NYBPM through the following steps:

- Updated the model to work with the new scenario manager toolbar available in TransCAD 8.
- Implemented the method for storing and updating parameters within the Flowchart Scenario Manger.
- Defined a folder structure used for storing model input, output, and other files associated with running a scenario.
- Created utilities to aid in managing and running the model, to aid in management and creation of transportation networks, and to create separate time period networks for the time periods used in the 2012 NYBPM.

5.1.3 Reporting

CS developed a customized reporting system that outputs an html file (NYMTC_Summary.html) for each scenario/alternative containing key model results and validation statistics. This file includes the following information:

- Summary of key highway network statistics
- Origin-destination trip summary by mode at the district level
- Vehicle-miles traveled and vehicle-hours traveled by facility type and area type at the district level, for each time period
- Transit boarding summary by transit mode and access mode
- Transit origin-destination trip summary by transit mode and access mode at the district level by time period
- Various summaries used in highway and transit assignment validation (usable only for base year validation since the observed data represent the base year)

5.2 PopGen

PopGen is an open source synthetic population generator. Arizona State University (ASU), where PopGen was developed, adapted PopGen for use in the 2012 NYBPM and revised the PopGen application to produce the synthetic populations needed to run the NYBPM. ASU prepared complete documentation (Arizona State University and Cambridge Systematics, 2021) of the NYBPM PopGen application, which describes the setup of PopGen, data and input file preparation, configuration files, scenario generation, post processing steps, output files. Data dictionaries are also provided.

While PopGen can be run in a standalone manner, it can also be run directly from the TransCAD user interface of the NYBPM, as described in the user guide. Running PopGen from the interface is a user option since some applications do not require generation of a new synthetic population. PopGen only needs to be run when the socioeconomic dataset has changed or been revised.

5.3 CEMSELTS/CEMDAP Code

Previous CEMSELTS and CEMDAP implementations used custom code created by the model developers at the University of Texas, Austin (UTA). For the NYBPM application, the code was substantially revised and updated by CS and UTA to run the newly designed and validated model. The code changes accomplished the following:

- Code revisions to accommodate the CEMDAP revisions that replaced the various “travel time to stop” models for workers, non-workers, children, and joint tours in the original CEMDAP structure with “travel distance to stop” models.
- Code revisions to accommodate revisions to the worker mode choice models that allow for mixed-mode tours for workers.
- New code to apply a new model added to CEMSELTS, the vehicle ownership model.
- Code changes to incorporate new variables that were not included in the original CEMSELTS/CEMDAP models. Some of these new variables were included in model estimation while others were added during model validation. Most of these were geographic-specific variables relevant for the NYMTC model region, such as indicator variables for specific origin-destination regions like Manhattan.

Details of the model changes noted above can be found in the model estimation report.

The CEMSELTS/CEMDAP code in some cases pulls the necessary outputs from previous steps from the database to use as input to subsequent components. In other cases, the legacy code that passes information directly from one component to the next was unchanged. As previously noted, the results of the CEMSELTS and CEMDAP components are outputted to the database by the code.

6.0 Model Validation

This chapter summarizes the validation of the NYBPM. The model validation report (Cambridge Systematics, Inc. et al., 2020) provides complete details of the validation process and results. Presented here are summaries of the validation process and some of the key validation results.

A model validation plan (Cambridge Systematics, Inc. and EA Harper Consulting, 2017) was developed prior to model development. This plan laid out the process for the model validation and specified the tests that were performed. A few tests changed slightly or were more specifically defined for the final model validation, but generally the plan was followed. The tests in the plan included checks of the results of all model components compared to the observed data, checks of the highway and transit assignment, and tests of the sensitivity of the model to changes in input data.

A major component of the validation process was the comparison of model results for the base year of 2012 to observed data (see Section 3.4 for a summary of the observed data used in validation). This consisted of the validation of the demand components (i.e., CEMSELTS and CEMDAP), which is summarized in Section 6.1, and the validation of highway and transit assignment, discussed in Section 6.2. The other major validation component was sensitivity testing, where the model results for scenarios where key inputs are changed are compared to the base year scenario results. This was done to test the sensitivity of the model to various key inputs related to the types of analyses the NYBPM will be used for. Sensitivity testing is summarized in Section 6.3.

6.1 Validation of Activity-Based Demand Components

To validate the CEMSELTS and CEMDAP components, the estimated models were applied in sequence in a "single pass" using the congested speed network data used for the model estimation process. Model parameters and constants were adjusted as necessary to better replicate expanded data from the observed survey data. The single-pass process involved the following steps:

- **Apply each estimated model component** using the skims and socioeconomic data used for model estimation. The applied model results were compared to the validation targets.
- **Calibrate individual CEMSELTS and CEMDAP components**. Based on the above step, each component was calibrated by adjusting parameter values and as necessary. In a few cases, models were re-estimated to include new variables.
- **Examine error propagations**. As all of the activity-based model components are linked to one another and applied in sequence, each subsequent model component is affected by models upstream. Doing a single-pass validation therefore helped to understand the magnitude and direction of error propagation through the model system.

The comparisons of model results to observed data were performed for market segments relevant to the particular component such as subregions, household characteristics such as income, and personal characteristics such as age and gender. These comparisons were done using Excel spreadsheet files. R scripts were used to export data from the model database and process the data to be imported into the Excel spreadsheets, which were populated in advance with the observed data summaries.

In some cases, model parameters were adjusted ("calibrated") to produce more reasonable results. There was not, however, a universal attempt to match all results from the observed for all market segments by adjusting model constants or other parameters. Calibration adjustments were made only when the uncalibrated model results appeared unreasonable and the survey data results were based on a substantial number of observations. The specific calibration adjustments are documented in the Excel files, which were provided to NYMTC along with the model validation report.

The purpose of the base year comparisons was to verify, to the extent possible, that the model produces reasonable estimates of travel behavior. While it is desirable for a model's base year scenario to reasonably reflect the observed data, a more important objective is for the model to react correctly when run for scenarios representing transportation system, policy, or land use changes that planners wish to study. It is usually possible to improve the match between model results and observed data by adding or making changes to the values of parameters pertaining to various travel market segments, but while such parameters are added for better prediction of variables that obviously need correction, increasing the effects of such parameters—such as constants—can make the model less sensitive to factors that affect travel in these scenarios.

The remainder of this section summarizes the validation results for CEMSELTS and CEMDAP. A few selected tables from the Excel spreadsheets for key models are presented as examples of the validation results.

6.1.1 CEMSELTS Components

The following is a summary of the base year model comparisons for CEMSELTS components.

- Education attainment – Regional model results are within one or two percentage points of the observed, and comparisons for all segments are close.
- School location – Average modeled home-school distances are within four percent of observed; the coincidence ratio for the distance frequency distribution is 89 percent. Average modeled home-school distance comparisons by subregion are shown in Table 6-1.
- College location – Average modeled home-college distances are within two percent of observed; the coincidence ratio for the distance frequency distribution is 85 percent.
- Labor force participation – Regional model results are essentially the same as observed; results by subregion, age group, and gender are all within five percent.
- Employer type – Regional model results are essentially the same as observed; results by subregion, age group, and gender are all within 10 percent (all but a few within five percent).
- Occupation industry – Regional model results are with five percent of observed; results by subregion, age group, and gender are all within 10 percent (most within five percent).
- Household income – Regional model results for all income groups are very close to observed (there is a slight shift from \$150K-\$200K to \$100K-\$150K in the model). Results by subregion and other segments are all close.
- Residential tenure – Regional model results are essentially the same as observed; results by subregion and other segments are all close.
- Housing type – Regional model results are within a few percent of observed; results by subregion, household size, and income level are also close.
- Employment location – Regional results show that the modeled average home-work distance is very close to the observed. The modeled averages for subregions are mostly close to the observed (shown

in Table 6-2) but are farther off for a few of the more remote subregions. The coincidence ratio for the distance frequency distribution is 84 percent. An additional comparison was performed by comparing the modeled subregion to subregion home-work distribution to the distribution from the ACS. This check showed a very close match between the two, as shown in Table 6-3.

- Weekly work duration – Regional model results are within three percent of observed; results by subregion, age group, and gender are mostly within five percent.
- Work flexibility – Regional model results are within two percent of observed; results by subregion, age group, and gender are mostly within five percent.
- Driver's license – Regional model results are within three percent of observed; results by subregion, age group, and gender are mostly within five percent (though license holding for Manhattan is somewhat overestimated).
- Parking pass – Regional model results are within one percent of observed; results by subregion, age group, and gender are mostly within five percent.
- Vehicle ownership – The regional modeled percentages of households by number of vehicles match the observed shares. Results by subregion, age group, and gender are mostly within two percent, with a few segments as much as five percent different.
- Annual mileage – The modeled average regional household mileage is within one percent of the observed from the NHTS data. The modeled percentages of households by mileage segment (generally 5,000 miles) are all within six percent of observed.
- Vehicle fleet composition – The model results match the observed regional distribution of vehicle types and ages closely.
- Primary driver allocation – The model matches well the observed distributions of vehicle types allocated to primary drivers across age and gender distributions.

Table 6-1. Average Modeled and Observed Home-School Distances (miles)

Subregion	Expanded RHTS data	Model	Difference (Model – Survey)
Manhattan	3.0	2.9	-0.1
Other NYC	4.2	4.3	0.1
Long Island	5.0	5.0	0.0
Westchester-Putnam-Dutchess	5.3	5.4	0.1
Rockland-Orange	8.8	5.0	-3.9
Bergen-Passaic	4.5	3.8	-0.7
Essex-Hudson-Union	3.1	3.5	0.4
Middlesex-Morris-Somerset-Mercer	4.1	3.9	-0.2
Monmouth-Ocean	4.4	3.7	-0.7
Hunterdon-Sussex-Warren	6.5	4.5	-2.0
Connecticut	4.6	3.8	-0.8
Region	4.5	4.2	-0.3

Table 6-2. Average Modeled and Observed Home-Work Distances (miles)

Subregion	Expanded RHTS data	Model	Percent Difference (Model - Survey)
Manhattan	5.1	5.1	0.0
Other NYC	8.3	9.2	0.9
Long Island	15.4	13.8	-1.6
Westchester-Putnam-Dutchess	15.2	16.0	0.8
Rockland-Orange	20.9	18.7	-2.2
Bergen-Passaic	11.9	10.4	-1.5
Essex-Hudson-Union	9.8	9.8	0.0
Middlesex-Morris-Somerset-Mercer	14.0	14.4	0.5
Monmouth-Ocean	18.0	19.8	1.8
Hunterdon-Sussex-Warren	21.2	23.1	1.9
Connecticut	12.0	12.7	0.8
Region	11.7	11.8	0.1
New York	10.7	10.8	0.0
New Jersey	13.4	13.7	0.2

Table 6-3. Modeled Subregion Level Home-Work Flows Compared to ACS

		ACS Journey to Work 2009-2013										
		1	2	3	4	5	6	7	8	9	10	11
1	Manhattan	85%	9%	1%	1%	0%	1%	1%	1%	0%	0%	1%
2	Other NYC	36%	57%	4%	2%	0%	0%	1%	0%	0%	0%	0%
3	Long Island	10%	11%	78%	0%	0%	0%	0%	0%	0%	0%	0%
4	Westchester-Putnam-Dutchess	15%	8%	1%	70%	2%	1%	0%	0%	0%	0%	4%
5	Rockland-Orange	8%	6%	0%	8%	68%	7%	1%	1%	0%	0%	1%
6	Bergen-Passaic	11%	3%	0%	1%	1%	64%	12%	6%	0%	0%	0%
7	Essex-Hudson-Union	15%	2%	0%	0%	0%	8%	61%	12%	1%	0%	0%
8	Middlesex-Morris-Somerset-Mercer	6%	1%	0%	0%	0%	4%	12%	71%	3%	2%	0%
9	Monmouth-Ocean	5%	2%	0%	0%	0%	1%	6%	12%	74%	0%	0%
10	Hunterdon-Sussex-Warren	2%	1%	0%	0%	1%	6%	8%	32%	0%	51%	0%
11	Connecticut	4%	1%	0%	3%	0%	0%	0%	0%	0%	0%	92%
	Total	23%	20%	12%	6%	2%	6%	9%	10%	4%	1%	8%

		Model 2012										
		1	2	3	4	5	6	7	8	9	10	11
1	Manhattan	86%	9%	1%	1%	0%	1%	1%	1%	0%	0%	0%
2	Other NYC	35%	56%	4%	2%	0%	1%	1%	0%	0%	0%	0%
3	Long Island	10%	10%	80%	0%	0%	0%	0%	0%	0%	0%	0%
4	Westchester-Putnam-Dutchess	15%	7%	0%	70%	3%	1%	0%	0%	0%	0%	3%
5	Rockland-Orange	9%	6%	0%	9%	65%	6%	1%	1%	0%	0%	1%
6	Bergen-Passaic	12%	3%	0%	1%	1%	66%	12%	5%	0%	0%	0%
7	Essex-Hudson-Union	15%	2%	0%	0%	0%	8%	61%	12%	1%	1%	0%
8	Middlesex-Morris-Somerset-Mercer	6%	1%	0%	0%	0%	4%	12%	73%	3%	1%	0%
9	Monmouth-Ocean	5%	2%	0%	0%	0%	2%	6%	12%	73%	0%	0%
10	Hunterdon-Sussex-Warren	2%	1%	0%	0%	1%	5%	7%	31%	1%	52%	1%
11	Connecticut	4%	1%	0%	3%	0%	0%	0%	0%	0%	0%	91%
	Total	23%	20%	12%	6%	2%	6%	8%	10%	4%	1%	8%

6.1.2 CEMDAP Components

The observed data for the CEMDAP comparisons come from NYMTC's RHTS. The GA series was validated first, and the four remaining series were validated in parallel. The following is a summary of the base year model comparisons.

GA Series

- GA1 – Child's decision to go to school – Regional model results are within one or two percentage points of the observed, and comparisons for grade levels are close. The RHTS data shows some variation by subregion which is not captured by the model (though it is unclear why attendance rates among subregion should vary much).
- GA2/GA3 – Child's school start and end times – The modeled average school activity duration is 6.9 hours, compared to 7.0 hours in the expanded RHTS data set. The coincidence ratios between the

modeled and RHTS diurnal distributions at the hourly level are 71 percent for start times and 63 percent for end times. The modeled and RHTS percentages of a.m. and p.m. peak period start and end times are shown in Table 6-4.

- GA4 – Adult's decision to go to work – Regional model results essentially the same as the observed, and comparisons for subregions, age levels, and work durations are close.
- GA5 – Adult's work start and end times – The modeled average work activity duration is 7.1 hours, compared to 7.5 hours in the expanded RHTS data set. The coincidence ratios between the modeled and RHTS diurnal distributions at the hourly level are 56 percent for start times and 58 percent for end times. The modeled and RHTS percentages of a.m. and p.m. peak period start and end times are shown in Table 6-5. The modeled percentage of work arrivals in the a.m. peak periods is low, with more peak spreading than in the RHTS data. The model's functional form made it difficult to produce a better match.
- GA6 – Adult's decision to go to school – Regional model results essentially the same as the observed, and comparisons by subregions and gender are mostly within five percentage points.
- GA7/GA8 – Adult's school start and end times – The modeled average school activity duration is 7.0 hours, compared to 7.0 hours in the expanded RHTS data set. The coincidence ratios between the modeled and RHTS diurnal distributions at the hourly level are 49 percent for start times and 60 percent for end times. The model underestimates the percentage of adult school start times in the p.m. peak period.
- GA9/GA10 – Child's travel mode to school and from school – At the regional level, the modeled shares for all modes are within one percentage point of the observed, as shown in Table 6-6. Observed trends of mode shares by income level and household size are reflected in the model results. In the subregional summaries, there are some differences between modeled and observed mode shares. The largest of these are in the splits between school bus and walk mode shares in Connecticut and most of New Jersey. These mode shifts do not affect trip assignment since neither walk nor school bus person trips are assigned.
- GA11/GA12 – Allocation of drop off and pickup episodes to parent – Regional model results are within six percent of observed.
- GA13 – Determination of households with non-zero out-of-home duration – The regional percentage of households with non-zero out-of-home activities is within one percent of observed. Comparisons by subregion, income level, and household size are mostly within five percent.
- GA14 – Determination of total OH time of a household – It was noted that the aggregate percentage of time spent inside the home as reported in the RHTS was likely too high to match observed regional travel counts. Calibration was performed to produce a lower percentage of time inside the home (61 percent) than observed (68 percent).
- GA15/GA16 – Independent and joint activity participation for households – The model tended to underestimate joint activity participate somewhat and to overestimate individual participation in work related and other activities. The model reflected observed trends by subregion, income level, household size, and auto ownership level.
- GA17 – Decision of adult to undertake other serve-passenger activities – Regional model results were essentially the same as observed in the expanded RHTS data set, as were model results by gender and employment status. The model reflected that serve-passenger activity participation was lower in Manhattan (though not quite as low as in the observed data).

Table 6-4. Modeled and Observed Percentages of Child School Start and End Times in Peak Periods

Peak Period	Expanded RHTS data		Model Results	
	Start	End	Start	End
AM (6:00-9:00)	81.5%	0.2%	85.6%	0.5%
PM (3:00-6:00)	1.0%	2.2%	3.8%	8.6%

Table 6-5. Modeled and Observed Percentages of Work Start and End Times in Peak Periods

Peak Period	Expanded RHTS data		Model Results	
	Start	End	Start	End
AM (6:00-10:00)	74.7%	1.8%	52.4%	0.1%
PM (3:00-7:00)	5.5%	69.6%	9.4%	64.2%

Table 6-6. Modeled and Observed Regional Child School Mode Shares

Tour Mode	Expanded RHTS data		Model Results	
	To School	From School	To School	From School
HOV - parent chauffeur	30.6%	24.4%	30.0%	24.3%
HOV - other chauffeur	6.9%	8.8%	6.9%	8.7%
Commuter rail/bus – walk access	0.5%	0.5%	0.5%	0.5%
Subway/ferry – walk access	3.6%	3.5%	4.0%	3.7%
Local bus – walk access	4.2%	4.0%	4.8%	4.1%
Walk	17.8%	21.1%	17.7%	20.7%
Bike	0.3%	0.4%	0.3%	0.4%
School bus	36.1%	37.3%	35.8%	37.6%

WSCH Series

- WSCH1 – Worker commute mode – As one of the key components of the entire model, significant attention was paid to the validation and calibration of this component. This included revisiting the validation after the initial highway and transit assignment results to better reflect observed travel conditions, including observations from transit rider surveys. As a result, some new "targets" for mode shares were established that differed from those observed in the RHTS data set. One of the most significant changes was revising the auto access and walk access split for the commuter rail/bus mode to match observed shares from commuter rail survey data.

  Table 6-7 compares the regional model results to the targets. The close match indicates that in the aggregate, the model is producing about the correct number of trips by mode. Because it was not possible to create new targets for the observed mode shares for all of the segments that are consistent with the revised regional targets, a direct comparison of modeled shares to observed is difficult. However, the trends in the model results track those in the RHTS data. For example:

  - For Manhattan residents, auto mode shares are very low (less than 10 percent) while transit shares exceed 60 percent, and non-motorized mode shares are around 20 percent.
  - The transit shares are slightly lower for residents of the other New York City boroughs while the auto shares are around 20 percent, and non-motorized mode shares are under 10 percent.

  - In the rest of the region, auto shares are in the 80 to 90 percent range, except in Essex/Hudson/Union Counties in New Jersey, where the auto share is under 70 percent. The highest auto shares are in the subregions farthest from New York City. Transit shares are in the 15 to 25 percent range in the nearest subregions to the city and are under 10 percent in the rest of the region. The non-motorized shares are under five percent outside New York City and are lower the farther from the city.
  - Travelers from households with annual incomes below $30,000 have auto mode shares of around 40 percent, transit mode shares around 30 percent, and non-motorized mode shares around 25 percent.
  - Travelers from households who own zero vehicles have transit shares around 55 percent range and non-motorized mode shares around 30 percent.

- WSCH2 – Number of before-work tours – The model closely matches the observed percentages of workers with zero, one, and two or more before-work tours.
- WSCH3 – Number of work-based subtours – The model closely matches the observed percentages of workers with zero, one, and two or more work-based subtours.
- WSCH4 – Number of after-work tours – The model closely matches the observed percentages of workers with zero, one, and two or more after-work tours.
- WSCH5/WSCH6/WSCH7 – Before-work/work-based/after-work tour mode – The models closely match the observed regional mode shares for these tours made by workers. Modeled mode shares for the various geographic and demographic segments are generally consistent with observed mode shares, for segments with large enough sample sizes to make worthwhile comparisons.
- WSCH8 – Worker number of stops on commute/before work/after work/at-work tours – The models closely match the observed regional mode shares for these tours. Modeled mode shares for the various demographic segments also match observed mode shares, for segments with large enough sample sizes to make worthwhile comparisons.
- WSCH9 – Worker home or work stay duration before tour – The model results match the observed average durations for all tour types (before-work/after-work/at-work).
- WSCH10 – Worker activity type at stop – The model overestimates the percentage of work-related activities and underestimates the percentages of maintenance, shopping, and social activities. This model is largely determined by the upstream models (GA15/GA16/GA17) which predict activity budgets.
- WSCH11 – Worker activity duration at stop – The model estimates the average activity duration at a stop at 48 minutes, compared to 49 minutes in the RHTS data. The modeled average duration is within a few minutes of the observed for most activity purposes (12 to 15 minutes different for the activities with the longest durations, recreation and social).
- WSCH12 – Worker travel distance to a stop – This is an interim model whose validation is effectively included in the WSCH13 model results described below.
- WSCH13 – Worker location of a stop – The modeled average trip distance is 5.2 miles, compared to the observed average of 5.4 miles. The modeled averages for subregions are all close to the observed (shown in Table 6-8). The coincidence ratio for the distance frequency distribution is 76 percent. The modeled percentage of intrazonal stops is 18 percent, compared to 15 percent observed.
- WSCH14 – Worker commute trip mode choice – The worker commute trip mode choice is closely related to the mode choice for the commute tour (model WSCH1, discussed above). Table 6-9 shows the regional modeled and observed trip mode shares, which are similar to those shown for model WSCH1 in Table 6-7. The model results show that, as is observed, trips on commute tours tend to use the same modes as the tour mode. Model results for geographic and demographic segments also match the observed mode shares well.

Table 6-7. Modeled and Observed Regional Worker Commute Mode Shares

Tour Mode	Commute to Work		Commute from Work	
	Observed*	Model	Observed*	Model
SOV	54.4%	54.1%	54.4%	53.6%
HOV 2	5.4%	5.4%	5.4%	6.3%
HOV 3+	1.0%	1.0%	1.0%	0.9%
Taxi	1.4%	2.0%	1.4%	1.7%
Commuter rail/bus – auto access	5.7%	5.0%	5.7%	5.0%
Commuter rail/bus – walk access	4.4%	4.5%	4.4%	4.6%
Subway/ferry – auto access	0.9%	0.6%	0.9%	0.7%
Subway/ferry – walk access	14.9%	15.8%	14.9%	16.0%
Local bus – walk access	4.4%	4.3%	4.4%	3.9%
Walk	5.4%	5.2%	5.4%	5.2%
Bike	0.7%	0.6%	0.7%	0.6%
School bus	1.6%	1.6%	1.6%	1.5%

* - Adjusted targets from RHTS

Table 6-8. Average Modeled and Observed Home-Work Distances (miles)

Subregion	Expanded RHTS data	Model	Difference (Model - Survey)
Manhattan	5.1	5.1	0.0
Other NYC	8.3	9.2	0.9
Long Island	15.4	13.8	-1.6
Westchester-Putnam-Dutchess	15.2	16.0	0.8
Rockland-Orange	20.9	18.7	-2.2
Bergen-Passaic	11.9	10.4	-1.5
Essex-Hudson-Union	9.8	9.8	0.0
Middlesex-Morris-Somerset-Mercer	14.0	14.4	0.5
Monmouth-Ocean	18.0	19.8	1.8
Hunterdon-Sussex-Warren	21.2	23.1	1.9
Connecticut	12.0	12.7	0.8
Region	11.7	11.8	0.1

Table 6-9. Modeled and Observed Regional Trip Mode Shares on Work Commute

	Observed*	Model
SOV	55.5%	54.1%
HOV 2	5.5%	6.7%
HOV 3+	1.0%	1.1%
Taxi	1.5%	1.9%
Commuter rail/bus – auto access	4.8%	4.0%
Commuter rail/bus – walk access	3.7%	4.0%
Subway/ferry – auto access	0.8%	0.5%
Subway/ferry – walk access	15.2%	16.1%
Local bus – auto access	0.0%	0.0%
Local bus – walk access	4.4%	4.5%
Walk	5.5%	5.3%
Bike	0.7%	0.6%
School bus	1.6%	1.6%

* - Adjusted targets from RHTS

NWSCH Series

- NWSCH1 – Non-worker number of independent tours – The model closely matches the observed percentages of non-workers with zero, one, and two or more independent tours.
- NWSCH2/NWSCH3 – Non-worker decision to undertake independent tour before/after pick-up or joint discretionary tour – The model closely matches the observed percentages of non-workers who choose to undertake an independent tour before a pick-up or joint discretionary tour and after a pick-up or joint discretionary tour.
- NWSCH5 – Non-worker number of stops in a tour – The model closely matches the observed percentages of non-workers with one, two, three, four, five, and six stops on tours.
- NWSCH6 – Non-worker number of stops following pick-up/drop-off – The model matches the observed percentage of non-workers (95 percent) who do not make any stops following pick-up/drop-off activities. Among the five percent of workers who make at least one stop, the model somewhat overestimates the percentage who make one stop and underestimates the percentage who make two or three stops.
- NWSCH7 – Non-worker home stay duration before tour – The model overestimates the home stay duration before the first tour made, and to a lesser extent, the home stay duration before the second tour. The model slightly underestimates the home stay duration before the third and fourth tours.
- NWSCH8 – Non-worker activity type at stop – The model matches the observed percentages for all activity types within five percentage points, except for the percentage of work-related activities, which is underestimated by about ten percentage points. Since work-related stops are special cases for non-workers, it was difficult to simulate many of these types of stops (the observed percentage of 15 percent seems a bit high in any case).
- NWSCH9 – Non-worker activity duration at stop – The model underestimates the average activity duration at a stop by a little over 25 minutes.
- NWSCH10 – Non-worker travel distance to a stop – This is an interim model whose validation is effectively included in the NWSCH11 model results described below.
- NWSCH11 – Non-worker stop location – The modeled average trip distance is 5.2 miles, compared to the observed average of 5.4 miles. The modeled averages for subregions are all close to the observed

(shown in Table 6-10). The coincidence ratio for the distance frequency distribution is 88 percent. The modeled percentage of intrazonal stops is 16 percent, compared to 15 percent observed.

- NWSCH4 – Non-worker trip mode – Table 6-11 shows the regional modeled and observed non-worker trip mode shares, which match well. Model results for geographic and demographic segments also match the observed mode shares well.

Table 6-10. Average Modeled and Observed Non-Worker Trip Distances (miles)

Subregion	Expanded RHTS data	Model	Percent Difference (Model - Survey)
Manhattan	2.6	2.6	0.0
Other NYC	3.9	3.9	0.0
Long Island	7.1	6.9	-0.1
Westchester-Putnam-Dutchess	5.4	5.4	-0.1
Rockland-Orange	8.5	8.5	0.0
Bergen-Passaic	5.6	5.5	-0.1
Essex-Hudson-Union	4.7	4.6	-0.1
Middlesex-Morris-Somerset-Mercer	6.5	6.4	-0.1
Monmouth-Ocean	7.4	7.4	0.0
Hunterdon-Sussex-Warren	11.3	11.3	0.0
Connecticut	5.1	5.0	0.0
Region	5.4	5.2	-0.2

Table 6-11. Modeled and Observed Regional Non-Worker Trip Mode Shares

	Observed*	Model
SOV	57.8%	57.3%
HOV 2	6.6%	7.2%
HOV 3+	1.5%	1.3%
Taxi	1.4%	2.1%
Commuter rail/bus – auto access	3.1%	2.9%
Commuter rail/bus – walk access	2.3%	1.9%
Subway/ferry – auto access	0.7%	0.7%
Subway/ferry – walk access	11.5%	12.0%
Local bus – auto access	0.1%	0.0%
Local bus – walk access	3.4%	3.3%
Walk	11.1%	10.3%
Bike	0.6%	1.0%
School bus	57.8%	57.3%

* - Adjusted targets from RHTS

JASCH Series

- JASCH2 – Joint activity start time – The overall coincidence ratio between the modeled and observed temporal distributions of joint activity start times is 55%. The main difference is that the model form tends to overestimate start times in the final period of the day (8:00 p.m. to 3:00 a.m.). Even considering that issue, the model tends to underestimate peak period activity start times. Of the joint activities that begin before 8:00 p.m., the modeled percentage of start times between 6:00

and 9:00 a.m. is 2.4 percent, compared to 5.7 percent observed, and the modeled percentage between 3:00 and 6:00 p.m. is 25.3 percent, compared to 30.9 percent observed.

- JASCH3 – Joint activity distance to stop – This is an interim model whose validation is effectively included in the JASCH4 model results described below.

- JASCH4 – Joint activity location – The modeled average trip distance is 4.3 miles, compared to the observed average of 4.5 miles. The modeled averages for subregions are all very close to the observed. The coincidence ratio for the distance frequency distribution is 77 percent.

- JASCH6 – Joint discretionary trip mode choice – Table 6-12 shows the regional modeled and observed joint trip mode shares, which match very closely. Model results for geographic and demographic segments also match the observed mode shares well.

- JASCH5 – Vehicle used for joint home-based tour – Since in the final overall model structure this component's results are not used downstream, its results were not validated.

Table 6-12. Modeled and Observed Regional Joint Trip Mode Shares

	Observed*	Model
HOV 2	45.0%	44.9%
HOV 3+	31.0%	30.9%
Taxi	1.0%	1.1%
Commuter rail/bus – auto access	0.3%	0.3%
Commuter rail/bus – walk access	0.4%	0.2%
Subway/ferry – auto access	0.1%	0.1%
Subway/ferry – walk access	3.0%	3.2%
Local bus – auto access	0.0%	0.0%
Local bus – walk access	2.7%	2.8%
Walk	16.3%	16.1%
Bike	0.3%	0.3%
School bus	45.0%	44.9%

* - Adjusted targets from RHTS

CSCH Series

- CSCH4 – Child departure time from home for independent discretionary tour – The overall coincidence ratio between the modeled and observed temporal distributions of child (non-school) activity start times is 61 percent. The model tends to underestimate a.m. peak period activity start times and overestimate p.m. peak period activity start times. The modeled percentage of start times between 6:00 and 9:00 a.m. is 11.4 percent, compared to 19.5 percent observed, and the modeled percentage between 3:00 and 6:00 p.m. is 53.5 percent, compared to 34.2 percent observed.

- CSCH5 – Child activity duration at independent discretionary stop – The model underestimates the average activity duration at a stop by about 45 minutes.

- CSCH6 – Child travel distance to independent discretionary stop – This is an interim model whose validation is effectively included in the CSCH7 model results described below.

- CSCH7 – Child location of independent discretionary stop – The modeled average trip distance is 2.8 miles, compared to the observed average of 3.1 miles. The modeled averages for subregions are all very close to the observed. The coincidence ratio for the distance frequency distribution is 53 percent.

- CSCH3 – Child mode for independent discretionary trip – Table 6-13 shows the regional modeled and observed joint trip mode shares, which match very closely except for underestimating walk trips and a corresponding overestimation of HOV 2 trips. Model results for geographic and demographic segments also match the observed mode shares well although the underestimation of walk trips and overestimation of HOV 2 trips is most noticeable in Manhattan.

Table 6-13. Modeled and Observed Regional Child Trip Mode Shares

	Observed*	Model
HOV 2	30.2%	46.7%
HOV 3+	50.6%	50.9%
Taxi	0.3%	0.3%
Subway/ferry – walk access	2.0%	0.1%
Local bus – walk access	1.3%	0.1%
Walk	15.3%	1.1%
Bike	0.4%	0.7%

* - Adjusted targets from RHTS

6.2 Highway and Transit Assignment Validation

6.2.1 Highway Assignment

The highway validation focused on three main classes of measures:

- Vehicle-miles of travel (VMT);
- Individual link traffic volumes; and
- Intra-regional traffic flows as defined by screenlines.

All of these measures are based on comparisons of assigned volumes from the model to observed traffic counts. Due to the large number of jurisdictions that maintain the roads in the network and the variety of roadway types, the counts are assembled from several sources. Generally, the highway assignment results match observed data reasonably well, with no major high or low biases compared to traffic counts.

It should be noted that during the validation process, some gaps and errors in the traffic count database were identified and corrected to the degree possible. Additionally, there appear to be some instances where traffic counts were performed at locations where the traffic loading points from TAZ centroid connectors would not well represent traffic on those links. These instances were also identified and corrected to the degree possible. While corrections to the traffic counts database and centroid connector locations improved the validation, neither of these issues has a major impact on model results though they do affect some of the comparisons between modeled volumes and counts—especially percentage root mean square error (%RMSE).

VMT Checks

For the region, the modeled VMT on links with traffic counts is about half a percent lower than the observed VMT computed from the counts. Table 6-14 shows the modeled and observed VMT by facility type, with the percentage difference compared to the targets from the model validation plan. All targets are met. There are no targets for the (generally low volume) local streets and ramp facility types, which comprise less than one percent of the VMT.

Table 6-14. Modeled and Observed Daily VMT by Facility Type

	Model VMT	Count VMT	Total	Target
Interstate/Freeway/Tollway	18,247,437	17,507,599	4.2%	7%
Principal Arterial	6,758,944	7,397,019	-8.6%	10%
Minor Arterial	3,422,622	3,743,737	-8.6%	10%
Major Collector	800,005	743,276	7.6%	15%
Minor Collector	188,273	198,300	-5.1%	15%
Local Street	29,682	55,979	-47.0%	
Ramp	173,104	126,806	36.5%	
Total	**29,620,067**	**29,772,716**	**-0.5%**	**1%**

The percentage differences between modeled and observed VMT for the four time periods used in highway assignment are:

- AM peak (6:00 AM - 10:00 AM): -4.0%
- Mid-day (10:00 AM – 3:00 PM): -1.8%
- PM peak (3:00 PM – 7:00 PM): -3.1%
- Night (7:00 PM – 6:00 AM): -3.2%

Note that this summary does not include all links included in the summary shown in Table 6-14; there are some links with daily counts but not counts by time of day.

Table 6-15 shows the modeled and count VMT for a set of districts that comprise the entire region. The VMT is within 5.5 percent for all subregions except Mercer County and Connecticut.

Table 6-15. Modeled and Observed Daily VMT by Subregion

	Model VMT	Count VMT	% Difference
Manhattan CBD	578,948	612,561	-5.5%
Upper Manhattan	727,844	718,358	1.3%
Other NYC	4,491,255	4,689,221	-4.2%
Long Island	3,052,757	3,013,531	1.3%
Mid-Hudson	6,057,722	6,285,356	-3.6%
NJTPA Core	3,789,973	3,979,602	-4.8%
NJTPA Other	8,245,810	8,128,133	1.4%
Connecticut	1,899,235	1,682,758	12.9%
Mercer County, NJ	776,522	663,196	17.1%
Total	29,620,066	29,772,716	-0.5%

Link Volume Checks

The overall fit between individual modeled and observed link volumes was examined using the percentage root mean square error (%RMSE) measure. Table 6-16 and Table 6-17 show the %RMSE grouped by facility type and volume group, respectively.

The %RMSE error for each segment, and for the entire set of all links with counts, does not meet most of the targets from the validation plan. This may be due to the issues with some count locations and network loading points discussed earlier. For example, the modeled volumes on roadways where zone centroid connectors meet the highway network may be high if actual network loading points for the zone are more dispersed; conversely, modeled volumes on roads where trips from a zone are actually loading may be low if the zone's centroid connectors are not nearby.

Table 6-16. %RMSE by Facility Type

Facility Type	Total	Target
Interstate/Freeway/Tollway	25%	20%-30%
Principal Arterial	54%	30%
Minor Arterial	77%	40%
Major Collector	131%	70%
Minor Collector	207%	70%
Local Street	64%	
Ramp	66%	
Total	**46%**	**40%**

Table 6-17. %RMSE by Volume Group

Volume Group	Links	% RMSE	Target
0 - 1,000	37	552%	100%-200%
1,000 - 5,000	321	150%	45%-100%
5,000 - 10,000	448	78%	36%-45%
10,000 - 20,000	655	54%	28%-34%
20,000 - 30,000	265	44%	24%-26%
30,000 - 50,000	239	38%	21%-24%
50,000 - 100,000	205	26%	12%-21%
100,000 and up	35	21%	12%
All Links	**2,205**	**46%**	**40%**

Table 6-18 shows the VMT, as estimated by the model and observed through traffic counts on 30 major routes that have at least 100,000 VMT on links with counts. Twenty-two of these 30 routes had modeled VMT within 25 percent of observed, and 18 routes had modeled VMT within 20 percent of observed. The model most notably overestimates volumes on the Long Island Expressway, the Taconic State Parkway, and the Palisades Interstate Parkway and underestimates volumes on Shore Parkway, Sunrise Highway, and Meadowbrook State Parkway.

Table 6-19 shows a comparison of volumes on the major crossings into and within New York City, grouped by waterway and location. With one exception, each group's modeled volume is within 15 percent of the traffic counts. The exception is the Kill Van Kull segment, which consists of only the Bayonne Bridge. The Modeled volume on the Bayonne Bridge is 23,000, compared to a count of 19,000.

Table 6-18. Modeled and Observed VMT on Major Routes

	Model VMT	Count VMT	% Diff.	No. of Counts
Garden State Pkwy	2,984,229	2,615,942	14%	81
NJ Turnpike	2,208,268	2,415,491	-9%	26
NYS Thruway	901,986	1,080,273	-17%	12
Southern Pkwy	621,345	603,541	3%	11
Long Island Expy	646,219	492,386	31%	11
I-84 in NY	333,044	373,683	-11%	5
I-84 in CT	460,893	370,743	24%	4
Shore Pkwy	218,822	312,825	-30%	7
Palisades Interstate Pkwy	391,391	309,329	27%	8
I-95 in CT	282,166	304,976	-7%	4
Northern State Pkwy	301,357	300,403	0%	8
I-87	225,942	278,489	-19%	14
Brooklyn Queens Expy	274,494	274,715	0%	10
I-684	300,627	246,085	22%	6
Cross Island Pkwy	198,472	245,611	-19%	9
FDR Drive	287,449	242,997	18%	15
Belt Pkwy	199,454	207,590	-4%	7
I-84	252,045	201,720	25%	10
Henry Hudson Pkwy	191,178	185,151	3%	7
State Hwy 440	148,866	182,235	-18%	11
Sunrise Hwy	126,221	182,235	-31%	5
I-95 In NY	129,302	162,316	-20%	11
State Hwy 17	117,256	149,002	-21%	4
Meadowbrook State Pkwy	80,060	143,504	-44%	6
Hutchinson River Pkwy	147,193	138,502	6%	4
Taconic State Pkwy	218,269	133,205	64%	11
US Hwy 9	97,222	130,592	-26%	15
Saw Mill River Pkwy	102,483	117,414	-13%	4
Gowanus Expy	139,305	104,992	33%	5
I-278	97,058	101,563	-4%	1

Table 6-19. Modeled and Observed Volumes on Major Crossings

	Links	Model	Count	% Diff.
1: Outerbridge Crossing	2	79,414	71,816	11%
2: Goethals Bridge	2	86,141	73,136	18%
Arthur Kill Subtotal	**4**	**165,555**	**144,952**	**14%**
3: Bayonne Bridge	2	23,304	18,755	24%
Kill Van Kull Subtotal	**2**	**23,304**	**18,755**	**24%**
4: Holland Tunnel	2	114,448	92,743	23%
5: Lincoln Tunnel	2	130,278	113,166	15%
6, 7: G Washington Bridge	4	304,266	276,647	10%
8: Tappan Zee Bridge	2	135,531	133,352	2%
9: Mountain Bridge Rd	1	38,003	19,999	90%
10: Newburgh Beacon Bridge	2	90,399	74,500	21%
Hudson River Subtotal	**13**	**812,925**	**710,407**	**14%**
11: Verrazano Bridge	4	218,419	193,100	13%
The Narrows Subtotal	**4**	**218,419**	**193,100**	**13%**
12: Hugh L Carey Tunnel	4	49,551	54,299	-9%
13: Brooklyn Bridge	2	86,723	100,288	-14%
14: Manhattan Bridge	5	103,692	89,087	16%
15: Williamsburg Bridge	2	131,825	112,696	17%
16: Queens Midtown Tunnel	2	72,950	87,938	-17%
17: Ed Koch Queensboro Bridge	5	194,104	178,188	9%
18: R.F. Kennedy Bridge (Queens/Bronx)	2	95,222	85,805	11%
19: R.F. Kennedy (Queens /Manhattan)	2	56,529	66,622	-15%
20: R.F. Kennedy (Bronx / Manhattan)	2	27,876	24,334	15%
21: Bronx Whitestone Bridge	2	128,176	105,719	21%
22: Throgs Neck Bridge	2	127,278	108,859	17%
East River Subtotal	**30**	**1,073,928**	**1,013,835**	**6%**
23: Willis Avenue Bridge - Nb	1	62,564	62,061	1%
24: 3rd Ave Bridge - Sb	1	54,638	59,054	-8%
25: Madison Avenue Bridge	1	43,577	41,782	4%
26: 145th St Bridge	1	24,520	27,918	-12%
27: Macombs Dam Bridge	1	48,108	39,020	23%
28: Cross Bronx Exp Bridge	2	147,857	185,308	-20%
29: Washington Bridge	2	65,124	57,011	14%
30: W 207th St Bridge	1	48,326	39,640	22%
31: Broadway Ave Bridge	1	47,535	35,410	34%
32: Henry Hudson Pkwy Bridge	2	28,466	63,435	-55%
Harlem River Subtotal	**13**	**570,716**	**610,639**	**-7%**

Screenlines

To examine how well the model reflects intra-regional traffic flows, a set of 29 screenlines was defined. The validation plan defined target percentages for the difference between the summed volumes and traffic counts based on the daily traffic across the screenline. Table 6-20 shows the modeled volumes and counts for both directions for these screenlines. The volume difference meets the targets for 25 of the 29 screenlines. Two of the four for which the targets are not met are single link screenlines with average daily volumes of around 10,000 per day per direction.

Some of the major regional trip movements were examined by summing volumes for multiple screenlines. This summary is shown in Table 6-21.

Table 6-20. Modeled and Observed Volumes on Screenlines

	Links	Model NB/EB	Count NB/EB	% Diff NB/EB	Model SB/WB	Count SB/WB	Diff SB/WB	Model Total	Count Total	% Diff Total	Target
Border betw Manhattan & Brooklyn	13	184,143	178,925	2.9%	187,648	177,445	5.7%	371,791	356,370	4.3%	20%
Border betw Manhattan & Queens	9	157,119	160,044	-1.8%	166,464	172,704	-3.6%	323,583	332,748	-2.8%	20%
Border betw Manhattan & Bronx	15	296,737	315,811	-6.0%	301,854	319,162	-5.4%	598,591	634,973	-5.7%	20%
Border betw NJ & Manhattan	8	259,588	238,204	9.0%	289,403	244,352	18.4%	548,991	482,556	13.8%	20%
Border betw CBD & upper Manhattan	17	282,263	325,315	-13.2%	330,648	311,941	6.0%	612,911	637,256	-3.8%	20%
Border betw Brooklyn & Queens	35	480,884	437,680	9.9%	423,593	385,655	9.8%	904,477	823,335	9.9%	20%
Border betw Staten Island & Brooklyn	4	112,639	100,991	11.5%	105,780	92,109	14.8%	218,419	193,100	13.1%	20%
Cross Bay Blvd betw Queens & Rockaway	1	16,164	11,140	45.1%	16,639	10,456	59.1%	32,803	21,596	51.9%	25%
Border betw Queens & Bronx	6	182,099	146,890	24.0%	168,577	153,493	9.8%	350,676	300,383	16.7%	20%
Border betw NJ & Staten Island	6	99,402	84,416	17.8%	89,456	79,291	12.8%	188,858	163,707	15.4%	20%
US202 Bridge betw Westchester & Orange	1	18,858	9,999	88.6%	19,146	10,000	91.5%	38,004	19,999	90.0%	25%
I-84 Bridge betw Dutchess & Orange	2	44,686	37,000	20.8%	45,713	37,500	21.9%	90,399	74,500	21.3%	22%
Border betw Westchester & Rockland (Cuomo Br.)	2	67,944	66,676	1.9%	67,587	66,676	1.4%	135,531	133,352	1.6%	20%
Border betw Bronx & Westchester	24	309,283	311,236	-0.6%	312,118	326,317	-4.4%	621,401	637,553	-2.5%	20%
Border betw Nassau & Suffolk	24	349,420	342,628	2.0%	338,712	343,032	-1.3%	688,132	685,660	0.4%	20%
Border betw Nassau & Long Beach/Jones Beach	8	35,278	66,950	-47.3%	33,527	67,446	-50.3%	68,805	134,396	-48.8%	20%
Border betw Putnam & Dutchess	12	70,752	76,922	-8.0%	72,794	74,973	-2.9%	143,546	151,895	-5.5%	20%
EW Border Betw Queens & Nassau	29	493,597	489,136	0.9%	473,552	468,320	1.1%	967,149	957,456	1.0%	20%
NS Border Betw Rockland & Orange	12	101,306	104,541	-3.1%	100,442	101,530	-1.1%	201,748	206,071	-2.1%	20%
EW Border betw Westchester & Putnam	18	94,189	73,398	28.3%	90,638	69,634	30.2%	184,827	143,032	29.2%	20%
NS Border betw Sussex NJ & Orange NY	8	28,005	24,895	12.5%	27,989	26,285	6.5%	55,994	51,180	9.4%	22%
NS Border betw Bergen NJ & Rockland NY	22	132,521	131,942	0.4%	131,194	131,787	-0.4%	263,715	263,729	0.0%	20%
EW Border betw Putnam & Fairfield CT	7	53,865	49,758	8.3%	54,299	53,863	0.8%	108,164	103,621	4.4%	20%
EW Border betw Westchester & Fairfield CT	21	131,058	121,588	7.8%	134,365	122,045	10.1%	265,423	243,633	8.9%	20%
EW Border of Dutchess & Litchfield CT	10	8,572	8,743	-2.0%	8,635	8,865	-2.6%	17,207	17,608	-2.3%	30%
NS Border of Dutchess & Columbia	14	16,755	16,455	1.8%	15,299	15,410	-0.7%	32,054	31,865	0.6%	25%
EW Border betw Ulster & Dutchess	4	30,871	30,999	-0.4%	30,936	31,000	-0.2%	61,807	61,999	-0.3%	22%
NS Border betw Orange & Sullivan/Ulster	15	75,819	74,044	2.4%	71,413	71,438	0.0%	147,232	145,482	1.2%	20%

Table 6-21. Aggregate Screenline Summary

	Links	Model NB/EB*	Count NB/EB*	% Diff NB/EB*	Model SB/WB*	Count SB/WB*	Diff SB/WB*	Model Total	Count Total	% Diff Total
From/to Manhattan	45	927,402	899,132	3.1%	915,554	907,515	0.9%	1,842,956	1,806,647	**2.0%**
Intra-Manhattan	8	282,263	325,315	-13.2%	330,648	311,941	6.0%	612,911	637,256	**-3.8%**
Other Intra-NYC	46	791,786	696,701	13.6%	714,589	641,713	11.4%	1,506,375	1,338,414	**12.5%**
Other Cross-Hudson	11	230,890	198,091	16.6%	221,902	193,467	14.7%	452,792	391,558	**15.6%**
Other Intra-NYS	127	1,453,825	1,464,811	-0.7%	1,421,783	1,451,252	-2.0%	2,875,608	2,916,063	**-1.4%**
Other NY-NJ	30	160,526	156,837	2.4%	159,183	158,072	0.7%	319,709	314,909	**1.5%**
NY-CT	28	184,923	171,346	7.9%	188,664	175,908	7.3%	373,587	347,254	**7.6%**
Regional cordon	43	132,017	130,241	1.4%	126,283	126,713	-0.3%	258,300	256,954	**0.5%**

* - For "From/to Manhattan," regardless of orientation, "NB/EB" represents to Manhattan and "SB/WB" represents to Manhattan.

6.2.2 Transit Assignment

The transit assignment validation was less straightforward because of gaps in and inconsistencies among observed data sources. Because of these, it was sometimes necessary to choose which measures to prioritize. In general, the goal was to make sure that total transit demand is reasonable and is consistent with areas of highest ridership. The specific checks discussed below provide some information on the results of some of these choices.

It should also be noted that some summaries reflect linked trips, which include transfers (and sometimes multiple modes) between the trip origin and destination, and some reflect boardings (unlinked trips).

Table 6-22 shows an overall summary of mode choice results from CEMDAP (reflecting linked trips), summarized by aggregate transit mode to show the overall mode shares. The mode share for subway is higher than the observed (from the RHTS) while the commuter rail share is lower. Overall, the transit share is 1.8 percentage points, or about 10 percent higher than the observed share. However, the total modeled linked trips (excluding subway/ferry, where there is not a good estimate of observed linked trips) is about 4.2 million daily and 1.3 million for the a.m. peak, compared to observed estimates of 3.8 million daily and 1.3 million for the a.m. peak, and the overall modeled (a.m. peak) subway boardings match the observed counts well (about 2.7 million in both cases) while modeled commuter rail boardings (about 500,000) are higher than observed (about 400,000). These discrepancies represent examples of the data inconsistencies mentioned above, which are not unusual for large metropolitan areas like New York.

Table 6-22. Mode Shares (trips)

Trip Mode	Expanded RHTS Data	Model Results
Commuter rail/bus	4.4%	3.4%
Subway/ferry	9.9%	12.5%
Local bus	3.5%	3.8%
TOTAL TRANSIT	17.7%	19.5%
Auto	65.0%	62.5%
Non-motorized/other	17.2%	18.1%

Linked Transit Trips

Table 6-23 shows that a.m. peak period linked transit trips are within three percent of observed trips. Note that these comparisons represent only local bus and commuter rail/bus. Other rail is not included because the observed data is not comparable with the observed data for local bus and commuter rail/bus (observed data for "other rail" is unlinked trips, not linked). The differences are small within New York City and modeled trips are lower than observed in New Jersey and northern suburbs.

Modeled trips are high in other time periods compared to observed, especially in the night periods. However, as noted above, the overall (daily) transit share is only about 10 percent higher than the observed. Because of this inconsistency and the sequencing of time of day and mode choice in CEMDAP, these differences could not be addressed without adversely affecting comparisons of a.m. boardings and auto trips in other time periods.

Station Groups

Four sets of station groups were defined for rail transit assignment validation. The groups correspond to modes: commuter rail, PATH, subway, and light rail. Nine major commuter rail terminals are defined as individual station groups of one station only. Other commuter rail groups are aggregations of established branches or lines. Subway station groups represent the four New York City boroughs that have subway service.

Table 6-24 shows the comparison of modeled and observed a.m. peak period boardings for aggregate station group segments for which observed data are available—namely, for the three commuter rail operators (MNR, LIRR, and NJT) for PATH, for the two Midtown commuter rail stations, and for the major stations in Brooklyn/Queens. While overall commuter rail boardings beyond the stations in or near New York City are reasonably consistent with counts (lower for the Long Island Railroad, higher for the others), the model overestimates boardings for nearby stations such as Jamaica, City Terminal, Secaucus, and Hoboken, as well as for the Hoboken light rail and outbound commuter rail ridership. Estimates for PATH are fairly close to observed.

Table 6-23. Linked Transit Trip Summary by Time Period and Subregion

Linked Transit Trips by Origin District - Model

	AM	PM	MD	NT	Daily
Manhattan CBD	77,657	365,430	137,077	192,777	**772,941**
Upper Manhattan	103,846	216,914	102,368	95,935	**519,064**
Bronx	122,012	111,623	114,512	52,602	**400,749**
Queens/Brooklyn	345,696	398,730	363,858	213,641	**1,321,925**
Staten Island	51,525	33,612	50,139	15,236	**150,512**
Long Island	120,279	101,005	146,112	115,605	**483,001**
Mid-Hudson East	102,326	84,722	115,294	78,434	**380,776**
Mid-Hudson West	103,257	81,161	101,266	56,573	**342,257**
NJ Essex/Hudson	129,206	128,900	130,089	81,484	**469,680**
NJ Northwest	80,312	73,685	95,850	53,313	**303,161**
NJ South Shore	23,201	19,002	26,193	9,996	**78,392**
Total	**1,259,316**	**1,614,785**	**1,382,760**	**965,597**	**5,222,458**

Linked Transit Trips by Origin District - Observed

	AM	PM	MD	NT	Daily
Manhattan CBD	74,787	420,393	91,844	66,487	**653,511**
Upper Manhattan	110,967	147,799	91,667	38,001	**388,434**
Bronx	143,516	123,187	120,011	58,731	**445,445**
Queens/Brooklyn	338,507	274,188	322,243	92,123	**1,027,061**
Staten Island	32,303	13,782	25,569	10,363	**82,017**
Long Island	125,104	23,053	31,793	19,141	**199,092**
Mid-Hudson East	121,736	48,587	63,438	15,832	**249,592**
Mid-Hudson West	94,549	36,608	37,631	8,034	**176,821**
NJ Essex/Hudson	144,687	103,622	83,824	27,480	**359,613**
NJ Northwest	93,774	33,353	25,643	10,099	**162,869**
NJ South Shore	23,167	3,613	7,052	2,369	**36,200**
Total	**1,303,098**	**1,228,183**	**900,714**	**348,659**	**3,780,655**

Table 6-23. Linked Transit Trip Summary by Time Period and Subregion (continued)

Linked Transit Trips by Destination District - Model

	AM	PM	MD	NT	Daily
Manhattan CBD	311,028	136,503	296,309	67,258	**811,098**
Upper Manhattan	197,703	133,201	180,011	49,200	**560,115**
Bronx	74,997	155,625	86,091	67,785	**384,498**
Queens/Brooklyn	258,102	468,636	306,098	257,786	**1,290,622**
Staten Island	25,414	50,945	32,936	28,756	**138,051**
Long Island	84,243	146,536	101,732	140,654	**473,165**
Mid-Hudson East	64,157	127,642	87,902	98,789	**378,491**
Mid-Hudson West	67,129	117,866	80,223	76,650	**341,867**
NJ Essex/Hudson	105,663	153,381	120,737	90,920	**470,701**
NJ Northwest	56,909	97,598	72,155	71,855	**298,517**
NJ South Shore	13,972	26,852	18,566	15,944	**75,334**
Total	**1,259,316**	**1,614,785**	**1,382,760**	**965,597**	**5,222,458**

Linked Transit Trips by Destination District – Observed

	AM	PM	MD	NT	Daily
Manhattan CBD	455,685	75,510	98,961	41,903	**672,060**
Upper Manhattan	115,563	136,743	95,639	37,781	**385,726**
Bronx	116,221	141,628	119,329	65,166	**442,343**
Queens/Brooklyn	312,551	308,126	312,906	103,080	**1,036,663**
Staten Island	13,202	28,000	28,400	8,297	**77,899**
Long Island	35,269	111,238	34,712	13,165	**194,384**
Mid-Hudson East	56,880	111,710	57,417	23,468	**249,476**
Mid-Hudson West	37,844	79,939	33,153	8,919	**159,855**
NJ Essex/Hudson	118,118	122,531	88,336	31,927	**360,913**
NJ Northwest	37,659	90,179	25,471	12,826	**166,135**
NJ South Shore	4,104	22,579	6,391	2,128	**35,202**
Total	**1,303,098**	**1,228,183**	**900,714**	**348,659**	**3,780,655**

Table 6-24. Station Group Transit Assignment Summary

ONS	Model	Observed	Difference	% Difference
Total	624,253	468,555	155,698	33%
NJT	114,503	91,642	22,861	25%
MNR	131,042	100,806	30,236	30%
LIRR	107,720	121,506	-13,786	-11%
Penn Station/GCT	22,593	31,738	-9,145	-29%
Jamaica/City Term/Newark/Secaucus	136,447	52,284	84,163	161%
PATH	111,948	70,579	41,369	59%
All commuter rail except NYC	353,265	313,954	39,311	13%
OFFS	**Model**	**Observed**	**Difference**	**% Difference**
Total	612,913	469,798	143,115	30%
NJT	114,503	91,642	22,861	25%
MNR	131,042	100,806	30,236	30%
LIRR	107,720	121,506	-13,786	-11%
Penn Station/GCT	22,593	31,738	-9,145	-29%
Jamaica/City Term/Newark/Secaucus	136,447	52,284	84,163	161%
PATH	111,948	70,579	41,369	59%
All commuter rail except NYC	353,265	313,954	39,311	13%
TOTAL ONS AND OFFS	**Model**	**Observed**	**Difference**	**% Difference**
Total	1,237,166	938,353	298,813	32%
NJT	114,503	91,642	22,861	25%
MNR	131,042	100,806	30,236	30%
LIRR	107,720	121,506	-13,786	-11%
Penn Station/GCT	22,593	31,738	-9,145	-29%
Jamaica/City Term/Newark/Secaucus	136,447	52,284	84,163	161%
PATH	111,948	70,579	41,369	59%
All commuter rail except NYC	353,265	313,954	39,311	13%

Hub-Bound Summary

The hub-bound summary (for a.m. peak boardings to the Manhattan CBD) is summarized in Table 6-25. The modeled results show fewer trips than observed inbound except for local bus. The outbound model results (where overall numbers are lower) are generally higher than observed. While there are some consistencies with other summaries (e.g., high outbound commuter rail summaries from the station group report), there are also inconsistencies. For example, overall a.m. subway boardings match counts well, as do modeled a.m. peak linked trips. Additionally, the work location model summary (see Table 6-3) shows a good fit between modeled and ACS commute patterns to Manhattan. It was felt that increasing work trips to Manhattan would result in a worse fit for other model measures including transit boarding totals and highway screenlines.

Table 6-25. Hub-Bound Transit Summary

Modeled INBOUND Hub-Bound Transit Flows

	Bus	Ferry	Rail	Subway/PATH	Tram	Total
60th St	38,690	0	49,111	189,591	0	**277,392**
Queens	33,970	435	62,196	117,022	0	**213,623**
Brooklyn	44,299	281	0	210,080	0	**254,660**
Staten Island	0	29,048	0	0	0	**29,048**
New Jersey	91,400	2,539	32,388	46,569	0	**172,897**
Total	**208,359**	**32,303**	**143,695**	**563,262**	**0**	**947,619**

Observed INBOUND Hub-Bound Transit Flows

	Bus	Ferry	Rail	Subway/PATH	Tram	Total
60th St	17,395	0	72,541	307,695	0	**397,631**
Queens	10,699	50	81,094	233,817	1,535	**327,195**
Brooklyn	18,028	454	0	382,237	0	**400,719**
Staten Island	0	16,373	0	0	0	**16,373**
New Jersey	110,502	11,128	49,696	67,700	0	**239,026**
Total	**156,624**	**28,005**	**203,331**	**991,449**	**1,535**	**1,380,944**

Modeled OUTBOUND Hub-Bound Transit Flows

	Bus	Ferry	Rail	Subway/PATH	Tram	Total
60th St	22,574	0	10,074	175,947	0	**208,596**
Queens	4,318	50	7,261	74,111	0	**85,741**
Brooklyn	1,369	102	0	101,527	0	**102,998**
Staten Island	0	8,735	0	0	0	**8,735**
New Jersey	42,976	269	5,257	9,496	0	**57,999**
Total	**71,238**	**9,156**	**22,592**	**361,082**	**0**	**464,069**

Observed OUTBOUND Hub-Bound Transit Flows

	Bus	Ferry	Rail	Subway/PATH	Tram	Total
60th St	5,777	0	7,474	153,059	0	**166,310**
Queens	191	68	4,975	53,932	294	**59,460**
Brooklyn	192	18	0	80,280	0	**80,490**
Staten Island	0	2,329	0	0	0	**2,329**
New Jersey	31,120	1,126	7,370	15,253	0	**54,869**
Total	**37,280**	**3,541**	**19,819**	**302,524**	**294**	**363,458**

6.3 Sensitivity Testing

Sensitivity testing involves adjusting key factors in the model and observing the effects on forecasted travel. These adjustments can be made to model parameter values (e.g., mode choice cost coefficients) and to model inputs (e.g., land use variables, socioeconomic conditions, fuel costs, etc.).

NYMTC defined the following sensitivity tests to be performed for the 2012 NYBPM:

- **Toll changes** – Increase tolls on all (tolled) crossings.
- **Parking cost changes** – Increase and decrease parking costs in the Manhattan CBD south of 60th Street.
- **Major development** – This consisted of coding a hypothetical major development in Long Island City, including expected changes in population and employment outside the development itself.
- **AV "package"** – While the effects of autonomous vehicles (AVs) on travel behavior are not truly known, there has been a fair amount of research that suggests assumptions about changes in the model to reflect an AV future. The following items were revised:
    - Roadway capacity increases, to reflect more efficient use of roadways by AV.
    - Revise in-vehicle time parameters for auto modes to reflect the utility of performing other activities in the car.
    - Change parking cost assumptions to reflect that AVs may not need to park at the destination.

The sensitivity tests are summarized in a technical memorandum (Cambridge Systematics, Inc., 2021b). Key findings are summarized below.

6.3.1 Toll change test

In this test, the toll rates for all tolled crossings (bridges and tunnels) in the region were increased by 25 percent compared to the base year scenario. The objective was to see how key demand measures were affected by increasing tolls.

There were slight decreases in volumes on major crossings with the increased tolls, on the order of about one percent. There was a decrease of about two percent on the crossings between New Jersey and Manhattan, where auto tolls cannot be avoided, and slightly larger increases for other crossings where auto tolls cannot be avoided. Changes in origin-destination patterns were very small.

In general, the sensitivity of the model to toll changes on major crossings appeared to be reasonable. Traffic volumes decreased by a modest amount, reflecting that in many cases, tolls cannot be avoided on these crossings or would require substantial travel time increases to do so. The toll increases do result in a small mode shift from auto to transit.

6.3.2 Parking cost change test

In this test, the parking costs – based on employment density - or the Manhattan CBD (below 60th Street) were revised. Two tests were performed, one where parking costs were increased by 25 percent compared to the base year scenario, and one where they were decreased by 25 percent. The objective was to see how key demand measures for travel to the CBD were affected by changing parking costs.

With the parking cost decrease, the changes in volumes varied across the four screenlines surrounding Manhattan (Brooklyn, Queens, Bronx, and New Jersey), resulting in a small overall increase in traffic volumes to and from Manhattan. While the sensitivity to parking cost is fairly low for the parking cost decrease scenario, it is near zero for the parking cost increase scenario, possibly because many of those who drive to the CBD have no better alternatives or because available parking is limited in the CBD. Consistent with the screenline results, the model showed very little change in transit trips under the parking cost increase scenario, but transit trips decreased by about two percent inbound under the parking cost decrease scenario and increased by four percent outbound.

In light of the relative insensitivity of the model results to parking cost changes and the near-complete insensitivity to cost increases these two tests indicate some areas where the model should be reexamined. The contribution of parking cost to mode choice utility functions should be examined, especially regarding parking cost increases. Additionally, further examination of outbound AM peak transit mode choice results could be undertaken.

6.3.3 Major development test

In this test, the modeled impacts of a simplified representation of a hypothetical major development with 15,000 employees in Long Island City were examined. A complete description of the assumptions regarding increases in service/retail employment supporting the additional workforce at the site, decreases in employment elsewhere in the region, and changes in population to supply the new workers are documented in the sensitivity test memo. The objective was to examine the sensitivity of the model to location-specific changes in socioeconomic data inputs.

Volumes on some of the major routes in the general area of the development site were examined; generally, the volumes in Brooklyn and Queens increased by about one percent under the major development scenario while major roadways in Nassau County increased by about half a percent. Volumes on the East River crossings near the site decreased by about three percent. Transit trips increased in Queens and Brooklyn while decreasing in the rest of New York City.

In general, the model results appear to change in the correct direction under the major development scenario. Traffic volumes and transit ridership increase near the development site and decrease in places where employment is assumed to decline. There seems to be a slight mode shift from transit to auto, reflecting that some employment is shifted from Manhattan, where transit service is the greatest, to Long Island City, which has transit service but at a lower level than Manhattan. The magnitude of the transit decreases in Manhattan, however, seems to be too high, and VMT decreases may be too large. As part of the next model update, It is worth looking further into the sensitivity of the model regarding trips to Manhattan, especially work trips.

6.3.4 AV “package” test

Autonomous vehicles (AVs) are a topic of great interest in transportation planning. While the pace of AV adoption in the future is unknown, many planners believe that AVs will become an important part of the transportation supply in the coming decades. While the effects of AVs on travel behavior are not truly known at this time, there has been a fair amount of research that allow making assumptions that would reflect a possible AV future.

A relatively simple “AV package” scenario was developed to test the sensitivity of the model to some of the impacts related to AVs. The changes made for this scenario include the following:

- <u>Increase highway capacity by 50% for arterials and 80% for freeways and expressways</u>. This would reflect that autonomous vehicles can operate more efficiently than human-driven vehicles, especially on limited access and other higher class roadways.

- <u>Reduce the sensitivity to auto travel to in-vehicle travel time by 25%</u>. This would reflect the decreased sensitivity to driving time in AVs. It was not a simple matter to model this assumption, but it is likely one of the most important impacts of AVs and needed to be included in some way.

The results of the AV test are briefly summarized below.

- Overall vehicle-miles traveled increased by 34 percent regionwide. There are some relatively minor geographic variations in the percentage increases. VMT increased more on freeways and expressways, likely because they saw the largest capacity increases.
- Volumes across the screenlines used for base year model validation increased by about the same percentage as VMT, with minor variations among some screenlines. Volumes on most major roadways increased by 30 to 50 percent.
- Transit trips increased by about 20 to 30 percent.

While it is impossible to know the true impacts of the introduction of AVs, the VMT and volume increases appear to be reasonable given the particular assumptions of the AV package scenario. It is apparent that the total increase in the number of tours and trips is a significant part of the increase because the transit trips also increased despite the relatively larger benefits to auto travel from the changes made in the scenario. It is counterintuitive that transit trips increased under the AV package scenario since transit level of service was assumed to remain the same as in the base scenario. However, the increase in the number of trips made due to increased accessibility appears to have increased trips by all modes, not just auto.

In conclusion, the sensitivity test was successful in that the AV package showed a significant increase in trip making, auto travel, and freeway volumes. The test results indicate some things to examine further as part of the next model update.. Specifically, the parameters of the model components relating tours and trip making to auto accessibility and the relative increase in transit trip making should be examined further during the model validation. Some additional sensitivity testing related to trip length as a function of highway impedance could also be considered.

7.0 Lessons Learned

In an effort of the magnitude of the 2012 NYBPM development, a number of lessons are learned. It is important to document these lessons so that they can help to improve future model development efforts. A summary of the major lessons learned is provided below.

- **Model complexity** – A desirable feature of the approach chosen for the 2012 NYBPM (PopGen/CEMSELTS/CEMDAP) was the attention to behavioral realism, which, at the time model development began, exceeded (and still exceeds) that available from other activity-based modeling approaches. Given the size and diversity of the population and the transportation system in the New York metropolitan area, simpler approaches would have great difficulty in understanding the ways that travelers behave and interact within the region, and how the temporal, spatial, and capacity constraints affect travel behavior.

  It was known from the beginning of the model development process, of course, that even a sophisticated modeling approach such as this is still a simplification of the exceedingly complex behavior exhibited by millions of travelers. The PopGen/CEMSELTS/CEMDAP process already included a number of simplifications required for practicality (for example, not explicitly modeling joint activities and travel for households with more than five persons). Additionally, for the sake of the feasibility of the process, the project team decided at the start not to take advantage of some available features in the PopGen/CEMSELTS/CEMDAP process, including vehicle tracking within households and the evolution of households and persons over time for forecasting.

  Even with these simplifications included in the NYBPM, the final model provides some lessons in terms of the tradeoffs between the useability of model results (and the run time needed to produce them) and the complexity of the real behavior being simulated. Looking at the final model, it is apparent that there are some additional simplifications that can be made that will make the model faster and/or easier to validate, use, and understand. These include some components that could be simplified or perhaps removed entirely, and some whose structures could be simplified.

- **Data inconsistencies** – As discussed in Chapter 3.0, developing, validating, and applying the 2012 NYBPM required a wide variety of data of different types and from varying sources. Many of these sources were independent of one another in terms of how they were collected or assembled. The various surveys used were conducted between 2007 and 2015, a fairly long period during which many factors affecting travel demand were changing (for example, the 2008 recession and its recovery). The surveys were conducted using different sampling and retrieval methods, resulting in inconsistencies among their results. The ability to impute or expand data for each survey depended on how the survey was conducted and the type of information collected.

  Traffic counts are a relatively stable data source, due to the frequency of data collection and the objective of collecting a 100 percent count (though in many cases, counts are taken only for a short period that needs to be adjusted to reflect average weekday conditions over the year). Individual counts are therefore often inconsistent with counts on nearby roadways when they are taken at different times. The unknown error rates for specific counts present challenges in producing a consistent traffic count database for assignment validation.

  Perhaps the data type that has the most uncertainty and inconsistency associated with it is the transit demand (count) data. There is a multitude of transit services and operators in the region, and the types of count data vary significantly. Boardings are easier to count for some types of services such as commuter rail, but some types of boardings, such as subway transfers, may not be able to be counted at all since they do not require passengers to pass through a gateway or pay an additional fare. Another issue that is unable to be considered is fare evasion, which is believed to represent a small but substantial number of riders for some services.

The discussion above has focused on the inconsistencies within data types, but there is even greater inconsistency between data types. While no two of the many types of data sources used counted *exactly* the same items, there are many overlaps between what is counted by different data sources. It was impossible to make all data sources completely compatible, and when model results were compared to observed data during validation, there was no way to make calibration changes that would improve all comparisons; a particular change might improve some results but make others worse.

The lesson is to spend time during the data assembly process to identify inconsistencies where possible and to make adjustments in the data to be used for validation. Additionally, since all inconsistencies can not be eliminated, it makes sense to determine which data sources used for validation comparison are considered the most reliable and to prioritize validation tests accordingly.

8.0 References

Arizona State University and Cambridge Systematics. *Synthetic Population Generation: An Application of PopGen*. Submitted to New York Metropolitan Transportation Council, 2020.

Cambridge Systematics, Inc. *New York Best Practice Model – Model Implementation Plan*. Prepared for the New York Metropolitan Transportation Council, March 2017.

Cambridge Systematics, Inc., University of Texas, Austin, and Arizona State University. *New York Best Practice Model – Model Design Plan*. Prepared for the New York Metropolitan Transportation Council, July 2017.

Cambridge Systematics, Inc. and EA Harper Consulting. *New York Best Practice Model – Model Validation Plan*. Prepared for the New York Metropolitan Transportation Council, August 2017.

Cambridge Systematics, Inc., David Rubin, EA Harper Consulting, Gallop Corporation, and Watchung Transportation, LLC. *New York Best Practice Model: Highway and Transit Network Development and Integration*. Prepared for the New York Metropolitan Transportation Council, September 2018.

Cambridge Systematics, Inc. and EA Harper Consulting. *Task 5B: Transit Validation Data Development*. Prepared for the New York Metropolitan Transportation Council, April 2018, revised December 2019.

Cambridge Systematics. "Visitor model estimation." Memo to NYMTC, October 5, 2018.

Cambridge Systematics, Inc., University of Texas, Austin, and Arizona State University. *New York Best Practice Model – Model Estimation Report*. Prepared for the New York Metropolitan Transportation Council, December 2018.

Cambridge Systematics, Inc. and EA Harper Consulting. *Task 5A: Highway Validation Data Development*. Prepared for the New York Metropolitan Transportation Council, April 2018, revised December 2019.

Cambridge Systematics, Inc., EA Harper Consulting, Florida International University, Gallop Corporation, and CDM Smith. *New York Best Practice Model – Base Year Model Validation Report*. Prepared for the New York Metropolitan Transportation Council, December 2020.

Cambridge Systematics, Inc. and EA Harper Consulting. *New York Best Practice Model 2012: Users' Guide*. Prepared for New York Metropolitan Transportation Council, 2021.

Cambridge Systematics, Inc. "NYBPM 2012 Update – Task 7 Summary." Memorandum to NYMTC, March 2021.

Cambridge Systematics, Inc. "NYBPM 2012 Sensitivity Test Results." Memorandum to NYMTC, July 2021.