APPENDIX I

Travel Demand Model Documentation

RTP

2026
Regional Transportation Plan

Chicago Metropolitan Agency for Planning

Table of contents

Overview 1

Overview of the regional model structure and process 2

Travel model data inputs 6

Travel survey data 6

Socioeconomic data 7

Highway network 9

Spatial and geometric accuracy 13

Highway project coding 14

Bus route coding 18

Future bus coding 22

Rail network 24

Link fields 25

Node fields 26

Rail route coding 27

Future rail coding 29

Zone systems 31

Trip generation zones 31

Modeling zones 32

Capacity zones 35

Analysis network preparation 37

Highway network 37

Transit network 40

Transfer links 41

Access and egress links 41

Zonal impedances 42

Ancillary data input files 43

M01 file 44

DISTR file 44

M023 file 46

CBD parking file 48

Monte Carlo simulation **50**
Population synthesis **51**
Trip generation **53**
Telework **54**
Trip generation processing steps **57**
Model control keywords 57
Study area geography 58
Household type table 59
Load synthetic households 60
Household vehicle availability 60
Household trip productions and attractions 63
Group quarters trip generation 68
Allocation of non-home trip ends 69
External trip ends 73
Create final vehicle trip output file 74
Mode choice-destination choice model **75**
Private auto costs **77**
Hired auto costs **77**
Taxi 77
Transportation network companies 78
Transit costs **79**
Utility calculations **80**
Visitor trips **90**
Time-of-day model **90**
Hired auto deadhead trips **92**
Final trip roster **92**
Traffic assignment **93**
Special trip handling **93**
Commercial vehicle trips **93**
Point-of-entry trips **96**
Tolling **97**

Assignment time periods **99**
Time-of-day factors **103**
Volume-delay functions **103**
VDFs for links ending at signalized intersections (vdf1 and vdf3) 104
VDFs for freeways and expressways (vdf2, vdf4 and vdf5) 106
VDF for metered freeway entrance ramps (vdf8) 107
VDF for links with tolls (vdf7) 107
Link speeds **108**
Emissions calculation **109**
Travel model validation **109**
Trip generation **110**
Household crosstabs 110
Vehicle availability crosstabs 114
Trip rates versus the NHTS 116
Employment 117
Work from home household analysis 118
Trip distribution **120**
Home-based work trip distances 120
County-to-county work trip flow comparison 122
Mode choice **123**
Home based work mode share 123
Distribution of transit origins and destinations 123
Bike trip distance analysis 124
Hired vehicle rates 125
Carpooling rates 125
Commercial vehicles trips 126
Traffic assignment (highway) **127**
Probe data analysis: travel speed 127
VMT shares by district and facility 130
Observed vs. modeled VMT 131
Transit assignment **134**
2026 validation **136**

Work-from-home households 136
Shortest path analysis 138
Model validation summary and areas for further research 144

Overview

The Chicago Metropolitan Agency for Planning (CMAP) is the primary agency responsible for the development and maintenance of travel forecasting methods for northeastern Illinois. Prior to CMAP, this responsibility resided at one of its predecessor agencies, the Chicago Area Transportation Study (CATS). CMAP/CATS has been developing and improving travel forecasting procedures regularly since 1956. CATS originally developed and employed travel demand models to assist in the development of regional transportation plans. The four-step modeling process (trip generation, trip distribution, mode split, and traffic assignment) was fundamental from the beginning. Early enhancements focused on making the process run more quickly on the computers available at the time and on the calibration of individual model components. As time passed and transportation questions changed, the model was updated, revised, and extended to answer them.

In the 1970s, in response to concerns about improving public transit, CATS concentrated enhancement activities on the mode split model and transit assignment techniques. In the late 1970s and early 1980s, efforts were focused on adapting the modeling process to subarea and project specific studies. For example, CATS developed a block-by-block zone system for downtown Chicago. Trips were generated based on zonal floor space from a building-by-building file of the area. Networks were coded with detailed pedestrian links. These techniques were employed to evaluate transit alternatives for the central business district. Similarly, zone sizes were reduced and more detailed highway networks were coded in suburban areas to evaluate freeway proposals.

When federal regulations were changed to require emissions estimates for air quality conformity analysis, the regional models were initially employed as they then existed. In 1994, the first significant model changes, explicitly motivated by conformity issues, were implemented. Since then, CATS (and now CMAP) has committed substantial resources to develop models that are responsive to the needs imposed by air quality requirements. CMAP continuously strives to improve its travel forecasting techniques in response to policy priorities.

This report documents the status of CMAP's regional travel demand model. Much of the text in this document is drawn from predecessor reports developed for the adoption of GO TO 2040[1] and ON TO 2050 plan[2] plans. Since then, several model coefficient updates have been implemented within the CMAP model.

---

[1] Chicago Metropolitan Agency for Planning, "Travel Model Documentation Final Report," October 2010, http://www.cmap.illinois.gov/documents/10180/29685/Travel-Model-Documentation_10-2010.pdf/526a71e3-62c2-450d-a34e-4cda1f81b1ad.

[2] Chicago Metropolitan Agency for Planning, "ON TO 2050 Travel Demand Model Documentation," October 2018, https://www.cmap.illinois.gov/documents/10180/911391/FINAL+Travel+Demand+Model+Documentation+Appendix.pdf/f3b1322c-2e60-2513-720f-38ee68b799d1?t=1538537837892.

Overview of the regional model structure and process

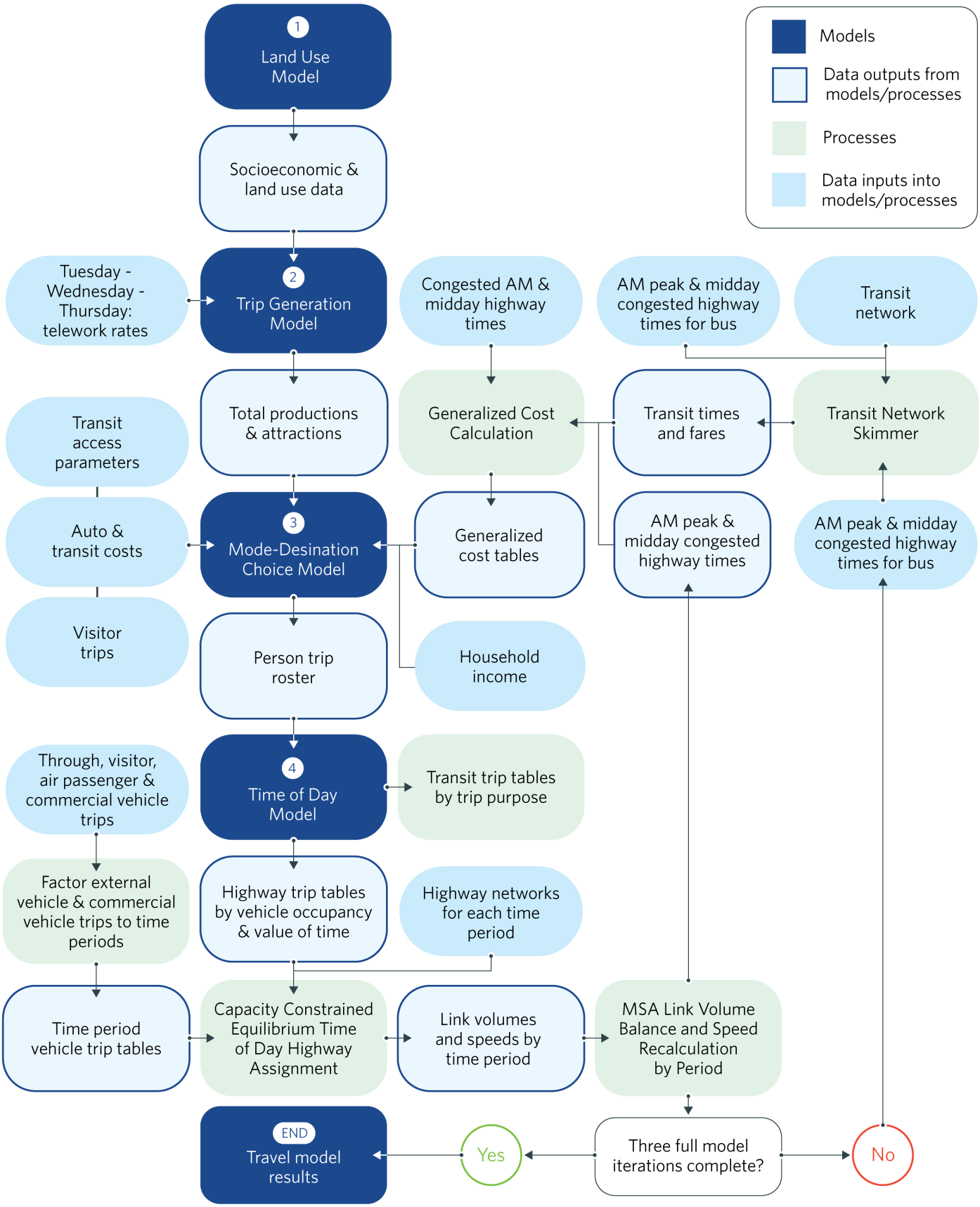
The CMAP trip-based travel demand model reflects the travel behavior reported in the 2018-2019 My Daily Travel survey. CMAP is currently administering a new household travel survey to capture post-COVID travel behavior, which will inform future updates to the model. Additionally, the model has been revised to include several modeling procedure improvements and enhancements, which will be discussed in this document. The present CMAP region, for analysis purposes, includes Cook, DuPage, Kane, Kendall, Lake, McHenry, and Will counties, as well as parts of other counties in Illinois, Indiana, and Wisconsin that buffer the region. Figure 1 contains a flow chart showing the general steps used in the travel demand modeling process, including data inputs that feed model procedures and processes, and the output data generated at each step. Many outputs then serve as inputs to subsequent procedures.

Socioeconomic and land use data are supplied to the travel model from the UrbanSim land use model. UrbanSim provides information on the spatial distribution of households, population, and employment throughout the modeling area, which is used to estimate where trips begin and end. UrbanSim provides a synthetic population for the region, which includes a variety of characteristics of both the households and the people who live in them.

Trip generation is the first step in the trip-based model. It is how land use planning and zoning quantities are converted into trip productions (beginnings) and attractions (endings) that serve as measures of transportation demand. The process enumerates of all households in the modeling area and matches them to households from the My Daily Travel survey to develop trips made by household members. For home-based trips (ones that start or end at a home location), trip ends located at the travelers' homes are defined as productions, and trip ends located at the non-home end are defined as attractions. The trip generation model includes special procedures to address households with at least one worker working from home. The output of trip generation is a complete set of trip productions and attractions. These are complete lists of trip starting and ending locations, segmented by specific trip purposes, but at this stage the trip ends are unconnected to one another.

The next model in the process is a joint mode choice-destination choice model, which simultaneously answers the questions of *where are trips going* (i.e., what are the origin and destination zones for each trip) and *what mode is being used to get there*. This model is implemented as a hierarchical nested logit model. Trips are estimated for five trip purpose categories, including: home-based work trips for residents of low-income households; home-based work trips for residents of high-income households; home-shopping trips; all other home-based trips; and all non-home-based trips.

Figure 1. Modeling process overview



The modal options included in the mode choice-destination choice model are single-occupant vehicles (driving alone), private vehicles with two occupants, private vehicles with at least three occupants, taxis, regular ride hail using a transportation network company (TNC), shared ride hail using a TNC, transit, walking, and biking. Procedures used to estimate auto and transit costs were ported over from the prior version of the CMAP model and retain the use of Monte Carlo simulation, which attempts to decrease the errors inherent in modeling when average values are used. By contrast, Monte Carlo simulation allows models to use knowledge of the distribution of attributes and probabilistically select them — a meaningful number of simulations is generally run to minimize variability in the results. Monte Carlo simulation is used to estimate parking costs, the traveler's income, and the access and egress times from the primary transit routes.

Following mode choice-destination choice, a time-of-day model is run for auto-based trips to determine which of the eight time-of-day periods modeled by CMAP the trip occurs in. These models are sensitive to road congestion during the day. Additionally, a set of deadhead trips is created for taxi and TNC trips. These represent no-fare trips the drivers make between dropping off a passenger and picking up the next one.

The final step of the travel demand procedures is the assignment models. The highway assignment and transit assignment models rely on two different algorithms. The highway assignment model uses the vehicle trips developed from the time-of-day model and a description of the transportation system to estimate the volume of trips on each segment of the road network. For the air quality analysis, the highway assignment procedure is essential for estimating the vehicle miles traveled (VMT) on each highway segment and estimating the speed of each highway segment. The highway assignment step has two significant features that are important for both transportation and air quality analysis. First, because it is a capacity-constrained equilibrium assignment, the level of service (in terms of travel time) worsens as additional volumes are assigned to each link. Second, the equilibrium procedure solution ensures that simulated travelers are not able to improve their level of service (i.e., travel time) by any alternate routing. For each individual simulated traveler, travel times are optimal to the supply and demand of transportation because the traveler cannot find a shorter route.

The transit assignment model is a multipath procedure that evaluates the potential times and costs of all the transit strategies between an origin and destination, identifies the good strategies, and creates zone-to-zone times and costs. This model is also used to assign person trips to the transit network. The transit assignment model is important because it generates the transit times and costs used by the mode choice-destination-choice model, while the assignment of trips to transit routes is needed for project studies.

As shown in Figure 1, the steps of mode choice-destination choice, time-of-day modeling, and time-of-day assignment are iterated through three times (iterations zero through two). Morning peak congested times and distances are used for the work-trip purposes and midday times and distances are used for the nonwork purposes to determine trip destinations and modes. To create these, the link volumes from each full model iteration time-of-day assignment are combined (the step termed “volume balancing and speed recalculation”) with the link volumes from the same period in the previous iterations using the Method of Successive Averages (MSA). For example, the link volumes resulting from the first and second iterations of the time-of-day highway assignment

for period three are combined using the MSA procedure, then skimmed to produce the highway travel information input to the generalized cost calculation for the next iteration of the process.

This process is enhanced through the inclusion of iterative feedback involving the transit system. During initial global iteration zero, the transit schedules for morning peak and midday service are used to feed the generalized cost calculation but are overwritten for links where estimated congestion would cause a bus to fall behind its schedule. Buses that operate on roads are obviously impacted by other traffic on the road (and vice versa). Once the congested roadway times are calculated at the end of global iteration zero, they are fed back into the appropriate transit schedules, which are adjusted to reflect the traffic conditions. These updated transit times are then used in the revised generalized cost calculations. Buses that have special operating priorities (such as bus-on-shoulders or traffic signal vehicle pre-emption) are only subject to the congested roadway times for the appropriate segments of their itinerary.

The time-of-day traffic assignment procedure more realistically matches travel demand to network supply and structure as these vary over the course of 24 hours. The time-of-day procedure relies on a multiclass traffic assignment, enabling the conformity emissions analysis to reflect link volumes by specific vehicle type rather than using regional or statewide averages. The traffic assignment also includes consideration of tolling where the separate vehicle classes experience different toll rates and toll rate weights, based on differing values of time for the vehicle classes. Separate assignments estimate highway vehicle-miles and travel speeds for eight time periods during the day:

- Overnight off-peak (8:00 p.m. to 6:00 a.m.)
- Preceding morning peak (6:00 to 7:00 a.m.)
- Morning peak (7:00 to 9:00 a.m.)
- Post morning peak (9:00 to 10:00 a.m.)
- Midday (10:00 a.m. to 2:00 p.m.)
- Preceding evening Peak (2:00 to 4:00 p.m.)
- Evening peak (4:00 to 6:00 p.m.)
- Post evening peak (6:00 to 8:00 p.m.)

Results of the separate period assignments are accumulated into daily volumes and tabulated into the vehicle-mile-by-vehicle-type-by-speed-range tables needed for the vehicle emission calculations. The assignment results are also used to support project, program, and policy analyses.

The remainder of this document discusses each component of the trip-based model and describes the various data inputs required to run the model.

Travel model data inputs

Model calibration ensures that the various components of the overall travel model replicate observed data for a base year. Calibration requires adjusting parameters for individual sub-models to ensure that they each produce realistic results. Calibration is complete once model results match the observed patterns found in Census data, regional household travel surveys, transit origin-destination surveys or other data sources. The CMAP trip-based model is calibrated to the base year 2019.

Several data inputs are required to provide the trip-based models with the information necessary to estimate travel patterns. Demand side information includes travel surveys and socioeconomic data on where people live and work. Supply side data include the physical road and transit networks. Various zone systems are used to aggregate data to meaningful geographies. This section briefly describes the data used to develop and apply the regional model.

Travel survey data

Travel models are behavioral models of travel choices made by people and require data describing observed travel behavior. These data come from household travel surveys. The original CATS home interview survey was taken in 1956 and consisted of almost 40,000 household interviews. CMAP's travel models were originally developed using a 1970 home interview survey, which obtained the daily travel patterns for over 21,000 households in the region.

In 1979, a much smaller home interview was conducted. This survey was combined with the 1980 Census Journey to Work data and was used to review and modify the agency's modeling procedures. Between 1988 and 1991, another large-scale home interview survey (over 19,000 households) was conducted. The information from this survey and the 1990 and 2000 censuses have been used to update and modify the travel demand procedures.

Starting in January 2007 and lasting one year, CMAP completed a comprehensive travel and activity survey for northeastern Illinois called Travel Tracker. A total of 10,552 households participated in either a one- or two-day survey, providing a detailed travel inventory for each member of their household on the assigned travel day(s). As a test of available technology, 460 Travel Tracker participants also volunteered to wear global positioning devices (GPS) or to use auto-based devices to track their travel.

The next regional travel survey completed by CMAP was called My Daily Travel. Data collection began in summer 2018 and concluded in spring 2019. Nearly 12,400 households participated, logging their detailed travel information for one day.[3] A quarter of these households recorded their itineraries via a GPS-enabled mobile app. At present, CMAP is conducting a multi-phase update to

[3] As with the Travel Tracker survey, the Northwestern Indiana Regional Planning Commission conducted the My Daily Travel survey in coordination with CMAP to collect the same data for three counties in northwest Indiana. The CMAP My Daily Travel survey data are available on the CMAP Data Hub.

My Daily Travel to capture post-COVID travel behavior. Phase 1 data were collected between September 2024 and June 2025 from 3,522 households.

In addition to the home interview surveys, there have been several other data collection efforts, including a 1986 Commercial Vehicle Survey, a 1963 Pedestrian Survey, a 1987 Survey of Parkers in the Chicago Central Business District, and a 1991 Survey of Parking Spaces in the Chicago Central Business District. All of these surveys were used to enhance the region's travel demand procedures.

Socioeconomic data

Socioeconomic data used for trip generation estimates are generated by UrbanSim and guided by the regional socioeconomic forecast. UrbanSim is a land use microsimulation model that predicts the activities of individual agents (households, employers) within the region. Several sub-models control the number and distribution of regional households and population including the employment and household transition models, employment and household relocation models, household tenure choice model, and employment and household location choice models. UrbanSim is used to predict growth patterns from the present to the year 2050. The current base year of the UrbanSim model is 2020.

A synthetic population is developed for 2020, as described in the population synthesis section of this document, and UrbanSim sub-models evolve this population over the years. 2020 base year data for households and residential units are acquired from the 2020 Census Demographic and Housing Characteristics (DHC) file and the 5-year 2018-2022 American Community Survey (ACS). Detailed socioeconomic and demographic information only available in the ACS at the block group and tract level are scaled to match the block-level totals in the DHC. These data are used as controls to synthesize households and units using 5-year 2018-2022 American Community Survey Public Use Microdata Sample (PUMS) data at the 2020 Public Use Microdata Area (PUMA) level. Totals will align with the DHC at block level and the more detailed ACS socioeconomic and demographic data will have their distributions align with the ACS at the block group or tract level.

Household control totals from the regional socioeconomic forecast are provided to UrbanSim to ensure reasonable household and population outputs for each year after 2020. The regional socioeconomic forecast consists of a demographic model and an economic model. The demographic model provides population projections for 2020 to 2050 using the cohort-component method. Results are provided at the sub-regional level. There are four sub-regions: the seven county CMAP region, external Illinois modeling areas, external Indiana modeling areas, and external Wisconsin modeling areas.

More information about UrbanSim and the regional socioeconomic forecast, as well as a discussion of their sources, is available in the Socioeconomic Forecast Appendix.

A few additional pieces of information are needed by the travel demand model, including:

Regional median household income: To obtain an accurate estimate of median household income for the 21-county modeling area, the 2017-2021 5-year ACS PUMS) was used, providing the number of households within 16 distinct income bins. Data was obtained at the tract level for all of Illinois, Indiana, and Wisconsin and filtered to use only tracts that exist within the 21-county modeling area. Grouped frequencies of the number of households across income bins were used to calculate the median income, assuming that households are weighted equally within a bin. The median income from the 2021 ACS PUMS was inflation-adjusted to 2019 dollars using the Bureau of Labor Statistics' Consumer Price Index (All Goods). Using this technique, the 2019 regional median household income was estimated to be $72,800 (in 2019 dollars).

Private vehicle occupancy rate of all worker trips to work: This is the ratio between the number of workers traveling to work in a zone (all workers, not just those in households) who commute by household vehicles (single-occupant vehicles and carpool) divided by the number of vehicles used for those trips. Estimates of modeling area resident worker flows at the census tract level were taken from the 2017-2021 Census Transportation Planning Products (CTPP) - table B302103, means of transportation (18) (workers 16 years and over). For each flow, workers were summed across auto modes and number of vehicles was calculated as vehicles = sov + (hov2/2) + (hov3/3) + (hov4/4) + (hov5_or_6/5.5) + (hov7_or_more/7.1). Workers and vehicles were allocated to zones using the workplace tract and average vehicle occupancy to each workplace zone was calculated as stated above.

Group quarters population: The travel model requires information on residents of non-institutionalized group quarters, namely persons in military barracks, college or university dormitories, and other group quarters. These data are based on the 2020 Census SF-1 block-level table P5, "Group Quarters Population By Major Group Quarters Type." At the time of computation, combined group quarter population totals were not available by age breakouts at the block-level. The 2010 Census age distributions were used to generate age breakouts by creating a 2010 block group file with proportions for each age group. Block-level populations were geocoded to subzones using block centroids. Base-year (2019) estimates by subzone are based on each subzone's share of the 2010 total for each group quarters type, excluding dorm group quarters population, and are scaled to match 2019 regional estimates provided by the regional socioeconomic forecast. Dormitory group quarter data are also based on 2020 Census SF-1 block-level table P5, "Group Quarters Population By Major Group Quarters Type." Since these data do not require age breakouts, the census data were used directly. Three rates were calculated for each sub-region using 2020 Census data: the percentage of the total population that lives in group quarters, the percentage of group quarters population that are non-institutionalized, and the percentage of non-institutionalized group quarter population that live in dorms. These rates were applied to the total population by sub-region from the regional socioeconomic forecast to determine dormitory population totals by sub-region and growth rates across the forecast period. Growth rates for each sub-region were applied to base-year dormitory population in subzones within the respective region for the forecasting period.

Highway network

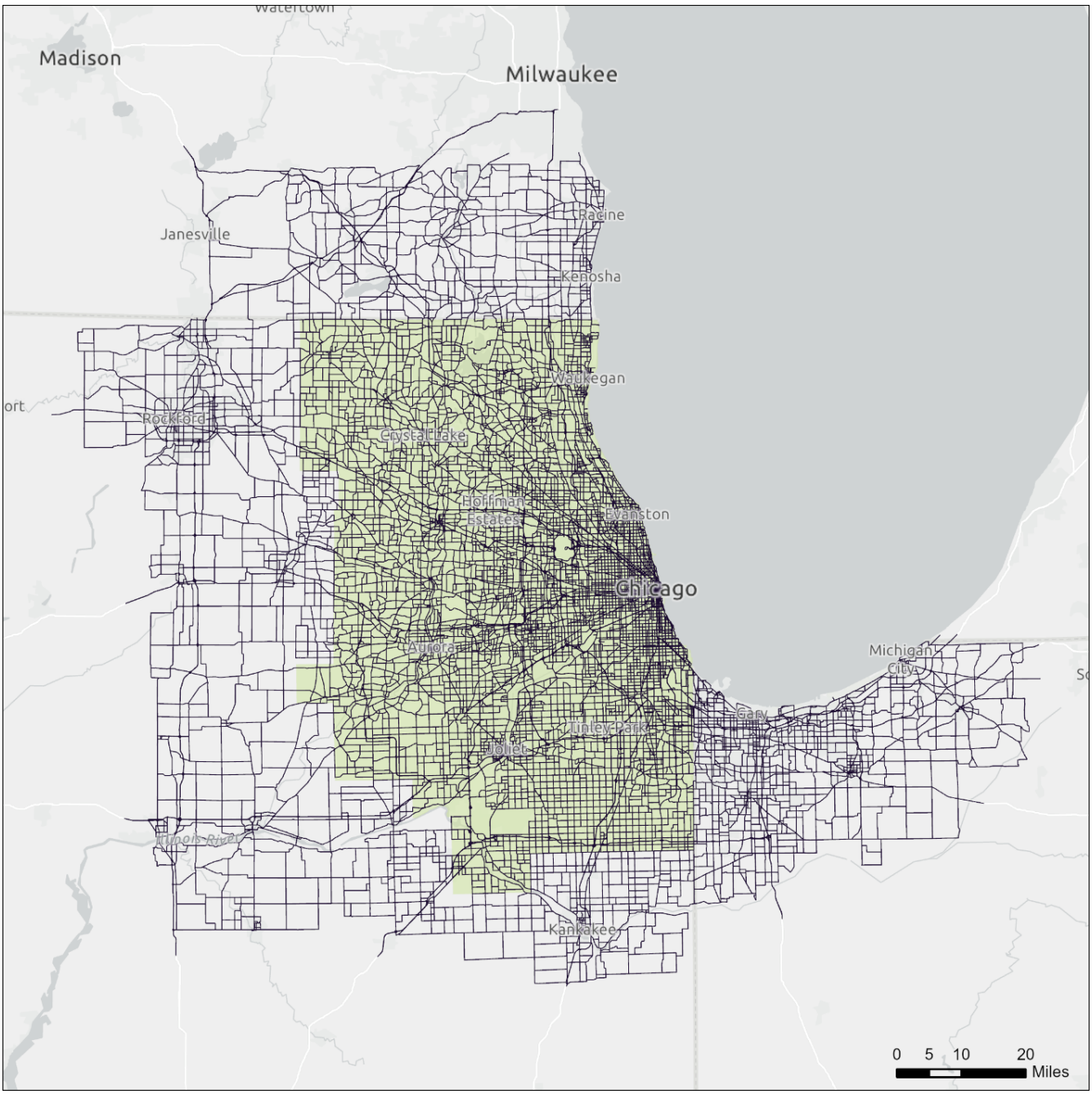
The Master Highway Network (MHN) is the official road network database used to develop travel demand model networks at CMAP. The MHN includes roads within northeastern Illinois that have a functional classification of minor collector or higher. In certain instances, additional local roads have been included in the MHN to provide connectivity within the network.

The MHN covers an area of more than 10,000 square miles and extends into northwestern Indiana and southeastern Wisconsin, as shown in Figure 2. The MHN includes roads for the following areas:

- **Illinois:** Twelve full counties (Boone, Cook, DeKalb, DuPage, Grundy, Kane, Kankakee, Kendall, Lake, McHenry, Will, and Winnebago) and three partial counties (LaSalle, Lee, and Ogle)
- **Indiana:** Three full counties (Lake, LaPorte, and Porter — corresponding to the Northwestern Indiana Regional Planning Commission's planning area)
- **Wisconsin:** Three full counties (Kenosha, Racine, and Walworth — the southern portion of the Southeastern Wisconsin Region Planning Commission's planning area) plus additional minimal road network extending into two other counties (Milwaukee and Rock)

The MHN is a collection of links and nodes representing road segments and intersections throughout the region. It contains information on more than 57,400 directional road segments and includes more than 21,200 nodes. The MHN is a comprehensive database for CMAP's regional travel demand modeling needs. It contains not only existing road segments and intersections, but also future planned facilities and improvements. The MHN itself is edited and maintained using Esri's ArcGIS Geographic Information System (GIS) software. The MHN has a current base year of 2015, meaning that the road attributes on existing facilities represent the conditions from that year. As other years are modeled, future project coding updates the network to make it consistent with the analysis year conditions. It is a relational database (specifically, an Esri file geodatabase) that maintains spatial and topologic relationships between features classes. The feature classes that define the road network in the MHN are lines and nodes. The projection of the data is State Plane Coordinate System, Illinois East zone, North American Datum of 1927. The unit of length is the U.S. survey foot.

Figure 2. CMAP Master Highway Network



Network line-node topology

MHN lines represent road segments located between intersections. Most lines in the MHN are digitized as bidirectional links with the appropriate direction-specific attributes coded to the link. Expressways are digitized as a set of parallel single-direction links to replicate their limited-access characteristics.

Table 1 lists the highway network link variables contained in the line attribute table that are relevant to CMAP's production modeling work. The attribute table also contains additional information not currently used for modeling purposes. As most links in the network represent bidirectional road segments, attributes must be included for each direction. Variable names ending in "1" describe attributes in the ANODE-BNODE direction of the link (this is the from-to direction of a link that is recognized by GIS software based on how the link was digitized). Those variable names ending in 2 represent attributes in the opposite direction. The directions variable indicates whether a link is a single or bidirectional segment and has three possible values, which determines how link attributes are coded:

1. A value of "1" indicates that the link represents a single direction of travel. No second direction variables are coded (i.e., they equal 0).
2. A value of "2" indicates that the link represents both directions of travel and all attributes are the same in both directions. No second direction variables require coding (except applicable parking restrictions).
3. A value of "3" indicates that the link represents both directions of travel and at least one attribute differs between the two directions. All second direction variables require explicit coding.

The BASELINK variable identifies whether a segment represents an existing facility (value of one) or a future facility (value of zero, referred to as skeleton links). If BASELINK=0, only ANODE, BNODE, MILES and DIRECTIONS are coded on highway links. All other link attributes are intuitively named. Each link in the MHN is identified by a unique ANODE-BNODE-BASELINK combination (variable ABB).

Table 1. Master Highway Network link attributes

Variable	Description
ANODE	Links "from" node
BNODE	Links "to" node
BASELINK	Link description flag: 0 = future project link ("skeleton" link), attributes added via highway project coding 1 = existing network link ("base" link), all attributes present
ABB	Unique arc ID, of the form "ANODE-BNODE-BASELINK"
MILES	Link length in miles
TYPE1 & 2	Facility type: 1=Arterial 2=Freeway (controlled-access) 3=Freeway-arterial ramp 4=Expressway (limited-access) 5=Freeway-freeway ramp 6=Centroid connector 7=Toll plaza 8=Metered ramp

TOLLDOLLARS	Toll amount in dollars for autos with I-PASS. If link type is 7 (toll plaza), this is applied as a fixed-cost toll; for other link types, it is applied as a per-mile rate.
AMPM1 & 2	Time period restrictions: 1=open all time periods (1-8) 2=open a.m. periods (2-5) only, e.g. Kennedy reversible lanes inbound 3=open p.m. periods (1, 6-8) only, e.g. Kennedy reversible lanes outbound 4=open off-peak periods (1, 5) only
SIGIC	Signal interconnect flag: 0=no, 1=yes
POSTEDSPEED1 & 2	Posted speed limit (mph)
THRULANES1 & 2	Number of driving lanes. (This represents the most-restrictive capacity on the link, i.e. the fewest number of lanes present at any single point.)
PARKLANES1 & 2	Number of on-street parking lanes
PARKRES1 & 2	Peak period parking restrictions, when on-street parking is not available and an extra through lane is available. *Coded separately for each direction on all 2-way links*. Code is text string of affected time periods (currently only 3 & 7). Default blank value means no peak period parking restrictions.
THRULANEWIDTH1 & 2	Average driving lane width (feet)
DIRECTIONS	Link directions flag: 1=one way 2=two way, attributes in both directions identical 3=two way, at least one attribute different in opposing direction
MODES	Modes permitted on link: 1=all vehicles 2=all vehicles (with truck restrictions from *TRUCKRES* applied) 3=trucks only 4=transit only (only called for transit networks) 5=HOV only
TRUCKRES	Detailed truck restriction codes, which translate into the following model coding: No restriction known/codeable: 0, 6, 15, 20, 22-24, 26, 28, 32, 33, or 36 No trucks: 1 or 18; also, 21 in time period 1 only No trucks except B-plates: 2-4, 9-11, 13, 25, 35, or 37; also, 12 in time period 1 only No medium or heavy trucks: 7, 8, 14, 16, 17, 19, 27, 29, 31, 34, 38-47, or 49 No heavy trucks: 5, 30, 45, 48
VCLEARANCE	Vertical clearance (inches). The following mode restrictions are applied for non-zero values: Clearance < 162”: no heavy trucks Clearance < 150”: no medium/heavy trucks Clearance < 138”: no light/medium/heavy trucks

Nodes in the MHN represent intersections between roads, junctions where road segments converge/diverge (such as an entrance ramp merging into an expressway through-lane), or where roadways otherwise significantly change (such as an interstate toll collection location). The line-node topology enforced in the MHN is such that nodes represent the end points of lines and lines with common end points are connected. CMAP’s modeling staff maintains a set of scripts that automatically update highway network topology after edits have been made and populate several attribute fields. These scripts (along with various other MHN processing scripts) are maintained in a publicly accessible GitHub repository at github.com/cmap-repos/mhn_programs.

Node attribute variables are listed in Table 2. These mostly serve to define the network lines by providing values for ANODE and BNODE. Values for the entire set of node variables listed are automatically populated through scripting. All node attributes are automatically populated by the scripts.

Table 2. Master Highway Network node variables

Variable	Description
NODE	CMAP network node ID
POINT_X	Auto-generated x-coordinate (NAD27 IL East State Plane feet)
POINT_Y	Auto-generated y-coordinate (NAD27 IL East State Plane feet)
SUBZONE17	Subzone ID from current CMAP modeling subzone system
ZONE17	Zone ID from current CMAP modeling zone system
CAPZONE17	2017 Capacity Zone code: 1=Chicago Central Business District (2009 subzones 1-47) 2=Remainder of Chicago Central Area (2009 subzones 48-80) 3=Remainder of City of Chicago (2009 subzones 81-976) 4=Inner ring suburbs where Chicago street grid is generally maintained 5=Remainder of Illinois portion of the Chicago Urbanized Area 6=Indiana portion of the Chicago Urbanized Area 7= Other Urbanized Areas and Urban Clusters within the CMAP Metropolitan Planning Area plus other Urbanized Areas in northeastern Illinois 8=Other Urbanized Areas and Urban Clusters in northwestern Indiana 9=Remainder of CMAP Metropolitan Planning Area 10=Remainder of Lake County, IN (rural) 11=External area 99=Points of Entry – not defined in the Capacity Zone system
IMAREA	Illinois Vehicle Inspection and Maintenance Program area flag: 0=no, 1=yes

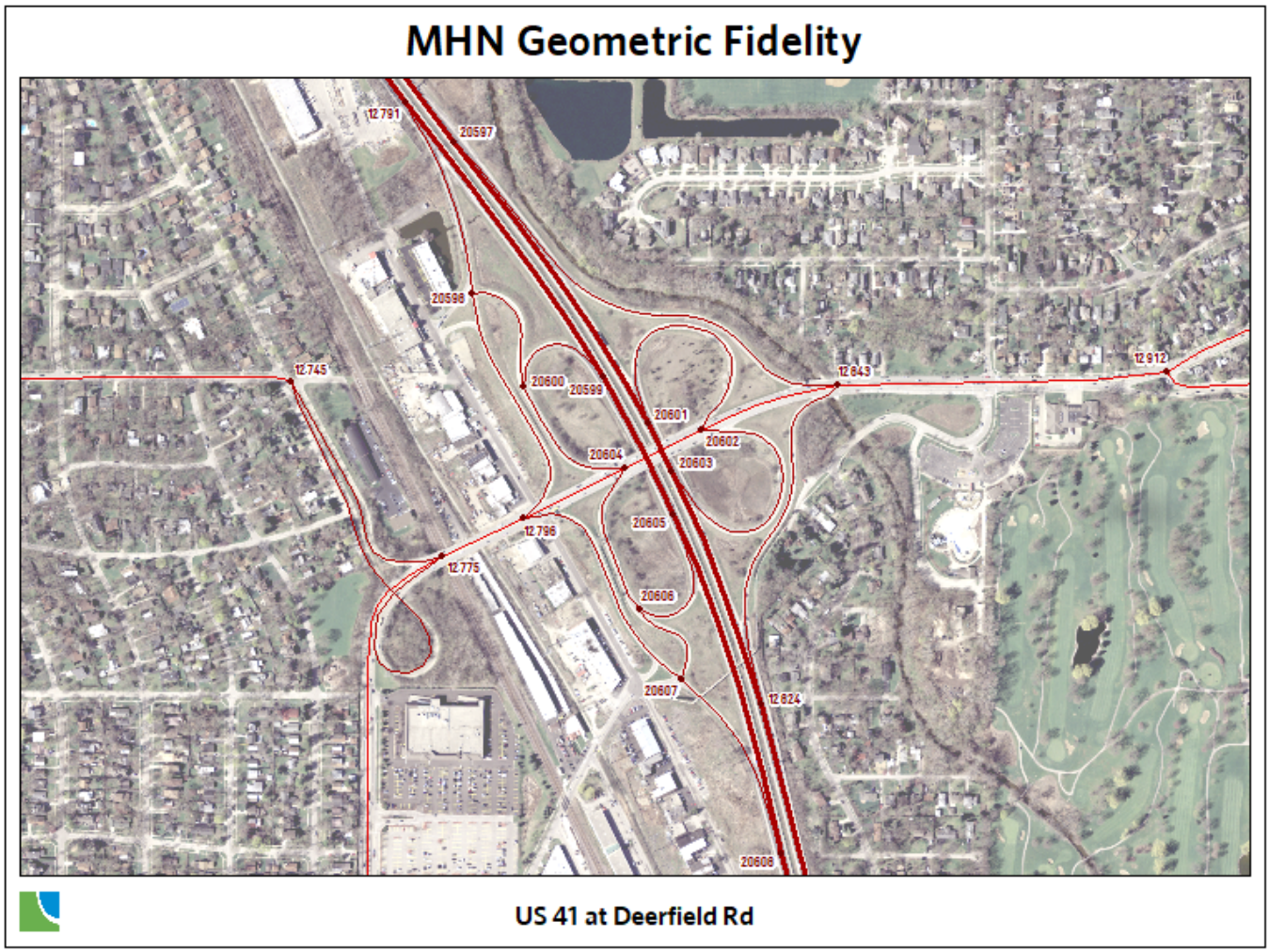
Spatial and geometric accuracy

CMAP ensures the MHN is spatially and geometrically accurate with respect to:

- Geocoding intersections to spatially accurate locations
- Applying vertex coordinates to links to replicate road geometry
- Ensuring expressway interchanges are fully expanded to include ramps representing all possible traffic movements

Figure 3 illustrates the network accuracy for one interchange in the MHN.

Figure 3. Master Highway Network accuracy



Highway project coding

The MHN includes links serving as placeholders for future planned facilities, identified by BASELINK=0. The Northeastern Illinois Transportation Improvement Program (TIP) database is the repository of project information for planned and programmed projects. Relying solely on the TIP database for project information provides a single direct link for reconciling model network coding with the planned improvement.

The MHN geodatabase stores highway project coding in a table containing detailed attributes and a line feature class containing the project ID, completion year and geographic extent. A group of network links is selected by their unique ABB values to define an individual highway project. The highway project line feature class contains summary information for each project, as shown in Table 3. Each project is represented by a single, multi-part line feature, which is automatically generated based on the links referenced in the detailed coding table. This data structure allows a single highway project to be associated with numerous MHN lines and, by extension, allows a single MHN line to be associated with multiple highway projects. The result is the project coding table contains one record for every line referenced by every highway project.

Table 3. Master Highway Network highway project line feature class attributes

Variable	Description
TIPID	TIP project identification number
COMPLETION_YEAR	Project completion year from the TIP. Inactive projects have a value of 9999.

The project coding table is used to store link attributes that will be updated or applied when the associated highway projects are completed. An action code determines how the attributes for each link are processed. The list of section table variables, shown in Table 4, most directly correspond to MHN line attributes. As with the line attribute table, variables ending in 1 apply to the "from-to" direction of the link, while those ending in 2 apply to the "to-from" direction. During network processing, data from the line attribute table are updated (overwritten) on the fly with project coding table entries to represent conditions after the project is implemented. Only those attributes changing due to project implementation are coded in the section table.

Project coding rules for parking lanes are slightly different than for other variables. The values for this attribute are added to (or subtracted from) line table coding to yield the final result. This allows for these attributes to be increased, decreased, or removed. This is necessary because there is no practical way to determine whether a zero in the section table represents no change in conditions or the removal of this attribute.

Four action codes control the link processing. Action code 1 modifies the coded attributes on links with existing attributes. Action code 4 is applied to new links (skeleton links), which have no attributes except MILES and DIRECTIONS. Action code 2 is used when new links replace an old link without any change in its attributes, such as when a new intersection is introduced into the network. This action code requires to fill in the REPLACE_ANODE and REPLACE_BNODE. These represent the nodes of the link where the attributes will be drawn from. Action code 3 deletes a link from the network.

Table 4. Master Highway Network highway project coding table fields

Variable	Description
TIPID	TIP project identification number
ABB	Reference to unique arc ID, of the form "ANODE-BNODE-BASELINK"
ACTION_CODE	CMAP action code: 1=modify (change an existing network link) 2=replace (replace an existing link with a new one but retain all attributes) 3=delete (remove a link from the network) 4=add (add a new link to the network)
NEW_TYPE1 & 2	New facility type number
ADD_SIGIC	Add signal interconnect to link (code=1)
NEW_THRULANEWIDTH 1 & 2	New average driving lane width (feet)
NEW_THRULANES1 & 2	New number of driving lanes
NEW_POSTEDSPEED1 & 2	New speed limit (mph)
REP_ANODE REP_BNODE	*ANODE* and *BNODE* of MHN link providing attributes, ONLY for action_code=2
NEW_TOLLDOLLARS	New I-PASS toll amount for autos (dollars). If link type is 7 (toll plaza), this is applied as a fixed-cost toll; for other link types, it is applied as a per-mile rate.

NEW_DIRECTIONS	New directions flag
ADD_PARKLANES1 & 2	Add/remove parking lanes, coded number will be added to number in MHN arc attributes to calculate final lanes (code positive to add, negative to remove)
NEW_AMPM1 & 2	New time period restrictions
NEW_MODES	New modes permitted
TOD	Time-of-day code indicating specific time periods when changes are applied. Default of blank or 0 means changes applied to all periods. Code is text string of affected time periods: 1=8 p.m. – 6 a.m. (overnight) 2=6 a.m. – 7 a.m. 3=7 a.m. – 9 a.m. (AM peak) 4=9 a.m. – 10 a.m. 5=10 a.m. – 2 p.m. (midday) 6=2 p.m. – 4 p.m. 7=4 p.m. – 6 p.m. (PM peak) 8=6 p.m. – 8 p.m.

Storing the existing and future highway network components in a single database allows the analyst to ensure project and base network reconciliation is handled comprehensively with all analysis networks for a particular application, at one step, existing in a single dataset. Storing the MHN in a GIS format also greatly simplifies project coding tasks. The MHN structure allows for:

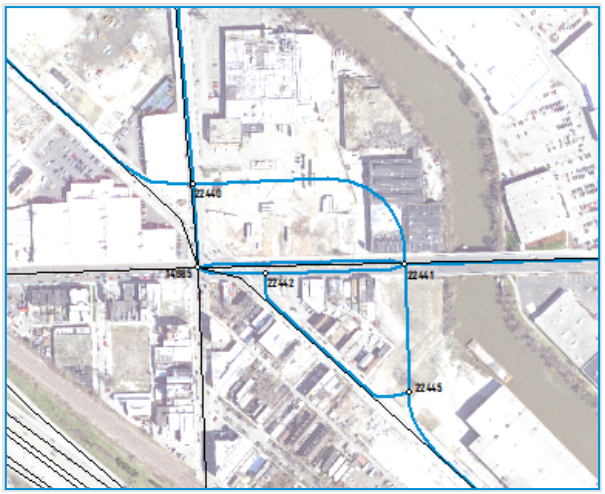
- **Analysis into multiple future years:** Assignable networks are produced that maintain consistent project coding into future years (e.g., a project that is built in an earlier year will be included in all subsequent networks).
- **Analysis across multiple scenarios:** Assignable networks are produced that maintain consistent project coding between differing analysis scenarios (e.g., a project that is included in one land use scenario will be identically coded in any other appropriate scenario).

This topology was in direct response to the types of comparative evaluations that were necessary under the air quality conformity baseline/action rules. With approval of a state implementation plan budget, conformity analysis no longer entails a baseline/action test, so a simpler hierarchy is used. Nonetheless, this ability is useful within any forecasting exercise where multiple time frames and scenarios are compared (e.g., land use/transportation interactions).

A list of modeled project TIP identification numbers and the year that they are to be constructed is all that is required to create a set of highway network files for the travel demand modeling software (Emme®). The completion year is attached to each project and stored in the highway project line feature class. As complete project coding information exists in the feature class and associated coding table, simple database queries can select only those records needed to prepare the desired analysis year network. A set of scripts process all the project coding information, apply the attribute updates to the set of links comprising the scenario network, and create a set of time-of-day link and node attribute files suitable for import into Emme. Figure 4 illustrates this process.

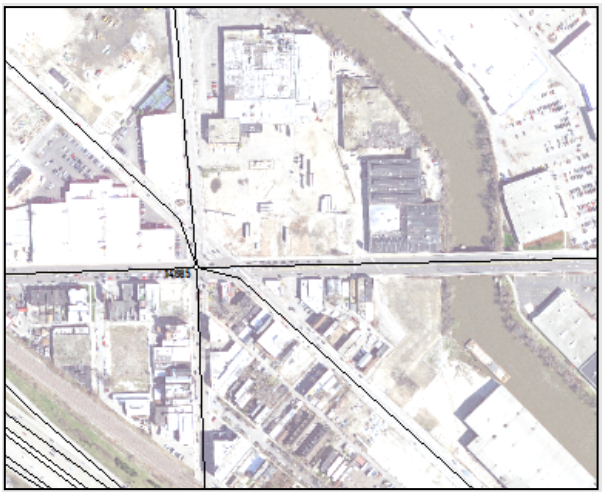
Figure 4. Highway network coding example

MHN network



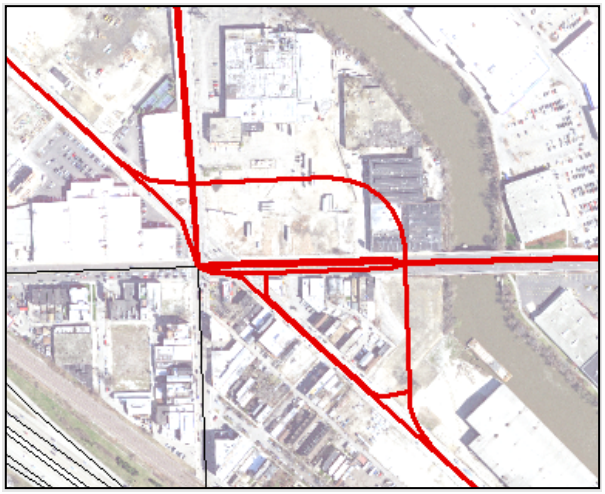
Step 1. All existing road facilities (black) and a planned realignment (blue) are coded into the MHN. The project will reroute the diagonal road northeast of the current six-way intersection and convert part of it to a service road.

Emme Network – base year scenario



Step 2. An Emme base year scenario network is created. Since the road realignment has not yet happened, it is not reflected in the model network.

MHN network project coding



Step 3. All links impacted by the project are associated with the project ID through a route system (red). The link coding contained in the route system's section table serves as a set of instructions on how to process each link. The coding identifies which links associated with this project to: (1) modify the characteristics of, (2) replace existing links, (3) delete from the network, and (4) add as new links.

Emme network – future year scenario

Step 4. A future scenario modeling network includes the realigned road and new intersections. Since this particular project was actually completed in 2018, it is displayed with current imagery showing the final result.

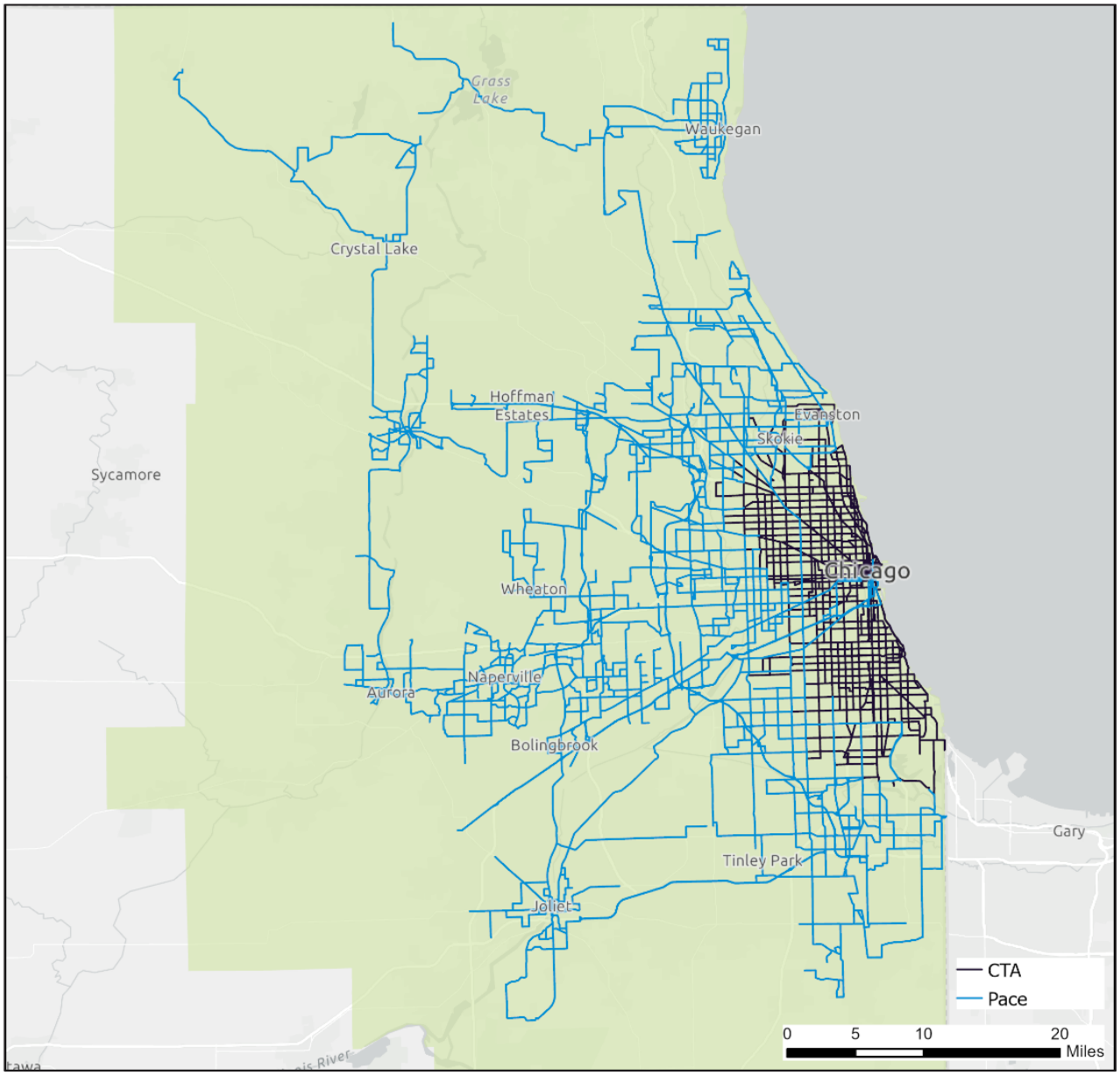
Bus route coding

Northeastern Illinois has one of the most extensive public transportation systems in North America. Bus and rail service is provided by three public operating agencies, including the Chicago Transit Authority (CTA), Metra commuter rail, and Pace suburban bus. Each of the three agencies has its own board, management, and operating personnel. The agencies' service areas overlap to varying degrees and many riders' trips involve transfers between services provided by different operators.

The CTA operates bus service within Chicago and several adjacent suburbs. Pace operates nearly exclusively in the suburbs, with some express service to downtown Chicago. Pace operates regular bus routes, feeder buses that provide connections to suburban Metra commuter rail and CTA rail stations, and all paratransit service in the region, as well as a vanpool program and some long-distance express buses. Bus route coding maintained in the model networks includes publicly operated, fixed-route service. It does not include vanpools, paratransit, or subscription service.

Like the highway project coding structure, existing bus service coding is maintained as two separate pairings of a route feature class and an itinerary table within the MHN. One pairing is for the base network, based on route data from 2015 (this is only used to reflect transit service in the 2019 model base year). Another is based on more current route data (currently December 2024), which forms the basis for most future model networks. The data structure ensures bus coding always reconciles with the underlying highway network links. Bus routes are forced to conform to the available MHN links. If a particular route uses local streets that are not included in the MHN, the coding for the route is altered accordingly so that it only uses MHN links. Figure 5 shows the extent of the region's bus service.

Figure 5. MHN general transit feed specification-based base bus coding, 2015



Bus coding in the MHN includes the complete itinerary (or node-by-node path) of the bus route and attributes associated with each itinerary segment. Bus routes are coded as single-direction runs. CMAP bus coding is derived from General Transit Feed Specification (GTFS) data files created by the transit operators. The GTFS files contain data on all runs of every bus during the entire week. As CMAP models weekday traffic, the bus coding data from the GTFS files are limited to a representative weekday (Wednesday).

Conversion of the GTFS data from its raw form into usable bus coding proceeds through the following steps:

- Python is used to transform source GTFS datasets for loading into an OpenPaths Emme project database.

- OpenPaths Emme API tools filter, load, and match schedule data to the model network, to import transit lines.
- Transit line geometry is extracted and loaded into an Esri file geodatabase for verification in ArcGIS Pro.
- ArcPy tools compare transit line geometry to shapes in the source GTFS datasets, to assess the network match quality. If necessary, adjustments are made to the match criteria and transit lines are imported again with the new parameters.
- OpenPaths Emme API tools transform and enrich transit line data for loading into the MHN and MRN geodatabases.
- Transit line data are extracted and loaded into the MHN and MRN geodatabases by custom geoprocessing tools that automatically generate the line feature classes and geodatabase tables that represent the route and itinerary for each scheduled service run.

As with the highway project coding, bus-run details are stored in a set of related itinerary tables. These tables relate to the link table in the same manner as the highway project coding tables. The bus-run-line features and itinerary tables are linked through the TRANSIT_LINE variable, which is a unique identifier given to each bus run.

Table 5 highlights the data fields maintained in the MHN that are used to describe bus route attributes. The variables shaded in gray correspond to header information Emme requires when reading in bus itineraries. Other attributes are merely informational. Note that at this point the HEADWAY value only represents the total minutes in the time-of-day period within which the bus run occurs. The actual bus service headway is calculated at a later stage, when individual runs are grouped into representative runs.

Table 5. Master Highway Network base and current bus run feature class attributes

Variable	Description
TRANSIT_LINE	Unique CMAP bus run identifier. (Mode + 5-digit number, starting at 00000 for base, 50000 for current.)
DESCRIPTION	Real-world description of bus route (format: "*ROUTE_ID LONGNAME*: *DIRECTION* TO *TERMINAL*")
MODE	Bus mode code: B=CTA regular service E=CTA express service P=Pace regular service Q=Pace express service L=Pace local service
VEHICLE_TYPE	Bus vehicle type code (based on mode code): 25-27=mode B 31-33=mode E 28=mode P 29=mode Q 30=mode L
HEADWAY	Length of the time-of-day period (in minutes) within which the bus run falls, since every run is represented individually
SPEED	Average bus route speed in MPH from GTFS data; minimum value of 15 allowed. [Not used in CMAP modeling but a non-zero value is required by Emme.]
FEEDLINE	Unique GTFS identifier for each run
ROUTE_ID	Bus number in route name (e.g. 52A, 112, X98)
LONGNAME	Proper name of bus in route name (e.g. Wentworth, Halsted/95th)
DIRECTION	Predominant direction of travel for bus run

TERMINAL	Location of final stop on bus run
START	Start time of bus run in seconds
STARTHOUR	Start hour of bus run

The actual itinerary information for bus routes is contained in the itinerary table variables, which are listed in Table 6. The itinerary provides the node-by-node path on the MHN that the bus follows. Again, most of these variables reflect information Emme expects to receive when bus routes are imported.

Two GTFS-based bus coding route systems exist simultaneously in the MHN database:

- **Bus_base:** GTFS–based coding that corresponds to CMAP's model base year of 2015 (reflecting service at that time).
- **Bus_current:** Coding built from the most recent GTFS data files. This represents up-to-date coding and is used as the basis for future modeling scenarios.

Each of the bus route systems listed above contains over 25,000 bus runs comprising roughly 750,000 itinerary segments, representing one weekday of service. For purposes of modeling travel demand, the bus runs are combined into representative bus routes. A script analyzes the runs of each bus route that occur during a time-of-day period and uses the stopping pattern to determine which are similar enough to be collapsed into a typical directional bus route. The start times of all the individual runs that are associated with a representative bus route are used to calculate the service headway that goes into the travel demand model.

Table 6. Master Highway Network base and current bus itinerary table fields

Variable	Description
TRANSIT_LINE	Unique CMAP bus run identifier. (Mode + 5-digit number, starting at 00000 for base, 50000 for current.)
ITIN_A	CMAP node number bus travels from
ITIN_B	CMAP node number bus travels to
ABB	Unique ID of the segment's corresponding MHN link
ITIN_ORDER	Order number of bus segment in itinerary, sequentially increasing from 1 for each run
LAYOVER	Layover time in minutes applied to *ITIN_B*. Default=3
DWELL_CODE	Code for stops (corresponding Emme code), applied to *ITIN_B*: 0=stop allowed (default time of 0.01 minutes) 1=no stop (#) *available for future use:* *2=alighting only (>)* *3=boarding only (<)* *4=boarding & alighting allowed (+)* *5=dwell time factor (*)*
ZONE_FARE	Incremental zone fare in cents
LINE_SERV_TIME	Itinerary segment travel time in minutes
TTF	Emme transit time function code: 0,1=1 2=2 (used for bus rapid transit/arterial rapid transit only)

DEP_TIME	Time departing node *ITIN_A* (seconds since midnight)
ARR_TIME	Time arriving at node *ITIN_B* (seconds since midnight)
LINK_STOPS	Number of stop locations from GTFS data that were combined into itinerary segment
IMPUTED	Flag indicating segment was imputed by shortest path algorithm during import: 0=not applicable. 1=itinerary segment created by shortest path algorithm 2=segment modified by logic to condense unreasonable vacillation in itinerary.

Table 7 lists the eight time-of-day (TOD) modeling periods used by CMAP for traffic assignment. It also includes the selection rules used to determine which TOD period a particular bus run falls in, and it shows the number of representative bus routes (from bus_current) used in the travel demand model. This information is used to provide background bus volumes on roadway links during traffic assignment.

For transit assignment, only four TOD periods are used: 1=Overnight/early AM (6 p.m. – 6 a.m.), 2=AM peak (6 a.m. – 9 a.m.), 3=Midday (9 a.m. – 4 p.m.) and 4=PM peak (4 p.m. – 6 p.m.). The AM peak and midday periods in this scheme are used to generate the transit level-of-service variables used to develop zonal generalized costs for the destination choice-mode choice model.

Table 7. Time-of-day base bus routes

Time period	Selection rule	Number of bus routes
1 – (8 p.m. – 6 a.m.)	*Starthour* >=20 or *Starthour* <=5	606
2 – (6 a.m. – 7 a.m.)	*Starthour* =6	631
3 – (7 a.m. – 9 a.m.)	7<= *Starthour* <=8	698
4 – (9 a.m. – 10 a.m.)	*Starthour* =9	497
5 – (10 a.m. – 2 p.m.)	10<= *Starthour* <=13	503
6 – (2 p.m. – 4 p.m.)	14<= *Starthour* <=15	645
7 – (4 p.m.– 6 p.m.)	16<= *Starthour* <=17	653
8 – (6 p.m. – 8 p.m.)	18<= *Starthour* <=19	563

Future bus coding

While the GTFS data provide for existing bus service, future bus routes also are coded in the MHN to represent planned or programmed service. Consistent with GTFS bus runs, future bus routes are generally coded as single-direction service. While GTFS routes are coded on existing network links (BASELINK=1), future routes are coded to run on future highway network links (i.e., links that will be in the network in the horizon year of the projects being modeled). Thus, it is convenient to have the highway project coding information stored in the same database as the bus coding.

Table 8 lists the future bus route information stored in the line-feature-classes-attribute table. Many of these variables provide the bus route information Emme requires to build transit routes and have the same definition as in the GTFS-based bus coding table. The notes field is used to store TIP project numbers or other useful information related to future bus service. The following variables provide instructions on how the future routes are processed:

- **Scenario:** Identifies all the specific modeling scenarios that individual bus routes should be included in

- **Replace:** Identifies the existing GTFS route(s) that will be replaced by the future route coding, if any
- **Reroute:** Identifies the existing GTFS route(s) that will be modified by the future route coding, if any
- **TOD:** Indicates which time-of-day networks will include the future bus service

Table 8. MHN future bus route attributes

Variable	Description
TRANSIT_LINE	Unique CMAP bus route identifier (Mode + 5-digit number, starting at 99000 for future)
DESCRIPTION	Real-world description of bus route
MODE	Bus mode code
VEHICLE_TYPE	Bus vehicle type code
HEADWAY	Average bus headway in minutes. This coded value will only be applied during time-of-day periods specified in TOD attribute, below. *A value of zero indicates that the headway for the existing route coding from route flagged in REPLACE will be used (i.e., there will be no change in service frequency).*
SPEED	Average bus route speed in MPH; default value of 12 used.[Not used in CMAP modeling but a non-zero value is required by Emme]
SCENARIO	Future scenarios bus line will be used in. Must include ALL scenarios that will contain route. May NOT be blank.
COMPLETION_YEAR	Completion year in which bus line will be implemented. May NOT be blank.
REPLACE	Identifier of the existing bus route coding that will be replaced by the future project. A blank indicates the future route is not replacing existing routes. Constructed as: Uppercase mode letter + "-" + route number. If multiple existing routes are being replaced by the same future route, separate the affected route numbers with colons. *Note: the replacement will only occur in the time periods identified in the TOD field.*
REROUTE	Identifier of the existing bus route coding that will be modified by the future project. A blank indicates the future route coding is not modifying existing routes. Constructed as: Uppercase mode letter + "-" + route number. If multiple existing routes are being modified by the same future route, separate the affected route numbers with colons. *Note: the replacement will only occur in the time periods identified in the TOD field.*
TOD	Transit assignment time-of-day periods when the new coding will be implemented. As with SCENARIO, all applicable time periods must be listed. *A value of zero indicates that the new coding will be applied to all time periods.*
NOTES	TIP ID number (and possibly other descriptive information). Entries must be separated by colons. 30 character limit.

As with the highway project coding, CMAP modeling staff maintain a set of scripts that generate scenario-specific TOD transit network input files for the travel demand model. While processing of the base/current year bus routes is relatively straight-forward, the future bus routes require additional logic to process all changes correctly. The following rules are used to generate future scenario TOD transit network input files:

- New routes only appear in the specific time periods identified in the TOD field, or in all periods if TOD=0.
- The REPLACE field identifies current bus routes that will be replaced by the future coding for the time periods in TOD. These current routes are deleted from the network. The routes are identified by the letter MODE code and the Route_id from the route table.

- The REROUTE field identifies current bus routes that will be modified by the future coding for the time periods in TOD. These current route itineraries are replaced with the future itinerary coding between the start and end points of the future itinerary coding. All other information for the modified routes remains the same. The routes are identified by the letter MODE code and the Route_id from the route table.

Several potential values are used to determine future headways:

- **Coded headway:** This is the headway coded for future bus routes. It applies only to AM and PM peak time periods.
- **Factored headway:** The value is coded headway (if greater than zero) times a TOD multiplier [3 for period 3; 4 for period 1].
- **Replaced headway:** The TOD headways for the bus routes being replaced by future service (this value is calculated when the bus runs are collapsed into representative routes). To avoid having to apply directional headways to the future service, this value is the minimum of the current directional headways.
- **Mode headway:** The average headway for a time period for a given bus mode (B,E,P,L,Q) based on the existing bus routes, excluding those being replaced.
- **Last chance headway:** A final future headway value if all other options fail; set to 90 minutes.

The logic used to determine the final TOD future headway for each future route is:

- **Priority 1:** If replaced headway is nonzero and is less than factored headway, use replaced headway. If that does not apply, go to Priority 2.
- **Priority 2:** If factored headway is nonzero, use factored headway. If that does not apply, go to Priority 3.
- **Priority 3:** The future headway is the maximum of [mode headway, last chance headway].

The outcome is that the future headway for non-peak periods cannot be less than the headway in the peak periods. Regardless of priority, headway is capped by the length of the time period the service is operating in.

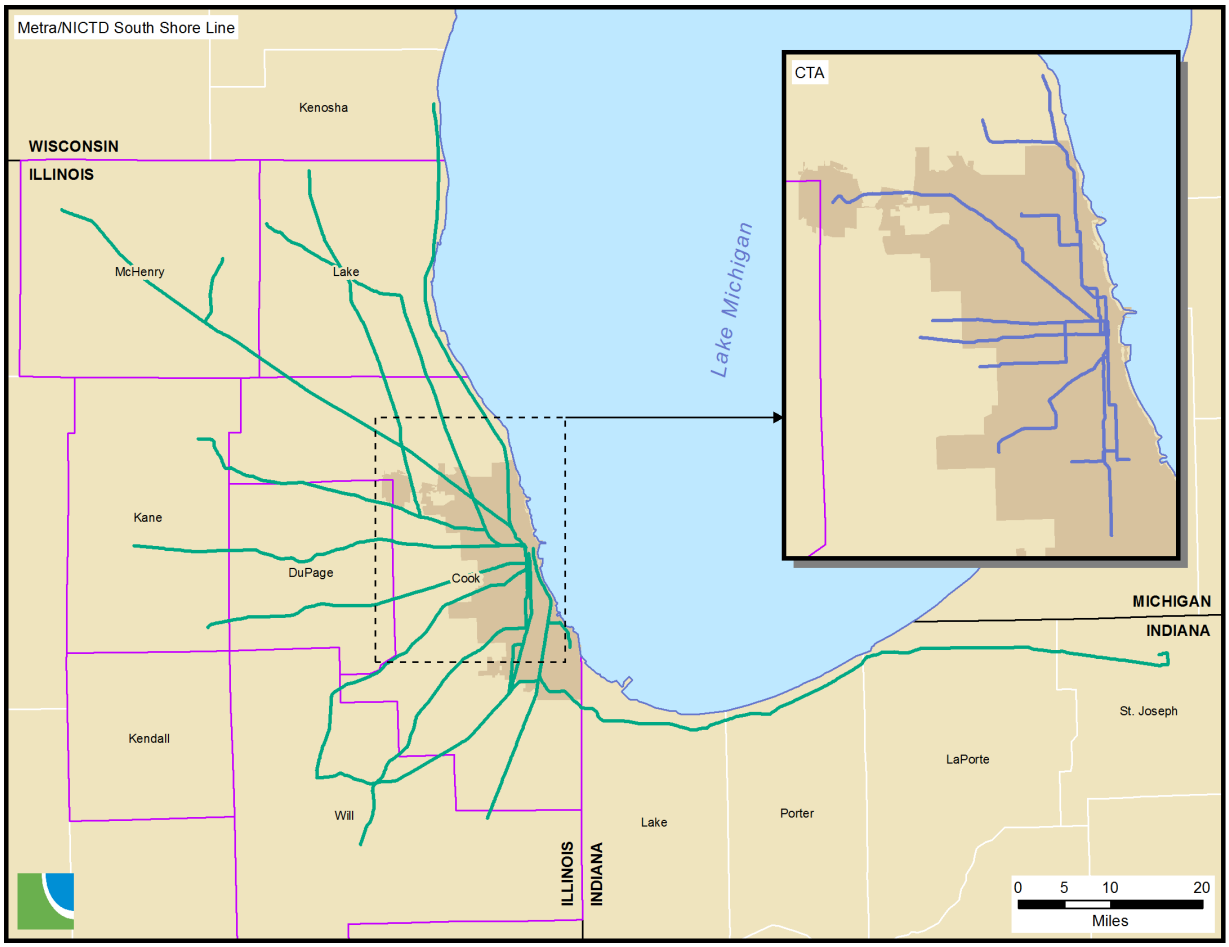
Rail network

The Master Rail Network (MRN) is stored in a file geodatabase, which contains all rail segments representing heavy and commuter rail service in northeastern Illinois. The geodatabase stores all feature class data needed to build rail transit networks for regional analyses, including links, nodes, rail routes and itineraries. While it would be possible to combine the MHN and MRN into one relational database, the rail network has few link attributes compared to the highway network, and it is much smaller in scope than the MHN. Thus, it is easier to maintain them separately. The MRN is stored in the same projection as the highway network database, so that they work together

seamlessly. Like the MHN, the MRN is maintained with a set of scripts stored in a publicly accessible GitHub repository.

In addition to bus service, the CTA operates heavy rail transit within the city of Chicago and several adjacent suburbs. Suburban commuters are served by Metra's radial rail services oriented between suburban areas and the central area. There are numerous Metra stations within the city of Chicago, and some Metra lines run parallel to CTA rail lines. Additionally, the Northern Indiana Commuter Transportation District operates commuter rail service between downtown Chicago and South Bend, Indiana. The extent of the MRN is illustrated in Figure 6.

Figure 6. Master Rail Network



Topology rules within the MRN are enforced programmatically by rebuilding the routes each time the network is updated or new routes are imported using the line geometry. In addition to being an efficient way to process the data, this procedure also ensures rail routes will always be coincident with the underlying lines (which is necessary for selecting scenario-specific route coding).

Link fields

Table 9 lists the rail network link variables contained in the line table and are the variables required by the travel demand modeling software. Most of the lines represent the rail line segments that connect stations (either mode "C" for CTA rail or mode "M" for Metra). Two other kinds of links are

included in the MRN on a limited basis — transfer links connecting different rail service and walk links providing access or egress to the service. These auxiliary links are discussed in detail later.

Table 9. Master Rail Network link attributes

Variable	Description
ANODE	CMAP "From" node
BNODE	CMAP "To" node
MILES	Link length in miles
MODES1	Modes permitted on ANODE-BNODE direction of link (string of mode letters)
MODES2	Modes permitted on BNODE-ANODE direction of link (string of mode letters); blank if link is only 1 direction
DIRECTIONS	Link directions flag: 1=one-way, 2=two-way

Node fields

Node variables are listed in Table 10. The nodes represent rail stations or, in a few instances, rail line junctions that are not actual stations. Rail service does not stop at junction locations. In addition to a unique identifier for each station, the node table contains information on the availability of parking at the rail stations. The value PSPACE indicates the number of parking spaces available at the station and the cost of parking is stored in PCOST. Both values represent conditions in the base year of the MRN. Thus, PSPACE equals zero if no parking is available and PCOST equals zero if there is no fee.

Table 10. Master Rail Network node attributes

Variable	Description
NODE	CMAP node number; used to assign ANODE and BNODE values in link table CTA rail: uses range 30000-39999 Metra: uses range 40000-49999
LABEL	Node label (4-character station name)
PSPACE	Number of parking spaces at node in base scenario; zero if not used in base scenario
PCOST	Parking cost at node in base scenario; zero if not used in base scenario
FTR_PSPACE	Number of parking spaces at node in future scenarios. This text string uses the format "s1:p1:s2:p2: ..." where s=the hundred's place value from CMAP's scenario numbering scheme and p=the number of parking spaces in the corresponding scenario. Each value must be separated by a colon.
FTR_PCOST	Parking cost at node in future scenarios; same format as *ftr_pspace*

Future scenario parking information is also stored in the node attribute table in variables FTR_PSPACE and FTR_PCOST. A value of "4:150:6:200" for FTR_PSPACE is interpreted as follows: The node will have 150 parking spaces, beginning in scenario 400 and 200 parking spaces beginning in scenario 600. The value is assigned through scenarios until a later scenario is specified, so the node will have 150 parking spaces in scenario 500, as well. While it is easier for the analyst to read this value if the scenarios are coded in chronological order, the processing programs do not require this to assign the correct value to each scenario. Values for FTR_PCOST are coded using the same format.

Rail route coding

As with bus coding stored in the MHN, rail service coding in the MRN is built from GTFS data files for a representative weekday (Wednesday). Each individual run of every rail line is stored in the database, representing a single direction of travel. Processing of the GTFS rail service data into usable model coding follows the same set of procedures and logical reviews as bus route data. Rail service coding is stored in the geodatabase as a pair of related tables containing information on the rail run and its itinerary.

Table 11 lists the variables in the rail route table. As with bus route coding, the variables are a combination of header fields that Emme requires when reading in rail itineraries (highlighted in gray) and GTFS fields maintained for clarity. These variables have the same definition in both the bus and rail route tables, although the values may differ. Special attention is given to the transit line name variable (TR_LINE) in the rail coding.

Table 11. Master Rail Network rail route attributes

Variable	Description
TR_LINE	Unique CMAP rail run identifier (6 characters)
DESCRIPTION	Real-world description of rail run (20 characters maximum – limit imposed by Emme)
MODE	Rail mode code: C=CTA rail M=Metra/NICTD
VEHICLE_TYPE	Rail vehicle type code (based on mode code): 1-5=mode C 6-24=mode M
HEADWAY	Rail headway (in minutes). A value of 99 indicates headway will be set to the length of the time-of-day period within which the rail run falls. In future routes, headway values can be different if headway changes throughout day. In this case the value is in a colon delimited format with the TOD periods followed by the effective headway for those periods. (e.g. 234678am:6.7:5:11.3:1:19.2)
SPEED	Average rail route speed in MPH from GTFS data; minimum value of 15 allowed. [Not used in CMAP modeling but a non-zero value is required by Emme]
FEEDLINE	Unique GTFS identifier for each run
ROUTE_ID	For CTA: lists line (Blue, Red, etc.) For Metra: lists train run number.
LONGNAME	Lists the proper name of the train line (Blue Line, Union Pacific Northwest, etc.)
DIRECTION	Lists final stop on train run
START	Start time of train run in seconds
STARTHOUR	Start hour of train run

To allow for simple identification of runs on a specific rail line, the following rail route naming scheme is applied to the six-character TR_LINE variable.

- **First character:** lowercase MODE letter.
- **Second and third characters:** two letter line identifier (lowercase).
- **Fourth through sixth characters:** unique counter for each MODE-line combination, starting with “001” for base runs and “401” for current runs (automatically generated).

The three-character rail line coding prefixes are summarized in Table 12.

Table 12. Rail line coding prefixes

Transit agency	Line	Coding prefix
CTA	Blue	cbl
	Brown	cbr
	Green (Ashland branch)	cga
	Green (Cottage Grove branch)	cgc
	Orange	cor
	Pink	cpk
	Purple	cpr
	Red	crd
	Yellow	cye
Metra	BNSF	mbn
	Heritage Corridor	mhc
	Metra Electric	mme
	Milwaukee District North	mmn
	Milwaukee District West	mmw
	North Central Service	mnc
	Union Pacific Northwest	mnw
	Rock Island District	mri
	SouthWest Service	msw
	Union Pacific North	mun
	Union Pacific West	muw
NICTD	South Shore	mss

The itinerary information for rail lines is stored in a related data table, and the contents are listed in Table 13. These variables provide the same information as their counterparts in the bus itinerary tables. One variable of interest is the zone fare variable (ZN_FARE). This value applies to commuter rail lines, and it is the marginal cost per ride (in cents) for traveling between fare zones. It is calculated as:

> [the difference between monthly pass costs from station zone to zone A] x [100 cents] / [40 (the average number of one-way rides for a monthly pass holder, assuming 20 workdays per month].

For example, a Union Pacific North (UP-N) line monthly pass from Kenilworth (zone D) to Ogilvie Transportation Center (zone A) is $181.25. A monthly pass from Wilmette (zone C) to Ogilvie is $159.50. The ZN_FARE on the link between Kenilworth and Wilmette is: [(181.25 - 159.50) x (100)]/40 = 54.38.

As with the bus coding, two GTFS-based rail coding route systems exist simultaneously in the geodatabase:

- **all_runs_base:** GTFS–based coding that corresponds to a base year of 2015 (reflecting service at that time)
- **all_runs:** Coding built from the most recent GTFS data files. This represents up-to-date coding, and it is used as the basis for future modeling scenarios.

Each of the rail route systems listed above contain more than 2,500 rail runs, made up of more than 76,000 itinerary segments, representing one weekday of service. For travel demand modeling purposes, the CTA rail runs are combined into representative rail routes. This is accomplished using the same script and collapsing logic that is applied to the bus run coding. All individual commuter rail runs are allowed to pass through to the travel demand model unchanged.

Table 13. Master Rail Network itinerary attributes

Variable	Description
TR_LINE	Unique CMAP rail line identifier (6 characters)
ITIN_A	CMAP node number of first node of link rail run travels on
ITIN_B	CMAP node number of second node of link rail run travels on
IT_ORDER	Order number of rail segment in itinerary
LAYOVER	Layover time in minutes applied to ITIN_B. Default=3
DWELL_CODE	Code for stops (corresponding Emme code), applied to ITIN_B: 0=stop allowed (default time of 0.01 minutes) 1=no stop (#) *available for future use:* *2=alighting only (>)* *3=boarding only (<)* *4=boarding & alighting allowed (+)* *5=dwell time factor (*)*
ZN_FARE	Incremental zone fare in cents
TRV_TIME	Itinerary segment travel time in minutes
DEP_TIME	Departure time at beginning of segment from GTFS data (in seconds)
ARR_TIME	Arrival time at end of segment from GTFS data (in seconds)
IMPUTED	Flag indicating segment was imputed by shortest path algorithm during import: 0=not applicable 1=itinerary segment created by shortest path algorithm

Future rail coding

As with bus coding, there is a need to store future rail project information for use by the travel demand model. The route coding table for future rail service includes the same set of fields as the existing coding, as well as a few additional fields shown in Table 14. The TOD variable identifies all specific time-of-day periods that individual future rail routes should be included in. The SCENARIO variable identifies all specific modeling scenarios that individual rail routes should be included in. The NOTES variable contains other descriptive information about the project. The ACTION variable requires a more detailed explanation.

Table 14. Future rail route additional attributes

Variable	Description
TOD	Time-of-day periods that rail route will be used in. A string of ALL TOD periods (1–4) that will contain the route. Applies to new service routes (ACTION=1) only.
SCENARIO	Future scenarios rail line will be used in. A string of ALL scenarios (first digit of three-digit code) that will contain route. May NOT be blank.
ACTION	Action code for the route
TIP_ID	TIP ID number
COMPLETION_YEAR	Year in which the project is to be completed
RSP_ID	ID for regionally significant projects

NOTES	Other descriptive information. Entries must be separated by colons. 30 character limit.
RCPnoBuild	Indicates if the route is included in the RCP No-build scenario
RCP_ID	ID for Regional Capital Projects

Implementation of the GTFS-based rail coding required a reimagining of how future rail project coding would be handled. The desired outcome was to maintain the simple, spreadsheet-based future service coding procedures CMAP has used for years. To achieve this, an ACTION code variable was added to the required attributes in the future route table. This value describes what type of service or improvement is being implemented and instructs the processing scripts on how to handle the data input. A brief description of the future rail ACTION codes is provided in Table 15.

Table 15. Future rail coding action codes

Action code	Meaning	Discussion
1	New line or service	The entire itinerary must be coded
2	Travel time reduction on selected links	ITIN_A and ITIN_B define the nodes between which the travel times will be reduced (only the end points need to be coded). Code both directions of travel if applicable. The TRV_TIME value represents the % reduction applied to the base year travel time (for instance 0.1 means a 10% reduction). To apply the travel time savings to the entire itinerary, code ITIN_A and ITIN_B as the beginning and ending nodes on the line and code LAYOVER=99. If the time reduction applies to multiple lines, each must be coded separately.
3	New station	ITIN_A and ITIN_B define the nodes between which the new station will be inserted. Store node number of new station in LAYOVER. Code both travel directions if applicable.
4	Line extension	Itinerary only contains the coding for the additional segments to be added to the base year runs. Code both directions of additional segments if applicable. For extension at ending terminal, code IT_ORDER values beginning with 1001. For extension at beginning station, use negative values.
5	Shift to different downtown station	The SWS will switch from Union Station to LaSalle Street Station – essentially this is just swapping one link for another at the CBD end of the itinerary. In itinerary coding: ITIN_A and ITIN_B are the original nodes defining the link. LAYOVER holds the new node number. IT_ORDER is used to identify which node is being replaced by the value in LAYOVER: 1=ITIN_A and 2=ITIN_B. Code both travel directions if applicable.
6	Placeholder for TIP identification number	When a future project reaches its time horizon and is implemented, it gets coded into the current routes. At this point the project could be removed from the future routes, except it is necessary to maintain the project's TIP ID in the future routes. To do this without affecting the network, the project is coded like an action code 2 with a 0% travel time reduction.
7	New consolidated rail station	Two existing rail stations are being consolidated and replaced with a new station. In itinerary coding: ITIN_A and ITIN_B are the remaining nodes between which the new station will be inserted LAYOVER holds the new station node number Code both travel directions if applicable

When the future rail coding is processed to create scenario transit networks, new lines/service (ACTION=1) are added to the set of existing ones to increase the total. For ACTION codes 2-5 and 7, the changes described in the coding are applied to the existing transit routes (no actual processing is performed for ACTION=6). Additionally, modified unique counters (characters four through six in TR_LINE) are used for future rail lines:

For ACTION code 1: the counter should be a 900 series (i.e., starting with “9” followed by two digits beginning with “01”).

For ACTION codes 2-7: the counter should start with two asterisks (one if the counter requires two digits) followed by a counter. For example, mri**1 identifies the coding as a general improvement that will apply to all runs on the specified line.

Future rail itinerary coding contains the same fields as the existing itinerary coding, except for the GTFS-derived fields. The use of the ACTION code allows for a great deal of flexibility in coding the itineraries. CMAP staff uses this flexibility to rely upon one future rail coding template where the definitions of the itinerary fields are dependent upon the ACTION code applied to the specific route. The benefit to the analyst of using this coding scheme is only minimal future rail coding input is required to implement the desired changes. Processing scripts perform all the painstaking work.

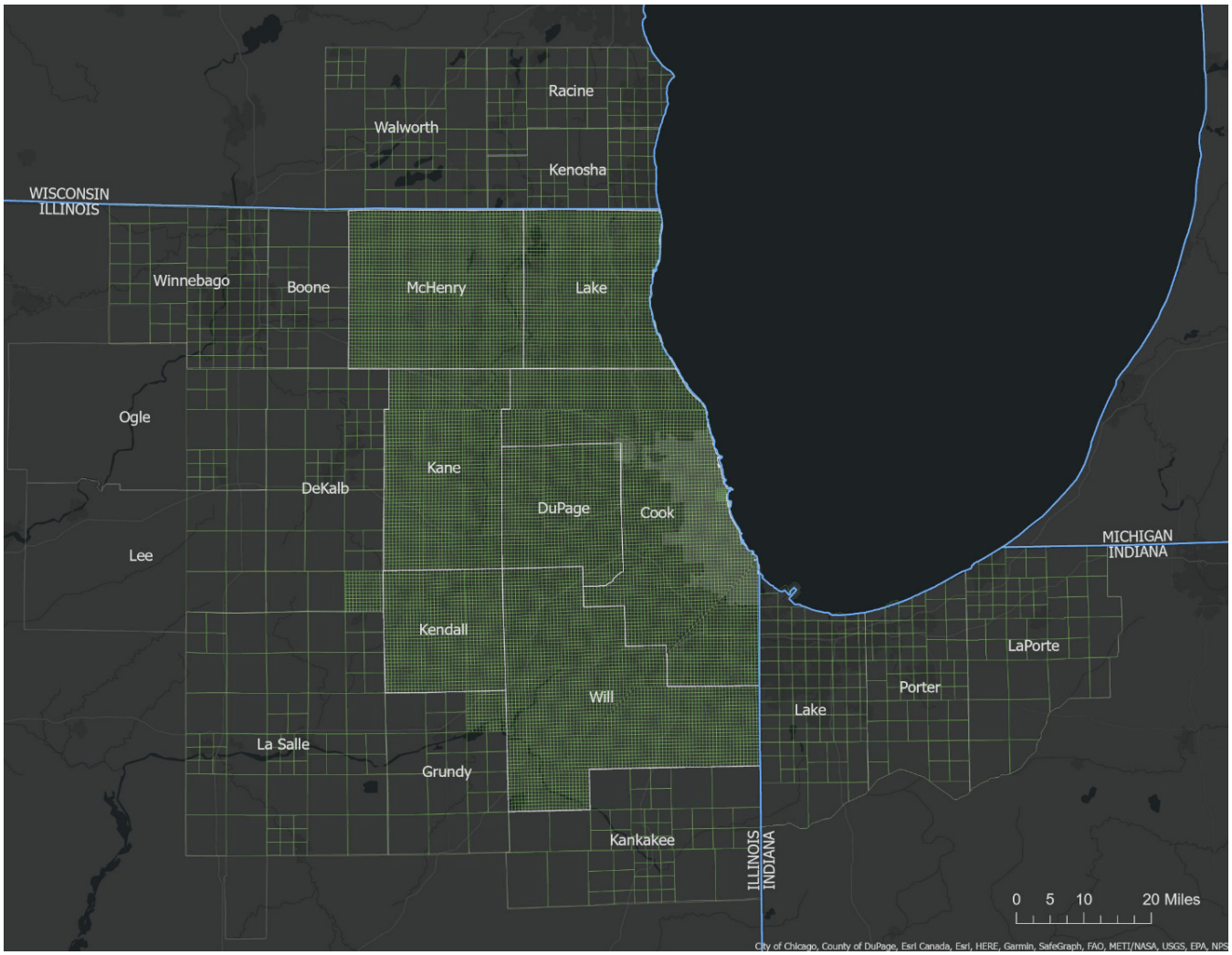
Zone systems

Three different zone systems are used in the regional travel demand model. The zone systems all serve different purposes within the regional travel demand model.

Trip generation zones

Trip generation zones (or subzones) are the smallest level of geography used in the travel demand model. Subzones are quarter-section-sized geographies that CMAP uses for household and employment forecasting. The current edition of the subzones is known as "Subzone17" (identifying the year in which it was developed). This most recent improvement to the subzones included adding a few hundred more subzones by replacing previous Census-based subzone boundaries with boundaries based on the Public Lands Survey Systems (PLSS). The CMAP modeling region comprises 17,418 subzones (shown in Figure 7).

Figure 7. CMAP trip generation zones



Quarter-sections are based on the PLSS subdivision of land into township and range, and then into sections. Two major benefits of using this system as the basis for the subzones are:

- The geometry does not change (unlike Census-based geography).
- The PLSS sections conform in most cases to state, county, and township boundaries (unlike other referencing systems, such as the U.S. National Grid).

As indicated by their name, the trip generation zones are used to aggregate socioeconomic data into geographic units suitable for providing input to the trip generation model to generate trip productions and attractions. Trip generation zones serve as the base level for CMAP's modeling zone systems. The two remaining zone systems are created by aggregating the subzones into larger geographies. Thus, the subzones always nest perfectly within the other zone systems.

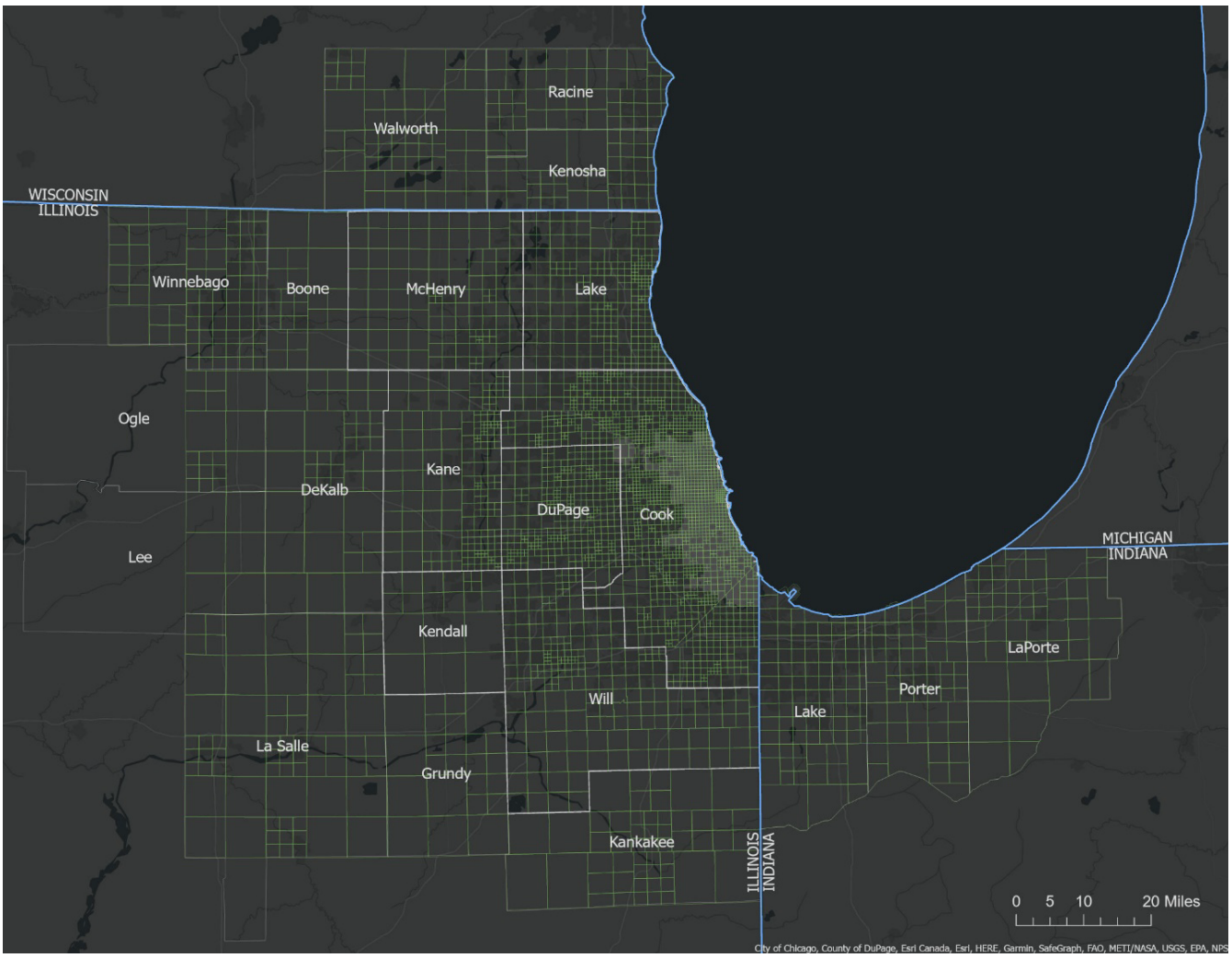
Modeling zones

While the trip productions and attractions are generated in a zone system based on survey quarter-sections, this level of detail is not used for the remaining modeling processes. At this time, the space and computing capabilities required to complete calculations on matrices composed of

over 17,000 trip generation zone origins and destinations (more than 300 million values) is prohibitive. Therefore, the subzones are aggregated into the CMAP modeling zone system for the remaining steps of the modeling process —destination choice-mode choice, and assignment.

Figure 8 shows the 3,632 modeling zones for the CMAP region. These zones generally follow the survey township geography. Zones are either sections (approximately one square mile) or regular subdivisions of townships (4-square-mile ninths of townships, 9-square-mile quarters of townships, or whole townships). The modeling zones are equivalent to their underlying subzones outside of the CMAP planning area. Additionally, there are 17 external zones, or points of entry, that are not shown. These are arrayed around the outside of the pictured zone system, representing trips on major highways entering the region.

Figure 8. CMAP modeling zones

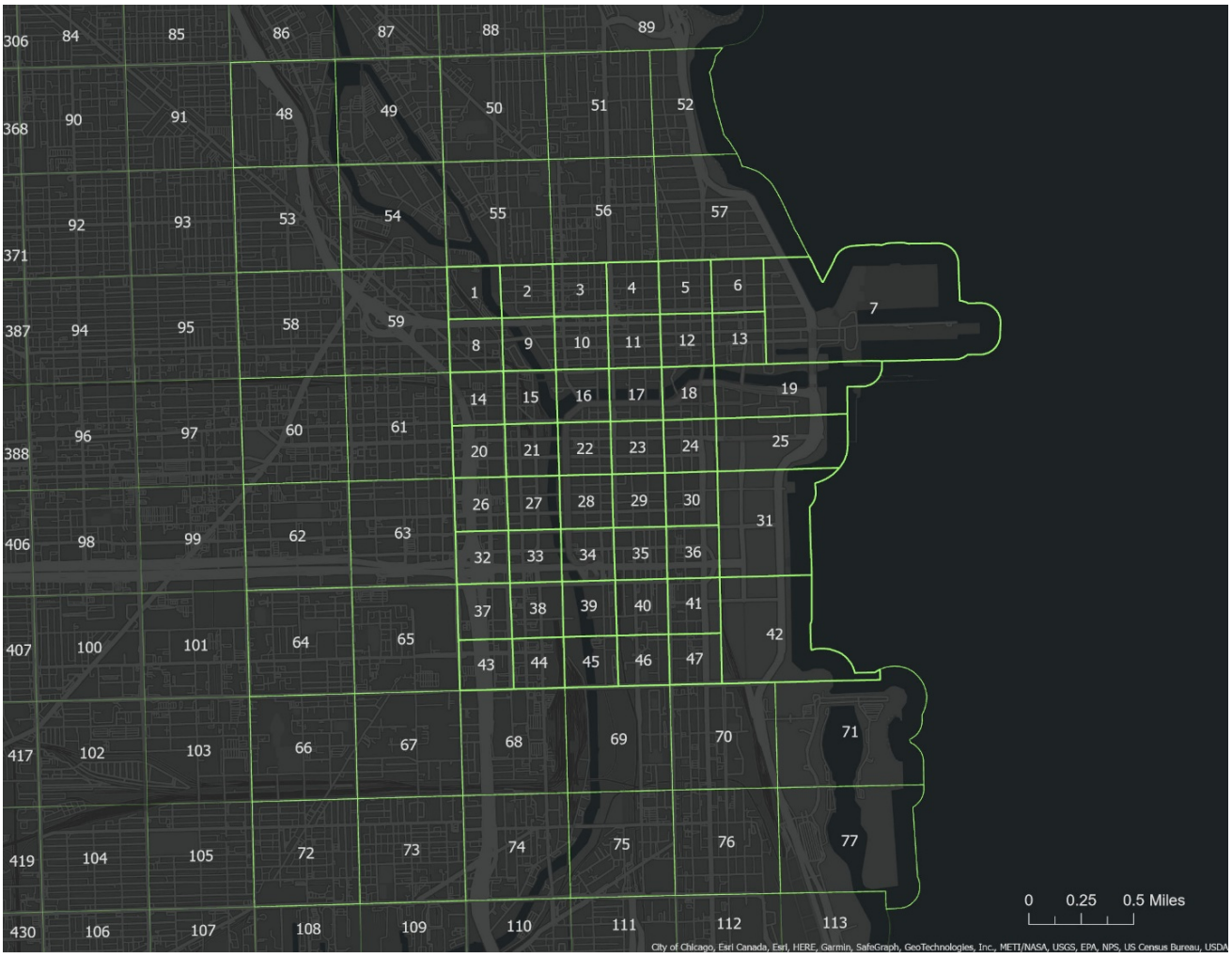


The density of the modeling zones (and by extension the subzones) increases within downtown Chicago. The Chicago Central Business District (CBD) is a pre-GIS convention established by CATS and the Northeastern Illinois Planning Commission (NIPC) that was based on boundaries set at Chicago Avenue, Halsted Street, and Roosevelt Road. It includes modeling zones 1 through 47. The larger Central Area also was established by CATS and NIPC prior to GIS to reflect the high density of

trips made in this area. It is based on the boundaries at North Avenue, Ashland Avenue, and Cermak Road. The Chicago Central Area is shown in Figure 9.

The Central Area includes modeling zones 1 through 77. Of the 77 zones, 30 are quarter section sized zones (one-half mile by one-half mile). Most remaining modeling zones (representing the CBD) are quarter-quarter-section sized zones (one-quarter mile by one-quarter mile).

Figure 9. CMAP central area zones



To simplify selecting discrete geographic areas, the modeling zones are numbered consecutively by county and township. Chicago is consecutively numbered, starting with the CBD (1-47), the Central Area (48-77), the Transit Hub (78-121) — which provides a one-mile buffer around the Central Area — and then by townships for zones within the city limit. Table 16 lists the correspondence between subzones, modeling zones, and geographic areas.

Table 16. CMAP subzone-zone correspondence

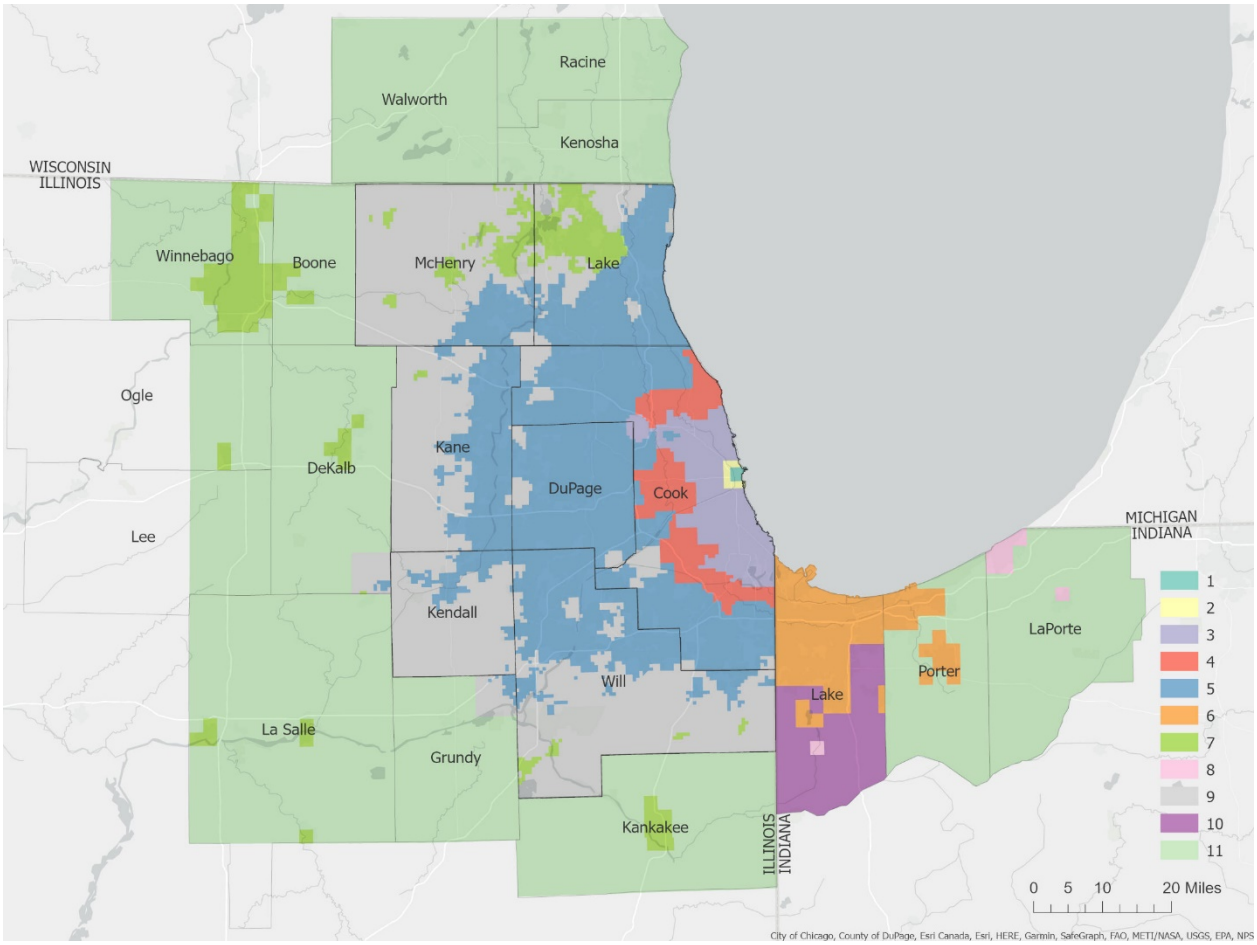
COUNTY		ZONE17		SUBZONE17	
FIPS	Name	First	Last	First	Last
17031	Cook	1	1732	1	3895
	Chicago (excluding DuPage portion)	1	717	1	983
	CBD	1	47	1	52
	Chicago Central Area	1	77	1	84
	Chicago Transit Hub	1	121	1	129
17043	DuPage	1733	2111	3896	5252
	Chicago portion	1733	1734	3896	3904
17089	Kane	2112	2304	5253	7406
17093	Kendall	2305	2325	7407	8702
17097	Lake	2326	2583	8703	10598
17111	McHenry	2584	2702	10599	13042
17197	Will	2703	2926	13043	16426
17063	Grundy	2927	2950	16641	16807
	Aux Sable Township	2949	2949	16664	16807
17007	Boone	2951	2975	16427	16451
17037	DeKalb	2976	3021	16452	16640
	Sandwich & Somonauk Townships	2977	2977	16454	16597
17091	Kankakee	3022	3073	16808	16859
17099	LaSalle (partial)	3074	3145	16860	16931
17103	Lee (partial)	3146	3151	16932	16937
17141	Ogle (partial)	3152	3168	16938	16954
17201	Winnebago	3169	3247	16955	17033
18089	Lake, IN	3248	3344	17034	17130
18091	LaPorte, IN	3345	3400	17131	17186
18127	Porter, IN	3401	3467	17187	17253
55059	Kenosha, WI	3468	3512	17254	17298
55101	Racine, WI	3513	3568	17299	17354
55127	Walworth, WI	3569	3632	17355	17418
	POEs	3633	3649	N/A	

Capacity zones

The final zone system used in the regional travel demand model is the capacity zone system (displayed in Figure 10). As with the modeling zones, these zones are built by aggregating the subzones. The capacity zones help estimate general road capacity for the highway assignment procedures. Some specific calculations that use the capacity zone value include:

- Calculation of an ordinal arterial functional class within the model
- Calculation of the number of inbound approaches into an intersection
- Estimation of traffic signal green-to-cycle ratios and signal cycle length for ramps connecting arterials and expressways

Figure 10. CMAP capacity zones



The values of the capacity zone system are listed below in Table 17. Within the travel demand model structure, the capacity zone values are stored as a node attribute in the highway network database. Thus, the values in the table below correspond to the values of the AREATYPE variable in the highway network node table.

Table 17. CMAP capacity zone codes

Capacity zone value	Description
1	Chicago Central Business District (2017 subzones 1-52)
2	Remainder of Chicago Central Area (2017 subzones 53-84)
3	Remainder of City of Chicago (2017 subzones 85-983 and 3896-3904)
4	Inner ring suburbs where Chicago street grid is generally maintained
5	Remainder of Illinois portion of the Chicago Urbanized Area
6	Indiana portion of the Chicago Urbanized Area
7	Other Urbanized Areas and Urban Clusters within the CMAP Metropolitan Planning Area plus other Urbanized Areas in northeastern Illinois
8	Other Urbanized Areas and Urban Clusters in northwestern Indiana
9	Remainder of CMAP Metropolitan Planning Area
10	Remainder of Lake County, IN (rural)
11	External area
99	Points of Entry – not defined in the Capacity Zone system

Analysis network preparation

Preceding sections briefly discussed the procedures used to take the highway network and transit service information from the GIS databases and process it for use in modeling networks. Processing programs export data from the model network databases and update the highway network and transit service characteristics based on the scenario network being created. After all characteristics are updated, a set of text files suitable for importing into the travel demand software is created, containing the data defining the transportation network.

Highway network

Separate import files are created for each of the TOD highway networks, as well as for one all-inclusive highway network. The TOD networks contain time-period specific changes to links, such as time-of-day parking restrictions and reversible lanes. Each TOD highway network has a set of four import files — two defining link attributes and two defining node attributes. The files contain the standard link and node attributes required by the travel demand software, as well as extra attributes (denoted by "@") used within the travel demand model. The files are imported into the modeling software to create the TOD highway networks.

Highway network link attributes are shown in Table 18 with required attributes highlighted in blue. Link modes are defined to enable a multiple vehicle class highway assignment that matches the vehicle types used for emission calculations. Mode "A" is the primary auto mode, and all other modes are secondary auto modes. No transit modes are included in the highway network because the transit network exists as a separate entity. This also means that no transit-only links (such as rail links or dedicated busways) are included in the highway network.

Secondary auto modes "S" for single occupancy vehicle (SOV) and "H" for high occupancy vehicle (HOV) allow high occupancy vehicle facilities to be represented in the network. For example, mode "S" would not be coded on HOV links. All links in the network allowing high occupancy vehicles would include mode code "H."

Table 18. Model Highway Network link attributes

Variable	Description		Source
i	From node		Batchin files
j	To node		
len	Link length in miles		
mode	Modes on link: A=primary auto S=single occupant auto (SOV) H=high occupancy auto (HOV) T=general truck	 b= B plate truck l=light truck m=medium truck h=heavy truck	
lanes	Number of driving lanes		
vdf	Volume-delay function code. 1=arterial street 2=freeway 3=freeway-arterial ramp	 5=freeway-freeway ramp 6=zone centroids connector 7=link where toll is collected	

	4=expressway 8=metered entrance ramp	
@speed	Speed limit or CMAP free speed	
@parkl	Number of parking lanes along roadway	
@sigic	Link with interconnected signals	
@width	Driving lane width in feet	
@toll	Toll amount in dollars	
@ftime	Uncongested link travel time in minutes	scripts
@emcap	Level of Service E lane capacity on link	
@artfc	Arterial link functional class: 1 = Principal Arterial 3 = Minor Arterial 2 = Major Arterial 4 = Collector	
@gc	Green time to cycle length ratio	

Secondary auto mode "T" is a general truck mode coded on all network links that allow trucks. By excluding truck modes, commercial vehicles can be prohibited from facilities, such as Lake Shore Drive, and the Kennedy and Dan Ryan express lanes. The additional truck modes "b," "l," "m," and "h" permit more specialized coding of truck prohibitions to represent local restrictions or the testing of truck-only facilities based on weight classes.

A link's volume-delay function (VDF) is based upon the five categories in CMAP's link capacity calculations, which include arterial, freeway, arterial-freeway ramp, expressway, and freeway-to-freeway ramps. Three additional volume-delay functions are included for links connecting zone centroids to the network, links where tolls are collected, and metered freeway entrance ramps.

In addition to these standard variables required by the modeling software, some additional link attributes are included in the network. Many supplemental variables come directly from the MHN database. Other attributes used in the scripts include the link's posted speed limit, as well as whether curb parking is allowed and the average width of driving lanes. For toll collection links, the amount of the toll also is included.

Table 19 lists the highway network node variables that are used. Standard node attributes are the node number and the x- and y-coordinates of the node. Node extra attributes are additional quantities associated with the node, including the zone number and area type (capacity zone value) at the node location. Area type definitions are listed in Table 17.

Table 19. Model Highway Network node attributes

Variable	Description	Source
i	Node number	Batchin files
xi	x-coordinate (NAD27 IL East State Plane feet)	
yi	y-coordinate (NAD27 IL East State Plane feet)	
@zone	Modeling zone node resides in	
@atype	Capacity Zone values (refer to Section 2.5)	
@napp	Number of approach links	scripts
@cycle	Traffic signal cycle length in minutes	

After the TOD highway networks are imported into the travel demand software, two scripts prepare the additional link and node attributes needed for the time assignments. The first script, *Ftime_capacity,* calculates link lane capacities in vehicles per hour, and uncongested speeds

based on link characteristics, such as functional class, lane width, and posted speed limit. The network database also includes variables to flag those links that change characteristics depending on the time, such as links that have peak period parking restrictions. These factors also are considered when link capacity is calculated.

The calculations in the *Ftime_capacity* script generally are consistent with the capacity procedures found in the 1985 Highway Capacity Manual and the 1994 update to the manual. The capacities of arterial street links reflect the type of signalized intersection at the link's j-node, or downstream node. The script logic first analyzes the links entering a node and then estimates the capacity for each approach link based on generalized signalized intersection characteristics. Capacities for ramps between freeways and arterial streets ending at signalized intersections are determined in the same manner as arterial streets.

The concept behind this process is link capacities. Uncongested travel times must always be recalculated before an assignment is run, rather than be maintained as static network variables in the database. The capacities and uncongested travel time for links ending at a signalized intersection depend on the characteristics of all approach links into the intersection, not just the link of interest. As a result, link capacities and uncongested travel times depend on network topology. Adding, removing, or modifying a link affects the capacities and uncongested travel times of all links that intersect it at a signalized intersection. Calculating these network quantities as part of the assignment procedure ensures they are current when the assignment is carried out. This approach simplifies the introduction of certain types of improvements into the modeled network. The effects of parking restrictions, traffic control device improvements, signal progression, and intersection improvements can be modeled in the script, eliminating lengthy manual adjustment of capacities and times on a link-by-link basis.

The *Ftime_capacity* script develops some extra link attributes, which briefly are described. Link uncongested travel time (@ftime) is calculated and is used in the volume-delay functions. It should be noted that this travel time does not contain any intersection delay, which is calculated separately by the volume-delay functions. Capacity values calculated by the script (@emcap) are hourly lane capacities at level-of-service E. Link capacity for the time-of-day period, referenced within the volume-delay functions, is later obtained by multiplying @emcap by the number of driving lanes on the link and the number of hours in the assignment time period.

The second script, *Arterial_delay,* repeats many of the same calculations as *Ftime_capacity*. It again evaluates approach links at signalized intersections and estimates signal cycle lengths at the j-nodes of arterial street links. It also estimates the proportion of the cycle length allocated to traffic on the link. These two quantities are retained in extra node and link attributes — @cycle and @gc, respectively — to be used later in the volume-delay functions that estimate intersection delays. An ad hoc functional class (@artfc) also is assigned to arterial street links based on the location of the link, its speed limit, and number of driving lanes. This functional class only is used to allocate green time at signalized intersections, which depends on the cycle length, and the number and types of conflicting approach links. The final link extra attribute in the table is the ratio of green time to cycle length, @gc, at the downstream node of a link. This value is used in the volume-delay functions.

Transit network

The travel model uses coded transit networks reflecting transit service in the morning peak period (6:00 to 9:00 a.m.) and the midday period (9:00 a.m. to 4:00 p.m.) to obtain transit cost and service information. A transit network contains over 12,000 bus and rail mode links, totaling nearly 5,700 miles in length. While the highway network data are all contained within the MHN database, the transit network comprises three separate components that must be integrated to create transit modeling networks, including:

- **Bus route coding:** All current and future bus route coding is stored in the route systems of the MHN geodatabase. This coding includes the following transit modes — B (CTA regular bus service), E (CTA express bus), P (Pace regular service), L (Pace local service), and Q (Pace express service).
- **Rail route coding:** All current and future rail route coding is stored in the rail route systems of the MRN geodatabase. This coding includes modes C (CTA rail) and M (Metra rail).
- **Auxiliary links:** In addition to the transit coding itself, model transit networks require a system of auxiliary links to provide needed connections. Auxiliary links provide transfer links between different transit modes or lines that do not pass through the same nodes, as well as walk access to transit service from zone centroids (trip beginning) and walk egress from transit service to zone centroids (trip end). A set of auxiliary links is created dynamically when the transit network files are generated. The procedures used to create the auxiliary links are discussed below. Zones requiring drive access to transit stations are not provided with specific auxiliary drive access links, rather drive access is handled by a matrix calculation that will be discussed later.

Table 20 lists the auxiliary link modes included in the transit networks. The transit network modes are case sensitive, which means all transit modes are uppercase and all auxiliary link modes are lowercase. The three types of auxiliary links are transfer links connecting transit lines to one another, access links to connect a zone centroid to transit service at the beginning of a trip, and egress links connecting transit service to a zone centroid at the end of a trip. Transfer links are bi-directional while access and egress links only serve one direction of travel. In practice, access and egress links are generally bi-directional links with the appropriate mode assigned to the appropriate direction. O'Hare International Airport's people mover transit service is a special auxiliary link type reflecting service moving travelers between O'Hare terminals and a multimodal center.

Table 20: Auxiliary link modes

Link Type	Mode	Description
Transfer	b	Bus-Bus walk
	c	Bus-CTA rail walk
	m	Bus-Metra walk
	d	Metra-Metra walk
	r	CTA rail-CTA rail walk
	t	CTA Rail-Metra walk
Access	u	Home-Bus walk

	v	Home-CTA rail walk
	w	Home-Metra walk
Egress	x	Bus-Work walk
	y	CTA rail-Work walk
	z	Metra-Work walk
Auxiliary	k	O'Hare people mover

Transfer links

There are six different transfer link modes, each identifying a connection between different types of transit service. Transfer links are needed when two services are physically separate because they don't share a station node. While Metra and CTA rail service are identified separately by the auxiliary links, no such distinction is made between CTA and Pace bus service. All bus service is combined into "bus." All transfer links are bi-directional, so passengers can move in either direction between the transit lines. Transfer links "r," "t," and "d" are hard-coded in the MRN and are read directly into the final transit network via the rail.network file. The remaining transfer links are created as follows:

Mode b: All highway network arterials (TYPE=1 in the MHN) located in modeling zones 1-78, representing an area just slightly larger than Chicago's Central Area, are assigned mode "b". This allows trips to use "sidewalks" along arterial links for transfers. These links are included in the input file bus.network.

Mode c: The Euclidean distance is calculated between CTA rail stops and all bus stops. A maximum distance of one-eighth mile (660 feet) is allowed between CTA rail stops in the CBD and bus stops. A maximum distance of a half mile (2,640 feet) is allowed between bus stops and the remaining CTA rail stops. The shortest mode "c" link available to connect a CTA rail stop to each bus route is retained. As CTA rail stations may be served by multiple bus routes, there may be instances where more than one mode "c" link connects a bus route to the same CTA rail stop. This only occurs if the shared stop was not the shortest link from the CTA rail stop for all the affected bus routes. These auxiliary links are stored in access.network.

Mode m: Straight-line distances are calculated between Metra stops and all bus stops. Maximum distances of a quarter mile (1,320 feet) and 0.55 miles (2,904 feet), respectively, are allowed between Metra stops and CTA bus stops, and Metra stops and Pace bus stops. Mode "m" links are attached to all bus stops determined to be within the allowable distance. There are no constraints on connecting a Metra stop to multiple stops on the same bus route or on connecting a bus stop to multiple Metra stops. These auxiliary links are stored in access.network.

Access and egress links

Access and egress links are specific to modes, so there are three access modes connecting zone centroids to different types of transit service and three egress modes connecting transit service to a centroid. Each access and egress mode applies to only one direction. For example, a centroid is connected to a Metra station by one mode "w" link from the centroid to the station, and one mode "z" link from the station to the centroid. In practice, this usually, but not always, results in a two-way link with different modes in each direction. The access and egress links are created as follows:

Modes u and x, generic bus access and egress: The Euclidean distance between centroids and transit stops is calculated to determine access and egress link length. Bus stops are separated into

CBD and non-CBD stops. A maximum distance of a quarter mile is allowed between centroids and CBD bus stops, while a maximum of 0.55 miles is allowed between centroids and the remaining bus stops. The access/egress links are merged with bus itineraries, connecting multiple links per transit line to the same centroid. This allows for numerous access/egress connections between the same bus route-zone centroid pair if the bus itinerary varies by direction. Redundant access and egress links are eliminated, and links are grouped by centroid and sorted by link length in ascending order. A maximum of eight mode "x" links are kept for each CBD centroid. No more than two mode "x" links are retained for each non-CBD centroid. A maximum of three mode "u" links are saved per centroid. These auxiliary links are stored in access.network.

Modes v and y, CTA rail access and egress: The straight-line distances are calculated between centroids and CTA rail stops. A maximum distance of 0.55 miles is allowed between CTA rail stops and zone centroids. By rule, each CTA rail station is connected to the centroid of the zone it resides in with an assigned distance of 0.55 miles if the link length exceeds 0.55 miles. The remaining access links are ranked in ascending order and are assigned to centroids until the maximum allowable number is reached. A maximum of three mode "v" links is assigned to each centroid in a zone with a CTA rail station. The same process is used for the egress links. Maximums of seven and three mode "y" links are assigned to centroids in the CBD and outside the CBD, respectively, in addition to the connection to the station zone. These auxiliary links are stored in access.network.

Modes w and z, Metra rail access and egress: Metra station access and egress links follow the same basic procedures as CTA rail station links. A maximum distance of 0.55 miles is allowed for these links. Metra stations are connected to the centroid of the zone they are within, and a length of 0.55 miles is assigned to the link if it exceeds the distance limit. Unlike with the other access and egress links, there is no constraint on the number of Metra access and egress links per centroid, so all are put in the final network. These auxiliary links are stored in access.network.

Table 21 summarizes the processing rules used to develop the auxiliary links. When the processing is completed, the result is a set of scenario transit network files that are formatted to be imported into a single scenario of the travel demand software.

Table 21: Auxiliary link processing rules

Mode	Maximum distance	Forced to connect to centroid?	Maximum number of links per centroid
c	⅛ mile in CBD ½ mile outside CBD	No	N/A
m	¼ mile for modes BE 0.55 miles for modes PLQ	No	N/A
u, x	¼ mile in CBD 0.55 miles outside CBD	No	u – 3 x – 8 in CBD, 2 outside CBD
v, y	0.55 miles	Yes	v – forced + max. of 3 additional y – forced + max of 7 additional in CBD, forced + max of 3 additional outside CBD
w, z	0.55 miles	Yes	N/A

Zonal impedances

A primary role of transit networks is to generate transit level of service variables for the generalized cost procedures used in the destination choice-mode choice model. Impedance matrices are

created for zone-to-zone in-vehicle times, fares, first wait time and remaining out-of-vehicle time. In the logic of the CMAP models, the zone-to-zone quantities are all measured from the point where transit service is first boarded rather than the actual trip origin. Access modes, times, and costs are generated using Monte Carlo simulation techniques. These techniques will be discussed later in the document.

Zone-to-zone impedances are built using the time and cost components of the transit network. Time components are weighted to reflect the relative disutility to the traveler. For instance, walking time is weighted at two times the rate of time spent within a transit vehicle. Similarly, fares are weighted so they can be combined with times to create an overall measure of the impedance of a particular path.

A multi-path transit assignment is completed to provide transit impedances for zones that have walk access to a transit station. For zones with no walk access to a transit station, highway impedances from a complementary highway assignment are used to index the origin zone to a station zone that minimizes drive access and transit impedance to the destination. All cost components in the impedance matrix between the auto access zone with no walk access and the destination are replaced with cost components from the selected station zone to the destination. The result is, instead of the zone being disconnected, the origin to destination times and costs are populated with times and costs reflecting the selected station. In this application, a generalized parking cost is calculated to reflect on- and off-street parking availability and cost.

The transit network scenario also is used to generate travel districts based on a hierarchy of services present in the zone. This is analogous to CMAP's historic use of first, last, and priority mode categorization. The mode matrices are constructed based upon the transit services likely to be used when moving between these zone groups.

The effects of congestion on bus travel times are included in the transit skimming procedures. The modeled bus travel times start with scheduled times from the GTFS files produced by the transit service providers. In iteration zero, the scheduled times for the morning peak and midday scenarios are compared to congested auto travel times from a comparable model run. The congested times are used if they are longer than the scheduled bus times. After each full model iteration, bus travel times are again updated with auto travel times when the auto time is longer than the scheduled time. The transit impedance matrices are recalculated with each iteration and maintain a consistent relationship with the auto travel times. This is important because the relationship between auto and transit travel times is an important determinant of the regional model results. An additional feature allows the schedule time to be retained by coding the transit travel time function as two where congested times should not be considered. This is helpful for scenarios that include bus rapid transit or other similar services that will not be impacted by prevailing traffic conditions.

Ancillary data input files

In addition to the network datasets, several ancillary data files contain information on transit service levels, park-and-ride availability, CBD parking, and auto operating costs. These files are briefly described in this section.

M01 file

The M01 file ("ALLPURPOSE_M01.TXT") stores several variables to provide the destination choice-mode choice model with zonal transit availability, and park and ride characteristic parameters. Some of the parameters are calculated using transit network characteristics and are specific to each scenario network. The contents of the M01 file are summarized in Table 22.

Table 22: M01 file attributes

Field name and position	Description
Zone number (1)	Modeling zone number
Zone type (2)	1 – Chicago Central Area zone (1-77) 2 – Chicago zone outside of the Central Area (78-309) 3 – dense suburban zone (more than 500 activities [population + employment] per square mile) 4 – remaining suburban zones
Park and Ride cost (3)	The cost of parking at the park and ride lot closest to the zone's geographic centroid. This represents the lowest rate of either the daily parking rate or the cost of a monthly parking pass divided by twenty workdays.
Median zone household income (4)	The median household income in the zone (in $100s) derived from POPSYN_HH.CSV (2019 dollars)
Park and ride availability (5)	A binary value indicating that the zone has park and ride access if there is a park and ride location within ten miles of the zone's geographic centroid
Average waiting time for bus service in zone for home-work trips (6)	The average wait time for bus work trip (in minutes) for modes BEPQ. Calculated as the mean of (headway*0.5) for all bus stops served by selected modes within the zone during the AM peak period. To account for spatial inaccuracy, stops are buffered by 0.1 miles. M*aximum value is 99, default value of 99 used for zones with no bus service*.
Average waiting time for bus service in zone for non-work trips (7)	The average wait time for bus non-work trip (in minutes) for modes BEPQ. Double the wait for bus work trip (as a proxy for midday service). *Maximum/missing values set to 99*.
Average waiting time for feeder bus service in zone for home-work trips (8)	The average wait time for feeder bus work trip (in minutes) for mode L. Calculated as the mean of (headway*0.5) for all bus stops served by mode L within the zone during the AM peak period. To account for spatial inaccuracy, stops are buffered by 0.1 miles. *Maximum/missing values set to 99*.
Average waiting time for feeder bus service in zone for non-work trips (9)	The average wait time for feeder bus non-work trip (in minutes) for mode L. No feeder bus service in off-peak period, so: *all values set to 99*.
Home-work trip auto work end auto occupancy (10)	Zone vehicle occupancy rate for commuters (i.e., workers per vehicle), measured at destination (work) location. Calculated using data from the 2017-2021 5-year CTPP data.

DISTR file

The DISTR files contain zonal transit approach distribution parameters, which are used to determine mode choice. The parameters are calculated using transit network characteristics and are specific to each scenario network. Table 23 describes the DISTR file fields.

Table 23: DISTR file attributes

Mode category	Field number	Description
	1	CMAP modeling zone number
Commuter rail (fields 2-4) CTA rail (fields 5-7) Park and Ride (fields 14-16)	2,5,14	Mean distance in miles to station (or park and ride) *Calculation*: A distance calculation is performed to find the rail station closest to each subzone centroid. The closest station distances are then aggregated to the zonal level to determine the mean distance weighted by subzone households. The maximum mean distance is set to 19.95 miles. The analysis is performed separately for the commuter rail stations, rapid transit stations and park and ride locations.
	3,6,15	Standard deviation of distance to station (or park and ride) *Calculation*: The standard deviation is calculated as the square root of the sum of [the zonal variance calculated above and a subzone variance, estimated to be 0.042].
	4,7,16	Type of Mode Choice distribution: 101 (default) = normal distribution 102 = exponential distribution
Bus (fields 8-10) Feeder Bus (fields 11-13)	8, 11	Minimum bus stop distance (miles). *Calculation*: A set of bus stop buffers (incremented by 0.1 miles from 0.1 up to 1.1 miles) are created. The buffers are overlaid with the zone system to determine the proportion of each zone covered by each buffer. The minimum bus stop distance is determined by the size of the smallest buffer that covers any part of the zone. The value is set to 999 for zones that are not covered by any portion of the largest buffer. The analysis is performed separately for bus (modes BEPQ) and feeder bus (mode L) stops.
	9, 12	Maximum bus stop distance (miles) *Calculation*: The maximum bus stop distance is determined by the size of the smallest buffer that covers the entire zone (for practical reasons the threshold is 97% of the zone). The value is set to 999 for zones with a minimum distance of 999 or is set to 1.1 if no buffer meets the coverage threshold and the minimum distance is not 999.
	10, 13	Proportion of the zone within minimum walking distance *Calculation*: This value is calculated as the area of the zone covered by the minimum distance buffer divided by the area of the zone covered by the maximum distance buffer. This value is set to 999 for zones with a minimum distance of 999.

Three separate DISTR files are used: HO_DISTR.TXT, HW_DISTR.TXT and NH_DISTR.TXT. As with the M01 file, two separate formats of the DISTR file are created, including one for home-based work trips and one for the other trip purposes. The only difference between the formats is that the feeder bus fields (11-13) are all set to 999 for the home-based other and nonhome-based files.

M023 file

The M023 file contains transit fare and auto operating cost data used by the pre-distribution and mode choice models. The cost data reflect 2019 values. The file is composed of six records, including:

- CTA fares
- Pace feeder bus fares
- Pace regional bus fares
- Auto operating costs in 5-mile-per-hour increments for speeds between 0-40 miles per hour
- Auto operating costs in 5-mile-per-hour increments for speeds between 40-80 miles per hour
- Average auto operating costs per-mile-by-zone-type used to estimate transit access/egress costs

The CTA fares used in the M023 file are:

- Bus boarding fare is $1.51
- Rail transit boarding fare is $1.70
- First transfer is $0.19
- Link-Up pass per ride is $1.38 (approximately equal to $55 monthly cost divided by 40 trips per month)

The Pace fares used in the M023 file are:

- Feeder bus boarding is $1.38 (assumes Link-Up pass)
- CBD feeder bus fare is $0.00 (fare calculations revised and no longer used)
- Pace current regular fare is $1.58
- Pace first transfer is $0.24

Auto operating costs were updated to reflect current fuel consumption and the current costs of tires, maintenance, and gasoline. These costs were derived from two sources. A 2016 publication by AAA was the source for the per-mile costs of auto maintenance and tires. These values were adjusted to 2019 using the Consumer Price Index. Figures on the average gasoline consumption per mile were obtained from the U.S. Energy Information Administration. A $3.00 per gallon gasoline cost was assumed to convert the gasoline consumption into a cost per mile. The resulting costs per mile for auto travel are listed in Table 24.

Table 24: M023 file auto operating costs

		Cents per mile			
Miles/hour	Fuel used gallons/mile	Gasoline (@$3.00/Gal)	Tires	Maintenance	Total
0-5	0.060	18.08	1.06	5.62	24.76
5-10	0.049	14.69	1.06	5.62	21.37
10-15	0.041	12.43	1.06	5.62	19.11
15-20	0.036	10.74	1.06	5.62	17.42
20-25	0.032	9.61	1.06	5.62	16.29
25-30	0.030	9.04	1.06	5.62	15.72
30-35	0.030	9.04	1.06	5.62	15.72
35-40	0.030	9.04	1.06	5.62	15.72
40-45	0.030	9.04	1.06	5.62	15.72
45-50	0.029	8.76	1.06	5.62	15.44
50-55	0.029	8.76	1.06	5.62	15.44
55-60	0.029	8.76	1.06	5.62	15.44
60-65	0.031	9.32	1.06	5.62	16.00
65-70	0.034	10.17	1.06	5.62	16.85
70-75	0.037	11.02	1.06	5.62	17.70
75-80	0.040	11.87	1.06	5.62	18.55

The last record in the file contains the average per mile auto operating cost for each of the four zone types (defined in the M01 file). The M023 data is collapsed into six records, which are shown in Table 25. All values are reported in cents.

Table 25: M023 file layout

	Field locations							
	1-5	6-10	11-15	16-20	21-25	26-30	31-35	35-40
CTA fares	151	170	19	138				
Feeder bus fares	138		0					
Pace fares	158		19					
Auto operating costs by 5 MPH	2476	2137	1911	1742	1629	1572	1572	1572
	1572	1544	1544	1544	1600	1685	1770	1855
Auto operating costs by zone type	2000	1700	1500	1500				

CBD parking file

CBD parking costs also are important to the destination choice-mode choice model. A database of selected central area parking facilities is used to provide parking cost distribution information to the composite cost and mode choice models. The specification of the variables and fields is described in CATS Working Paper 95-01. The parking supply database currently is treated as a fixed input unless a scenario is testing the effect of downtown parking costs on regional mode choice.

There are two different CBD parking files. The first file identifies the parking supply characteristics of each Central Area zone that contains parking. Each zone in this set has five records with the following information:

- Central Area parking zone number
- The probability of finding parking within the zone at or below the threshold parking cost (this value must be 100 percent in each zone's fifth record)
- The threshold parking cost in cents per hour
- The savings in parking costs in cents per hour determined by subtracting the threshold parking cost from the maximum parking cost in the zone
- The amount of time needed to walk one block in the CBD (180 seconds)

A sample of the parking supply records for CBD zone 5 is displayed in Table 26. Note the maximum cost to park in this zone was identified as $6.00 ($48.00 per eight-hour day).

Table 26: CBD parking file 1 — sample parking supply records

CBD parking zone	Parking probability (0.0001)	Threshold cost (cents/hour)	Cost savings (cents/hour)	Walk speed (seconds/block)
5	2000	153	447	180
5	3000	154	446	180
5	6000	178	422	180
5	8000	311	289	180
5	10000	600	0	180

The parking supply records were updated using the following procedures. A website for downtown parking availability (bestparking.com) provided data for these values. Off-street parking facilities were organized by Central Area parking zones. The least costly daily rate for each facility was determined using either the early bird daily rate, daily rate, hourly rate multiplied by nine hours, or a monthly rate divided by 20 workdays. Each parking facility in a zone was assumed to have the same selection probability. Probabilities of facilities in the same zone and with the same best daily rate were combined. Cumulative rate probabilities for each zone were then totaled from least to most expensive parking facility and the cost savings compared to the most expensive facility in each zone were calculated. Five parking supply records were created from the parking probabilities and

threshold costs. In zones with many off-street parking locations, threshold costs were selected to yield nearly equal probability intervals.

User characteristics are important to the cost paid for parking, and they were updated in 2020 using data from the My Daily Travel survey. These characteristics include the percentage of people who have access to free parking, and the percentage of trips by auto occupancy. Both characteristics are stratified by income. While it may seem curious to input quantities that are estimated by the models, these travel characteristics are needed to compute the free versus paid CBD parking and the parking costs per person. Given their role in the cost calculations, they only need to be rough estimates based on observed travel.

The user characteristics are included in the second CBD parking file. There are only five records in this file. These records are ordered by household income ranges and include the following variables for Central Area commuters:

- The upper value of the household income quintile range (last record is the lower bound of the highest quintile)
- The percentage of Central Area auto commuters with free parking
- The percentage of all Central Area workers taking transit to the Central Area
- The percentage of Central Area auto commuters in single-occupant vehicles
- The percentage of Central Area auto commuters ridesharing in two-person vehicles
- The percentage of Central Area auto commuters carpooling in three-person vehicles
- The percentage of Central Area auto commuters carpooling in four-or-more-person vehicles.

The user characteristics are shown in Table 27. Most of the values do not change with income. This is due to the limited sample of Central Area parkers found in My Daily Travel.

Table 27: CBD parking file 2 – user characteristics

Income quintile	Park free	Transit	Auto occupancy			
			One	Two	Three	Four or more
29999	32.6	41.3	85.4	10.1	0.8	3.7
59999	32.6	43.5	85.4	10.1	0.8	3.7
99999	32.6	53.9	85.4	10.1	0.8	3.7
149999	32.6	53.6	85.4	10.1	0.8	3.7
150000	32.6	49.1	85.4	10.1	0.8	3.7

Monte Carlo simulation

A major source of inaccuracy in travel demand modeling is the use of average values, such as the average cost of parking in a traffic analysis zone or the average income of the traveler. The CMAP/CATS travel demand analysts recognized this potential source for inaccuracy early, perhaps before anyone else was aware of the problems that could be generated by using average values. The solution was to identify the major areas affected by average values and use a method that would convert the average values into individual values. This methodology is called a Monte Carlo simulation technique, and, after the Chicago application, the technique also was used in the Dallas-Fort Worth region and the Cleveland region. Presently, a Monte Carlo simulation is used widely in travel demand modeling practices.

A Monte Carlo simulation focuses on selecting a representative value for a measure with this value being selected at random from a distribution of potential values. Table 28 shows an example where 1,000 parking spaces are spread out over six parking garages in an area, with each garage having the following characteristics.

Table 28: Monte Carlo example

Parking lot	Spaces	Daily cost	Probability
A	150	$3.50	0.150
B	175	$3.75	0.175
C	275	$3.25	0.275
D	75	$1.25	0.075
E	150	$3.50	0.150
F	175	$3.25	0.175

In this case, the weighted average parking cost for the 1,000 spaces is $3.26. But a few lucky people (7.5 percent) could park for $1.25 and some unfortunate people (17.5 percent) have to pay $3.75. The difference between the average cost and the low cost is $2.00 while the difference between the average cost and the high cost is 50 cents. These differences are substantial given that the cost of using a mode is a major determinant of which mode is selected. In a Monte Carlo simulation, a specific parking lot would be selected using a random number draw. The probability of being selected is a function of a relative parameter — in this case, the number of spaces. Therefore, the inexpensive parking lot would be selected 7.5 percent of the time, while the most expensive lot would be selected about 17.5 percent of the time.

Throughout the CMAP trip-based model, Monte Carlo simulation is used to determine several components of the submodels. Table 29 summarizes how Monte Carlo simulation is used.

Table 29: Monte Carlo simulation applications

Model	Monte Carlo applications
Trip generation	Selection of household workers who work-from-home Work-from-home frequency for workers who telecommute less than five days a week (for the 2019 base year) Selection of survey household to use for trip enumeration
Mode choice-destination choice	Traveler's household income Distances to available approach modes for transit access Distances to available departure modes for transit egress Parking costs Walk distance from parking lot to destination Final selection of destination and mode
Time-of-day	Selection of time period for trip

Population synthesis

CMAP uses an open-source platform for population synthesis called PopulationSim to generate a synthetic population for the modeling region for 2020 — the UrbanSim base year. PopulationSim generates household weights that satisfy marginal control distributions using entropy maximization-based list balancing. It then expands households using these weights to create a full synthetic population. The file, HH_IN.TXT (Table 30), provides the marginal control distributions for various attributes at the household and person level. The number of households is specified at the subzone level. Household size bin counts are specified at the regional level. The remaining attributes are specified at the zone level. Data from the 2018-2022 ACS 5-year PUMS is used for the disaggregate population sample, or seed data. A geographic crosswalk is used to provide a correspondence between subzones, zones, and PUMAs.

Once the synthetic distribution has been developed, an output file, called POPSYN_HH.CSV, is created to provide the relevant information needed by the trip generation model. The variables in this file are listed in Table 31. This file provides the household information the trip generation model will use to create trips for each enumerated household.

Populations for model years 2019 to 2050 come from UrbanSim. A script converts UrbanSim household and person output tables to the tables required by the trip-based model. To guide the evolution of the 2020 population through 2050, household control tables are developed and provided as inputs to UrbanSim for each year between 2020 and 2050. To produce the household counts at the disaggregate level required for the travel model, synthetic populations are created for every scenario year needed to support Regional Transportation Plan analyses and to meet vehicle emissions budgets out to the year 2050. Marginal control distributions for these future-year populations come from the CMAP demographic model. All attributes are specified at the sub-regional level except for the number of households, which is specified at the subzone level, as required by the population synthesizer. Since the household totals are only provided by the demographic model at the sub-regional level, the base-year distribution of subzone households is preserved through 2050.

Table 30. HH_IN.TXT input file

Variable	Description
SUBZONE	The trip generation subzone, which must be in sequence in the file from low to high values
SUBZONE HOUSEHOLDS	Number of households in the trip generation subzone
ADULTS	Total adults (sixteen and older) in the subzone
WORKERS	Total adult workers in the subzone
CHILDREN	Total children (fifteen and younger) in the subzone
INCOME CATEGORY 1	Census income data in 2019 dollars: Households in category 1 (under $30,000)
INCOME CATEGORY 2	Households in category 2 ($30,000 - $59,999)
INCOME CATEGORY 3	Households in category 3 ($60,000 - $99,999)
INCOME CATEGORY 4	Households in category 4 ($100,000 and over)
AGE OF HOUSEHOLDER CATEGORY 1	Households in subzone where head of household is under 35
AGE OF HOUSEHOLDER CATEGORY 2	Households in subzone where head of household is 35-64
AGE OF HOUSEHOLDER CATEGORY 3	Households in subzone where head of household is 65 and older
HOUSEHOLD SIZE 1	Number of 1-person households
HOUSEHOLD SIZE 2	Number of 2-person households
HOUSEHOLD SIZE 3	Number of 3-person households
HOUSEHOLD SIZE 4	Number of 4-person households
HOUSEHOLD SIZE 5	Number of 5-person households
HOUSEHOLD SIZE 6	Number of 6-person households
HOUSEHOLD SIZE 7	Number of 7-or-more-person households
PRIVATE AUTO COMMUTE MODE SHARE	The ratio between the workers in the subzone who commute by household vehicle (auto, truck, van, sport utility vehicle, motorcycle) either as a single occupant or in a carpool divided by the workers in the subzone developed from the 2017-2021 CTPP
SIDEWALK DENSITY	Miles of sidewalk per square mile from CMAP's regional sidewalk inventory – a measure of pedestrian friendliness

Table 31. POPSYN_HH.CSV output file

Variable	Description
SUBZONE	The trip generation subzone
HHTYPE	Household type code for the synthetic household – one of 624 values based on the combination of adults-workers-children-age of householder and income index (see Trip Generation Section).
VEHICLES	Number of vehicles available in the synthetic household (PUMS data)
SERIALNO	Unique identifier assigned to the household in the ACS data
STPUMA5	Concatenation of the State FIPS code and the 5% PUMA number

ROWCOL	Variable identifying the location of the synthetic household in the region: 1 = Inner Chicago 2 = Outer Chicago and Inner Suburbs 3 = Mid Suburban 4 = Fringe and External Areas
ADULTS	Number of adults in the synthetic household
WORKERS	Number of workers in the synthetic household
CHILDREN	Number of children in the synthetic household
INCOME CATEGORY	Income category for synthetic household (2019 dollars): 1 = under $30,000 2 = $30,000 - $59,999 3 = $60,000 - $99,999 4 = $100,000 and over
AGE CATEGORY	Age of head of householder for synthetic household: 1 = householder age 16-34 2 = householder age 35-64 3 = householder age 65+
HHVTYPE	Household vehicle type code for the synthetic household – one of 624 values based on the combination of adults-workers-children-age of householder and vehicles available (see Trip Generation Section).
INCOME	Household income of the synthetic household (in 2019 dollars)

Trip generation

Trip generation is the first model step used by CMAP to forecast travel behavior. It is how land use planning and zoning quantities, such as households and employment, are converted into trip origins and destinations that serve as transportation demand measures. The trip generation process links the region's current and forecasted socioeconomic characteristics — the variables that drive travel demand — with the remaining sequential steps used to estimate choices of a trip destination, its mode, and route.

Trip generation is based upon an enumeration of all households in the study area. Each trip generation subzone is populated with synthetic households whose information on daily trips is drawn from CMAP's My Daily Travel survey data based on specified criteria. Since the household sample in the survey is small relative to total regional households, a survey household may appear multiple times in the same subzone. Trips reported by these households are used in the travel model, instead of the typical trip generation methodology, which is based upon trip generation rates. This approach eliminates the intermediate step of estimating trips generated per person or household.

As the trip generation model software executes, it writes files of households tabulated by composition, income, and vehicle ownership. These files have value beyond their role in trip generation. For example, these household files might prove useful in studies dealing with issues of social and economic justice related to alternative transportation investments.

Telework

CMAP's trip-based model explicitly accounts for workers who telework (or work from home). In 2021, it became apparent that the widespread occurrence of workers teleworking may not be a fleeting phenomenon and should be addressed in the model. Data from the 2018-19 My Daily Travel and 2007-08 Travel Tracker surveys showed the share of workers who telework at least part time had been increasing over the past decade. A telework allocation method is implemented prior to trip generation to identify households as having no workers teleworking or at least one worker teleworking. Household trip enumeration then applies trips from the appropriate work from home (WFH) or non-WFH household category.

Due to the large difference in telework rates observed in 2019 (the model base year) compared to current post-pandemic rates, separate telework allocation methods are used to model the base year and future years. For the base year, two targets are set for telework behavior: the overall share of workers who usually telework (usualwfh) and the overall share of workers who telework between one and four days per week (tc14). These targets are implemented through a process that uses person and household attributes from the synthetic population to flag specific workers as teleworking five or more days a week, teleworking one to four days a week, or not teleworking. For the 2019 base year, usualwfh is set to 0.0510 and tc14 is set to 0.1031.

Attributes used in determining work-from-home status are means of transportation to work, household income, worker industry, and worker education level. Inputs into the process define relationships between variables (see Table 32). Data for these files came from analysis of work-from-home workers in travel survey data and 2015-2019 PUMS data.

Table 32. Telework allocation inputs

Name	Description
incdist	Defines distributions of workers across income bins for 'low' and 'high' income industries
indusmix	Defines distribution of workers across industries for each work-from-home group (usualwfh and tc14)
indp_naics	Crosswalk file to attach NAICS codes to workers in the UrbanSim output files
lowlist	List of NAICS2 industries where <60% of work-from-home workers are in the highest income group. Corresponds to the 'low' distribution in incdist.
highlist	List of NAICS2 industries where >=60% of work-from-home workers are in the highest income group. Corresponds to the 'high' distribution in incdist.
eduw	List of education weights for the usualwfh group in order of ascending education level.
eduwtc	List of education weights for the tc14 group in order of ascending education level
tcportions	Distribution of work-from-home days within tc14 group in ascending day order. (1 day per week at home, 2 days, etc.)

The main steps in the telework allocation process are:

- The person files output by UrbanSim are read and all workers are identified. The industry-NAICS crosswalk file is used to attach a two-digit NAICS code to each worker.
- Numeric targets are established for each industry/income group within the two telework categories (usualwfh and tc14).

- Workers are sampled from both telework categories until the target values are reached.
- For tc14 workers:
    - Assign each worker the number of days they telework (1, 2, 3 or 4).
    - Determine if each tc14 worker is teleworking on the simulated model day.

The final output file, HH_WFH_STATUS.CSV shown in Table 33, assigns a work-from-home status to each household. Since the relevant information for the travel model is whether workers worked from home on the modeled day, workers who work-from-home one to four days a week are assigned work-from-home probabilities based on the tcportions input, and then a random number is used to determine final assignments.

Table 33. HH_WFH_STATUS.CSV output file

Variable	Description
SERIALNO	Unique identifier assigned to the household in the ACS data.
FLAG	Work-from-home status for the household. 1 = at least one worker with 'usualwfh' status in household 2 = at least one worker with 'tc14' status in household that is working on the modeled day, and no workers with 'usualwfh' status in household
WFHWORKERS	Number of people working from home in the household on the modeled day
TC14NW	Flag to indicate if there are tc14 workers in the household that are not working on the modeled day 0 = no people in this category 1 = at least one tc14 worker present who is not working

The telework allocation process for post-pandemic analysis years uses a modified methodology. Observed data show that Monday and Friday currently have higher rates of teleworking, while Tuesday through Thursday see more workers commuting to the office. The future year allocation method uses rates that reflect Tuesday-Wednesday-Thursday telework.

The Current Population Survey (CPS) publishes monthly data on national telework rates. Data from January 2025 through August 2025 were analyzed to determine mean monthly telework rates by NAICS code. The NAICS codes were then grouped into low, medium and high telework categories (shown in Table 34) based on the following thresholds:

- Low – under 20% telework rate
- Medium – between 20% and 30% telework rate
- High – more than 30% telework rate

Table 34. Telework categories

Telework category	NAICS code	Description
Low	11	Agriculture, Forestry, Fishing and Hunting
	21	Mining
	23	Construction
	31-33	Manufacturing
	44-45	Retail Trade
	48-49	Transportation and Warehousing
	55	Management of Companies and Enterprises
	56	Administrative and Support and Waste Management and Remediation Services
	61	Educational Services
	62	Health Care and Social Assistance
	72	Accommodation and Food Services
	81	Other Services (except Public Administration)
Medium	22	Utilities
	42	Wholesale Trade
	71	Arts, Entertainment and Recreation
	92	Public Administration
High	51	Information
	52	Finance and Insurance
	53	Real Estate Rental and Leasing
	54	Professional, Scientific and Technical Services

My Daily Travel Phase One data (2024-25) were analyzed in a similar fashion to determine average Tuesday-Wednesday-Thursday telework rates by NAICS code. The survey data required some additional processing as there are people who commute to work for part of the day and telework during the same day. Based on how the household trip enumeration process is applied, for travel model purposes anyone who commutes to work is deemed to be not teleworking (even if they do both) so that the appropriate travel behavior can be applied within the model.

While the My Daily Travel data showed a pattern similar to the CPS data, ultimately the CPS data were used due to the household travel survey data having small sample sizes for certain NAICS categories. The rates used for workers in the telework categories are:

- Low – 0.143
- Medium – 0.232
- High – 0.552

As with the base year telework allocation, household income and the worker's level of education are used to influence the sample of workers selected for telework. Data from the My Daily Travel Phase One survey were used to develop the shares in Table 35, as the CPS data do not provide the detailed data to support this analysis. Household income is simplified into two categories: under $100,000 and $100,000 and up, and education level reflects workers who have not earned a bachelor's degree and those who have earned at least a bachelor's degree. The work-from-home status of each household is written to HH_WFH_STATUS.CSV.

Table 35. Future allocation teleworker shares

Telework category	Household income		Education level	
	Under $100,000	$100,000 or more	Less than bachelor's degree	At least bachelor's degree
Low	0.353	0.647	0.305	0.695
Medium	0.418	0.582	0.268	0.732
High	0.231	0.769	0.202	0.798
Overall	0.295	0.705	0.251	0.749

Trip generation processing steps

The main processing steps in the trip generation model are briefly summarized below.

Model control keywords

Several important parameters (displayed in Table 36) are supplied to the trip generation model code through the file trip_generation_model.yaml. These keywords are read from the file and control the operation of the program.

Table 36. Trip generation input file parameters

Variable	Description	Model run values
MODEL VERSION	A text description identifying the model version being run	
SUBZONES	Number of trip generation subzones in the study area	17418
PUMA5	Number of Census 2010 five percent sample PUMAs in the modeled study area	74
ZONES	Zones used in the remaining CMAP models for trip distribution (linking of trip ends into trips between zones), mode choice (allocation of trips to travel modes), and assignment (allocation of trips to highway and transit routes)	3632
COUNTIES	Number of counties in the study area	21
MEDIAN HOUSEHOLD INCOME	The regional median household income	72800
SIDEWALK MAXIMUM DENSITY	Maximum value used by the vehicle availability submodel for subzone sidewalk density (miles of sidewalk per square mile) – this value replaces the Pedestrian Environment Factor	40.0
REPLICATE MAXIMUM	Maximum number of times a specific survey household can be matched to a synthetic household before resampling occurs	1900
RESAMPLE MAXIMUM	Maximum number of times resampling will occur if a survey household has reached the replicate maximum	100
MILITARY WORKERS PER PERSON	Number of workers per group quarters resident of military barracks	1.00
UNIVERSITY WORKERS PER PERSON	Number of workers per group quarters resident of college or university dormitories	0.385

OTHER 16-64 WORKERS PER PERSON	Number of workers per other group quarters resident ages 16-64	0.306
OTHER 65+ WORKERS PER PERSON	Number of workers per other group quarters resident ages 65 and older	0.12
MILITARY WORK TRIP RATE	Work trips per military worker	2.00
UNIVERSITY WORK TRIP RATE	Work trips per dormitory resident worker	0.77
OTHER 16-64 WORK TRIP RATE	Work trips per other group quarters resident worker ages 16-64	0.612
OTHER 65+ WORK TRIP RATE	Work trips per other group quarters resident worker ages 65 and older	0.24
RANDOM SEED	The program uses random numbers in the process that selects households to populate subzones. This keyword allows the user to repeat the same sequence of random numbers when set to a positive value.	211

Study area geography

The trip generation model uses the geographic input file GEOG_IN.TXT, to define various geographies. The fields in the file are listed in Table 37. GEOG_IN.TXT input fileTable 37.

Table 37. GEOG_IN.TXT input file

Variable	Description
SUBZONE	The trip generation subzone
COUNTY	A five digit code identifying the county where the subzone is located. The first two digits are the state Federal Information Processing Standards (FIPS) code for the state and the remaining three digits are the county FIPS code. (Example: DuPage County is 17043).
COUNTY NAME	Ten character county name enclosed in double quotes
STATE	Two character state code enclosed in double quotes
FIVE PERCENT PUMA	Five digit numeric code for five percent PUMAs (2010 Census)
ZONE	Modeling zone
CHICAGO	Variable is set to 1 when trip generation subzone is inside Chicago; 0 elsewhere
CBD	Variable is set to 1 when trip generation subzone is inside the Chicago CBD; 0 elsewhere. The first forty-seven trip generation subzones presently make up the Chicago CBD.
ROW-COLUMN	Variable identifies the portion of the region that previously provided the tables of cross-classified households for the procedure to disaggregate households within smaller trip generation subzones: 1 = Inner Chicago 2 = Outer Chicago and Inner Suburbs 3 = Mid Suburban 4 = Fringe and External Areas
AREA	Area of the trip generation subzone in square miles
CMAP	Variable is set to 1 when trip generation subzone is inside the CMAP planning area; 0 elsewhere

Household type table

Within the trip generation model households are classified by one of 624 different categories (HHTYPE) defined by the composition of the household. Households initially are cross classified by:

- **Adults –** four categories [1, 2, 3, 4 or more adults]
- **Workers –** four categories [0, 1, 2, 3 or more workers]
- **Children 15 years and younger –** four categories [0, 1, 2, 3 or more children]
- **Income categories –** four categories
- **Age of householder –** three categories [16-34, 35-64, or 65 and older categories].

Workers include all employed persons (classes of worker 1 through 8 in the 2018-2022 PUMS person file), a small number of family business, or farm unpaid workers. Six hundred and twenty-four different types of households are possible with this scheme. This is less than all possible combinations of the household stratifying variables because of the constraint that households must have a number of adults that is equal to or greater than the number of workers. Thus, there are only 13 possible combinations of adults and workers (rather than 16) as shown in Table 38.

Table 38. Adult-worker household types

		Adults in household			
		1	2	3	4 or more
Workers in household	0	1	2	3	4
	1	5	6	7	8
	2		9	10	11
	3 or more			12	13

The 13 adult-worker categories provide a key to the numbering of all 624 categories. The adult-worker categories are nested within the other stratifying variables of age of householder, income quartile and number of children. Table 39 shows the complete list of household-type definitions.

Table 39. Household-type definitions

Householder ≤ 34			35 ≤ Householder ≤ 64			Householder ≥ 65		
Household type	Children	Income category	Household type	Children	Income category	Household type	Children	Income category
1-13	0	1	209-221	0	1	417-429	0	1
14-26	1	1	222-234	1	1	430-442	1	1
27-39	2	1	235-247	2	1	443-455	2	1
40-52	3+	1	248-260	3+	1	456-468	3+	1
53-65	0	2	261-273	0	2	469-481	0	2
66-78	1	2	274-286	1	2	482-494	1	2
79-91	2	2	287-299	2	2	495-507	2	2
92-104	3+	2	300-312	3+	2	508-520	3+	2
105-117	0	3	313-325	0	3	521-533	0	3
118-130	1	3	326-338	1	3	534-546	1	3
131-143	2	3	339-351	2	3	547-559	2	3
144-156	3+	3	352-364	3+	3	560-572	3+	3
157-179	0	4	365-377	0	4	573-585	0	4
170-182	1	4	378-390	1	4	586-598	1	4
183-195	2	4	391-403	2	4	599-644	2	4
196-208	3+	4	404-416	3+	4	612-624	3+	4

The file tg_hhtype_lookup.csv (shown in Table 40) is used to define the 624 household types. Household income categories were derived from the household income options available in the My Daily travel survey. A similar file (tg_hhvtype_lookup.csv) is used to define the 624 household types that rely on household vehicles available rather than household income category. This file has the same format shown in Table 40.

Table 40. tg_hhtype_lookup.csv input file

Variable	Description
HOUSEHOLD TYPE	Code indicating cross-classification of household (as per numbering in Table 37).
ADULTS IN HOUSEHOLD	Coded 1, 2, 3, or 4 with code 4 equal to 4 or more.
WORKERS IN HOUSEHOLD	Coded 0, 1, 2, or 3 with code 3 equal to 3 or more.
CHILDREN IN HOUSEHOLD	Coded 0, 1, 2, or 3 with code 3 equal to 3 or more.
HOUSEHOLD INCOME QUARTILE	Coded 1, 2, 3, or 4 from low to high income categories (in 2019 dollars). 1 = under $30,000 2 = $30,000 - $59,999 3 = $60,000 - $99,999 4 = $100,000 and over
AGE OF HOUSEHOLDER CODE	Age of head of householder index: 1 = householder age 16-34 2 = householder age 35-64 3 = householder age 65+

Load synthetic households

The synthetic households from UrbanSim are read into the program, populating each subzone with a list of unique households identified by household type. These are stored in POPSYN_HH.CSV.

Household vehicle availability

The household vehicle ownership sub-model is applied to estimate the vehicle ownership levels for each household. This effectively adds an additional dimension (vehicles available in the household) to the cross-classification of households. Household types are redefined with vehicle availability, replacing income quartile.

The household vehicle availability sub-model is a discrete choice logit model similar to models that predict mode choice behavior. There are three or four possible vehicle level choices for each household, depending on the number of adults (workers plus nonworking adults) in the household. Single-adult households may have zero, one, or two or more vehicles. Larger households have the alternatives of zero, one, two, or three or more vehicles.

Each vehicle availability level has an associated utility. In these logit models, utilities are weighted linear combinations of household and subzone variables that have been entered or developed in earlier steps. Model estimation consists of determining which variables best explain observed vehicle availability levels and the relative importance of these variables in the utility expressions.

The utilities also may include bias constants that indicate preferences toward certain vehicle availability levels that are not otherwise accounted for by the utility expressions.

The vehicle availability sub-models have the following general form:

$$\text{Prob\{Vehicle Availability Level } i\} = \frac{e^{u_i}}{\sum_i e^{u_i}}.$$

Where u_i is the utility of household vehicle availability level *i*, which is computed by the linear equation:

$$u_i = \sum_j \alpha_{ij} H_j + \beta_i.$$

For these utility equations:

- The model coefficient α_{ij} is the weight attached to the *j*'th independent household variable for household vehicle availability level *i*.
- H_j is the *j*'th independent household variable (number of workers in household, household income quartile, etc).
- The constant β_i is the bias toward vehicle availability level *i*, and it must equal zero for at least one alternative.

Model coefficients for the sub-models were estimated with data from the CMAP Travel Tracker survey. After completing estimation, several incremental changes to the sub-models were introduced. This includes some model coefficients that were similar for different levels of vehicle availability and the utility equations were modified to replace these multiple coefficients with a single coefficient across utility equations. Also, a few marginally significant model coefficients became insignificant and were dropped. Lastly, a complete set of bias coefficients for location and age of householder were estimated.

Model calibration was completed by adjusting bias coefficients, so that estimated and observed levels of household vehicle ownership matched within the study area. Additional bias constants also were included to account for the new age of householder household variable. The modified vehicle availability sub-models finally were coded into the CMAP trip generation model code. The original pedestrian environment factor (PEF) has been replaced with the sidewalk density index (SDI), which was capped at the same maximum value as the PEF. The PEF was a walkability metric estimated from street network data provided by NAVTEQ. The SDI is derived from CMAP's 2018 Regional Sidewalk Inventory and measures the miles of sidewalk per square mile. However, only 86 percent of the modeling subzones, located exclusively in the seven-county area, have data from this inventory. For those subzones, SDI is equal to of subzone area. For the remaining subzones, the SDI value is estimated based on a linear regression model using the former walkability metric.

The vehicle availability sub-model was not re-estimated, because it provides a good match to observed data. The following table (Table 41) lists the coefficients in the logit equation utilities that comprise the household vehicle availability sub-model. Separate models were estimated and calibrated for three different sized households defined by the total adults (workers plus nonworking adults) in the household. Each line in the table lists the model coefficients for estimating the utility attached to that level of vehicle availability for one of the three household type models.

Table 41. Vehicle availability sub-model coefficients

	One adult household vehicle availability			Two adult household vehicle availability				Three or more adult household vehicle availability			
Variable	0	1	≥ 2	0	1	2	≥ 3	0	1	2	≥ 3
Sidewalk density index (Maximum=40)	0.0616	0.0319		0.1280	0.0631	0.0336		0.1703	0.0659	0.0659	
Workers?											
≥ 1 (0/1)		0.4731	0.4731			0.6940	0.6940		1.114	1.114	1.114
≥ 2 (0/1)						0.5198	0.5198			0.7934	0.7934
≥ 3 (0/1)											1.389
Income quartile?											
≥ 2 (0/1)		1.182	1.766		1.702	2.466	2.466		0.9492	1.487	1.487
≥ 3 (0/1)		0.9910	1.690			0.8650	0.8650			0.8723	1.571
= 4 (0/1)			0.467			0.4517	0.8827			1.390	1.834
Commute auto mode share		4.677	4.677			5.284	5.284			4.959	4.959
Nonworking adults											0.1491
Children?											
≥ 1 (0/1)						0.2218					
Household location bias											
1. Inner Chicago		-2.600	-5.077		2.018	-2.827	-4.393		2.806	-1.836	-1.631
2. Rest of Chicago and inner suburbs		-2.676	-4.823		2.259	-2.637	-3.944		2.552	-2.139	-1.789
3. Mid-suburbs		-2.869	-4.914		2.151	-2.728	-4.126		1.547	-2.783	-2.668
4. Far suburbs and fringe		-3.082	-4.984		1.925	-3.144	-4.302		2.272	-2.430	-2.278
Age of householder bias											
16-34		0.392	0.394		0.392	0.394	0.403		0.392	0.394	0.403
35-64		0.401	0.465		0.401	0.465	0.574		0.401	0.465	0.574
≥ 65		0.249	0.218		0.249	0.218	0.007		0.249	0.218	0.007

The vehicle availability sub-model is applied to each of the enumerated households. The household type variable is then revised to include household vehicle availability (HHVTYPE). It has the same structure as the original household type variable listed in Table 39, except that four levels of vehicle availability (0, 1, 2 or 3 or more vehicles) replace the income quartile value. At the conclusion of the vehicle availability model, the file, SIMULATED_HHVEH.TXT, is written to store the category of simulated vehicles available for each household, so that it can be used in the trip enumeration process. The contents of this file are displayed in Table 42 below.

Table 42. SIMULATED_HHVEH.TXT

Variable	Description
PUMS SERIAL NUMBER	PUMS household serial number for the enumerated household
VEHICLE AVAILBILITY CATEGORY	Simulated vehicle availability category for the enumerated household: 0=0 vehicles, 1=1 vehicle, 2=2 vehicles, 3=3+ vehicles
RECORD NUMBER	The record number of the household in POPSYN_HH.CSV

Once the number of household vehicles has been simulated, the next step is separating work-from-home households from households that have no workers working from home. The file, HH_WFH_STATUS.CSV, which was created by the work-from-home allocation model, is used to determine which group each household belongs to.

Household trip productions and attractions

The subzones have been populated with the households from the population synthesis. The next step is to pair each one with a comparable household from the My Daily Travel survey, so that actual trips can be assigned to specific households. Household trip generation can then be accomplished without the intermediate step of computing trip generation rates. Table 43 below defines the 49 trip purposes recognized in the trip generation model.

Table 43. CMAP trip purposes

Origin activity	Destination activity	Trip code
Worker	Workplace-low income households (a)	1
	Workplace-high income households (a)	2
	Work-related (a)	3
Home (P)	School (a)	4
	Non-home or work at residences [visit] (a)	5
	Non-home or work not at residences (a)	6
	Shop (a)	7
Work (P)	Non-home or work at residences [visit] (a)	8
	Non-home or work not at residences (a)	9
	Shop (a)	10
Work (O)	Work (d)	11
Non-home or work at residences [visit] (O)	Non-home or work at residences [visit] (d)	12
	Non-home or work not at residences (d)	13
	Shop (d)	14
Non-home or work not at residences (O)	Non-home or work at residences [visit] (d)	15
	Non-home or work not at residences (d)	16
	Shop (d)	17
Shop (O)	Non-home or work at residences [visit] (d)	18
	Non-home or work not at residences (d)	19
	Shop (d)	20
Nonworking adult	School (a)	21
	Non-home at residences (a)	22
Home (P)	Non-home not at residences (a)	23
	Shop (a)	24

Non-home at residences [visit] (O)	Non-home at residences (d)	25
	Non-home not at residences (d)	26
	Shop (d)	27
Non-home not at residence (O)	Non-home at residences (d)	28
	Non-home not at residences (d)	29
	Shop (d)	30
Shop (O)	Non-home at residences (d)	31
	Non-home not at residences (d)	32
	Shop (d)	33
Child 12-15	School (a)	34
Home (P)	Non-home at residences (a)	35
	Non-home not at residences (a)	36
	Shop (a)	37
School (P)	Non-home at residences (a)	38
	Non-home not at residences (a)	39
	Shop (a)	40
Non-home at residences (O)	Non-home at residences (d)	41
	Non-home not at residences (d)	42
	Shop (d)	43
Non-home not at residences (O)	Non-home at residences (d)	44
	Non-home not at residences (d)	45
	Shop (D)	46
Shop (O)	Non-home at Residences (D)	47
	Non-home Not at Residences (D)	48
	Shop (D)	49

A file of travel survey households is read from HI_HHENUM_IN.TXT (shown in Table 44). This file has a serial number identifying the household, the household type, characteristics of the household, and trips made by the household during the weekday surveyed. Trips in this input file were tabulated in the same manner as earlier versions of the program. Trip files that were the basis for trip generation rates simply were resumed by household. Trip purpose definitions are unchanged and rules for the linking of trips are the same as in previous versions.

The end result is 48 different trip purposes, including 19 for workers, 13 for nonworking adults, and 16 for children between the ages of 12 and 15, as listed in the preceding table. Note that in the input file HI_HHENUM_IN.TXT, 49 trip purposes are allowed for the splitting of home productions-work attractions into home-work trips made by low and high income households by the trip generation model. The first trip purpose in the file includes all home productions to work attractions, while the second trip purpose serves as an empty placeholder.

Each non-work-from-home synthetic household is matched to a survey non-work-from-home household using the following process, which encompasses four mutually exclusive methods. These include:

- If there are seven or more survey households with the same vehicle-based household type (HHVTYPE) within the resident PUMA, one of these households is selected randomly as the match (match category 1).

- If there are less than seven survey households in the PUMA with the correct HHVTYPE, they are combined with households of the same type in the ring of adjacent PUMAs around the central PUMA. If this results in at least seven households of the specific type, one of these households in this larger group is selected randomly as the match (match category 2).
- If there are less than seven survey households of the same type in the adjacent geography, but seven or more in the full study area, one of these households is selected randomly as the match (match category 3).
- If there are less than seven survey households of the same type in the full study area, a final match category (match category 4) ensures all remaining households find a match in the survey households, regardless of how small the likelihood is that the household occurs. Households initially are categorized into 13 groups, which correspond to the combinations of adults and workers in the households. These groups then are subdivided based on the number of children in the household (0, 1, or 2+). Large groups (those with at least 500 households) are further subdivided based on the number of household vehicles (0, 1, or 2+). At this stage, groups with fewer than 20 households are recombined based on the adult-worker-children combinations. The end result is 43 household categories that all have a minimum of 20 households to select from.

A similar process is used to match synthetic work-from-home households to work-from-home survey households. A notable difference is that work-from-home households are matched based on the resident PUMA and the household type based on income (HHTYPE) not the number of vehicles. Work-from-home matches use HHTYPE because a strong income component was discovered in analyzing these households for the work-from-home allocation model. These households use the same match priority as the non-work-from-home households. For the final match category, households are divided into groups based on the adult-worker combinations, the presence or absence of children, and the household income category. Groups with fewer than 20 households are aggregated based on the adult-worker-children combinations to achieve that minimum number of members. The result is 24 household categories to select survey households from.

Table 44. HI_HHENUM_IN.TXT input file

Variable	Description
PUMA	A seven-digit code with the first two digits equal to the FIPS state code and the last five digits equal to the PUMA
HHVTYPE	The revised household type code based on vehicle availability in the survey household
SERIAL NUMBER	The eight-digit household code from the My Daily Travel
ADULT	Number of adults in survey household
WORKER	Number of workers in survey household
NONWORKER	Number of non-workers in survey household
CHILD	Number of children aged 0-15 in survey household
CHILD 12-15	Number of children aged 12-15 in survey household
VEHICLES	Number of vehicles available in survey household
TRIPS1	Household's home production-work attraction trips

TRIPS2	Empty placeholder
TRIPS3	Household's home production-work related attraction trips
….	Additional columns for TRIPS4 through TRIPS48
TRIPS49	Household's shop-shop trips by child 12-15.

The processing of the survey households into match categories is handled outside of the trip generation model. The results are read into the model via files. Work-from-home households use the information in HHID_choices1.csv and HHID_choices2.csv. The first file, whose fields are shown in Table 45, provides the following information: the resident PUMA, the household type code (HHVTYPE), the match category used (for tracking purposes), and the positions of the first and last potential households that can be matched against in the array in the second file. The second file (HHID_choices2.csv) is an array of household serial numbers that correspond to the set of households available within each match category. For this final match category, it lists the households available within each subgroup discussed above. Households within each selection category do not have an equal probability of being chosen. Household expansion weights from the My Daily Travel survey are used to develop a cumulative probability distribution within each selection category to better reflect observed household data and minimize the likelihood that extremely rare households replicate many times during the enumeration process. The file also contains the probability of each household being selected from its match category.

Comparable files are used to enumerate trips for the WFH households (HHID_wfh1.csv and HHID_wfh2.csv). These contain the same fields as their counterparts; however, the household type identifier in the work-from-home version is HHTYPE. Two values in trip_generation_model.yaml help prevent individual survey households from being replicated an excessive number of times during the enumeration process. A maximum replicate value is provided. This is not a hard ceiling on the value but once a survey household has been selected this many times, it will trigger resampling of the synthetic household, up to the maximum number of iterations identified.

Table 45. HHID_choices1.csv input file

Variable	Description
PUMA	A seven-digit code with the first two digits equal to the FIPS state code and the last five digits equal to the PUMA
HOUSEHOLD TYPE	Household vehicle type code (HHVTYPE)
MATCH CATEGORY	A numeric code identifying the household match methodology used to match the enumerated household with a survey household: 1 = household in PUMA 2 = household in adjacent PUMA 3 = household in region 4 = final match category
CHOICES START	Identifies the start location in HHID_choices2.csv of the first potential household to match against based on PUMA-HHVTYPE
CHOICES END	Identifies the end location in HHID_choices2.csv of the final potential household to match against based on PUMA-HHVTYPE

Table 46. HHID_choices2.csv input file

Variable	Description
SERIAL NUMBER	The eight-digit household code from the My Daily Travel survey
PROBABILITY	The cumulative probability (based on My Daily Travel household weights) of the specific household being selected within the PUMA-HHVTYPE category

After the household trip enumeration process has finished, two files are created that summarize the enumeration selections. The file households_selected_noWFH_enumeration_count.csv lists the travel survey identification number of each potential household that could have been chosen during enumeration for households with no teleworkers and lists the number of times each was selected. A similar file (households_selected_WFH_enumeration_count.csv) provides the same information for survey households used to complete the enumeration for households with at least one teleworker.

Optionally, enumerated synthetic households can be written to the output file HI_HHENUM_TRIP_OUT.CSV (shown in Table 47). Trip productions and attractions are summed by subzone, organized, and reported as in previous versions of the trip generation model.

Table 47. HI_HHENUM_TRIP_OUT.CSV output file

Variable	Description
HOUSEHOLD RECORD	Record number of household in POPSYN_HH.CSV
SUBZONE	The trip generation subzone, which must be in sequence in the file from low to high values
SURVEY NUMBER	Eight digit household serial number assigned to the household in the My Daily Travel survey
PUMA	A seven digit code with the first two digits equal to the FIPS county code and the last five digits equal to the PUMA
HHTYPE	The original income quartile household type code
HHVTYPE	The revised vehicle availability household type code
VEHICLES	Modeled household vehicles available
ROW-COLUMN	Geographic identifier: 1 = Inner Chicago 2 = Outer Chicago and inner suburbs 3 = Mid suburban 4 = Fringe and external areas
PUMS SERIAL NUMBER	Thirteen digit household serial number from ACS data
MATCH CATEGORY	Matching method used to enumerate household trips
WORK-FROM-HOME	Work from home flag for household
INCOME	Household income (2019 dollars)
CHILDREN	Number of children aged 0-15 in household
TRIPS1	Household's home-work trips by low income worker
TRIPS2	Household's home-work trips by high income worker

TRIPS3	Household's home production-work related attraction trips
….	Additional columns for TRIPS4 through TRIPS48
TRIPS49	Household's shop-shop trips by child 12-15

Optionally, a summary of the subzone-level trip productions by household type can be written to PRODS_HH_OUT.CSV. The productions can be further segmented by trip type and by households with and without teleworkers. The format of this file is listed in Table 48.

Table 48. PRODS_HH_OUT.CSV contents

Variable	Description
SUBZONE	Trip generation subzone
WORK-FROM-HOME	Work from home flag
HOUSEHOLD CATEGORY	Household category: 1 = 0 children and 0 vehicles 2 = 1 or more children and 0 vehicles 3 = 0 children and 1 or more vehicles 4 = 1 or more children and 1 or more vehicles
TRIP TYPE	Type of trip production (1-49)
TRIP PRODUCTIONS	Total trip productions

Group quarters trip generation

The CMAP survey did not specifically gather travel data from group quarters residences. The ACS does collect some limited information on persons in group quarters, including institutionalized and non-institutionalized persons. For trip generation, it is assumed institutionalized persons do not travel independently, which means group-quarters trip generation applies only to individuals in dormitories, military barracks, group homes, and the like. A second assumption is all children in group quarters are in institutions and do not travel independently.

As noted, four types of group quarters residents remain to be considered, including:

- Persons in military barracks
- Persons in college/university dorms
- Persons aged 16 to 64 in other types of group quarters
- Persons aged 65 or more in other types of group quarters

A trip generation subzone input file, called GQ_IN.TXT, containing workers and nonworking adults in non-institutionalized group quarters, is read by the program. It contents are listed in Table 49.

Table 49. GQ_IN.TXT input file

Variable	Description
SUBZONE	The trip generation subzone
GQ MILITARY	Persons in military barracks

GQ UNIVERSITY	Persons in college/university dormitories
GQ OTHER (16-64)	Persons in other group quarters aged 16 to 64
GQ OTHER (≥ 65)	Persons in other group quarters aged 65 and older

Group-quarters trip rates for non-work trips were developed using data from the My Daily Travel survey. These rates are read into the model via an input file (GQ_TRIPRATES.CSV). The format of the file is listed in Table 50 below. Home-work trip rates were developed from the 2014-18 ACS PUMS person file, which provided the number of workers per person in group quarters and the home-work trip rate. Trip rates for group-quarters, non-work trips are the household survey rates for workers and non-workers in single person, low income, zero vehicle households.

Table 50. Group quarters trip rates

Variable	Description
TRIP TYPE	Type of trip production (3-33) – [home-work trip rates are read from trip_generation_model.yaml]
TRIP PRODUCTION RATE	My Daily Travel trip generation rate

Once group quarters trip productions are calculated, an optional summary file of the data (PRODS_GQ_OUT.CSV) can be written. The contents of this file are listed in Table 51.

Table 51. PRODS_GQ_OUT.CSV file contents

Variable	Description
SUBZONE	The trip generation subzone
GQ TYPE	Type of group quarters (military, college, other16-64 or other65)
GQ PERSONS	Number of persons in non-institutionalized group quarters
GQ WORKERS	Number of workers in non-institutionalized group quarters
GQ NONWORKERS	Number of nonworkers in non-institutionalized group quarters
TRIP TYPE	Type of trip production (1, 3-33)
TRIP PRODUCTIONS	Total trip productions

Allocation of non-home trip ends

At this point in the logic of the CMAP trip generation model, all trips generated by persons residing inside the modeled study area (both in households and group quarters) are accounted for. Due to the synthetic household information, the location of the home end of each trip also is known. What remains to be determined is where the non-home trip ends are located. These are allocated to trip generation subzones in this step of the model.

To allocate an attractiveness share, a function of employment, households, or school enrollment (depending on the trip's purpose) is calculated for each subzone. Since Chicago's central business district is atypical in its mix of employment, subzones within the central business district are weighted differently from non-central-business-district subzones. After totaling the shares for all subzones, trip ends then are proportionally allocated to subzones by these shares. The input ATTR_IN.TXT (shown in Table 52) is read into the model to supply the employment values necessary for the allocation.

Table 52. ATTR_IN.TXT input file

Variable	Description
SUBZONE	The trip generation subzone
RETAIL EMPLOYMENT	Retail employment in the trip generation subzone
TOTAL EMPLOYMENT	Total employment in the trip generation subzone
FRACTION OF HIGH EARNERS	The fraction of workers working in the subzone with earnings above the regional median worker earnings, developed from the 2017-2021 CTPP

The two employment quantities in ATTR_IN.TXT are derived from the employment data provided by UrbanSim. The fraction of higher income workers working in the subzone was estimated from the 2017-21 CTPP. Weights for allocating non-home productions or trip origins are listed in Table 53 below. The tables display coefficients that have been normalized since the values are only used to create relative weights for the allocations.

Table 53. Allocation weights for origin non-home trip ends

				Employment category					
		Households		Retail		Non-Retail		Total	
Origin activity	**Destination activity**	**CBD**	**Non-CBD**	**CBD**	**Non-CBD**	**CBD**	**Non-CBD**	**CBD**	**Non-CBD**
Worker (non-WFH)	Non-home/Work at residence (A)							1.000	3.048
Work (P)	Non-home/Work not at residence (A)							1.000	0.765
	Shop (A)							1.000	1.901
Work (O)	Work (D)							1.000	1.837
Non-home/work at residence (O)	All destinations (D)	1.000	1.401						
Non-home/work not at residence (O)				1.000	0.401	0.050	0.097		
Shop (O)				1.000	1.129				
Worker (wfh)	Non-home/work at residence (a)							1.000	6.545
Work (P)	Non-home/Work not at residence (A)							1.000	2.525
	Shop (A)							1.000	4.900
Work (O)	Work (D)							1.000	2.059
Non-home/work at residence (O)	All destinations (D)	1.000	1.010						
Non-home/work not at residence (O)				1.000	0.963	0.038	0.195		
Shop (O)				1.000	1.227				
Nonworking adult									
Non-home at residence (O)	All destinations (D)	1.000	1.281						
Non-home not at residence (O)				1.000	0.734	0.046	0.176		
Shop (O)				1.000	1.441				
Child 12-15									
School (P)	All attractions (A)	1.000	2.599						
Non-home at residence (O)	All destinations (D)	1.000	22.00						
Non-home not at residence (O)					1.000		0.137	0.065	
Shop (O)				1.000	1.000				

The formulas used to determine the attractiveness shares for non-home productions are stored in coefficients_nonhome_productions.csv. The contents of the file are listed in Table 54. This input file simplifies the process of updating the coefficients used to allocate non-home trip productions. A separate file (coefficients_attractions.csv) stores the weight formulas to allocate trip attractions and has the same format shown in Table 54.

Table 54. coefficients_nonhome_productions.csv input file

Variable	Description
TRIP TYPE	Type of trip production (8-20, 25-33, 38-49)
WFH FLAG	Telework flag: 0=no, 1=yes
CBD FLAG	CBD flag: 0=no, 1=yes
WEIGHT FORMULA	The formula to develop subzone allocation weights (e.g., 0.065 * subzone total employment)

Similar to Table 53, Table 55 lists the weights used for allocating attractions or trip destinations. These weights were updated using My Daily Travel survey data and SCHOOL_IN.CSV, an input file of combined high school and university enrollment totals by subzone. SCHOOL_IN was developed specifically to guide the allocation of home-school trips for adults. It is comprised of data from the National Center for Education Statistics and the Illinois State Board of Education. It includes both public and private schools.

Table 55. Allocation weights for destination trip ends

				Employment category						
		Households		Retail		Non-Retail		Total		
Origin activity	**Destination activity**	**CBD**	**Non-CBD**	**CBD**	**Non-CBD**	**CBD**	**Non-CBD**	**CBD**	**Non-CBD**	**Enrollment**
Worker (non-WFH)	Workplace (A)							1.000	0.755	
Home (P)	Work-related (A)			1.000	3.103	0.241	0.634			
	School (A)									1.000
	Non-home/work at residence (A)	1.000	1.132							
	Non-home/work not at residence (A)			1.000	0.623	0.042	0.180			
	Shop (A)			1.000	1.430					
Work (P)	Non-home/work at residence (A)	1.000	2.309							
	Non-home/work not at residence (A)			1.000	0.383	0.095	0.106			
	Shop (A)			1.000	1.882					
Work (O)	Work (D)							1.000	1.550	
All origins (O)	Non-home/work at residence (D)	1.000	1.414							
	Non-home/work not at residence (D)			1.000	0.401	0.051	0.099			
	Shop (D)			1.000	1.106					

Worker (WFH)										
	Workplace (A)							1.000	0.755	
	Work-related (A)			1.000	3.103	0.241	0.634			
	School (A)									1.000
Home (P)	Non-home/work at residence (A)	1.000	0.567							
	Non-home/work not at residence (A)			1.000	0.941	0.024	0.208			
	Shop (A)			1.000	0.938					
Work (P)	Non-home/work at residence (A)	1.000	0.738							
	Non-home/work not at residence (A)			1.000	0.641	0.027	0.158			
	Shop (A)			1.000	0.556					
Work (O)	Work (D)							1.000	2.722	
All origins (O)	Non-home/work at residence (D)	1.000	0.990							
	Non-home/work not at residence (D)			1.000	0.938	0.036	0.189			
	Shop (D)			1.000	1.217					
Nonworking adult										
	School (A)									1.000
Home (P)	Non-home at residence (A)	1.000	1.173							
	Non-home not at residence (A)				1.000		0.263	0.088		
	Shop (A)			1.000	1.634					
	Non-home at residence (D)	1.000	0.985							
All Origins (O)	Non-home not at residence (D)			1.000	0.754	0.047	0.180			
	Shop (D)			1.000	1.530					
Child 12-15										
	School (A)	1.000	1.511							
	Non-home at residence (A)	1.000	36.33							
Home (P)	Non-home not at residence (A)			1.000	2.501	0.031	0.324			
	Shop (A)			1.000	1.000					
	Non-home at residence (A)	1.000	16.44							
School (P)	Non-home not at residence (A)				1.000		0.922	0.058		
	Shop (A)			1.000	1.000					
	Non-home at residence (D)	1.000	16.44							
All Origins (O)	Non-home not at residence (D)				1.000		0.922	0.058		
	Shop (D)			1.000	1.000					

Certain non-home trip ends (residences, shops, schools) are restricted in subzones located entirely within airport boundaries. The file airport_sz.csv identifies which subzones are within airport boundaries and their opening year. This allows flexibility for the trip generation model to accommodate future airport expansion or the opening of new airports. After workplace attractions are allocated, they are factored into workplace attractions for workers with above and below median earnings by the factors in the ATTR_IN.TXT input data set.

Optionally the file FIRST_PA_OUT.CSV can be created to summarize all preliminary household and group quarters productions and attractions by trip type. The contents of the file are listed in Table 56. Work trip productions and attractions will proceed through one final processing step.

Table 56. FIRST_PA_OUT.CSV contents

Name	Description
SUBZONE	The trip generation subzone
ZONE	Modeling zone
WORK-FROM-HOME	Work from home flag for household
TRIP TYPE	Type of trip production (1-49)
HOUSEHOLD PRODUCTIONS	Total trip productions of household residents
HOUSEHOLD ATTRACTIONS	Total trip attractions of household residents
GQ PRODUCTIONS	Total trip productions of non-institutionalized group quarters residents
GQ ATTRACTIONS	Total trip attractions of non-institutionalized group quarters residents

External trip ends

After accounting for all trips made by study area residents, the next step is to factor out trip ends for trips with one trip end outside the modeled study area. In the model, only home-workplace trips for households are factored in this manner (e.g., trip types 1 and 2). This is largely due to data limitations. The CMAP survey sample is far too small for any reliable estimation of these factors and only the CTPP commuting data are available. Additionally, home-workplace trips are lengthier than other trip purposes and more likely to have one trip end outside the area modeled. Trips by persons residing in group quarters are not factored for external trip ends.

Summary home-workplace tables were first developed from CTPP Table B302103, using data from the five-year (2017-2021) ACS estimates. These tables contain the home to workplace commutes by mode for all workers. Tract-to-tract journey-to-work flow totals were scaled to match county-to-county journey-to-work flow totals to adjust for data suppression at the tract geography level. After removing workers who work at home, the tables have three components. These components include workers who travel to work and who (1) live and work inside (internal-internal) the modeled study area; (2) workers who live outside (external) the study area but work inside (internal) the study area; and workers who (3) live inside (internal) the study area but work outside (external) the study area. Total workers for each group by tract are aggregated to the 5 percent PUMAs geography prior to calculating the external trip factors.

External trip factors to adjust the home productions and workplace attractions in a 5 percent PUMA were developed in the following manner. The home production factor is the fraction of all trips from internal households (internal-internal plus internal-external) linked to external workplaces (internal-external), and the workplace attraction factor is the fraction of all trips to internal

workplaces (internal-internal plus external-internal) that are linked to external households (external-internal). Home productions in 5 percent PUMS are factored first followed by workplace attractions. These external trip factors are stored in an input file named EXT_IN.TXT (shown in Table 57).

Home Production Factor = [internal-external] / ([internal-internal] + [internal-external])

Workplace Attraction Factor = [external-internal] / ([internal-internal] + [external-internal])

Table 57. EXT_IN.TXT input file

Variable	Description
FIVE PERCENT PUMA	Seven digit numeric code for five percent PUMAs (two digit county FIPS code plus five digit code for five percent PUMA)
EXTERNAL PRODUCTION FACTOR	Fraction of home productions in home-workplace trips linked to external workplace attractions
EXTERNAL ATTRACTION FACTOR	Fraction of workplace attractions in home-workplace trips linked to external home productions

The factors in EXT_IN.CSV are applied to the household productions and attractions based on the PUMA value. External home-work attractions are then balanced to external productions and adjustments are distributed proportionally within PUMAs. An optional file (EXTERNAL_PA_OUT.CSV) of the adjusted external home-work productions and attractions can be written, and its contents are listed Table 58. The final step is subtracting the external productions and attractions from the subzone-level home-work trip productions and attractions.

Table 58. EXTERNAL_PA_OUT.CSV contents

Name	Description
SUBZONE	The trip generation subzone
ZONE	Modeling zone
WORK-FROM-HOME	Work from home flag for household
TRIP TYPE	Type of trip production (1-2)
EXTERNAL HOUSEHOLD PRODUCTIONS	Total external trip productions of household residents
EXTERNAL HOUSEHOLD ATTRACTIONS	Total external trip attractions of household residents

Create final vehicle trip output file

Lastly the trip generation model code creates two fixed-width output files (TRIP49_PA_OUT.CSV and TRIP49_PA_WFH_OUT.CSV) for use in the mode choice-destination choice model. These files contain the final productions and attractions for internal trips for all residents in the modeled study area. Both files have the same structure, shown in Table 59.

Table 59. TRIP49_PA_OUT.CSV output file

Variable	Description	Width
SUBZONE	Trip generation subzone	6
ZONE	Modeling zone number	6

TRIP TYPE	Trip type	2
HH PRODUCTIONS	Household trip productions/origins for trips within the modeled study area	9
HH ATTRACTIONS	Household trip attractions/destinations for trips within the modeled study area	9

A final file created by the trip generation model is TG_HHENUM_OUTPUT.CSV (shown in Table 60), which lists the subzone and zone of each enumerated household, as well as the HHVTYPE. This file is subsequently used in the mode choice-destination choice model.

Table 60. TG_HHENUM_OUTPUT.CSV output file

Variable	Description
SUBZONE	Trip generation subzone
ZONE	Modeling zone number
HOUSEHOLD VEHICLE TYPE	Household type code based on vehicle availability

Mode choice-destination choice model

Unlike past versions of CMAP's trip-based model, the current model does not execute trip distribution and mode choice as two separate, sequential models. Rather, an integrated mode choice and destination choice model is implemented as a hierarchical nested logit model. This model was estimated using full-information, maximum-likelihood methods.

The 49 types of trips created during trip generation are aggregated into the following trip purposes, which are modeled by the full mode choice-destination choice model.

- Home-based work trips for residents of low-income households (HBWL). The regional median household income value is used to determine whether households fall into the low- or high-income categories.
- Home-based work trips for residents of high-income households (HBWH).
- Home-based shopping trips (HBS).
- All other home-based trips not included in the first three categories (HBO).
- All non-home-based trips (NHB).

Nine discrete modal alternatives are included in the model, listed in Table 61. This represents a major improvement from the binary auto-transit mode choice alternatives included in prior versions of the model. A private auto group nest includes single- and multiple-occupancy vehicles. The hired auto group nest includes three modal alternatives not previously available in CMAP's trip-based model, including taxi and transportation network company (TNC) options for passengers using either the regular or the shared-ride services. The final group of alternatives (which is not implemented as a nest) includes the non-auto modes of transit, walking and biking. Past versions

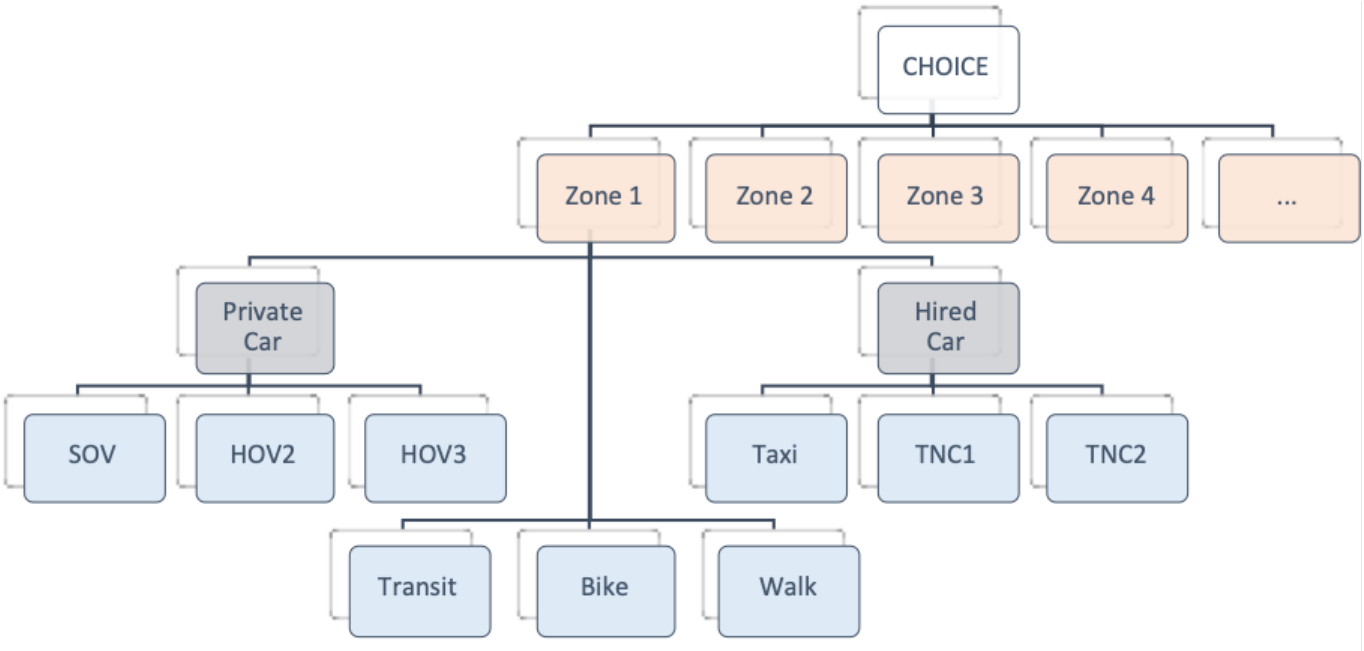
of CMAP's trip-based model did not include non-motorized modes of transportation (walking and bicycling) beyond the trip generation model.

Table 61. Modal alternatives

Modal alternative	Group	Description
1	Private Auto Nest	Single occupant private vehicles
2		High occupancy private vehicles – 2 occupants
3		High occupancy private vehicles – 3 or more occupants
4	Hired Auto Nest	Taxi
5		Transportation Network Company – regular ride hail
6		Transportation Network Company – shared ride hail
7		Transit
8		Walk
9		Bicycle

Figure 11 shows the structure of the hierarchical nested logit model. The top of the hierarchy is the attraction zone for each trip. This is the non-home end for all home-based trips and the trip destination for all other trip purposes. No sampling of alternatives is implemented during model application. So, for each trip, the total mode and destination alternatives is 32,688 (9 modes x 3632 zones).

Figure 11. Mode-destination choice model logit structure



Each modal alternative considered has its own cost components that factor into the utility calculations. These cost components and utility calculations are discussed in the following sections. Table 62 lists the general cost attributes included in the calculations that apply across modes. The procedures to calculate costs for private auto trips, and transit access and egress were ported over from the prior version of CMAP's trip-based model.

Table 62. Mode-destination choice cost components

Value	Description
50	Number of trips used to determine average impedance
3	Walk speed (MPH)
14	Driver's value of time per minute (cents)
2.0	Ratio of out of vehicle time (OVT) to in-vehicle time (IVT)
35	Auto fixed cost for driver (cents)
20	Auto fixed cost for passenger (cents)
2	Walk time from park and ride lot to platform (minutes)
Zonal approach speeds	**Area definition**
7 (auto), 5 (bus)	Chicago Central Area zone
15 (auto), 10 (bus)	Chicago zone outside of the Central Area
20 (auto), 12 (bus)	dense suburban CBD zone
30 (auto), 17 (bus)	remaining suburban zones

Private auto costs

Auto operating costs are calculated using the information in the M023 file. Auto times and distances are pulled using zone-to-zone assigned times and distances. Matrices of morning peak conditions are applied to home-based work trips, and matrices of midday conditions are applied to all other trip types.

Parking costs are estimated using different methods depending on the trip purpose and destination. As noted, home-based work trips to the central area use Monte Carlo simulation and the central business district parking file data to estimate parking costs, including the possibility of free parking. Parking costs for destination zones outside of Chicago's central area are applied using the rates Table 63. Currently, these hourly rates are the same for all trip purposes, but the flexibility exists to alter the rates by trip purpose.

Table 63. Private auto costs

Hourly parking costs (cents)	Area definition
80	Chicago Central Area zone
10	Chicago zone outside of the Central Area
20	dense suburban CBD zone
0	remaining suburban zones
Parking duration (hours)	
10 – home-based work trips	
6 – all other home-based trips	
3 – non-home based trips including visitors	

Hired auto costs

The same time and distance skims used for private autos are also used for the utility calculations for hired autos. In addition, these modal options have additional costs reflected in fares and surcharges, which are discussed below. These cost components are stored in the file cmap_trip_config.yaml, which is used by the mode choice-destination choice model.

Taxi

A single set of taxi rates based on Chicago medallion rates for in-city trips is used to estimate rates for these trips. Because fares for taxi trips outside of Chicago have similar rates and are far less numerous, the same single set of rates is used. Table 64 below lists the taxi rates and wait times that are used in the cost estimation. The airport departure surcharge is applied to specific zones for O'Hare and Midway airports.

Table 64. Taxi costs

Cost (cents)	Description
325	Initial base fare
225	Per-mile cost
41.667	Cost per minute, regulated at 25 cents per 36 seconds
400	Airport departure surcharge
Wait time (minutes)	**Area definition**
1 (peak), 2 (offpeak)	Chicago Central Area zone
5 (peak), 7 (offpeak)	Chicago zone outside of the Central Area
10	dense suburban CBD zone
15	remaining suburban zones

Transportation network companies

The TNC cost structure is similar to that used for taxis, however separate costs are applied for pooled versus regular service, which are summarized in Table 65. TNCs implement a downtown Chicago surcharge applied to specific zones. In addition, a special surcharge is applied to Chicago's airports and major attractions.

Table 65. Transportation network company costs

Cost (cents)		Description
TNC	***Pooled***	
136 (peak), 163 (offpeak)	68 (peak), 128 (offpeak)	Initial base fare
109	106 (peak), 101 (offpeak)	Per-mile cost
28 (peak), 26 (offpeak)	0	Cost per minute
485	0	Minimum fare
257	257	Booking fee not included in minimum fare (in 2019 $1.85 to the company, 72¢ taxes)
500	500	Surcharge for O'Hare and Midway airports, Navy Pier and McCormick Place
0 in 2019, 175 thereafter	0 in 2019, 175 thereafter	Downtown Chicago surcharge applied 6am – 10pm
Wait time (minutes)	**Area definition**	
1 (peak), 2 (offpeak)	Chicago Central Area zone	
5	Chicago zone outside of the Central Area	
5	dense suburban CBD zone	
15	remaining suburban zones	

Transit costs

Unlike the auto modes, the utility calculations for transit modes must include estimates of the costs passengers incur when accessing and egressing the transit service. As noted, the FORTRAN-based transit access and egress calculation procedures used in past versions of CMAP's trip-based model were transferred into Python. The underlying logic of the level of service calculations is unchanged, and it is based on a random distribution of transit access characteristics based on the geography and service levels of each zone.

The concepts of first mode, priority mode, and last mode are used to identify available transit access and egress options. These mode identifiers are based on the following hierarchy of transit service, listed from highest to lowest priority: Metra, CTA rail, and bus. First mode is identified by skimming the transit network for zonal interchanges by using the transit access links listed in Table 20 based on the priority hierarchy. Similarly, the transit egress links are used to identify last mode for zone pairs. The priority linehaul mode is determined based on the combinations of first and last mode.

There are five potential transit approach (access) modes: 1-Walk, 2-Bus, 3-Park and Ride, 4-Kiss and Ride, and 5-Feeder Bus. Monte Carlo simulation is used to determine the access and egress distance for each of the approach options, given that the DISTR files contain different distances by zone for Metra, CTA rail, bus, feeder bus, and Park and Ride. Fewer options are available for transit egress, based on the following rules:

- Walk is always available as an access and egress option.
- For Bus priority mode, walk is the only available access and egress option.
- For Rail priority modes:
    - Park and Ride is only available as an access option for home-based trips.
    - Kiss and Ride is only available as an access option for HBW trips.
    - Park and Ride, Kiss and Ride, and Feeder Bus are not available as egress options.

Table 66 below summarizes the logic of the transit access and egress cost calculations.

Table 66. Transit access and egress cost logic

Approach option	Cost components
Walk (first or last mode)	• The only approach cost is the time to walk to the bus. For work trips in the CBD, this cost is increased by 20 percent. • The walk time is multiplied by the driver's value of time and the OVT factor
Bus (first or last mode) Feeder Bus (first mode only)	• Walk time – Simulated distance to walk divided by walk speed (increased by 20% in the CBD to account for congested walking conditions). Value multiplied by the OVT factor and the driver's value of time. • Drive time – Approach distance divided by the zone type auto speed multiplied by the driver's value of time • Wait time – Zonal average wait time for bus multiplied by the OVT factor and the driver's value of time • Cost – Based on the zone location the appropriate fares and transfer costs are included

Park and Ride (first mode only)	• Walk time – Default walk time from parking lot to platform multiplied by the OVT factor and the driver's value of time • Drive time – Approach distance divided by the zone type auto speed multiplied by the driver's value of time • Cost – Driving distance multiplied by the per-mile auto operating costs based on zone type. The fixed driver auto costs and half of the parking costs are added to this.
Kiss and Ride (first mode only)	• Walk time – Default walk time from parking lot to platform multiplied by the OVT factor and the driver's value of time • Drive time – Approach distance is divided by the zone type auto speed and multiplied by the driver's value of time. This is multiplied by three (to cover the time traveled by the driver and passenger, as well as an assumed trip by the driver back home) and by the driver's value of time. • Cost – Driving distance multiplied by the per-mile auto operating costs based on zone type. Double the auto passenger fixed costs are added to this.

Utility calculations

This section shows the utility formulas for each modal alternative in the mode choice-destination choice model. The formulas quantify the benefits and costs associated with the different travel choices. Two placeholder variables are listed in the equations to reflect where the specific value varies based upon the specific conditions:

- [*PK|OP*] – reflects that either PEAK or OFFPEAK values will be used. PEAK values are applied for home-based work trips (HBWH, HBWL) and OFFPEAK values are applied for all other trip purposes.
- [*PURPOSE*] – reflects the variable value specific to each trip purpose.

Each of the utility equations includes the same formula for utility at the trip destination, which is shown in Table 67 and referenced by name (utility destination) in the subsequent modal utility formulas. "P" values represent the model coefficients while the "X" variables are the actual trip values.

Table 67. Utility_destination

Formula
P.samp_af * X('log(1/actualdest_samp_wgt)') + P.log_attraction * X.actualdest_log_attractions_[*PURPOSE*] + P.intrazonal * X('o_zone == actualdest') + P('distance[1]: up to 5') * X('piece(actualdest_auto_dist_OFFPEAK,None,5)') + P('distance[2]: 5 to 10') * X('piece(actualdest_auto_dist_OFFPEAK,5,10)') + P('distance[3]: over 10') * X('piece(actualdest_auto_dist_OFFPEAK,10,None)')

In the following tables shaded blocks represent groups of conditional logic and identify which components are applied when specific conditions are met.

Table 68. SOV (1) utility

Formula	Conditions
P.cost * 0.01 * X.actualdest_auto_opcost_[*PK|OP*] + P.totaltime * X.actualdest_auto_time_[*PK|OP*] + P.cost * 0.01 * X.actualdest_auto_parking_cost_[*PURPOSE*] + P.unavail * X('1-actualdest_auto_avail_[*PURPOSE*]') + utility_destination	
+ P.cost * 0.01 * X.actualdest_auto_toll_hiinc_PEAK + P.cost * 0.01 * X.actualdest_auto_toll_loinc_PEAK + P.cost * 0.01 * X.actualdest_auto_toll_OFFPEAK	If HBWH If HBWL Not HBW
+ P.AUTO_ozone_autopropensity * X.ozone_autopropensity + P.AUTO_dzone_autopropensity * X.actualdest_autopropensity + P.AUTO_no_veh * X('hhveh==0') + P.AUTO_sufficient_veh * X('hhveh>=hhadults')	If NHB If NHB Not NHB Not NHB

Table 69. HOV2 (2) utility

Formula	Conditions
P.Const_HOV2 + P.cost * 0.005 * X.actualdest_auto_opcost_hov_[*PK|OP*] + P.totaltime * X.actualdest_auto_time_[*PK|OP*] + P.cost * 0.005 * X.actualdest_auto_parking_cost_[*PURPOSE*] + P.unavail * X('1-actualdest_auto_avail_[*PURPOSE*]')	

+ P('HOV2_distance[1]: up to 5') * X('piece(actualdest_auto_dist_OFFPEAK,None,5)') + P('HOV2_distance[2]: 5 to 10') * X('piece(actualdest_auto_dist_OFFPEAK,5,10)') + P('HOV2_distance[3]: over 10') * X('piece(actualdest_auto_dist_OFFPEAK,10,None)') + utility_destination	
+ P.cost * 0.005 * X.actualdest_auto_toll_hov_hiinc_PEAK + P.cost * 0.005 * X.actualdest_auto_toll_hov_loinc_PEAK + P.cost * 0.005 * X.actualdest_auto_toll_OFFPEAK	If HBWH If HBWL Not HBW
+ P.AUTO_ozone_autopropensity * X.ozone_autopropensity + P.AUTO_dzone_autopropensity * X.actualdest_autopropensity + P.AUTO_no_veh * X('hhveh==0') + P.AUTO_sufficient_veh * X('hhveh>=hhadults')	If NHB If NHB Not NHB Not NHB

Table 70. HOV3+ (3) utility

Formula	Conditions
P.Const_HOV3 + P.cost * 0.0033 * X.actualdest_auto_opcost_[*PK|OP*] + P.totaltime * X.actualdest_auto_time_[*PK|OP*] + P.cost * 0.0033 * X.actualdest_auto_parking_cost_[*PURPOSE*] + P.unavail * X('1-actualdest_auto_avail_[*PURPOSE*]') + P('HOV3_distance[1]: up to 5') * X('piece(actualdest_auto_dist_OFFPEAK,None,5)')	

+ P('HOV3_distance[2]: 5 to 10') * X('piece(actualdest_auto_dist_OFFPEAK,5,10)') + P('HOV3_distance[3]: over 10') * X('piece(actualdest_auto_dist_OFFPEAK,10,None)') + utility_destination	
+ P.cost * 0.0033 * X.actualdest_auto_toll_hov_hiinc_PEAK + P.cost * 0.0033 * X.actualdest_auto_toll_hov_loinc_PEAK + P.cost * 0.0033 * X.actualdest_auto_toll_OFFPEAK	If HBWH If HBWL Not HBW
+ P.AUTO_ozone_autopropensity * X.ozone_autopropensity + P.AUTO_dzone_autopropensity * X.actualdest_autopropensity + P.AUTO_no_veh * X('hhveh==0') + P.AUTO_sufficient_veh * X('hhveh>=hhadults')	If NHB If NHB Not NHB Not NHB

Table 71. Taxi (4) utility

Formula	Conditions
P.Const_TAXI + P.cost * 0.01 * X.actualdest_taxi_fare_[*PK|OP*] + P.ovtt_dist * X('actualdest_taxi_wait_time_[*PK|OP*]/actualdest_auto_dist_[*PK|OP*]') + P.totaltime * X.actualdest_taxi_wait_time_[*PK|OP*] + P.totaltime * X.actualdest_auto_time_[*PK|OP*] + P.unavail * X('1-actualdest_auto_avail_[*PURPOSE*]') + utility_destination	

+ P.cost * 0.01 * X.actualdest_auto_toll_hiinc_PEAK	If HBWH
+ P.cost * 0.01 * X.actualdest_auto_toll_loinc_PEAK	If HBWL
+ P.cost * 0.01 * X.actualdest_auto_toll_OFFPEAK	Not HBW

Table 72. TNC (5) utility

Formula	Conditions
P.Const_TNC1 + P.cost * 0.01 * X.actualdest_tnc_solo_fare_[*PK|OP*] + P.ovtt_dist * X('actualdest_tnc_solo_wait_time_[*PK|OP*]/actualdest_auto_dist_[*PK|OP*]') + P.totaltime * X.actualdest_tnc_solo_wait_time_[*PK|OP*] + P.totaltime * X.actualdest_auto_time_[*PK|OP*] + P.unavail * X('1-actualdest_auto_avail_[*PURPOSE*]') + utility_destination	
+ P.cost * 0.01 * X.actualdest_auto_toll_hiinc_PEAK	If HBWH
+ P.cost * 0.01 * X.actualdest_auto_toll_loinc_PEAK	If HBWL
+ P.cost * 0.01 * X.actualdest_auto_toll_OFFPEAK	Not HBW

Table 73. TNC shared ride (6) utility

Formula	Conditions
P.Const_TNC2 + P.cost * 0.01 * X.actualdest_tnc_pool_fare_[*PK|OP*] + P.ovtt_dist * X('actualdest_tnc_pool_wait_time_[*PK|OP*]/actualdest_auto_dist_[*PK|OP*]')	

+ P.totaltime * X.actualdest_tnc_pool_wait_time_[*PK|OP*] + P.totaltime * X.actualdest_auto_time_[*PK|OP*] + P.unavail * X('1-actualdest_auto_avail_[*PURPOSE*]') + utility_destination	
+ P.cost * 0.005 * X.actualdest_auto_toll_hiinc_PEAK + P.cost * 0.005 * X.actualdest_auto_toll_loinc_PEAK + P.cost * 0.005 * X.actualdest_auto_toll_OFFPEAK	If HBWH If HBWL Not HBW

Table 74. Transit (7) utility

Formula	Conditions
P.Const_Transit + P.cost * 0.01 * X.actualdest_transit_fare_[*PK|OP*] + P.totaltime * X.actualdest_transit_ovtt_[*PK|OP*] + P.cost * 0.01 * X.actualdest_transit_approach_cost_[*PK|OP*] + P.totaltime * X.actualdest_transit_approach_drivetime_[*PK|OP*] + P.totaltime * X.actualdest_transit_approach_walktime_[*PK|OP*] + P.totaltime * X.actualdest_transit_approach_waittime_[*PK|OP*] + P.unavail * X('1-actualdest_transit_avail_[*PURPOSE*]') + P.ovtt_dist * X('actualdest_transit_ovtt_[*PK|OP*]/actualdest_auto_dist_[*PK|OP*]') + P.ovtt_dist * X('actualdest_transit_approach_walktime_[*PK|OP*]/actualdest_auto_dist_[*PK|OP*]')	

+ P.ovtt_dist * X('actualdest_transit_approach_waittime_[*PK|OP*]/actualdest_auto_dist_[*PK|OP*]') + P.transit_intrazonal * X('o_zone == actualdest') + utility_destination	
+ P.transit_areatype2 * X('fmin(ozone_areatype, actualdest_areatype)==2') + P.transit_areatype3 * X('fmin(ozone_areatype, actualdest_areatype)==3') + P.transit_areatype4 * X('fmin(ozone_areatype, actualdest_areatype)==4') + P.metra_longtrip * X('actualdest_auto_dist _[*PK|OP*]') [*where commuter transit is available and drive time to the station is less than walk time or waiting for a feeder bus and the trip is 15-35 miles long between area types 1 and 3]* + P.transit_walk_is_short * X('hard_sigmoid(actualdest_transit_approach_walktime_OFFPEAK, 4.0, 2.0)') + P.transit_areatype3) * X('fmax(ozone_areatype, actualdest_areatype)==3')	if HBW if HBW if HBW if HBW not HBW not HBW
+ P.totaltime * X('piece(actualdest_transit_ivtt_OFFPEAK, None, 20)') + P.ivtt_longtransit * X('piece(actualdest_transit_ivtt_OFFPEAK, 20, None)') + P.totaltime * X.actualdest_transit_ivtt_PEAK	if NHB if NHB not NHB

Table 75. Walk (8) utility

Formula
P.Const_WALK + P('walk_time[1]: up to 0.5') * 20.0 * X('piece(actualdest_auto_dist_OFFPEAK,None,0.5)') + P('walk_time[2]: 0.5 to 1.0') * 20.0 * X('piece(actualdest_auto_dist_OFFPEAK,0.5,1.0)') + P('walk_time[3]: over 1.0') * 20.0 * X('piece(actualdest_auto_dist_OFFPEAK,1.0,None)')

+ P.walk_intrazonal * X('o_zone == actualdest') + P.walk_areatype2 * X('fmax(ozone_areatype, actualdest_areatype)==2') + P.walk_areatype3 * X('fmax(ozone_areatype, actualdest_areatype)==3') + P.walk_areatype4 * X('fmax(ozone_areatype, actualdest_areatype)==4') + utility_destination

Table 76. Bicycle (9) utility

Formula
P.Const_BIKE + P.bike_time * 5.0 * X.actualdest_auto_dist_OFFPEAK + P.bike_intrazonal * X('o_zone == actualdest') + utility_destination

Table 77 contains the parameter coefficient used by the mode choice-destination choice models. Within the model setup, these are stored in the file choice_model_params.yaml.

Table 77. Mode-destination choice utility coefficients

Parameter name	Coefficient value				
	HBWH	HBWL	HBS	HBO	NHB
AUTO_no_veh	-2.0207	-2.6528	-3.9228	-2.9707	
AUTO_sufficient_veh	0.5191	0.6889	0.8557	0.7343	
AUTO_dzone_autopropensity					2.6097
AUTO_ozone_autopropensity					7.9026
Const_BIKE	-1.0735	-1.094	-1.848	-1.62	3.6597
Const_HOV2	-0.3669	-0.4461	-0.62	-0.2296	-0.3537
Const_HOV3	-0.4181	-0.629	-1.7535	-0.7619	-1.0022
Const_TAXI	-2.4311	-3.8304	-5.2536	-2.9427	4.042
Const_TNC1	-2.0083	-2.4175	-3.8162	-2.1496	4.506
Const_TNC2	-2.4553	-2.8384	-4.5589	-2.9871	3.1428
Const_Transit	0.601	0.48	-0.9632	0.1118	6.2099
Const_WALK	1.9303	3.5174	6.5967	4.6708	9.2578
"HOV2_distance[1]: up to 5"	-0.0649	-0.1107	0.009	0.0179	-0.0103
"HOV2_distance[2]: 5 to 10"	-0.1154	-0.0492	0.0493	-0.0002	-0.0767
"HOV2_distance[3]: over 10"	-0.0132	-0.0028	-0.0235	-0.0166	-0.0181

"HOV3_distance[1]: up to 5"	-0.1252	-0.1545	0.1013	0.0398	0.0018
"HOV3_distance[2]: 5 to 10"	-0.1476	-0.1264	0.0233	-0.0128	-0.0687
"HOV3_distance[3]: over 10"	-0.0171	-0.0172	-0.0269	-0.0177	-0.0403
HOV_no_veh	-2.1033	-2.039	-2.6988	-2.3244	
HOV_sufficient_veh	0.4762	0.3912	0.8003	0.7002	
HOV_dzone_autopropensity					2.2062
HOV_ozone_autopropensity					7.9581
Mu-Dest	0.5835	0.8404	1	0.9173	1
Mu-HiredCar	0.4376	0.6302	0.75	0.688	0.75
Mu-PrivateCar	0.4376	0.6302	0.7489	0.688	0.75
bike_intrazonal	-1.6637	-0.994	-0.4423	-0.6768	0.6138
bike_time	-0.0278	-0.0633	-0.147	-0.0755	-0.0297
cost	-0.0119	-0.0294	-0.0305	-0.0254	-0.107
"distance[1]: up to 5"	-0.2988	-0.4452	-0.7298	-0.6121	-0.4686
"distance[2]: 5 to 10"	-0.251	-0.2557	-0.5109	-0.4018	-0.3335
"distance[3]: over 10"	-0.1017	-0.0913	-0.1317	-0.1082	-0.0641
intrazonal	0.3947	0.0217	-0.8717	-0.4601	-0.0319
log_attraction	1	1	1	1	1
metra_longtrip	9.375	5.625			
ovtt_dist	-0.1598	-0.0764	-0.2468	-0.2748	-0.2685
samp_af	1	1	1	1	1
totaltime	-0.017	-0.0147	-0.03	-0.0254	-0.03
ivtt_longtransit					-0.011
transit_areatype2	-1.3747	-1.2441			
transit_areatype3	-3.1399	-3.8502	-0.5	-1.1	-1.1
transit_areatype4	-2.5397	-1.7423			
transit_intrazonal	1.6254	1.2836	2.8339	2.1563	3.5821
transit_walk_is_short			1.6524	0.5258	0.7232
unavail	-999	-999	-999	-999	-999
walk_areatype2	0.149	0.6465	0.0794	0.4343	0.5752
walk_areatype3	-0.2521	0.0891	-1.4849	-0.0197	0.9757
walk_areatype4	-0.6561	0.0183	-1.7738	-0.3177	1.0862
walk_intrazonal	-0.3987	-1.113	-1.2366	-0.8354	-0.3277
"walk_time[1]: up to 0.5"	-0.1508	-0.2681	-0.4763	-0.3086	-0.1878
"walk_time[2]: 0.5 to 1.0"	-0.0505	-0.0588	-0.0959	-0.0835	-0.0905
"walk_time[3]: over 1.0"	-0.0451	-0.0789	-0.1057	-0.0848	-0.0938

During execution of the mode choice-destination choice model, the Python code directly reads skim matrix values to perform the utility calculations. Table 78 summarizes the matrices used to determine the utility of each modal option. Note that matrices applied for HBW trips reflect morning peak conditions, while the matrices applied for non-work trips reflect midday travel conditions. The utility calculations include the effects of tolling. The presence of tolls increases the cost to travelers of using specific routes, so it is accounted for within the traffic assignment procedures. The increased travel cost also impacts the choice of travel mode and is included in the utility calculations.

Table 78. Utility calculation matrices

	Matrix	
Auto		
	HBW	**Non-work**
Travel time - SOV	44	46
Travel time - HOV	76	46
Travel distance – SOV	45	47
Travel distance – HOV	77	47
Toll – SOV low income	111	NA
Toll – SOV high income	114	NA
Toll – HOV low income	112	NA
Toll – HOV high income	115	NA
Toll	NA	117
Transit		
	HBW	**Non-work**
In-vehicle time	822	922
Out of vehicle time	823	923
Headway	838	938
Fare	828	928
First mode	829	929
Priority mode	830	930
Last mode	831	931

A file of the propensity of home-based trips to be attracted to zones based on a private auto mode share (SOV or HOV) is used to adjust the probability that non-home-based trips produced in these zones also will use a private auto mode. This file is created during each global iteration of the model and is used in the subsequent iteration of the mode-destination choice model. A default file (default_auto_propensity.csv.gz) is used during the initial model iteration.

As noted in trip generation, modeled households are divided into two groups — those with no workers working from home and those with at least one worker who works from home with some frequency. This results in separate sets of trip productions and attractions. This bifurcation of the trip data continues through mode choice-destination choice. The model is first run on the trip data for households with no workers who work-from-home and then is run again for the other group of trips. While the parameters of the mode choice-destination choice models are the same between the two groups, the trip productions and attractions do differ, resulting in unique utility functions and different behavior patterns.

The trip-based model is not a full microsimulation. Only a representative set of trips is modeled for each origin zone. No destination sampling is used in the mode choice-destination choice model. Typically, such sampling is used to ensure the model provides good coverage of less attractive destinations. A weighted sampling approach was used to estimate the model, which used fully disaggregated trips from the My Daily Travel survey. After testing several sizes, a sample size of 25 destinations ultimately was used and provided parameter estimates that are statistically indistinguishable from larger sample sizes.

Visitor trips

CMAP's trip-based model includes daily demand for visitor trips — trips with both ends located within the modeling area that are made by people whose home location is outside of the modeling area. This demand was developed using 2019 LOCUS mobile location data to create a base year zone-to-zone visitor trip table. The mode choice model for non-home base trips is used to determine the travel mode for visitor trips, with constants shifted so visitors have a higher likelihood of using taxis and TNCs than residents would for non-home-based trips.

Visitor trips assigned to the SOV mode must also be categorized by value of time level for traffic assignment. The value of time distribution for visitor trips is assumed to mirror that for resident non-home-based trips. The VOT distribution is shown in Table 87.

As the visitor trip table reflects 2019 trips, a set of growth factors were developed to represent increased visitor demand in future scenarios. These growth factors were developed based on historical trends in the growth in air passengers at O'Hare and Midway airports, and the growth in tourism nationally and for Chicago. The growth factors assume visitor demand in northeastern Illinois returns to pre-pandemic levels (2019) in 2023. The visitor trip growth rates are shown in Table 79.

Table 79. Visitor trip growth factors

2019	2026	2030	2035	2040	2045	2050
1.00	1.05	1.18	1.34	1.51	1.70	1.92

Time-of-day model

After the mode-destination-choice model runs, the time-of-day model is applied to auto trips (for both private and hired autos). This model separates auto trips into the eight time-of-day periods modeled by CMAP. The time-of-day model is sensitive to congested travel times throughout the day. It is a simple multinomial logit model that has alternative specific constants that reflect behavioral preferences and the duration of each period. The utility functions are also sensitive to mode choice as TNC use is significantly different than private auto use. Many TNC trips happen in the evening and overnight periods.

The time-of-day utility formula is listed in Table 80. Two placeholder variables are listed in the equations to reflect where the specific value varies based on the specific conditions:

- [*TOD label*] – the time period label refers to the following: 'EA'=evening/early morning; 'AM1'=pre-AM peak shoulder; 'AM2'=AM peak; 'AM3'=post-AM peak shoulder; 'MD'=midday; 'PM1'=pre-PM peak shoulder; 'PM2'=PM peak; and 'PM3'=post-PM peak shoulder.
- [*TOD period*] – the numeric time-of-day period (1-8). MF461-468 are skimmed highway times for SOVs for each period.

Table 80. Time-of-day model utility formula

Formula	Conditions
P.ASC_[*TOD label*] + P.time * X.mf46[*TOD period*] + P.hiredcar_[*TOD label*] * X('mode9 in ('TAXI','TNC1','TNC2')')	
+ P.ASC_[*TOD label*]_r * X.paFlip	not NHB
+ P.hiredcar_[*TOD label*]_r * X('(mode9 in ('TAXI','TNC1','TNC2'))*paFlip')	if HBW, HBO

The following table contains the parameter coefficient used by the time-of-day model. Within the model setup, these are stored in the file tod_model_params.yaml. Parameters ending with "_r" represent the reverse trip. These are only applied to home-based trips, meaning they represent the attraction-production trip.

Table 81. Time-of-day model parameters

Parameter name	Coefficient value				
	HBWH	HBWL	HBS	HBO	NHB
ASC_AM1	-1.1259	-1.1102	-1.8298	-1.2922	-1.8283
ASC_AM1_r	-1.1231	-0.9133	-0.3437	-0.126	
ASC_AM2	0	0	0	0	0
ASC_AM2_r	0	0	0	0	
ASC_AM3	-2.5156	-1.8888	-0.2645	-1.14	-0.8498
ASC_AM3_r	0.7526	0.7235	0.7867	0.6787	
ASC_EA	-1.9622	-1.6429	-1.5805	-1.5318	-1.348
ASC_EA_r	1.6611	2.1924	2.47	2.7275	
ASC_MD	-1.6343	-0.9985	1.2134	0.1653	1.1168
ASC_MD_r	1.9892	1.6542	1.4565	1.3714	
ASC_PM1	-2.2204	-1.8976	0.2849	-0.4321	0.5517
ASC_PM1_r	3.7918	3.4383	2.2642	1.8336	
ASC_PM2	-2.3658	-2.4245	0.3984	0.1484	0.2782
ASC_PM2_r	3.654	3.6741	2.1922	1.5952	
ASC_PM3	-3.9189	-4.0392	-0.0874	-0.9093	-0.6503
ASC_PM3_r	4.0171	4.3707	2.2669	2.5158	
hiredcar_AM1	-0.2335	-1.1946	-11.7505	1.00E-04	0.8795
hiredcar_AM1_r	-12.1216	-10.706		0.189	
hiredcar_AM2	0	0	0	0	0
hiredcar_AM2_r	0	0		0	
hiredcar_AM3	1.0709	0.9707	-0.1476	0.0347	0.4629
hiredcar_AM3_r	-14.0951	0.0756		-0.1058	

hiredcar_EA	-0.1723	0.5339	-17.5549	0.9544	2.8669
hiredcar_EA_r	3.1961	1.6282		1.5105	
hiredcar_MD	0.3483	0.4738	0.0168	0.2375	0.7199
hiredcar_MD_r	0.6849	0.262		-0.378	
hiredcar_PM1	-1.5686	0.7303	0.0839	-0.2086	0.9617
hiredcar_PM1_r	2.5324	-0.0527		0.2851	
hiredcar_PM2	-1.3687	-0.497	0.6512	0.6264	1.2375
hiredcar_PM2_r	3.3293	1.1441		-0.642	
hiredcar_PM3	0.7728	-16.158	1.209	1.6335	2.0767
hiredcar_PM3_r	1.5941	17.7847		-0.5611	
time	-0.0551	-0.0367	-0.0273	-0.0552	-0.0200

Hired auto deadhead trips

Within the model, a set of deadhead trips are generated for taxis and TNCs to represent trips connecting revenue trips. These are the trips that drivers make once a passenger is dropped off to reach the location where the next passenger will be picked up. The method links together the destinations of taxi and TNC trips (when passengers are dropped off and thus, the productions for deadhead trips) with the origins of taxi and TNC trips (when the new passenger is collected and thus, the attractions for deadhead trips) within each time period. A simplifying assumption is made that these productions and attractions are balanced within each period.

Distribution of the deadhead trips is accomplished using a simple gravity model with congested travel time skims for the time period as the impedance, meaning that drivers travel close to the minimum distance needed to serve the revenue trips. This reflects the rational behavior of drivers trying to minimize the distance of trips traveled without a fare. For traffic assignment, these trips are added to the SOV trip tables using the value of time shares listed in Table 87.

Final trip roster

Completion of the time-of-day model results in the creation of a set of parquet files, which use columnar data storage to reduce file sizes. These parquet files collectively contain the roster of trips created by the models and can be used to develop specific trip demand tables. The labeling for these files is "choice_simulator_trips_[*range of zones*]_HBWH_HBWL_HBS_HBO_NHB_[*typical*|*wfh*].pq" where:

- *Range of zones* lists the origin zones included in the file
- *typical|wfh* identifies whether the trips are from households with at least one worker who works from home or from households with no workers working from home.

Table 82 lists the fields contained in the parquet files.

Table 82. Parquet file fields

Variable	Description
trips	Number of trips
purpose	Trip purpose (HBWH, HBWL, HBS, HBO, NHB, VISIT, DEAD)
mode	Trip mode (numeric value)
o_zone	Origin zone (the beginning of the actual trip – may not be the home/production zone)
d_zone	Destination zone (the end of the actual trip – may not be the attraction zone)

a_zone	Attraction zone (either the production or attraction zone, depending on trip directionality)
hh_autos	Number of household vehicles available (0,1,2+) – applicable only for home-based trips
hh_inc5	Household income category of the trip maker (in 2019 dollars): 1=less than $30,000, 2=$30,000-$59,999, 3=$60,000-$99,999, 4=$100,000-$149,999, 5=$150,000+
timeperiod	Time period labels; NA applied to transit, walk and cycling trips

During execution of the time-of-day model, home-based trips (which are modeled in production-attraction format rather than origin-destination format) are converted to origin-destination trips using the following factors to create the attraction-production trips.

- Home-based work high income: 0.459
- Home-based work low income: 0.443
- Home-based shopping: 0.647
- Home-based other: 0.513

Traffic assignment

The final step in the trip-based model is traffic assignment. This step uses all the travel demand created and routes trips over the highway network. A final step after completion of the time-of-day model is the creation of time-period-specific demand matrices that are used for the highway assignment. Person trips are converted to vehicle trips for these matrices, meaning auto occupancy is accounted for. The following vehicle occupancy rates are used for HOV trips with at least three people.

- Home-based work trips: 3.36
- All other home-based trips: 3.31
- Nonhome-based trips and visitor trips: 3.39

Special trip handling

Traffic assignment includes several other classes of vehicle trips in addition to auto vehicle trips. These require special data-handling procedures to create the demand matrices.

Commercial vehicle trips

CMAP models commercial vehicle trips for four truck classes: B-plate, light trucks, medium trucks, and heavy trucks. B-plate trucks are vans, pickup trucks, and sport utility vehicles with performance characteristics similar to passenger cars and carrying “B” license plates. These license plates are issued to vehicles weighing less than 8,000 pounds. Light trucks are “step vans” and smaller delivery vans, which carry weight plates D-J and MD-MJ. Medium trucks are defined as heavy fixed-wheelbase trucks, such as concrete mixers, scavenger trucks, double rear axle refrigerator units, and some other lighter weight articulated vehicles carrying weight plates K-T and

MK-MT. This covers trucks weighing between 28,001 and 64,000 pounds. Finally, heavy trucks comprise the 73,280- and 80,000-pound maximum load vehicles, which are tractor-trailer combinations. These carry weight plates of V-Z.

In the past, CMAP used vehicle registration files from the Illinois Secretary of State's office to develop "base year" trip totals for each of the truck classes. The relationship between registered vehicles and actual trips was always somewhat tenuous, but it provided the best available information at the time. CMAP currently uses more robust data to develop the base year trip totals that inform the trip-based model, including:

B-plate trucks: These license plates are issued to either personal or commercial vehicles. However, for the purpose of modeling b-plate trucks as a specific vehicle class, CMAP is only interested in commercial vehicle demand. Personal vehicle demand is included in household trips. An analysis of b-plate vehicle registrations was conducted for the Illinois counties in the CMAP modeling area using data from the Illinois Secretary of State's office. Using the registrations and total population in these counties, an average rate of b-plate vehicles per person was developed. This rate was applied to the remaining counties in the CMAP modeling area to determine total b-plate vehicles registered within the modeling area. It was factored by 10 percent to reflect external b-plates that operate on the road network within the modeling area.

National auto sales information from the last two decades was used to determine the increased share of trucks, minivans and sport utility vehicles in use. Sales shares were combined with Illinois data from the Vehicle Inventory and Use Survey on the share of miles by these vehicle types used for commercial purposes. Finally, an analysis of b-plate vehicles included in the My Daily Travel survey provided the average number of trips made daily. This value multiplied by the number of b-plate vehicles provides an estimate of daily commercial trips for this vehicle class.

Heavy trucks: CMAP purchases data on heavy truck trips within the modeling area from the American Transportation Research Institute (ATRI). Agency staff analyzed this dataset to determine the number of trips heavy vehicles make within the CMAP modeling area and to identify the specific locations they visit.

Light and medium trucks: The data purchased from ATRI does not include these truck classes. Instead, the number of trips for these trucks is derived from vehicle registration data and established relationships in the number of trips these truck classes take relative to heavy trucks. As an extra level of data verification, the reasonableness of these trip values was confirmed by reviewing the results of a special truck demand model developed for the Illinois Department of Transportation (IDOT) in 2012 to support the Illiana Expressway analysis.

Once the truck trips were developed, they were converted to a year 2000 "base value" to provide a set of trip values consistent across all truck classes. These base values were developed assuming a growth rate of 10 percent per decade, the same growth used to forecast future truck trips. Table 83 presents the base trip totals by vehicle class.

Table 83. Truck trip totals by vehicle class

Truck type	Base year 2000 total
B-plate trucks	1,468,500
Light trucks	246,500
Medium trucks	229,500
Heavy trucks	395,000

Once the total number of trucks per class is determined, the non-heavy truck trips are allocated to production and attraction zones based on development patterns that come from UrbanSim. The measure of development is represented by nonhome-based trip productions. These are most closely related to total development with an emphasis on employment density. The process is a simple allocation of trips to zones based on the zonal share of the total regional development. The distribution of trips is then created based on trip length distribution parameters derived from data on light and medium truck trips in the region purchased from INRIX.

For b-plate trucks, the trip length distribution was enhanced using odometer readings collected by the Illinois Environmental Protection Agency as part of the vehicle inspection and maintenance program. These readings were obtained for the seven CMAP counties and Grundy County. Specific vehicles were matched to the Illinois Secretary of State registration data to isolate b-plate trucks from the resulting dataset, and the distribution of average daily vehicle miles traveled was calculated for b-plates.

The distribution of heavy truck trips is also weighted using information from UrbanSim. Base year employment by NAICS category for each building in the seven-county CMAP region is supplemented with average heavy commercial vehicle trip rates per 1,000 square feet by building type, as identified in NCHRP Report 298, to estimate regional trips per employee rates by NAICS code. UrbanSim provides employment by NAICS for each forecasted year, so total trips are estimated by multiplying the trips per employee by the number of employees by NAICS code for each zone. Employment-based estimations of trip frequencies tend to favor employment dense regions, such as the CBD, so the estimated trips are scaled by the share of non-residential building area per zone to produce the final heavy truck trip weights.

Since these final weights are not used as actual productions and attractions, but merely as weights for the distribution process, scaling the final estimated trips by the share of non-residential building area helps differentiate between pickup and delivery needs of different types of land uses. Non-residential building area zonal share was calculated from the UrbanSim base year buildings file and shares are held constant in all future years. The buildings file only includes buildings in the CMAP seven-county region, so an estimated zonal share of non-residential building area was assigned to the remaining modelling zones by matching the zone with an unknown share to a zone with a known share based on similar observed trip rates from the ATRI dataset.

The size and operating characteristics of commercial vehicles require them to be treated differently than automobiles during traffic assignment. Prior to the traffic assignment process, truck vehicle trips are converted to trips measured in vehicle equivalents. The truck vehicle trips are converted using the following factors:

- B-plate and light trucks equal one vehicle equivalent.
- Medium trucks equal two vehicle equivalents.
- Heavy trucks (and buses) equal three vehicle equivalents.

Point-of-entry trips

Point-of-entry trips represent three categories of travel: auto travel entering or leaving the modeled region on major expressways, heavy truck travel entering or leaving the region on major expressways, and auto passenger travel supporting the region's airports entering or leaving the region. Point-of-entry locations are external zones (numbered 3633 through 3649) and are not modeled in the same way as the rest of the region's travel because there is little knowledge about the traveler, the trip purpose, or the trip origin or destination located outside of the region. These trips are created based on observed traffic counts at the locations in question and some assumptions about the travel behavior of the trip maker, including an assumption that external travelers are indifferent about the actual length of the trip within the region (i.e., their destination is fixed).

Base year trip production totals for highway point of entry are derived from expressway traffic counts at locations around the region. Base year trips for airport point of entry are based on an analysis of observed enplanements. To create future productions and attractions, the base year number of total trips is factored using the same growth rates as commercial vehicles: 10 percent per decade. The year 2015 POE productions are presented in Table 84.

Table 84. Point-of-entry base year productions

Truck type	Base year 2015 total
Auto external	255,600
Truck external	109,500
Air traveler	69,500

All point-of-entry trips are handled at the same time using a gravity model. To begin, an impedance file based on a gamma function was created. To accomplish this, a destination vector of non-work trip attractions plus a weighted number of point-of-entry trips was calculated. Again, this information is used as a measure of development density with an emphasis on employment density. The impedance matrix is proportional to the productions multiplied by the attractions and inversely proportional to the square of the midday travel distance (capped at 60 miles):

$$Impedance = \frac{(.0001 * (POE\ Productions * Destination\ Development))}{(60.max.travel\ distance)^2}$$

The impedance matrix is balanced using the original productions at the origin, and trip attractions apportioned to destinations based on zonal shares of non-work attractions and zonal point-of-entry totals as the attractions.

At this point, the balanced trip matrix must be separated into its component pieces to be used within the traffic assignment procedures. Trips with origins at the expressway points-of-entry are extracted to a matrix, which is summed with its transpose matrix. This represents the total external expressway daily trip table. Thirty percent of the trips in this daily trip table are apportioned as external truck trips. The remaining 70 percent are allocated to external auto trips. To determine air traveler trips, all trips with origins in the region are extracted to another matrix, which is also summed with its transpose matrix. External truck trips are assumed to be heavy commercial vehicles. Thus, this demand matrix is factored by three vehicle equivalents prior to the traffic assignment.

Tolling

Within the traffic assignment procedures, tolling is reflected in the generalized cost of a road segment. While all the user classes perceive the same travel time on a link, they may perceive differing generalized costs. This scheme allows different vehicle classes to be assessed with different toll amounts. It also allows for differing toll amounts to be charged to the user classes based on the time-of-day. CMAP's traffic assignment includes seven user classes:

- SOVs with a low value of time (1)
- SOVs with a medium value of time (2)
- SOVs with a high value of time (3)
- HOVs (4)
- B-plate and light trucks (5)
- Medium trucks (6)
- Heavy trucks (7)

The generalized cost on toll links reflects travel time and a fixed link cost. The fixed link cost is the traveler's value of time multiplied by a perception factor, which reflects that users may not perceive 100 percent of the cost of a toll during a trip. Table 85 lists the hourly values of time used for each user class. Values of time for commercial vehicle drivers were developed using information published by the Bureau of Labor Statistics on the mean hourly wages for drivers of heavy and light trucks in the Chicago region. Values of time for household vehicles were developed using a method described by Lemp and Rossi.[4]

Table 85. Vehicle value of time and perception factor by user class

User class	Hourly value of time	Perception factor
SOV value of time bin 1	$2.30	2.25
SOV value of time bin 2	$7.20	2.25
SOV value of time bin 3	$24.00	2.00

[4] Lemp, J. and T. Rossi, (2018) Practical Framework to Incorporate Value of Time Heterogeneity in an Aggregate Travel Model, Proceedings of the Annual Transportation Research Board Conference, Washington, D.C.

HOV2	$9.00	2.25
HOV3+	$14.00	2.25
B-Plate truck	$24.00	1.00
Light truck	$32.00	1.00
Medium truck	$47.00	1.00
Heavy truck	$52.00	1.00
External auto	$18.00	1.00
Air traveler	$30.00	1.00

The generalized cost process converts the value of time into a minutes per dollar value that is multiplied by the toll amount. This straight-forward calculation fails to account for all of the other elements that individuals consider when determining whether the cost of a tolled route is "worth it." Drivers may not perceive the cost of paying a toll as 100 percent of the actual amount. Transponder users, for instance, are not paying cash out-of-pocket for each toll and may only "see" the cost when they view a monthly credit card statement (at which point the cost has already been incurred). Additionally, the expected travel time savings from using the tollway may more than offset the perceived cost of the toll. Similarly, commercial vehicle drivers may be reimbursed for toll expenses or their Just-in-Time delivery responsibilities may play a much larger role in routing decisions than tolls. The perception factors in Table 85 are an attempt to quantify these individual decisions and effectively increase the user classes' value of time, lowering their minutes per dollar and the effective impedance of tolling. The perception factors were calibrated using vehicle class volumes on tollway facilities.

As noted, three value of time bins (low, medium, and high) are used to categorize SOV trips. The value of time for a specific trip is not solely a function of income but is also related to the trip purpose and the perceived "cost" of not arriving on time. Work trips and attending a concert may have high values of time associated with them, while a routine shopping trip may have a low value of time. Table 86 shows how the value of time bins are applied to SOV trips across trip purposes and household income categories.

Table 86. SOV trip VOT distribution

Trip purpose	Income category	Low VOT	Mid VOT	High VOT
HBWL	1	0.648	0.333	0.018
HBWL	2	0.383	0.537	0.08
HBWH	3, 4	0.197	0.605	0.198
HBWH	5	0.036	0.425	0.53
HBO	1	0.648	0.333	0.018
HBO	2	0.383	0.537	0.08
HBO	3, 4	0.197	0.605	0.198
HBO	5	0.036	0.425	0.53
HBS	1	0.648	0.333	0.018
HBS	2	0.383	0.537	0.08
HBS	3, 4	0.197	0.605	0.198
HBS	5	0.036	0.425	0.53

NHB	1	0.648	0.333	0.018
NHB	2	0.383	0.537	0.08
NHB	3, 4	0.197	0.605	0.198
NHB	5	0.036	0.425	0.53
VISIT	N/A	0.3	0.55	0.15
DEADHEAD	N/A	0.7	0.25	0.05

Table 87 shows the value of time factors applied to categorize the remaining trip demand into the VOT bins. Drive to transit trips are assumed to be mostly work trips, which is why their VOT distributions are similar. The VOTs for external and air passenger trips are assumed to be higher than average. Implementation of this VOT scheme ensures that route choice decisions are sensitive to changes in toll amounts.

Table 87. Value of time distribution

VOT category	Drive to transit trips	External trips
Low VOT bin	15 percent	10 percent
Medium VOT bin	50 percent	45 percent
High VOT bin	35 percent	45 percent

Assignment time periods

The principal objective behind multiple period highway assignments is to develop more accurate estimates of vehicle-miles by different speed ranges and vehicle classes for air quality conformity analyses. Separate assignments estimate highway vehicle-miles and travel speeds for eight time periods during the day:

- Overnight off-peak (8:00 p.m. to 6:00 a.m.)
- Preceding morning peak (6:00 to 7:00 a.m.)
- Morning peak (7:00 to 9:00 a.m.)
- Post morning peak (9:00 to 10:00 a.m.)
- Midday (10:00 a.m. to 2:00 p.m.)
- Preceding evening peak (2:00 to 4:00 p.m.)
- Evening peak (4:00 to 6:00 p.m.)
- Post evening peak (6:00 to 8:00 p.m.)

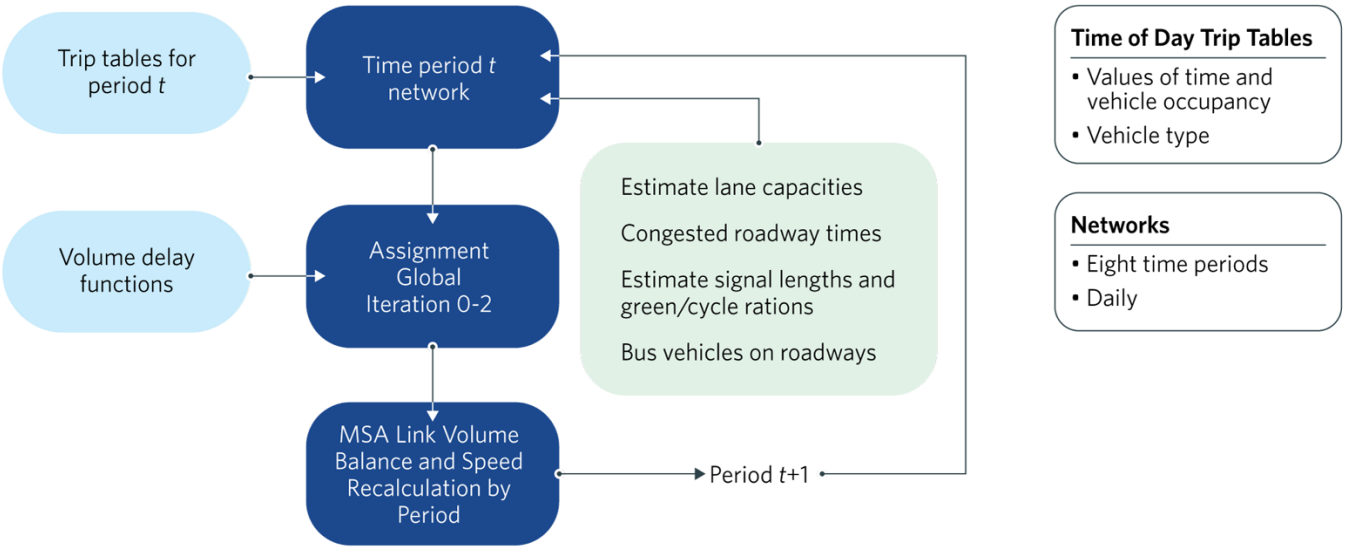
Figure 12 is a schematic diagram that shows the sequence of steps in the multiple time period assignment. Nine highway network scenarios are first assembled (eight time-of-day specific networks and a ninth all-inclusive network to hold the sum of daily information). The presence of time-of-day restrictions on highway network links allows for variation between the TOD networks. At present, these restrictions are modeled on the Kennedy Expressway reversible lanes and an on-

ramp from eastbound IL Route 38 to eastbound I-290, as well as on arterials with peak period parking restrictions. In practice, the eight TOD periods use a total of four unique highway networks:

- A morning peak network used in the morning peak and its shoulders.
- A midday network used during time period 5.
- A evening peak network used in the evening peak and its shoulders.
- An overnight network used in time period 1.

The travel model proceeds through three global iterations. During each iteration, the time period assignments are executed, and the assignment results are averaged with the results of the TOD assignment for the same period from the previous global iteration using the Method of Successive Averages (MSA). This results in a final link volume for each time period. These are used to estimate the morning peak and midday travel times, which are fed back into the rest of the modeling process. The effects of bus operations on other traffic also are accounted for in the assignment process, as buses operating on shared-use facilities are included in the volume-delay function calculations.

Figure 12. Multiple time period highway assignment process



After three passes through the time-of-day modeling process, the results of the separate MSA period assignments are accumulated into daily volumes. The results also are tabulated into the vehicle-mile-by-vehicle-type-by-speed range tables needed for the vehicle emission calculations. The completion and summarizing of the eight period assignments is highly simplified by using scripts to automate repetitive processes.

Traffic assignment is accomplished using a Second-Order Linear Approximation (SOLA) algorithm in Emme. This algorithm is an extremely efficient method to reach network equilibrium and allows for parallelization for shorter run times. Vehicle trips are assigned to the road network and the zone-to-zone travel costs (generalized cost including travel time and tolls) are determined. The

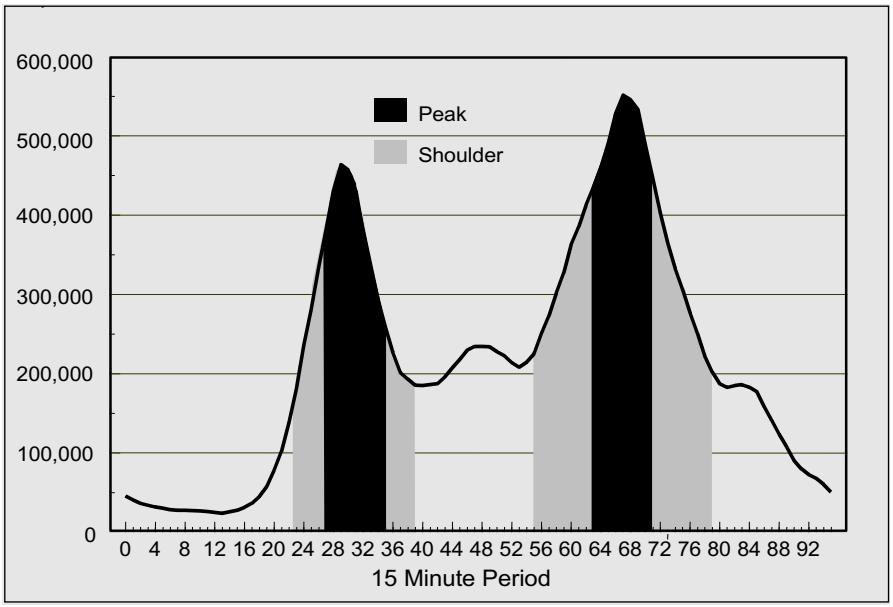
algorithm seeks equality in costs among alternative paths between the same origin-destination zone pair. The behavioral representation in the SOLA algorithm is that drivers find the best choice of travel route to minimize their travel costs. If a less costly route is available, drivers will choose to take it. This is Wardrop's user optimal principle.[5] This also means that an individual's route choice impacts all other travelers.

The process begins by calling the script for the first period assignment (the overnight period). The appropriate network scenario is selected and the corresponding demand matrices for the time period are called to assign the trips. For the first time period, the standard set of volume-delay functions (discussed below) are loaded, the scenario is prepared for assignment, and a full equilibrium assignment is completed. The procedure repeats through the remaining time periods. When the eight time periods are completed, the link volumes and travel times are successively averaged with the same time periods from previous global iterations. This occurs for the second and third global iterations since iteration 0 has no previous iteration to be averaged with. The result is a set of eight modeled scenarios representing each time period and containing the final MSA volumes and speeds.

The travel data that led to selecting the eight time periods is illustrated in Figure 13. This is a plot of the auto driver and auto passenger trips in motion reported in CATS' 1990 household travel survey. Trips were accumulated at the end of 96, 15-minute periods throughout the day. The plot shows a moving average of these accumulated trips calculated over four consecutive 15-minute periods. The moving average smooths out the irregularities in the plot that are caused by the tendency of surveyed travelers to report trip start and completion times to the nearest quarter-hour or half-hour. Analysis of the Travel Tracker and My Daily Travel data confirmed that the eight time periods are still relevant, thus, they continue to be used.

[5] Wardrop, J. (1952) "Some theoretical aspects of road traffic research," *Proceedings of the Institution of Civil Engineers,* Part 2: 325-378.

Figure 13. Time distribution of auto driver and passenger trips



The plot shows the distinct peaking of auto travel during the morning and evening peak periods. The large number of trips in motion during peak travel periods is due not only to increased trip making during these time periods. Peak period auto trips also stay in motion longer because they are more likely to be lengthy work trips subjected to slower congested peak period travel speeds.

The plot in Figure 13 is not symmetric because the evening peak period is longer and slightly worse than the morning peak. The two peak periods are separated by a midday period that has a uniform number of trips in motion, except for a bulge in trip making around the noon lunch period. Trips in motion plateau between 8:00 and 9:00 p.m. after the evening peak period, and then quickly decline during the late-night period.

The two assignment peak periods are defined differently because of these auto travel characteristics. The shading under the Figure 13 curve shows the peak and shoulder periods used in the multiple time period assignments. A two hour morning peak (7:00 to 9:00 a.m.) and two, one-hour morning peak shoulder periods (6:00 to 7:00 a.m. and 9:00 to 10:00 a.m.) effectively cover the morning peak period. Six hours are needed to capture the evening peak period, including: a two-hour early evening peak (4:00 to 6:00 p.m.) plus two hour afternoon and evening peak shoulder periods on either side of the PM peak (2:00 to 4:00 p.m. and 6:00 to 8:00 p.m.). This leaves a nearly uniform four-hour midday period between the two peaks (10:00 a.m. to 2:00 p.m.), and an off-peak period (8:00 p.m. to 6:00 a.m.) covering the late evening and early morning hours.

The time period assignments provide a more detailed and accurate picture of congestion effects in the highway network, which is advantageous for several reasons. While daily estimated traffic volumes may just be marginally improved compared to volumes produced by average daily assignments, estimates of network speeds are substantially improved and regional vehicle-miles of travel agree more closely with state estimates of daily vehicle-miles. Since congestion is more correctly modeled, impacts from proposed highway improvements that reduce congestion are also more accurately reproduced by the time period assignments.

Time-of-day factors

Auto trips processed through the mode choice-destination choice model proceed to the time-of-day model, where they are assigned to one of the eight TOD periods. Truck trips and the external auto trips are not part of this process, so factors are applied to convert this demand into TOD trips. These factors are listed in Table 88.

TOD factors for the other trips were developed using the following methods:

Heavy trucks: Time-of-day factors for these vehicles were derived directly from the ATRI truck trip dataset. These same factors are applied to external truck trips.

Light and medium trucks: Factors for these vehicles were developed using transaction data from the Illinois Tollway; specifically focusing on Tier 2 and Tier 3 transactions (which correspond to CMAP's light and medium truck categories, respectively). These data were combined with hourly count data of single unit trucks on Cook County arterials provided by IDOT. The final factors represent an averaging of these two data sources.

All other vehicles: The TOD factors for all other vehicles were developed from the model validation time-of-day traffic analysis.

Table 88. Commercial vehicle and external auto time-of-day factors

	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	Period 7	Period 8
external auto	0.163	0.052	0.127	0.051	0.211	0.133	0.155	0.109
Air Passenger to/from airports	0.163	0.052	0.127	0.051	0.211	0.133	0.155	0.109
B trucks	0.230	0.007	0.016	0.016	0.238	0.165	0.176	0.152
L truck	0.230	0.007	0.016	0.016	0.238	0.165	0.176	0.152
M trucks	0.230	0.007	0.016	0.016	0.238	0.165	0.176	0.152
H trucks	0.213	0.044	0.121	0.067	0.266	0.121	0.103	0.065
external trucks	0.213	0.044	0.121	0.067	0.266	0.121	0.103	0.065

Volume-delay functions

The volume-delay functions (VDFs) are used to represent the congestion that occurs on links as traffic volumes increase. The volume-delay functions include estimated traffic signal characteristics for links that end at signalized intersections. This means that assignments are sensitive to signal characteristics and can reflect major signal modernization programs. In addition to more accurately representing the characteristics of the network, these signal sensitive volume-delay functions allow the emission reductions from signal improvements to be evaluated.

CMAP's volume-delay functions have evolved from their initial versions because of the previous functions' limitations when they were used for time period assignments. Their most severe limitation was that freeways and expressways tended to be over-assigned in the congested peak time periods. Several factors contributed to this peak period over-assignment including: (1) an

unrealistic initial peak period assignment since paths were built using uncongested travel times; (2) the inability to model bottlenecks in the freeway network that occur during peak periods, and; (3) not restricting freeway on-ramps whose peak period capacities were controlled by metering. The approach taken was to alter the volume-delay functions for freeways, expressway and metered freeway entrance ramps so that travel times increase far more quickly after capacity is reached. The capacities of metered on-ramps are also set to maximum metered flow rates.

Note that the link volume included in the volume-delay function calculations includes all assigned auto and truck traffic (in VEQs), as well as buses operating on the roadway links, represented as three vehicle equivalents. The capacity values in CMAP's VDFs represent lane capacities at level of service E traffic conditions except for arterials (vdf1) and ramps connecting freeways (vdf5), which reflect level of service C.

VDFs for links ending at signalized intersections (vdf1 and vdf3)

Intersection delays in the volume-delay functions are based upon the Webster equation.[6] In this equation, intersection delay has uniform and incremental components, and both are rather complicated to calculate. For CMAP's volume-delay functions, simpler regression equations were fit to calculated uniform and incremental delays for a range of signal cycle lengths and green time-to-cycle length ratios.

The regression equations for uniform and incremental signal delays are combined with link travel time estimates in the first (arterial) and third (freeway exit ramp to arterial) volume-delay functions as follows:

Link travel time between intersections is:

$$T_{link} = T_0 \times \left(1 + 0.15 \times \left(\frac{volume}{capacity}\right)^4\right)$$

This is the widely used BPR (Bureau of Public Roads) function where T_{link} equals the link's travel time without any intersection delay and T_0 is the uncongested link travel time without intersection delay. The uncongested link travel time is computed using the maximum speed permitted on the link. Quantity *volume* is the link's traffic volume for the time period in auto equivalents. Capacity represented within the link travel time function is approximately the service volume at level of service C. It is calculated as 75 percent of the level of service E time period link capacity. Note that link capacity is calculated by multiplying the hourly lane capacity by the number of lanes and the number of hours in the assignment time period.

Uniform intersection delay is the maximum of zero or:

$$D_u = 5.96 \times \left(\frac{volume}{capacity}\right) - \left(0.234 \times cycle \times \frac{green}{100}\right) + (0.21 \times cycle) - 4.47$$

[6] F. V. Webster and B. M. Cobbe, *Traffic Signals*. Road Research Laboratory, Ministry of Transport Road Research, Technical Paper No. 56, 1966.

Where D_u is the average uniform intersection delay at the link's j-node in seconds. *Green* is the green time allowed the link at the j-node intersection and *cycle* is the cycle length at the intersection. Both quantities are in seconds. The uniform delay is restricted to positive values in the volume-delay functions.

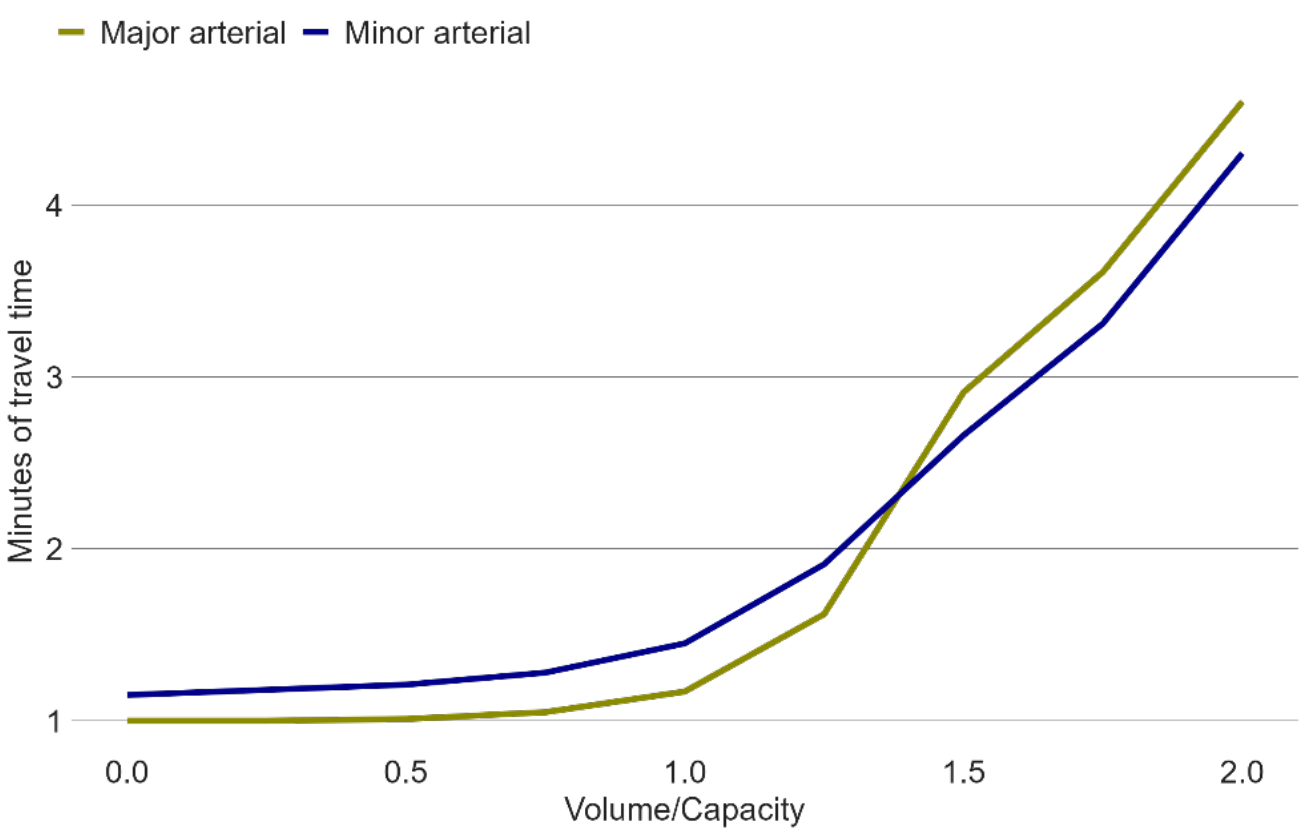
Incremental delay at intersections is the maximum of zero or:

$$D_i = 2.65 \times \left(\frac{volume}{capacity}\right)^8 - \left(7.3 \times \frac{green}{100}\right) + 0.338$$

Where D_i is the average incremental intersection delay at the link's j-node in seconds. Incremental delay is also restricted to positive values in the volume-delay functions.

Figure 14 shows the estimated travel times for a major and minor arterial including intersection delay. Both arterials have an uncongested travel time of one minute between intersections. For the minor arterial the signal cycle length at the j-node is ninety seconds, and the link receives thirty seconds of green time in the cycle. The signal cycle length at the j-node is 120 seconds for the major arterial, with ninety seconds of green time in the cycle. The major arterial is allowed more green time at the j-node than the minor one and intersection delays on the major link are less than on the minor link until the volume to capacity ratios exceed 1.25. Both volume-delay relationships have a kink in them because the maximum combined uniform and incremental intersection delay is limited to one cycle length.

Figure 14. Example volume-delay functions for two arterial links



VDFs for freeways and expressways (vdf2, vdf4 and vdf5)

The second (freeway), fourth (expressway) and fifth (freeway-freeway ramps) volume-delay functions start with a variation on the BPR function. Additional adjustments were made to the expressway function: uncongested link travel times on freeway links were reduced 15 percent to reflect drivers' tendency to exceed speed limits on high-volume facilities at low traffic volumes, and the link capacity value was increased by 300 vehicles to reflect traffic management and operations strategies implemented on the region's Interstate system. For all three of these VDFs the exponent was increased from the BPR value to increase travel times more dramatically for volume to capacity ratios greater than one. The quantity T_0 is determined by the maximum legal speed

For freeways (vdf2):

$$T_{link} = \left(\frac{T_0}{1.15}\right) \times \left(1 + 0.15 \times \left(\frac{volume}{capacity + 300}\right)^8\right) \times \left(1 + 0.15 \times Minimum\left(\frac{volume}{capacity}\middle|1.0\right)\right)$$

For expressways (vdf4):

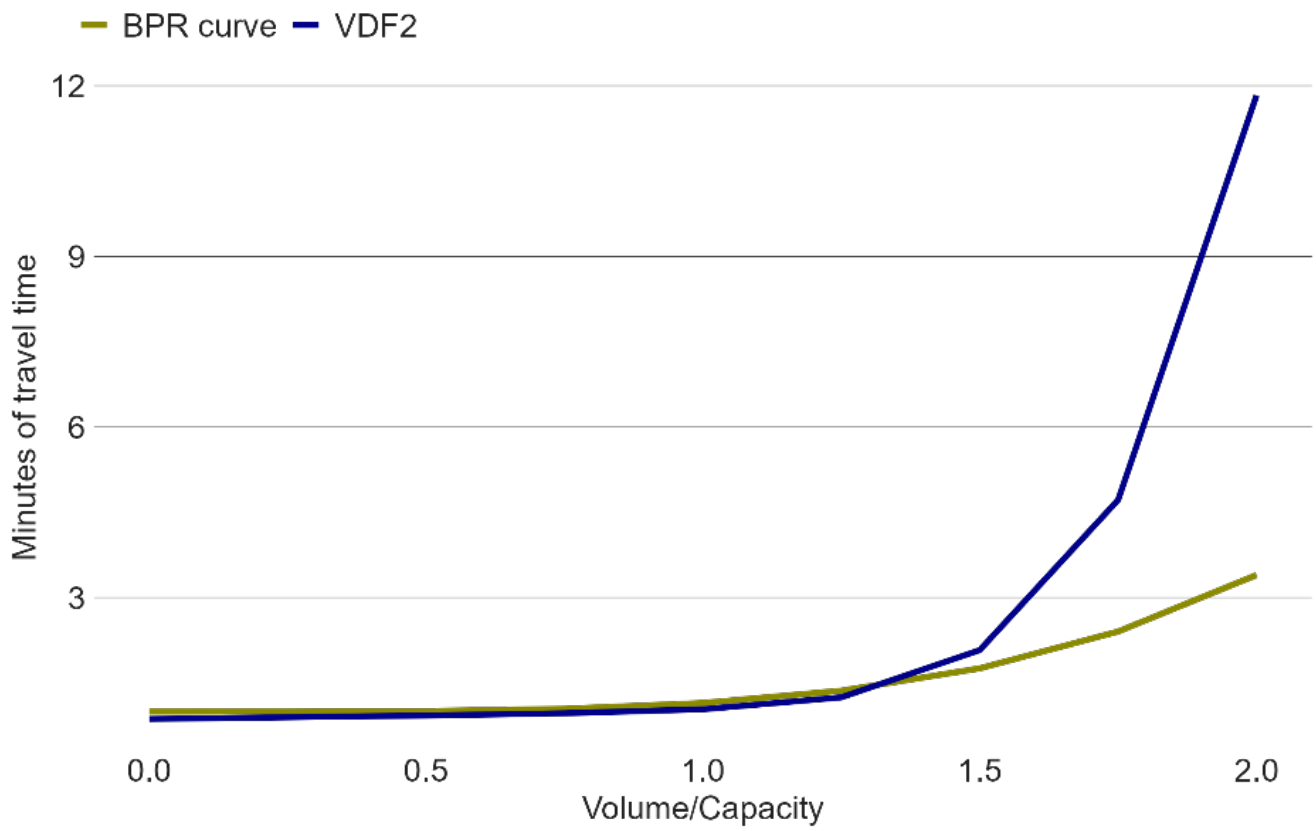
$$T_{link} = T_0 \times \left(1 + 0.15 \times \left(\frac{volume}{capacity}\right)^8\right) \times \left(1 + 0.15 \times Minimum\left(\frac{volume}{capacity}\middle|1.0\right)\right)$$

For freeway-to-freeway ramps (vdf5)

$$T_{link} = T_0 \times \left(1 + 0.15 \times \left(\frac{volume}{capacity}\right)^8\right)$$

Figure 15 compares expressway link travel times using CMAP's VDF to those from the original BPR volume-delay function for a one mile link with a maximum speed of 60 miles per hour. At volume to capacity ratios below 1.25, both functions predict very similar link travel times; in fact, the CMAP value is slightly below that of the BPR curve due to the lower initial uncongested travel time. For volume to capacity ratios greater than 1.25, the travel time predicted by CMAP's function is higher and rapidly increases because the volume to capacity ratio is exponentiated to a higher power.

Figure 15. Revised BPR volume-delay function for freeway links



VDF for metered freeway entrance ramps (vdf8)

For metered freeway entrance ramps, the original BPR function is revised so that travel time greatly increases when the link volume exceeds the maximum metered flow rate. This effectively restricts the ramp's volume to the metered flow rate. The adjusted BPR function is:

$$T_{link} = T_0 \times \left(1 + 0.15 \times \left(\frac{volume}{metered\ flow}\right)^{10}\right)$$

The maximum metered flow rate is taken as 720 vehicles per hour per lane, or an average vehicle delay at the ramp metering signal of five seconds.

VDF for links with tolls (vdf7)

The original toll collection link volume-delay function implemented in the trip-based model decades ago attempted to measure delay at the toll collection facility itself through the generalized cost (time and money) of using the facility. This method of reflecting the impact of tolls on individuals' route choice makes little sense today given the state of toll collection technology. The entire system operated by the Illinois Tollway uses electronic toll collection technology: mainline plazas require no reduction in operating speed and no-stop tolling is available at ramp tollbooths. Further, most toll transactions in the CMAP region use transponders, thus only a small share of tollway drivers even stop at plazas.

The current toll collection link volume-delay function merely reflects the travel time on the link (generally coded as 200 feet long) based on the average speed of the incoming link. These links are of negligible length within the larger overall network. Thus, no attempt is made to constrain their capacity. The true impact of tolling on route choice is measured through the generalized cost procedures described earlier.

Link speeds

Traffic volume on every link for each period of the day is one product of the time-of-day network assignment. The speed of travel for each link is calculated by an equation that uses the volume-capacity ratio for the link as the independent variable. The following equations are used to produce the final link speed.

Freeways:

$$S = S_0 \frac{1}{1+0.15\,(V/C)} \times \frac{1}{1+0.15\,(V/C)^8} \qquad \text{for } V/C \leq 1$$

$$S = S_0 \frac{1}{1+0.15\,(V/C)^8} \qquad \text{for } V/C > 1$$

Arterials:

$$S = S_0 \frac{1}{(\ln(S_0) * 0.249) + 0.153\,(V/(C * .75))^{3.98}}$$

Where:

S = Speed on link used for emission calculation

S_0 = Initial speed on link

V/C = Volume-capacity ratio for the link

These curves represent modifications to the BPR curves that have been used at CMAP and other agencies for many years. Consistent with a national trend for agencies to use modified curves based on local data, these curves are based on the information gathered from local empirical data. The freeway curve is the same as used in the volume delay functions in the time-of-day assignment iterations. The arterial curve is slightly modified to better correlate with the empirical data. The data used to develop the modification is from IDOT's traffic sensor system for the expressway system as well as CATS-conducted speed runs for the arterial system. This database is documented in CATS Working Paper 95-09: Travel Time Database and Structure Chicago Area Expressway System (September 1995), and CATS Working Paper 97-09: 1994, 1995 and 1996 Combined Travel Time

Database Documentation: Arterial Highway System (July 1997)[7]. The methodology for the curve development is presented in CATS Working Paper 97-12: Method for Adjusting Modeled Speeds Based on Empirical Speed Data (August 1997).

Emissions calculation

As northeastern Illinois does not meet federal air quality standards for ozone, CMAP must show the region's long range transportation plan and Transportation Improvement Program meet established emission budgets. This is the air quality conformity process. Final link volumes and link speeds from the travel demand model serve as inputs to the vehicle emissions model, which is used to make the conformity determination. Please refer to the Air Quality Conformity Documentation Appendix for more information.

Travel model validation

An important step following the development of a travel model is model validation, demonstrating that model results are reasonable when compared to observed data. Model validation is focused on comparing the results of a calibrated travel model to other data sources such as daily traffic counts, daily transit boardings, and more recent Census data. Best practice within travel demand modeling is to use different datasets for model calibration and validation activities to confirm that the results reasonably replicate observed travel patterns. In instances where comparison data are limited, a dataset may be used for calibration and validation if the data records are separated into distinct subsets for each purpose.

There are no specific regulations defining what a well-validated travel model looks like, and the federal government has not developed specific benchmarks that must be met to show that a travel model is well-validated. The Federal Highway Administration has published the Travel Model Validation and Reasonableness Checking Manual to offer guidelines and examples on how travel model validation can be evaluated. It is important to note that no single evaluation measure can be used to determine if travel model results are reasonable. Rather, the model should be evaluated across an array of measures to establish an overall pattern consistent with observed data.

Travel demand model results are not expected to perfectly match observed data, and in fact travel demand modelers are wary of "over-fitting" models. This refers to adjusting model parameters to a point where the results nearly match observed data. While this might sound like a worthwhile goal, it is often the case that this level of fine-tuning leads to a loss in the model's sensitivity to key variables and thus diminishes the model's explanatory and predictive capabilities.

[7] Note: All CATS Working papers are available on the CMAP Data Hub.

The following sections, divided into trip generation, trip distribution, mode choice, and travel assignment, compare various aspects of the model against real-world measurements.

Trip generation

Household crosstabs

This section compares the travel models' synthetic households to the 2017-2021 5-year ACS PUMS data. The households used in the trip generation model are developed during the population synthesis process within the UrbanSim land use model. Household characteristics are defining factors used within trip generation to develop daily household productions and attractions, which are the building blocks for the remainder of the travel model.

Data are summarized for the seven counties in CMAP's Metropolitan Planning Area (MPA) and the external area that comprises the remainder of the modeling geography. The PUMS data were summarized to be consistent with the household type definitions used in CMAP's trip generation model (e.g., adults were defined as anyone age 16 and older).

In total, the travel model's household characteristics match PUMS data very closely, with differences of less than one percent in most categories. The largest differences between modeled and observed household characteristics occur in very large or very small household sizes. Figures 16 through 18 below describe these differences in more detail.

Figure 16 compares households by number of adults and children. The single largest categories are "1 Adult, 0 Children" households and "2 Adult, 0 Children" households within CMAP's seven counties, which combined represent nearly half of the total households in the model. The model slightly underrepresents "1 Adult, 0 Children" households (by 1.4 percentage points) but matches the share of "2 Adult, 0 Children" households almost exactly. All other categories show a close correlation between the modeled and observed data.

Figure 16. Share of households by number of adults and children

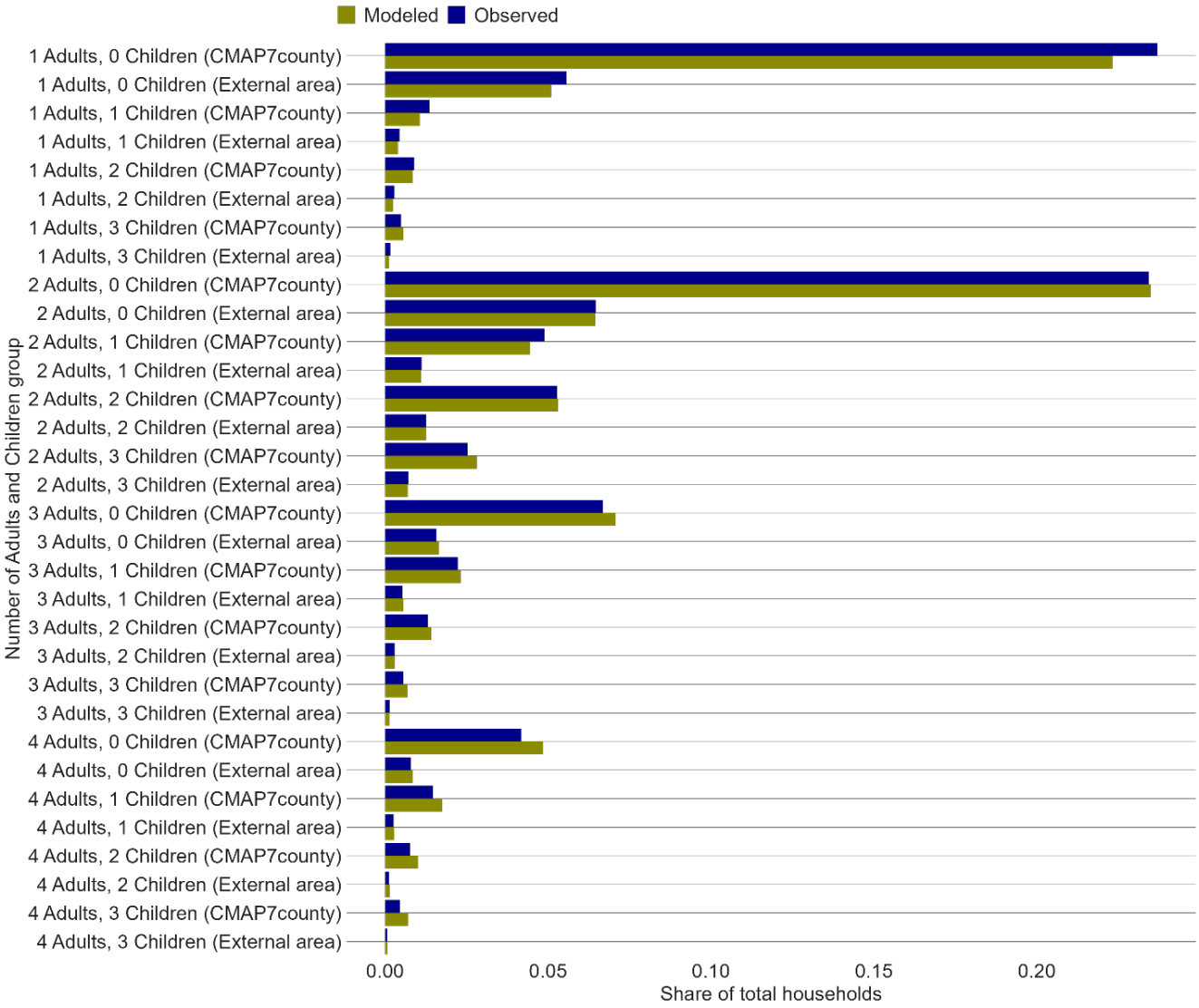


Figure 17 compares households by number of adults and workers. The largest category is “2 Adult, 2 Worker” households in the CMAP seven counites, and the modeled share of total households nearly matches the observed share exactly. The model slightly underrepresents both categories of “1 Adult” households in the CMAP area and overrepresents households with at least four adults and at least three workers in the seven counties. Overall, the modeled data closely matches the PUMS data.

Figure 17. Share of households by number of adults and workers

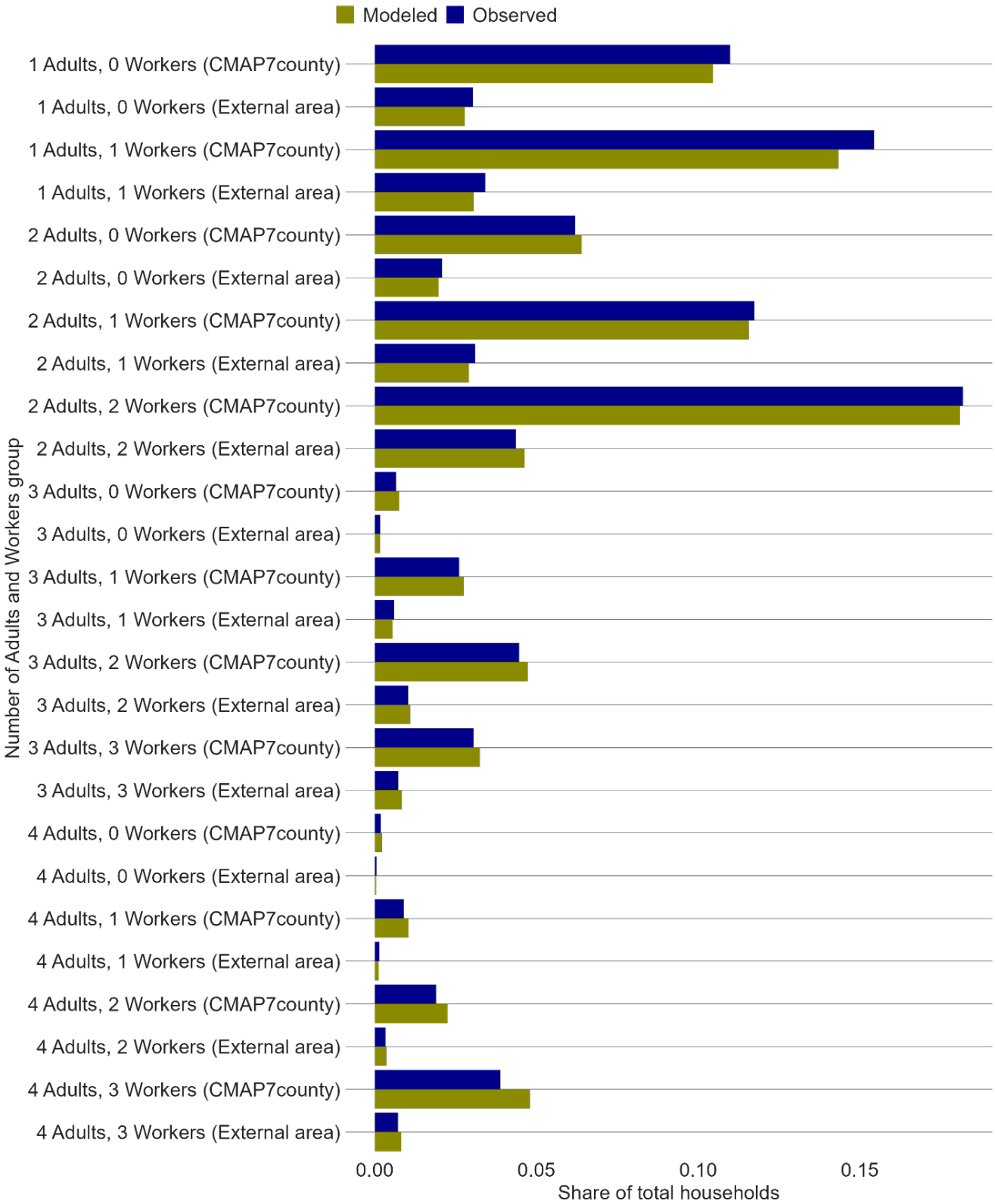
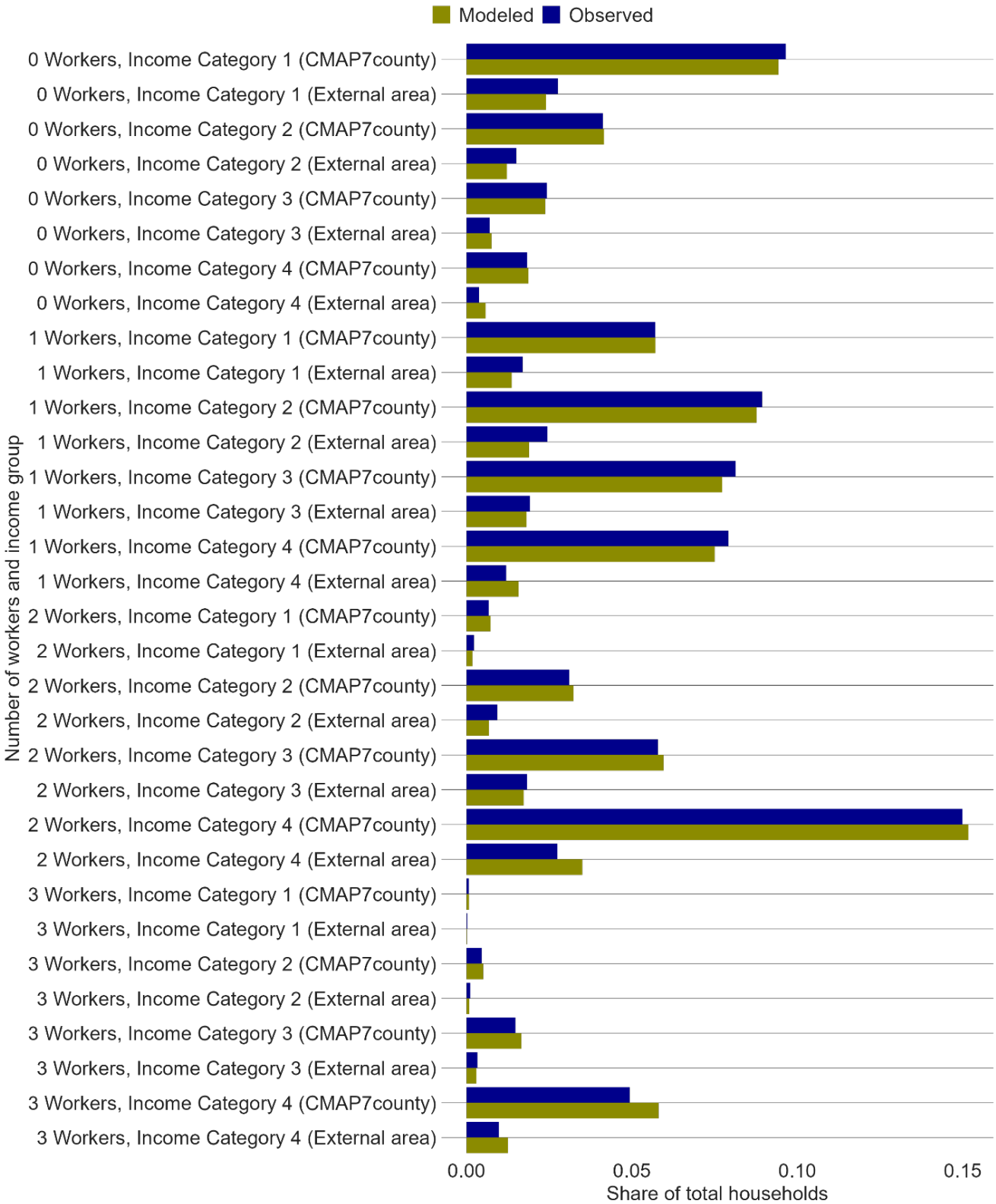


Figure 18 compares households by number of workers and household income category. These reflect the income categories listed in Table 30. Overall, the modeled households reflect the observed data. Within the CMAP seven counties, the model slightly underpredicts the share of "1

Worker" households for all but the lowest income category and overpredicts the share of households with at least three workers in the highest income category by around one percentage point.

Figure 18. Share of households by number of workers and income category



Vehicle availability crosstabs

Household vehicle ownership plays an important role in the model's trip generation and destination choice calculations. This section provides a comparison of vehicles available to synthetic households in the travel demand model to the 2017-2021 5-year ACS PUMS data for vehicle ownership per household.

Overall, the travel model data exhibits the same pattern as the PUMS data. Figure 19 below shows the share of total households in both CMAP's seven-county and the external area by number of vehicles available and adults per household. Households with 2 adults and 2 vehicles make up the largest share of households in the 7-county region (20 percent for both the PUMS data and the CMAP model data). The second largest group is households with 1 adult and 1 vehicle in the 7-county region, which the travel model slightly underestimates.

Figure 19. Household share by number of vehicles and adults

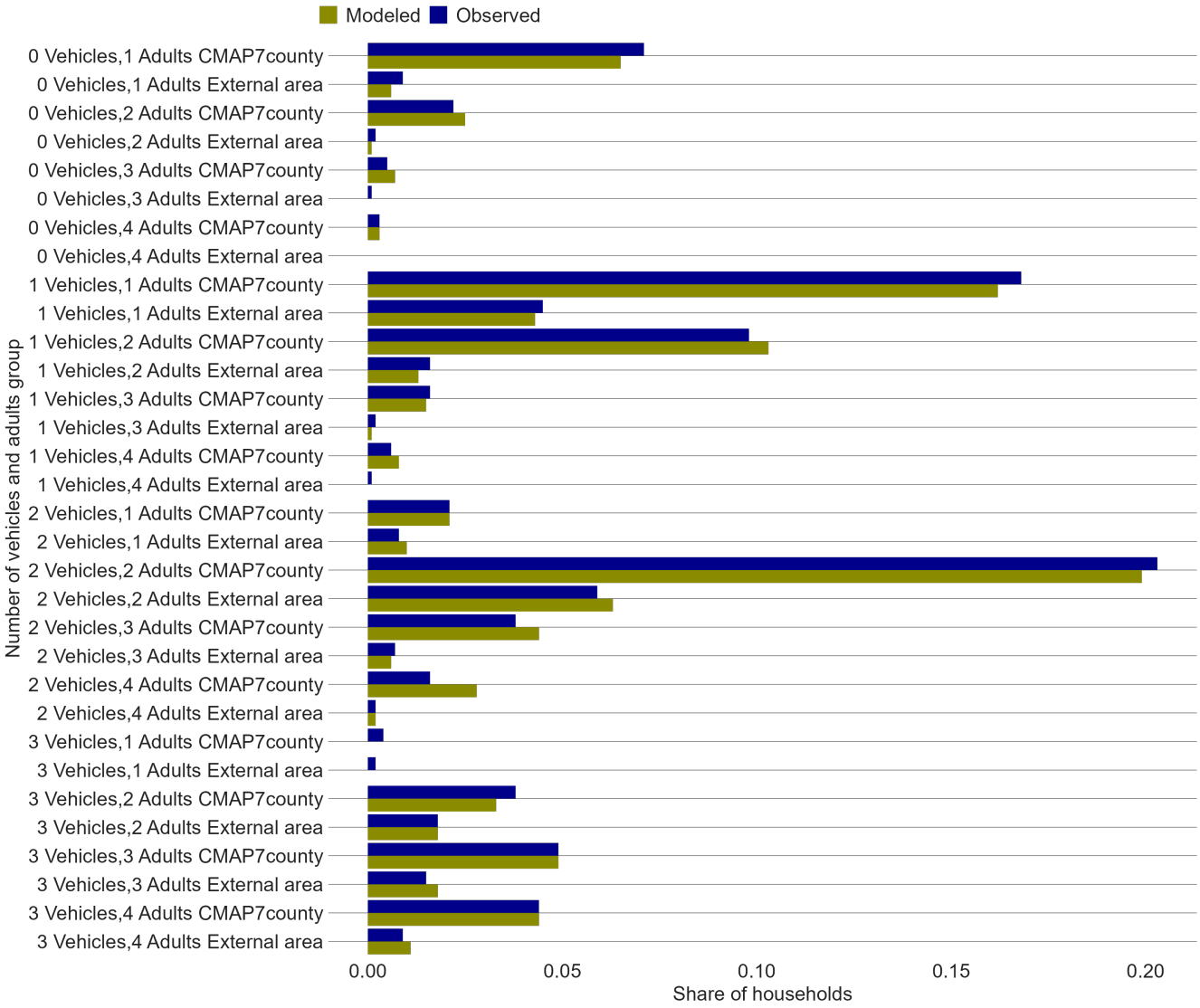


Figure 20 illustrates the share of total households by number of vehicles and number of workers. The model shows a pattern that closely matches the observed data. The two largest categories are 1 vehicle-1 worker and 2 vehicles-2 workers in the 7- county region. The discrepancies between the modeled and observed data for these categories are less than one percentage point.

Figure 20. Household share by number of vehicles and workers

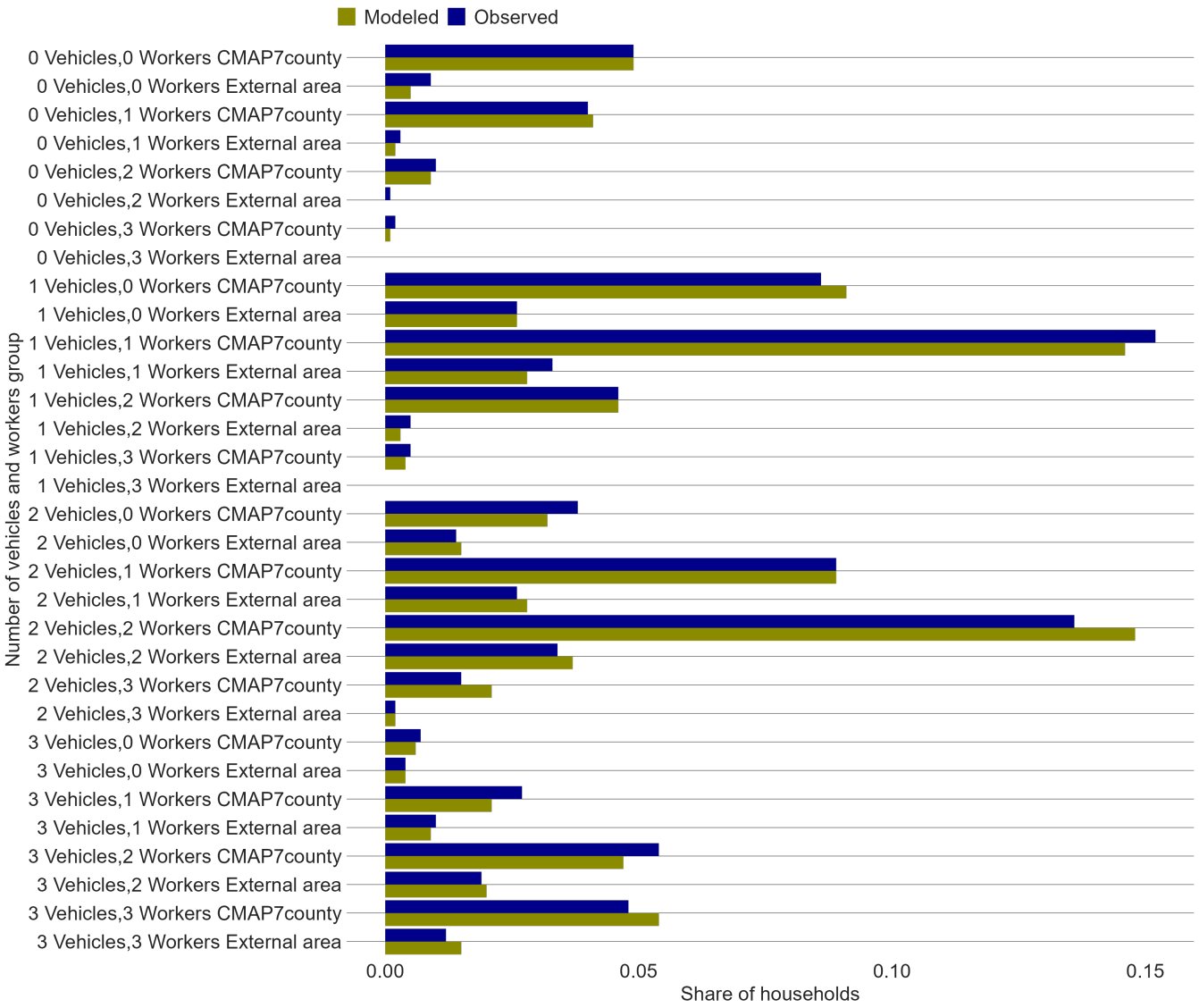
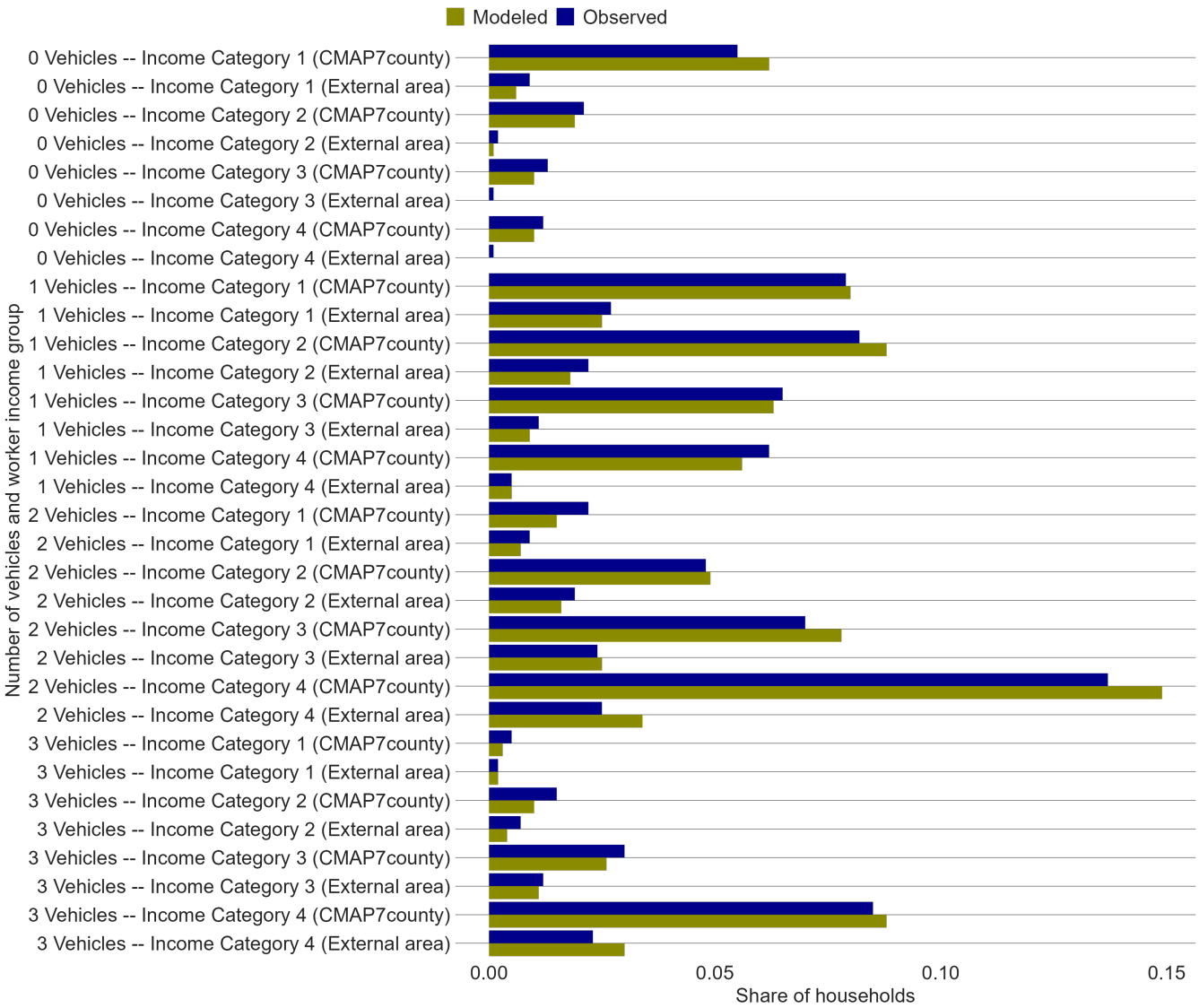


Figure 21 shows the share of total households by number of vehicles and household income category. Comparing households based on these attributes, the largest model overestimation occurs in the two-vehicle category in the highest income level for both the CMAP seven-county and external areas. Overall, the model data support the general pattern reflected in the PUMS data.

Figure 21. Household share by number of vehicles and income category



Trip rates versus the NHTS

Once household information is imported and processed, the trip generation model converts household data into individual trips. To verify the reasonableness of the number of trips generated in the CMAP travel demand model, data are compared to the 2017 National Household Travel Survey (NHTS). The NHTS data have been processed to make trip purpose definitions as comparable to the CMAP model definitions as possible, and trips with a primary purpose of "change mode" or "serve passenger" have been linked with the following trip.

Table 89 displays average household trip rates in the model and in the 2017 NHTS. Home-based shopping trips and non-home based trips contain the largest difference between the model and the

NHTS, with a daily trip difference of 0.3 trips per household for each purpose. Home-based work and home-based other trips are within 0.1 trips per household compared to the NHTS.

Table 89. Average household trip rates

Trip purpose	CMAP model	2017 NHTS	Difference
Home-based work	1.7	1.6	0.1
Home-based shop	0.7	1.0	-0.3
Home-based other	2.9	2.8	0.1
Non-home based	1.9	2.2	-0.3
All trips	7.2	7.6	-0.4

Table 90 provides a trip rate comparison for households with and without teleworkers. The model data show patterns like the NHTS data: fewer home-based work trips in households with teleworkers and more home-based other trip in households with teleworkers. Note that the NHTS data for home-work trips for households with teleworkers seems a bit questionable. It shows virtually no difference in the average number of home-work trips for households without teleworkers, which is counterintuitive and conflicts with other observed data. The largest discrepancy between the model data and the NHTS data occurs for non-home based trips in households with teleworkers, with the model showing a small decrease in values while the NHTS shows a small increase. Non-home based trips are one of the more difficult trip purposes to estimate because the origin and destination of these trips are both unknown at the beginning of modeling, while for home-based trips either the origin or destination is the home location. The model underestimates non-home based trips for households with at least one teleworker by less than one trip, which is not large enough of a discrepancy to cause concern.

Table 90. Average household trip rates based on the presence of teleworkers

Teleworkers	Trip purpose	CMAP model	2017 NHTS	Difference
Households with zero teleworkers	Home-based work	1.8	1.6	0.2
	Home-based shop	0.7	1.0	-0.3
	Home-based other	2.8	2.7	0.1
	Non-home based	2.0	2.2	-0.2
	All trips	7.3	7.5	-0.2
Households with at least one teleworker	Home-based work	1.1	1.5	-0.4
	Home-based shop	0.9	1.0	-0.1
	Home-based other	3.7	3.4	0.3
	Non-home based	1.6	2.5	-0.9
	All trips	7.3	8.4	-1.1

Employment

In addition to reviewing household data, it is important to look at employment data, another driver of activity within the trip-based model. Table 91 presents a comparison between the work trip attractions and jobs, and between the employment used as input to the trip generation model and observed Longitudinal Employer-Household Dynamics (LEHD) employment. Within the CMAP

region, the 2019 total employment considered by the trip generation model is virtually identical to the 2019 LEHD regional employment. This is expected given that both are derived from data provided by the Illinois Department of Employment Security (IDES). The model uses IDES data that has been improved by CMAP staff to properly allocate employment inaccurately reported at headquarters locations to their actual branch locations. The LEHD data do not reflect these updates.

If all workers traveled to their jobs each day, the model would produce two attractions for every job, since the production-attraction format of the trip data in the model reflects both the home-to-work and work-to-home trips as ending at work. However, many people do not go to work every day due to job schedules, vacations, appointments, illness, or a host of other reasons, so the true number of attractions per job should be less than two. A benchmark of between 1.20 and 1.55 work attractions per job is commonly cited as a reasonable range based on work from the Florida Department of Transportation. All counties within the CMAP area have a modeled ratio within the benchmark range and compare reasonably well with the observed ratios, with the largest discrepancies in Kendall and McHenry counties.

Table 91. Modeled home-based work attractions per observed employee

	Attractions	Model input employment	Observed employment	Attractions / Model input employment	Attractions / Observed employment	Employment difference (model – observed)
Chicago - Cook	2,064,068	1,409,162	1,444,392	1.46	1.43	-2.5%
Suburban Cook	1,528,962	1,220,829	1,221,801	1.25	1.25	-0.1%
DuPage	781,790	617,695	639,614	1.27	1.22	-3.6%
Kane	267,367	213,415	218,456	1.25	1.22	-2.4%
Kendall	40,368	32,972	28,366	1.22	1.42	14.0%
Lake	436,010	346,755	334,367	1.26	1.30	3.6%
McHenry	130,675	104,639	93,794	1.25	1.39	10.4%
Will	331,743	267,366	260,203	1.24	1.27	2.7%
CMAP region	5,580,983	4,212,833	4,240,993	1.32	1.32	-0.7%

Sources: LEHD Origin-Destination Employment Statistics (LODES7) data by county, 2019 Workplace Area Characteristics for All Jobs accessed via the 'lehdr' R package.

Work from home household analysis

Many workers in the region work from home for one or more days of the workweek. It is important to accurately model work-from-home behavior since working from home directly impacts the volume of traffic on the roadway. Observed data was obtained from 2017-2021 5-year ACS PUMS data. Work-from-home rates were compared at the worker level and at the household level. Table 92 shows the modeled share of work-from-home workers compared to the observed share of work-from-home workers by county.

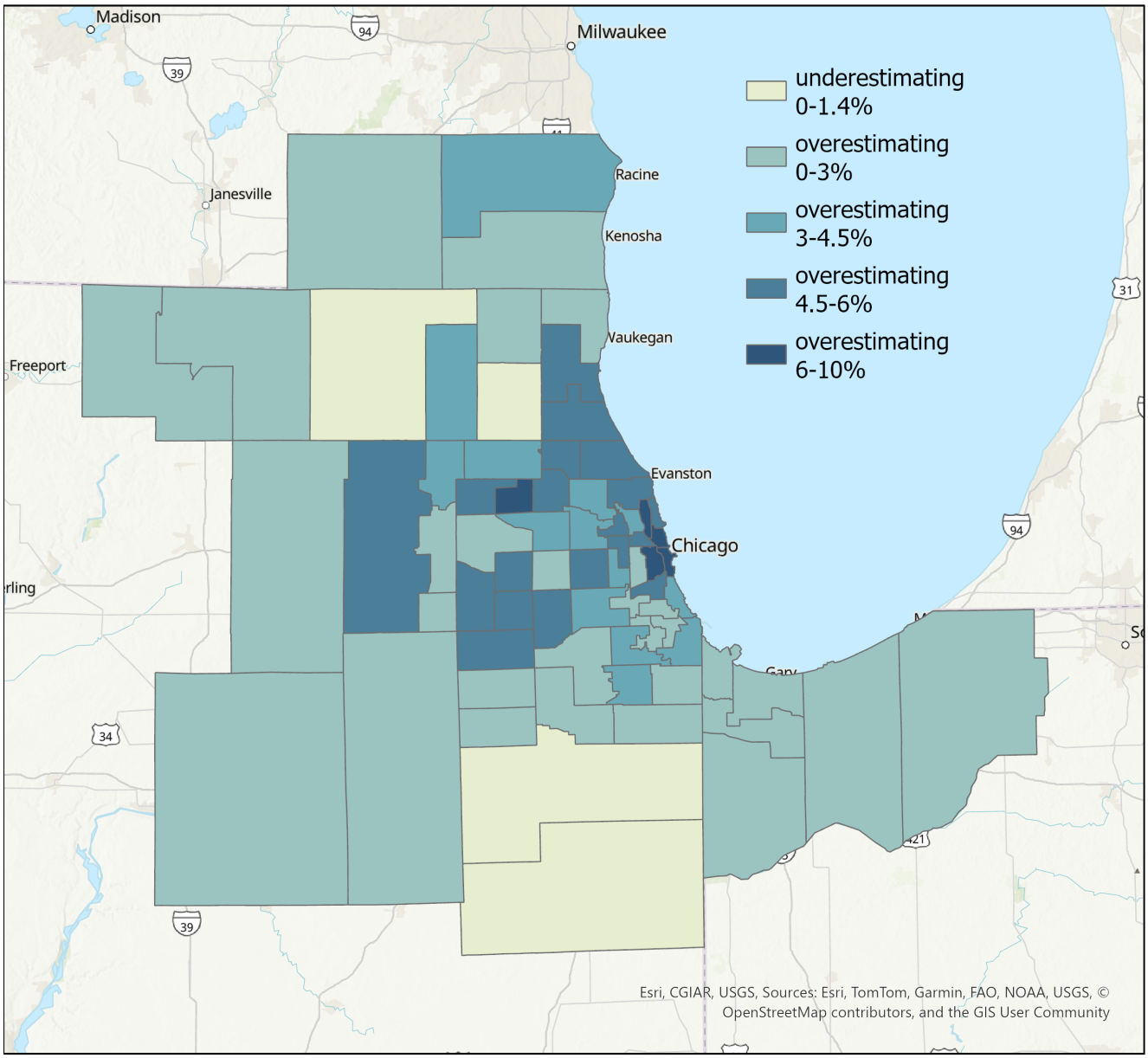
Table 92. Modeled vs observed share of work from home workers by county

County	Modeled	Observed
Cook	0.088	0.026
DuPage	0.106	0.041
Kane	0.076	0.027
Kendall	0.065	0.038
Lake	0.087	0.035
McHenry	0.085	0.035
Will	0.076	0.027

Regionally, modeled work-from-home households are 3.64 percent higher than is reported in the observed data. This is expected, because the ACS questionnaire asks survey respondents to describe their "typical" workday, missing individuals who work from home one or two days a week. As noted in the trip generation section, the CMAP model provides a more holistic accounting of teleworkers by including those workers who usually work from home (the PUMS share) and those who work from home on a less frequent basis.

Figure 22 depicts the differences in telework rates by PUMA. Modeled telework household rates range from 1.4 percentage points lower than the observed rate up to 10.2 percentage points higher than the observed rate. While the PUMS data tends to slightly underestimate work-from-home households due to its more limited definition of teleworkers, most PUMAs have a modeled rate that falls within 5 percentage points of the observed share.

Figure 22. Modeled versus observed household work-from-home rates, 2019



Trip distribution

Home-based work trip distances

After trips are generated, the travel model determines where and how individual trips are distributed throughout the region. Table 93 compares average work trip distances in the travel model against 2017-2021 ACS CTPP data, sorted by county of residence. Modeled roadway distances were applied to CTPP tract-to-tract trip flows to estimate the distance traveled on the transportation network. Regionwide, the model underestimates work trip distances from most counties by about 4.85 miles.

Table 93. Average work trip distances (miles) by residence county

County	Modeled	Observed
Cook	9.32	13.55
DuPage	11.18	14.77
Kane	12.44	17.42
Kendall	15.60	22.57
Lake	11.60	16.96
McHenry	13.52	19.60
Will	13.77	19.87
CMAP Region	10.69	15.54

Figure 23 displays the comparison between modeled and observed work trip distances by TAZ. Overall, modeled distances from trips originating within the MPA tend to match more closely to observed data compared to distances from trips originating outside the MPA.

Figure 23. Difference in average work trip distance

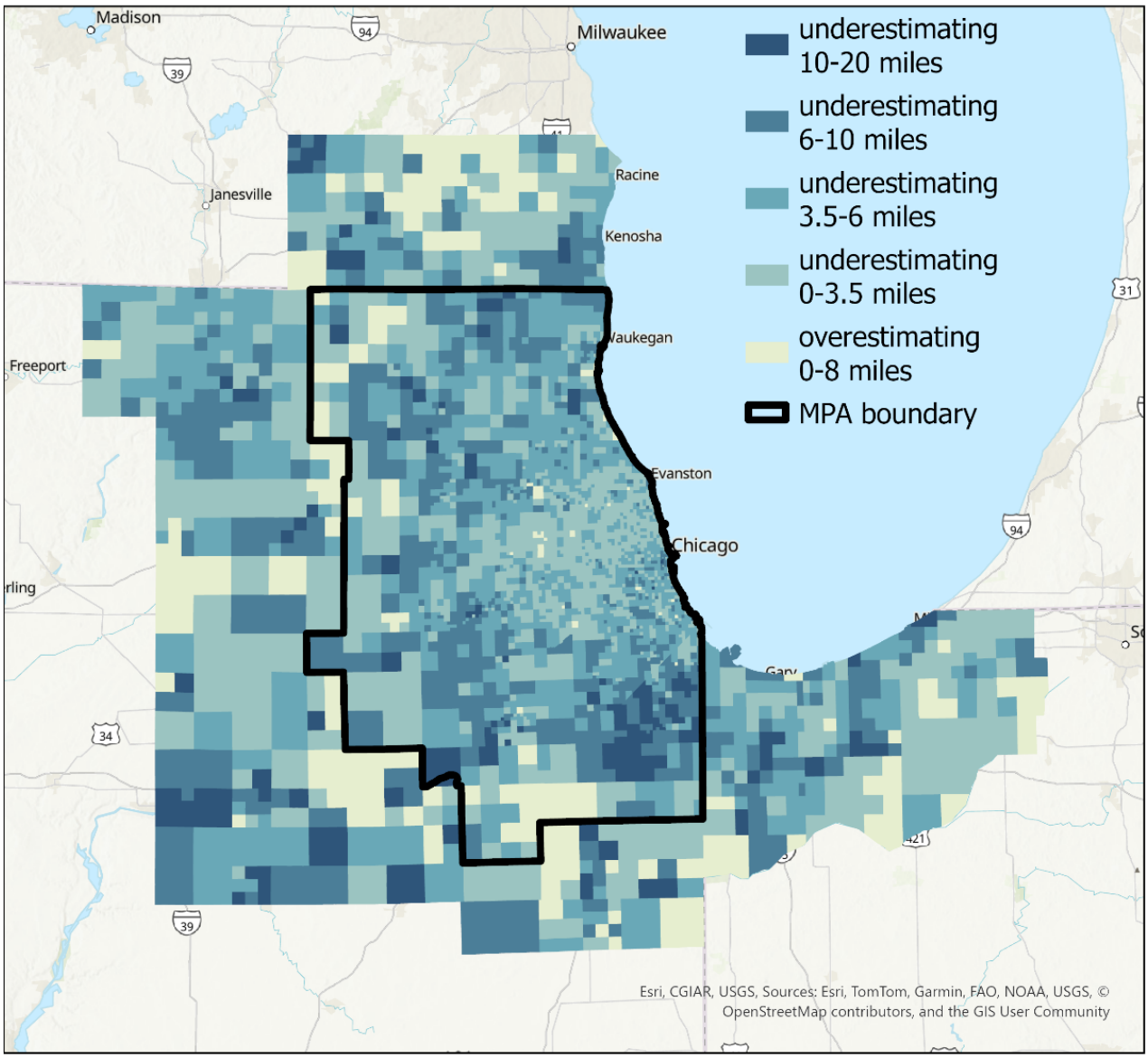
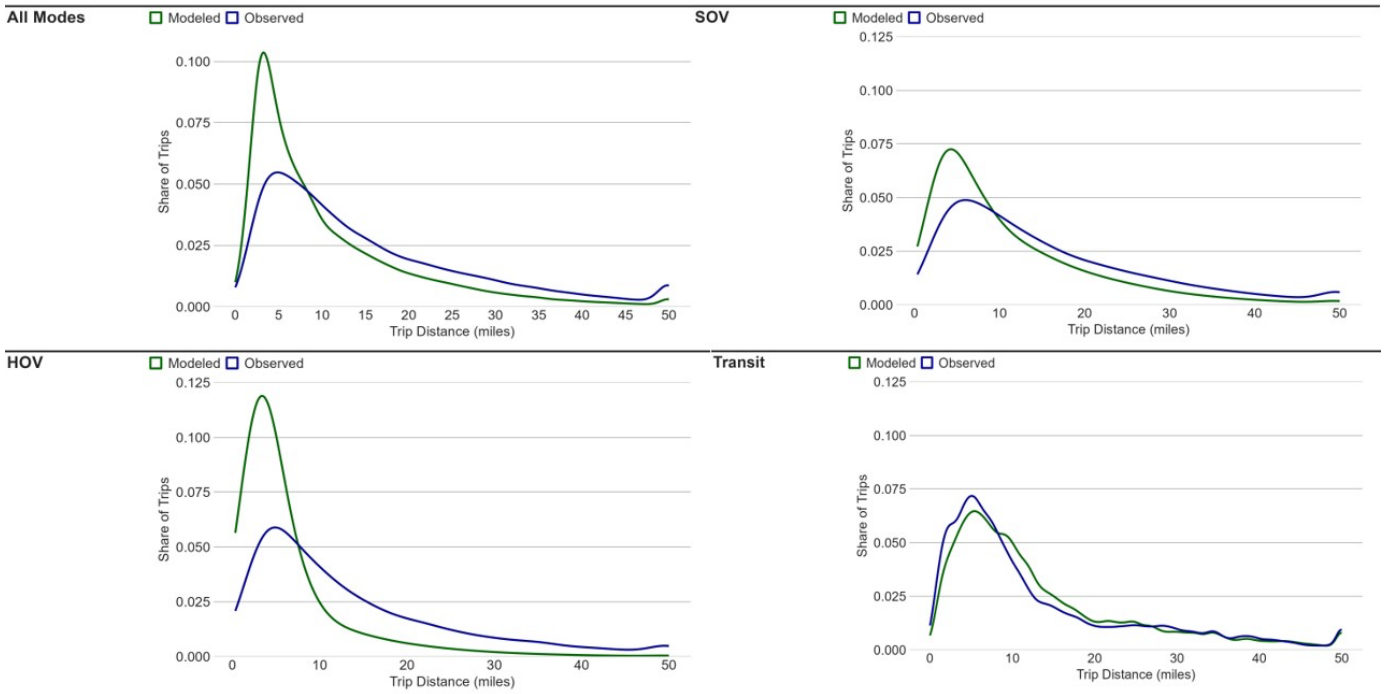


Figure 24 illustrates observed versus modeled trip distances by travel mode. For SOV and HOV modes, the model tends to assign shorter work trip distances compared to observed data. For transit trips, trip distances match relatively well with observed data.

Figure 24. Share of observed and modeled trips by distance for each mode



County-to-county work trip flow comparison

While the model tends to underestimate SOV and HOV trip distances, the number of modeled work trips strongly correlates to observed data. Table 94 illustrates a ratio of observed to modeled county-to-county work trips.

Observed trips comprise all primary jobs within the CTPP data derived from the 2017-2021 ACS. County-to-county ratios were calculated as the total number of modeled work trips divided by the total number of CTPP work trips. A ratio value of 1.00 reflects perfect agreement between the modeled and observed flows; values greater than 1.00 indicate more modeled trips than observed, while values less than 1.00 indicate less modeled trips than observed.

Overall, modeled and observed work trip flows for all county pairs have a correlation coefficient of 0.928, indicating a strong positive relationship between the sets of values. Most county pairs have modeled work trips within 30 percent of observed trips. County-to-county flows to, from, and between counties outside of the seven-county region tend to diverge from observed data more greatly; of the twelve county-to-county flows that are more than 50 percent different from observed data, eight include counties outside of the region. Large over- and under-estimations occurring in external counties are not as concerning as the travel model is primarily calibrated on the seven-county MPA.

Table 94. Work trip flow ratios

Origin	Cook	DuPage	Kane	Kendall	Lake	McHenry	Will	Ext. IL	Ext. IN	Ext. WI
Cook	0.88	0.74	0.93		0.51	0.73	0.63	0.25	1.70	0.30
DuPage	0.97	0.84	0.96	1.06	0.49		0.91	0.31		
Kane	0.81	0.94	0.95	0.68	0.92	1.28	1.16	1.00		
Kendall	0.73	0.91	1.01	1.00			1.38	0.77		
Lake	0.87	0.84	1.55		0.93	1.63				1.61
McHenry	0.65	0.68	1.20		1.07	1.01		1.22		1.49
Will	0.82	0.94	1.34	1.34	0.28		0.97	0.49	1.90	
External IL	0.67	0.73	0.59	0.73		0.79	0.74	1.17		
External IN	1.16	0.50					0.67		1.13	
External WI	0.57				0.85	1.98				1.15

Note: County-to-county interchanges with fewer than 1,000 observed work trips are excluded. Interchanges shown represent 99% of observed work trips.

Mode choice

Home based work mode share

Table 95 compares the mode share of modeled HBW trips within the MPA to observed commuter mode share from CTPP data derived from the 2017-2021 ACS. The model underestimates the share of SOV work trips by nine percentage points, and the model overestimates HOV trips by five percentage points. This means that the model represents the same number of trips with fewer vehicles, potentially leading to slightly lower localized traffic volumes and congestion than one would expect. The region-wide mode share for HBW trips is modeled quite well for all other modes relative to the CTPP 2017-2021 data.

Table 95. Modeled and observed home-based work trip shares

Mode	Modeled trips	Observed trips	Modeled share	Observed share	Share difference (modeled – observed)
SOV	2,413,850	2,657,285	0.65	0.74	-0.09
HOV	505,496	309,462	0.14	0.09	0.05
Transit	442,124	463,490	0.12	0.13	-0.01
Taxi-TNC	72,613	16,861	0.02	0.01	0.01
Bicycle	111,506	28,266	0.03	0.01	0.02
Walk	181,374	128,894	0.05	0.04	0.01
All	3,726,963	3,604,258	1	1	--

Source: CTPP Means of Transportation to Work 2017-2021 (Table B302103).

Distribution of transit origins and destinations

Table 96 and

Table 97 illustrate how closely the model reproduces observed transit work trip patterns at the TAZ level. For each origin and destination zone, the modeled share of work trips made by transit is

compared to the observed share from CTPP data. TAZs are grouped into bins based on the net percent difference between modeled and observed transit mode shares. These tables report the percentage of zones falling within each variance range, providing an overall measure of model calibration performance across the region.

Table 96. Transit work trip distribution by origin zone

Variance bin	Share of zones
±0-5%	61.7%
±5-10%	22.7%
±10-20%	14.5%
±20-50%	1.1%
±50-100%	0.1%

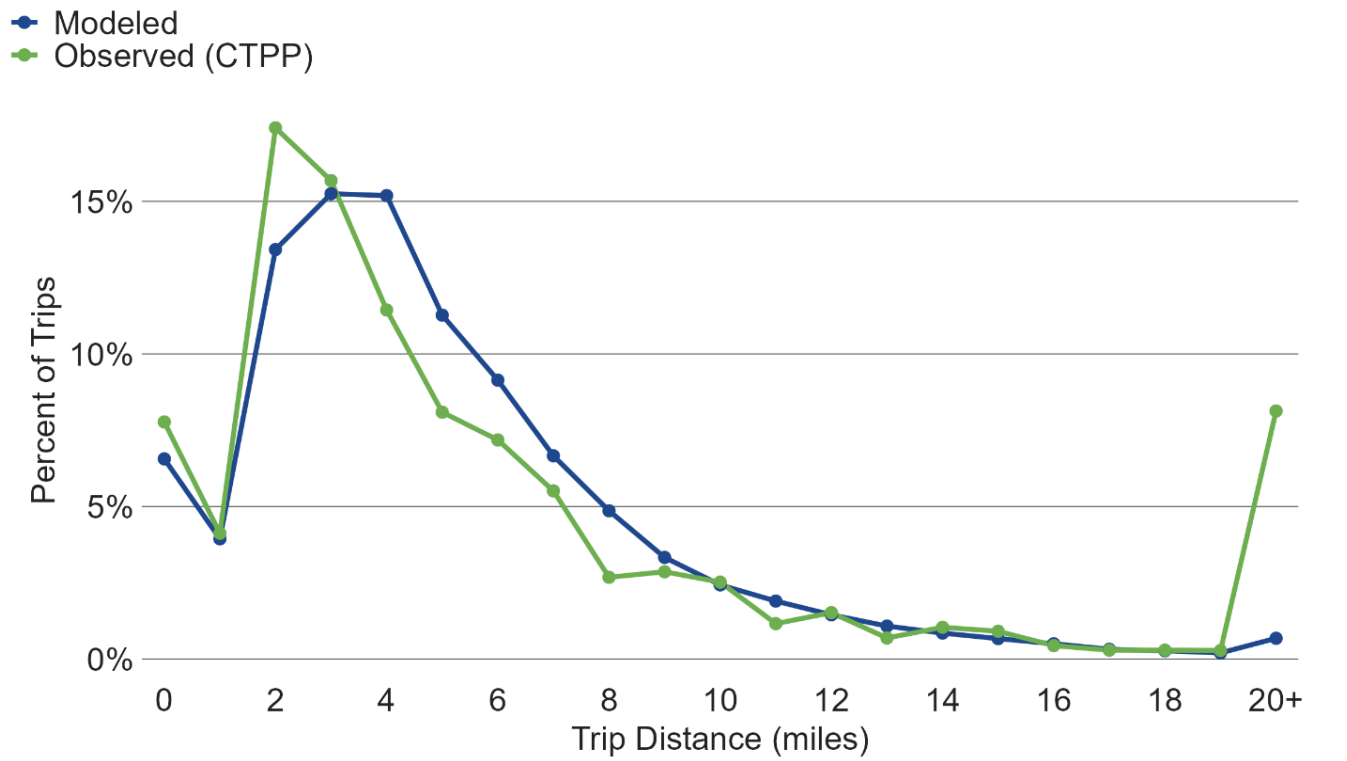
Table 97. Transit work trip distribution by destination zone

Variance bin	Share of zones
±0-5%	71.0%
±5-10%	14.9%
±10-20%	11.6%
±20-50%	2.4%
±50-100%	0.1%

Bike trip distance analysis

Figure 25 compares the trip distances of modeled bicycle work trips to those from the CTPP data derived from the 2017-2021 ACS. The distance distributions between the two are very similar. The model tends to overestimate bicycle commute trips between 3 and 8 miles, while understating the number of longer bicycle work trips (over 20 miles). It is worth noting that bicycle commute trips longer than 19 miles represent an extremely small portion of total commute trips.

Figure 25. Modeled versus observed bicycle work trip distance



Sources: CMAP travel demand model; CTPP 2021 Table B302103

Hired vehicle rates

Table 98 below compares hired vehicle trip rates from the trip-based model with CTPP data. Hired vehicle trips are defined as trips using one of the following modes: taxi, TNC shared rides and TNC solo rides. The modeled data tends to overstate the percentage of TNC trips across all counties within the region when compared to observed TNC trips. However, the difference is less than one percentage point for each county except DuPage, which is just over one percentage point.

Table 98. Hired vehicle home-based work trip mode share by origin county

Origin	Share of modeled work trips	Share of observed work trips	Difference (modeled – observed)
Cook	2.2%	1.4%	0.8
DuPage	1.9%	0.8%	1.1
Kane	1.9%	0.9%	1.0
Kendall	1.0%	0.7%	0.3
Lake	1.7%	1.0%	0.7
McHenry	1.1%	0.8%	0.3
Will	1.3%	1.1%	0.2

Carpooling rates

Table 99 below compares 2-person and 3-or-more person carpooling rates with CTPP data for commute trips. This table shows the trip-based model's tendency to overestimate HOV trips by vehicle occupancy levels. In most instances, modeled HOV work trips rest within a few percentage points of observed data.

Table 99. HOV home-based work trip mode share by origin county

Origin	Modeled data		Observed data		Difference (modeled – observed)	
	HOV2	HOV3+	HOV2	HOV3+	HOV2	HOV3+
Cook	8.5%	4.4%	5.8%	1.9%	2.7	2.5
DuPage	9.5%	4.6%	4.9%	1.5%	4.6	3.1
Kane	9.6%	4.8%	6.9%	2.4%	2.7	2.4
Kendall	8.7%	4.1%	4.4%	2.4%	4.3	1.7
Lake	10.0%	5.2%	6.3%	1.9%	3.7	3.3
McHenry	9.6%	5.0%	4.7%	1.9%	4.9	3.1
Will	9.3%	4.7%	5.3%	1.6%	4.0	3.1
Region	8.9%	4.6%	6.5%	2.1%	2.4	2.5

Commercial vehicles trips

Table 100. Heavy commercial vehicle trips by origin county shows a comparison of modeled and observed heavy truck trips by origin county. The observed data reflect July 2019 weekday conditions and are sourced from INRIX. Overall, results suggest the travel model slightly underrepresents heavy truck trips originating outside the CMAP region and traveling into the region. In the model, 9.1 percent of heavy truck trips begin in external Illinois counties, compared to 12.4 percent in the observed data. Grundy County represents over half of this variance.

Table 100. Heavy commercial vehicle trips by origin county

Origin	Modeled share	Observed share	Difference (modeled – observed)
Cook	34.2%	41.3%	-7.0
DuPage	14.0%	11.2%	2.8
Kane	5.3%	4.8%	0.5
Kendall	1.4%	1.5%	-0.1
Lake	8.7%	2.6%	6.1
McHenry	2.9%	1.0%	1.8
Will	14.2%	18.1%	-4.0
External IL	9.1%	12.4%	-3.3
External IN	5.9%	5.5%	0.4
External WI	4.3%	1.5%	2.8

Table 101 shows a comparison of modeled and observed light and medium truck trips by origin county. The observed data reflect July 2019 weekday conditions and are sourced from INRIX. Overall, the CMAP model aligns reasonably well with observed origin shares for most CMAP counties, with the largest difference occurring in Will and Cook counties. In contrast, the model overrepresents trips originating outside the region, particularly from external Indiana and Wisconsin.

Table 101. Light and medium truck trips by origin county

Origin	Modeled share	Observed share	Difference (modeled – observed)
Cook	47.5%	56.9%	-9.4
DuPage	11.7%	11.2%	0.6
Kane	4.3%	5.2%	-0.9
Kendall	0.8%	0.8%	0.0
Lake	7.2%	6.3%	0.9
McHenry	2.4%	2.9%	-0.5
Will	5.7%	8.6%	-2.9
External IL	7.7%	5.6%	2.2
External IN	7.7%	1.8%	5.9
External WI	4.9%	0.7%	4.2

Traffic assignment (highway)

The highway assignment validation analysis examines two aspects of highway assignment: vehicle miles traveled (VMT) and corridor travel speeds.

Probe data analysis: travel speed

The National Performance Management Research Data Set (NPMRDS) provides vehicle probe data, including average travel time and speed by directional Traffic Message Channel (TMC) Code segments. TMC segments are unique geographies that are used to communicate traffic and travel data to and from motorists and are generally integrated into vehicle navigation systems. TMC segments used in this analysis range in length from less than 100 feet to up to 22 miles, with the average TMC having a length just under one mile. NPMRDS data used in this analysis reflect 2019 values and are reported for 15-minute periods.

The TMC analysis compares modeled travel speeds against observed speed information from probe measurements. As the modeled and observed data originate from independent networks with different geographies, CMAP developed a program for conflating the networks to transfer travel speed data from the model network to the TMC network.

Data were analyzed for four different time-of-day periods reflected in the trip-based model: Overnight (8 PM–6 AM), AM peak (7 AM–9 AM), Midday (10 AM–2 PM), and PM peak (4 PM–6 PM). To ensure the quality of the probe data used in this analysis, minimum thresholds of probe data were applied to the different time-of-day periods to ensure data quality, as shown in the table below (Table 102). For example, for a given TMC during the overnight period, data must exist for at least fifteen of the 40 total 15-minute time slices that comprise the overnight period for the data to be included in the analysis.

Table 102. Minimum data analysis threshold for NPMRDS sensor data

Time period	Total 15-minute time slices	Minimum data analysis threshold
1: Overnight (8 PM – 6 AM)	40	15
3: AM Peak peak (7 AM – 9 AM) 7: PM Peak peak (4 PM – 6 PM)	8	6
5: Midday (10 AM – 2 PM)	16	12

The following analysis compares modeled and observed travel speeds using TMCs as the unit of analysis. The difference between modeled travel speeds and observed travel speeds were examined for each TMC to show the distribution of travel speeds across the entire network.

Figure 26 shows the distribution of travel speed difference (calculated as modeled speed minus observed speed) for all roadway TMCs by time-of-day period. Overall, the distribution of values is mostly negative, showing that modeled speeds are generally lower than observed speeds. This is most pronounced in the overnight period with the distribution peaking near -5 mph. Differences between observed and modeled speeds may be influenced by real-world conditions not captured by our model, such as speeding or other risky driving behaviors. Risky driving may contribute to the especially high differences observed during the overnight periods when there is minimal traffic congestion.

Figure 26. Frequency distribution of speed differences on all roadways, by time of day

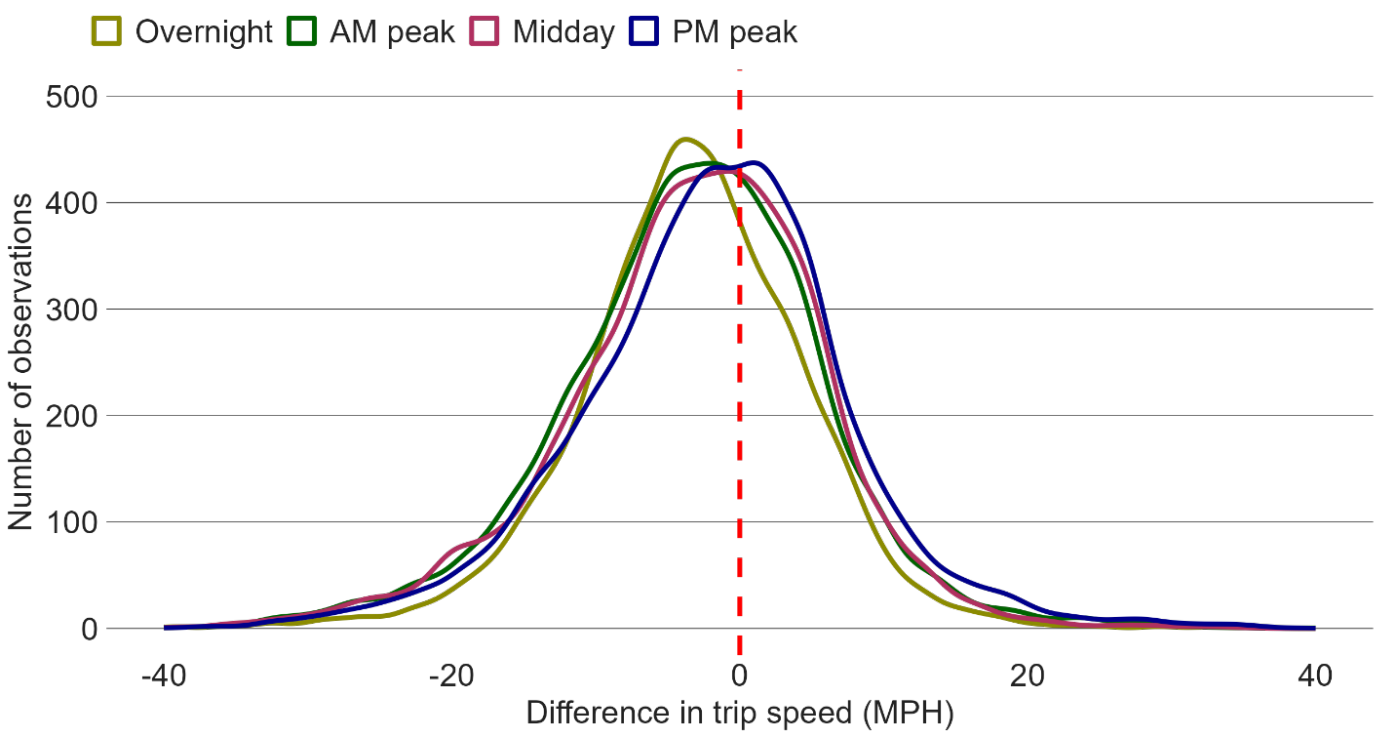


Table 103. Modeled vs observed speed differences on all roadways, by time of day

Time period	Mean difference (modeled - observed)	Standard deviation	Median difference (modeled – observed)
Overnight (8 PM – 6 AM)	-3.43 mph	7.98	-3.34 mph
AM peak (7 AM – 9 AM)	-3.49 mph	9.10	-3.02 mph
Midday (10 AM – 2 PM)	-3.34 mph	9.06	-2.63 mph
PM peak (4 PM – 6 PM)	-1.97 mph	9.34	-1.58 mph

Figure 27 shows the same distribution of travel speeds for all expressway (freeway and tollway) TMCs. While the distribution of differences is larger, modeled speeds are still generally lower than observed speeds for all time-of-day periods.

Figure 27. Frequency distribution of speed differences on expressways, by time of day

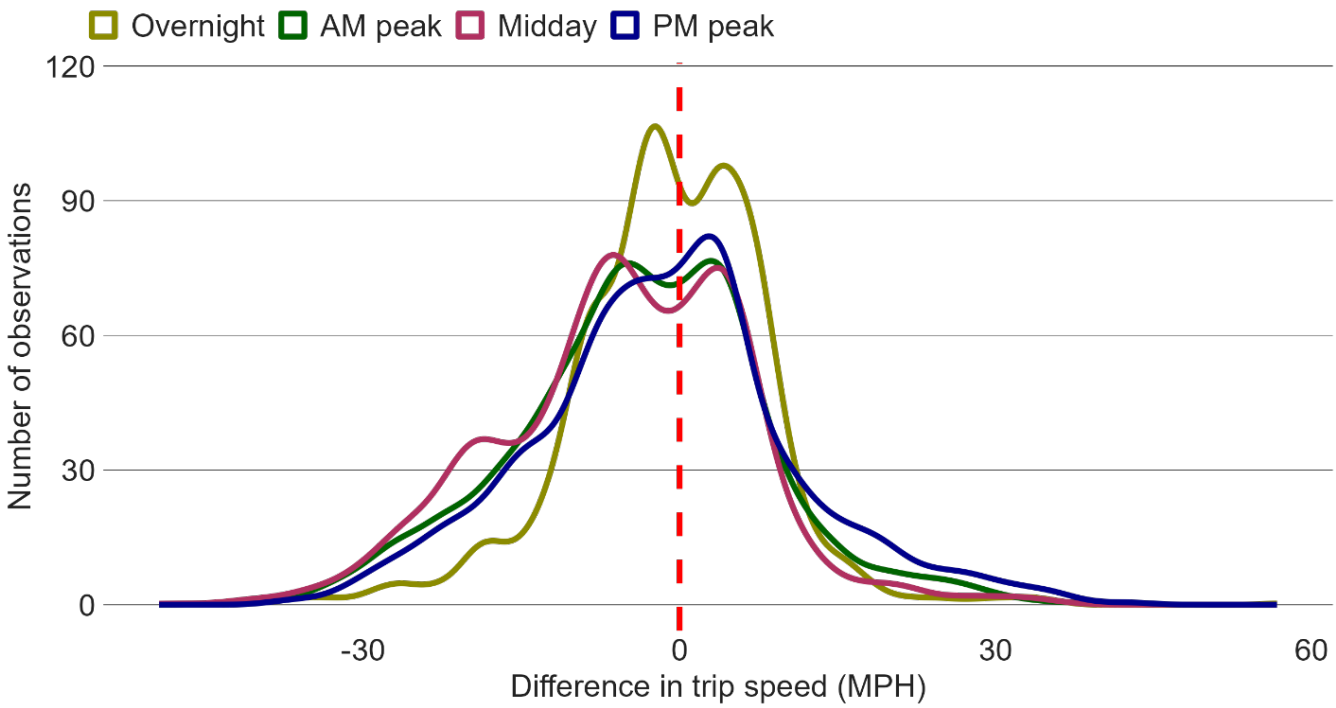


Table 104. Modeled vs observed speed differences on expressways, by time of day

Time period	Mean difference (modeled - observed)	Standard deviation	Median difference (modeled – observed)
Overnight (8 PM – 6 AM)	-1.43 mph	9.77	-1.19 mph
AM peak (7 AM – 9 AM)	-4.19 mph	12.03	-3.82 mph
Midday (10 AM – 2 PM)	-6.01 mph	11.95	-5.42 mph
PM peak (4 PM – 6 PM)	-2.05 mph	12.59	-1.83 mph

Finally, Figure 28 shows the same distribution of travel speeds for TMCs on arterial and major collector roads. Modeled speeds are generally within two to three miles per hour of observed speeds, with the overnight period seeing observed speeds about four miles per hour higher than modeled.

Figure 28. Frequency distribution of speed differences on arterial roadways, by time of day

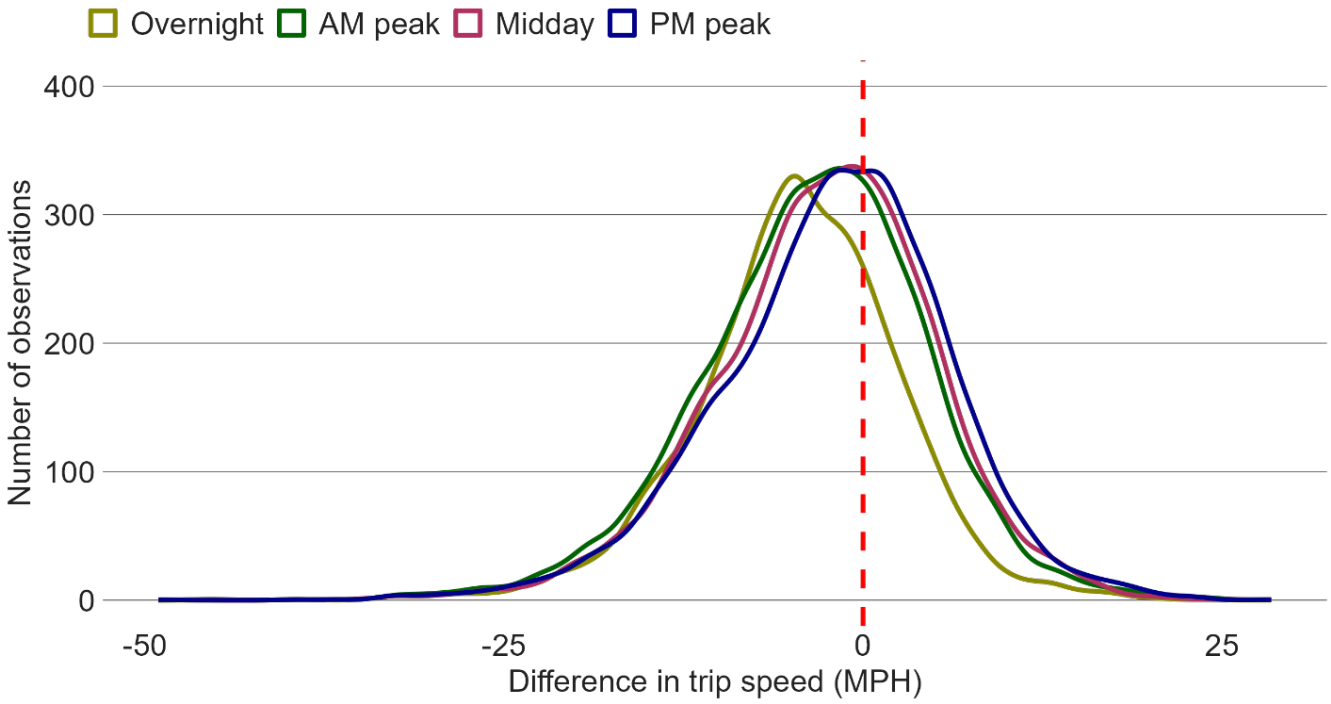


Table 105. Modeled vs observed speed differences on arterial roadways, by time of day

Time period	Mean difference (modeled - observed)	Standard deviation	Median difference (modeled – observed)
Overnight (8 PM – 6 AM)	-4.48 mph	7.05	-4.27 mph
AM peak (7 AM – 9 AM)	-3.51 mph	7.82	-3.05 mph
Midday (10 AM – 2 PM)	-2.64 mph	7.63	-2.21 mph
PM peak (4 PM – 6 PM)	-2.11 mph	7.93	-1.62 mph

VMT shares by district and facility

The following analysis compares the VMT for all links in the model network against data published by the Illinois Department of Transportation (IDOT). Results from traffic assignment are summarized by geographic area and facility type. Table 106 compares the regional totals and shares of daily average modeled VMT to observed VMT for the six CMAP counties that comprise IDOT District 1. (Corresponding values by facility type for Kendall County are not published by IDOT.) For convenience, data are reported separately for the City of Chicago and the remainder of Cook County. The observed daily VMT values were developed using annual VMT data from the 2019 Illinois Travel Statistics published by IDOT.

Table 106. Daily VMT share by county and facility type

County	Roadway class	Observed	Modeled
Cook-Chicago	Expressway	6.5%	8.6%
	Arterial	9.9%	7.9%
	Local	2.1%	1.8%
	Subtotal	18.5%	18.3%
Cook-Suburban	Expressway	13.6%	14.3%

County	Roadway class	Observed	Modeled
	Arterial	17.5%	15.8%
	Local	4.3%	3.3%
	Subtotal	35.4%	33.4%
DuPage	Expressway	4.7%	6.2%
	Arterial	7.7%	7.7%
	Local	1.9%	1.6%
	Subtotal	14.3%	15.5%
Kane	Expressway	1.4%	1.8%
	Arterial	4.2%	4.6%
	Local	1.2%	0.9%
	Subtotal	6.8%	7.3%
Lake	Expressway	1.9%	3.1%
	Arterial	6.6%	6.1%
	Local	1.5%	1.2%
	Subtotal	10.0%	10.4%
McHenry	Expressway	0.3%	0.4%
	Arterial	2.8%	3.2%
	Local	1.0%	0.6%
	Subtotal	4.1%	4.2%
Will	Expressway	3.7%	4.3%
	Arterial	5.0%	5.2%
	Local	2.1%	1.3%
	Subtotal	10.8%	10.8%
TOTAL	Expressway	32.1%	38.7%
	Arterial	53.7%	50.5%
	Local	14.1%	10.7%
	Subtotal	99.9%	99.9%

Overall, the distribution of modeled VMT shares corresponds very well with the observed VMT across counties and roadway types. The model overestimates total VMT on expressways and underestimates total VMT on arterial and local facilities.

Observed vs. modeled VMT

Table 107 displays a comparison of modeled versus observed daily VMT for both arterials and expressways. Within each facility type, the links are separated into volume bins based on their observed volumes rounded to the nearest 10,000 VMT. For instance, the 10,000-vehicle bin includes links with Annual Average Daily Traffic (AADT) values ranging from 5,000 to 14,999. The general trend of the data in the table shows that CMAP's model overestimates traffic on low-volume links and underestimates volumes on high-volume links.

Table 107. VMT by volume range for expressways and arterials

	Volume range	Links	Observed VMT	Modeled VMT	Percent difference
Arterial	0	13,017	21,466,941	27,354,693	27.4%
	10000	11,867	48,619,148	49,414,176	1.6%
	20000	1,776	14,256,522	14,398,304	1.0%
	30000	84	811,673	660,375	-18.6%

	Volume range	Links	Observed VMT	Modeled VMT	Percent difference
	Subtotal	26,744	85,154,284	91,827,548	7.8%
Expressway	0	900	730,841	1,523,978	108.5%
	10000	1,197	6,980,670	9,407,880	34.8%
	20000	516	6,668,426	8,674,859	30.1%
	30000	302	6,456,117	7,561,546	17.1%
	40000	140	4,659,290	5,410,343	16.1%
	50000	116	3,554,405	3,787,440	6.6%
	60000	211	8,232,885	8,057,550	-2.1%
	70000	218	9,314,330	9,190,572	-1.3%
	80000	173	9,897,080	9,484,206	-4.2%
	90000	118	5,709,158	5,614,425	-1.7%
	100000	60	3,108,286	3,072,884	-1.1%
	110000	45	1,326,477	1,324,037	-0.2%
	120000	63	1,704,562	1,660,446	-2.6%
	Subtotal	4,059	68,342,526	74,770,166	9.4%
TOTAL		30,803	153,496,811	166,597,714	8.5%

Figure 29 shows a scatterplot of the daily link volumes, with observed values (AADT) plotted on the horizontal axis and modeled volumes plotted on the vertical axis. The linear regression line and equation are also included. The graph shows an extremely strong relationship between the datasets, which is verified by the R^2 value of 0.9. This coefficient of determination indicates that the regression line is a nearly perfect fit with the observed data and accounts for 90% of the variation in the data.

Figure 29. Observed versus modeled average daily link volumes

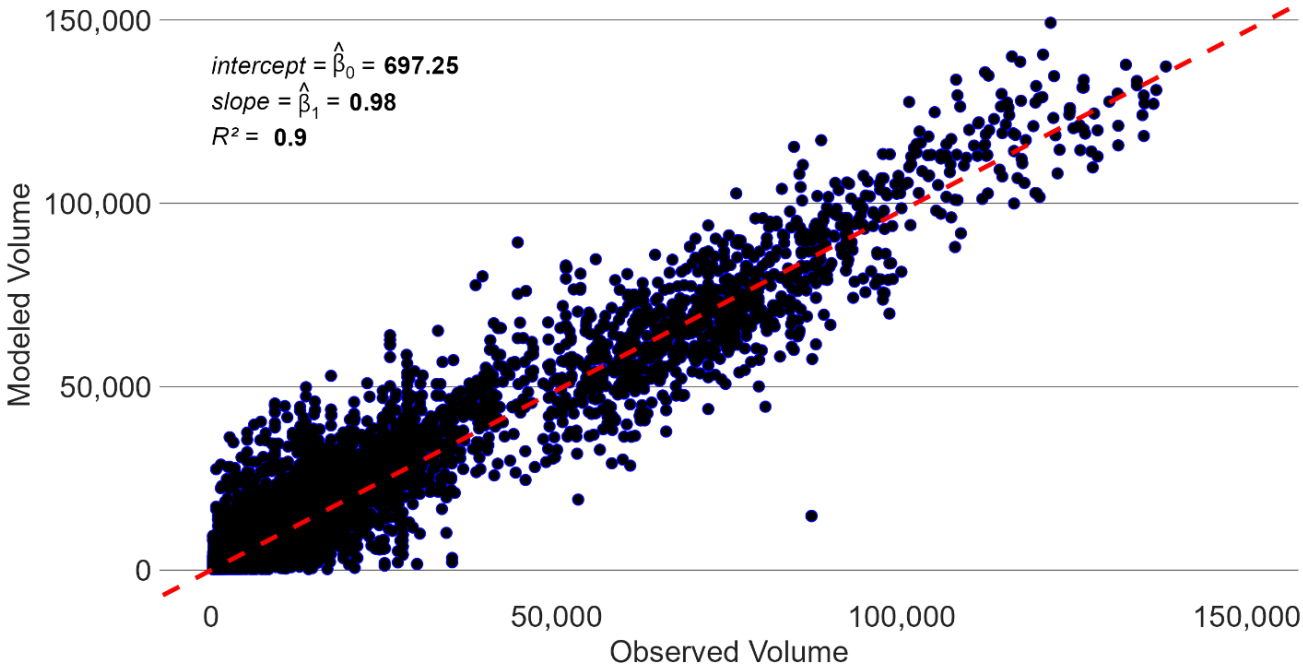


Table 108. Root mean squared error analysis of average daily link volume shows a comparison of modeled and observed link volumes. In this table, links were stratified into volume bins based on AADT and linear regression analyses (weighted by link length) were completed for each stratum. The root mean squared error (RMSE) measures the average difference between the observed values and the modeled volume predicted by the linear regression. The percent RMSE standardizes the RMSE by dividing it by the mean of the AADT (observed value). Generally, within the field of travel demand modeling, a percent RMSE of 30 percent or less is considered a good result. The Florida DOT suggests an aggregate RMSE of 35 percent, and by volume groups:

- 0 – 4,999 AADT: 45% to 100%
- 5,000 – 14,999 AADT: 27% to 45%
- 15,000 – 24,999 AADT: 15% to 35%
- 25,000 – 34,999 AADT: 15% to 27%
- 35,000 – 59,999 AADT: 10% to 25%
- 60,000 AADT or more: 10% to 20%

Based on these standards, the results show a generally favorable comparison and similar pattern, with the percent RMSE declining as the volume categories increase.

Table 108. Root mean squared error analysis of average daily link volume

Volume range	Links	AADT	Model volumes	Root mean squared error	% root mean squared error
0	13,917	2,442	3,200	2,398	98.2%
10000	13,064	8,713	8,957	4,442	51.0%
20000	2,292	18,290	18,970	6,594	36.1%
30000	386	29,050	30,788	11,228	38.7%
40000	140	39,626	45,936	10,023	25.3%
50000	116	50,870	53,395	13,234	26.0%
60000	211	60,124	57,960	10,450	17.4%
70000	218	70,277	67,543	9,692	13.8%
80000	173	79,430	75,627	10,646	13.4%
90000	118	89,197	86,572	14,969	16.8%
100000	60	99,192	99,551	11,109	11.2%
110000	45	109,660	112,795	10,538	9.6%
120000	63	124,070	122,558	9,920	8.0%

Transit assignment

Table 1099 shows how the number of modeled boardings by transit mode compares to observed boardings (from ridership data reported by the region's transit operators). The modeled heavy rail (CTA rail) boardings are extremely close to the observed counts, while commuter rail (Metra) and bus boardings are overestimated in the model.

Table 109. Daily transit boardings by mode

	Modeled	Observed	Difference
Commuter rail	334,045	236,062	42%
Heavy rail	677,861	708,149	-4%
Bus	1,242,966	850,235	46%
Total	2,254,872	1,794,446	26%

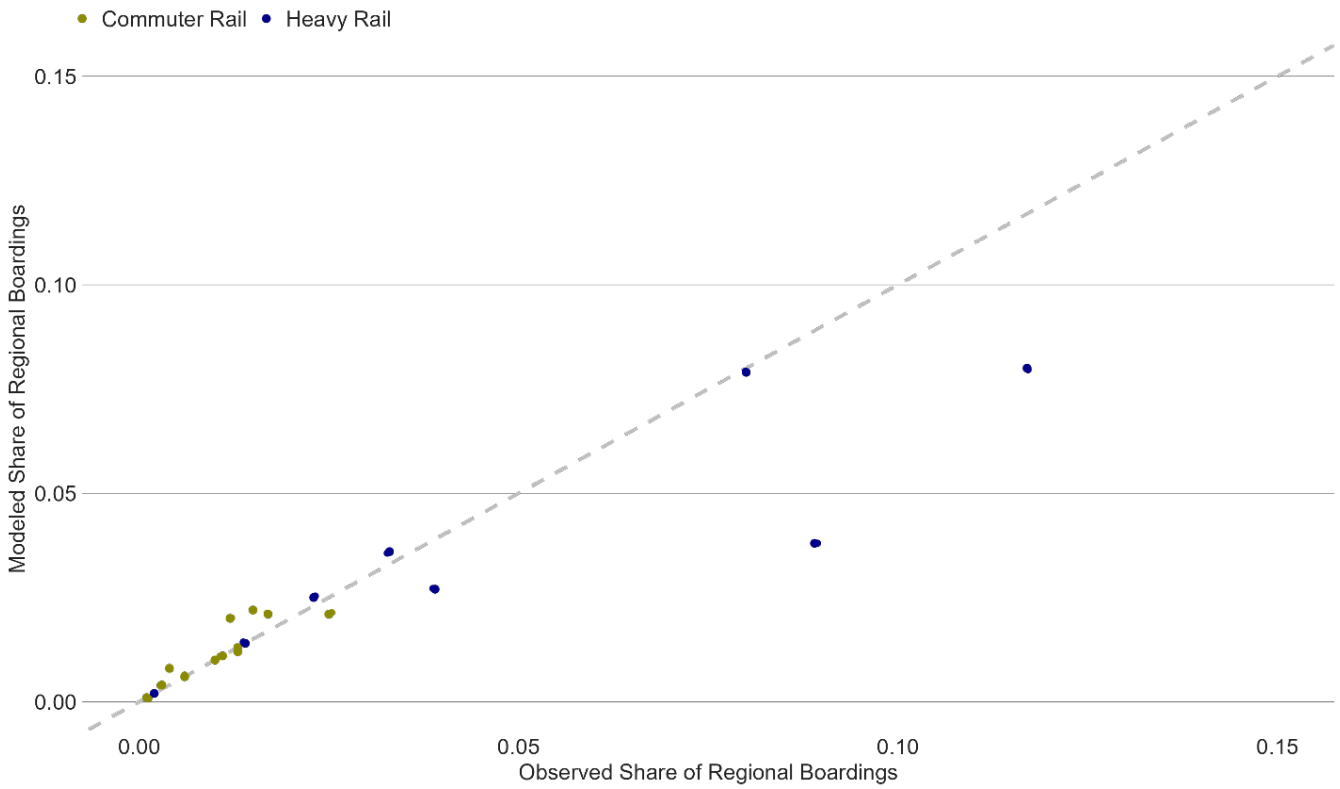
Table 110 provides more detail on transit boardings, with rail boardings broken out by individual lines. While the model overestimates total transit boardings, it does do a good job overall of matching each rail line's share of transit boardings. Figure 30presents the same information in a scatterplot. Unsurprisingly, there is a significant amount of variability in how well the modeled boardings on individual lines match up with the observed boardings, but they are generally quite close with the exception of two CTA lines. In most instances the difference between modeled and

observed transit boarding shares is less than one percentage point. The largest discrepancies are on the CTA Brown and Red lines, with the model underestimating the number of boardings.

Table 110. Share of daily transit boardings by rail line and bus

	Modeled	Observed
Commuter rail (Metra)		
BNSF	2.1%	2.5%
Heritage Corridor	0.1%	0.1%
Metra Electric	1.3%	1.3%
Milwaukee District North	1.1%	1.1%
Milwaukee District West	1.0%	1.0%
North Central Service	0.4%	0.3%
Rock Island District	2.0%	1.2%
SouthWest Service	0.8%	0.4%
Union Pacific North	2.2%	1.5%
Union Pacific NorthWest	2.1%	1.7%
Union Pacific West	1.2%	1.3%
South Shore (NICTD)	0.6%	0.6%
Subtotal	14.8%	13.2%
Heavy Rail (CTA)		
Blue Line	7.9%	8.0%
Brown Line	3.8%	8.9%
Green Line	3.6%	3.3%
Orange Line	2.7%	3.9%
Pink Line	2.5%	2.3%
Purple Line	1.4%	1.4%
Red Line	8.0%	11.7%
Yellow Line	0.2%	0.2%
Subtotal	30.1%	39.5%
Bus (CTA and Pace)		
All Routes	55.1%	47.4%

Figure 30. Share of daily transit boardings by rail line



2026 validation

The COVID-19 pandemic caused drastic changes to the travel behavior of the region's residents. As the pandemic has subsided, many employers continue allowing employees to work from home at least part of the time. Since the travel model is calibrated to 2019, special procedures have been introduced to more accurately reflect current work-from-home rates. It is natural to ask how well a model calibrated to 2019 behavior reflects current post-pandemic travel patterns. While current observed data can be difficult to find due to a lag in publication time, this section provides a snapshot comparing modeled results for 2026 to available data. Note that CMAP will undertake a full recalibration of the travel model following completion of the next phase of the My Daily Travel survey.

Work-from-home households

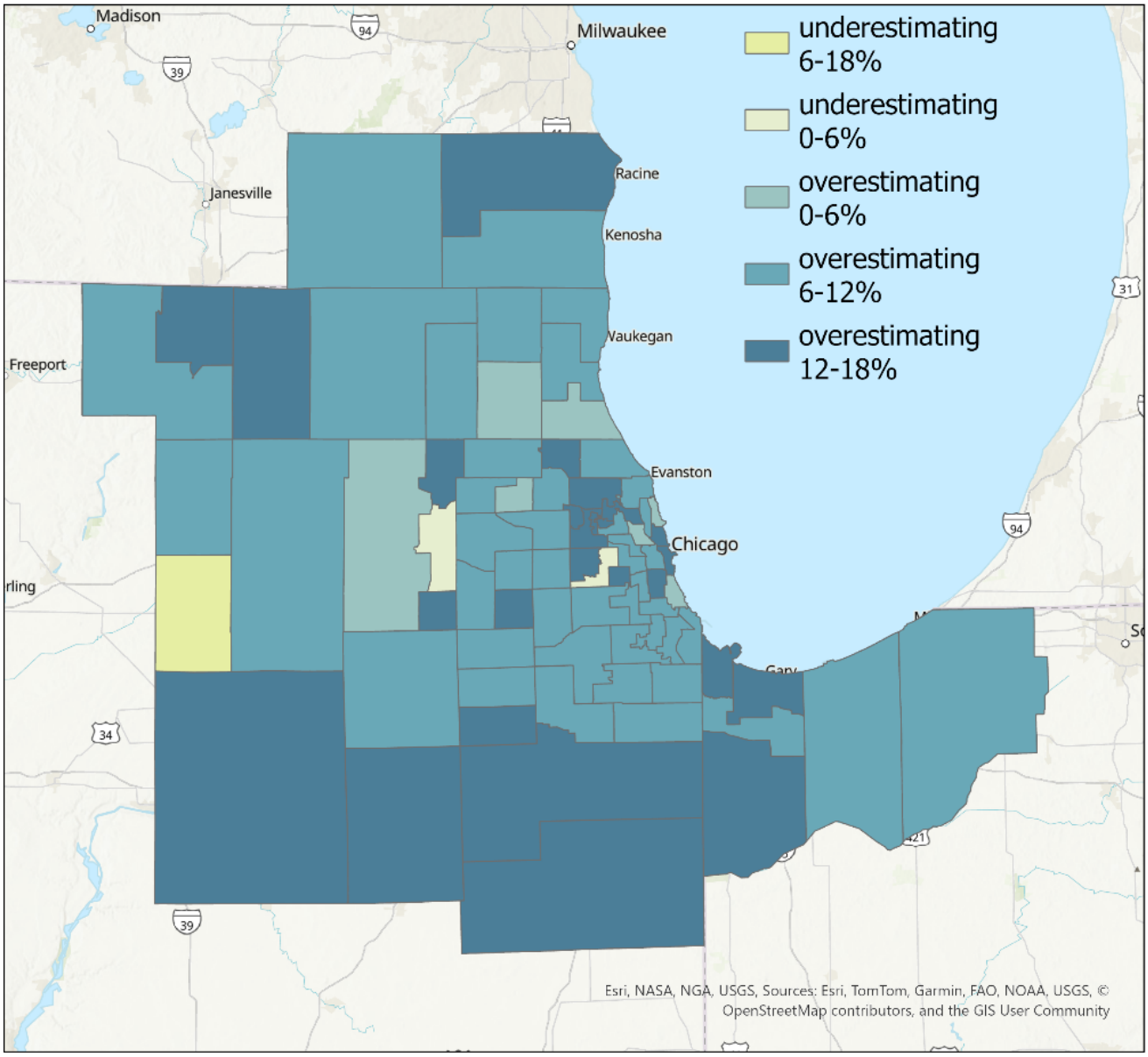
Table 111 shows the modeled shares of workers who work-from home compared to the observed share, by county using data from the 2024 ACS PUMS data. On average, modeled work-from-home rates are roughly 10 to 15 percent higher than observed data. Similarly to 2019 work-from-home rates, this discrepancy is expected. The ACS questionnaire asks survey respondents to describe their commuting patterns on a "typical" workday, missing individuals who work from home one or two days a week. By contrast, the travel model forecasts work-from-home activity for workers on any given day of the week. Current Population Survey data from 2025, which use a telework definition consistent with CMAP's definition, show an overall telework rate for the CMAP region of 22.4 percent.

Table 111. Modeled vs observed share of work from home workers by county

County	Modeled	Observed
Cook	0.240	0.081
DuPage	0.244	0.084
Kane	0.209	0.077
Kendall	0.199	0.073
Lake	0.224	0.080
McHenry	0.214	0.074
Will	0.217	0.059

Work-from-home household shares for 2026 show similar trends to 2019. About 24.7 percent of modeled households have at least one individual who works from home; the ACS PUMS reports about 14.6 percent. Figure 31 depicts these differences in observed and modeled work from home rates by PUMA.

Figure 31. Modeled versus observed household work-from-home rates, 2026



Shortest path analysis

During traffic assignment, the travel model assigns each trip to its "best route," considering roadway congestion, tolling, and total travel time, among other factors. To verify these routes are realistic, the following section compares modeled shortest paths to routes generated by the Google Routes API.

A sample of 40 individual points in the travel model were defined, and shortest path calculations were generated for each possible pair. Then, these shortest path calculations were compared to Google Routes API's "best guess" paths. "Best guess" paths represent the trip duration, distance, and "best path option" based on expected congestion at the specified date and time. For this analysis, travel paths in both the AM and midday time-of-day periods were selected for comparison to the modeled shortest paths for the same time-of-day periods.

Figure 32 shows the observed and modeled share of trips by trip duration for the selected paths. Generally, the model has a higher share of longer trips (in minutes) compared to the observed data. The travel model estimates congestion iteratively based on input data about roadway attributes and iterative information about changing traffic flows as drivers seek to find shorter travel times. Conversely, Google Routes estimates congestion based on both live and historical traffic conditions on the roadway at that time of day. These different methods to determine the impact of congestion on travel duration are likely to drive the difference between observed and modeled trip durations.

Figure 32. Modeled vs observed share of trips by trip duration.

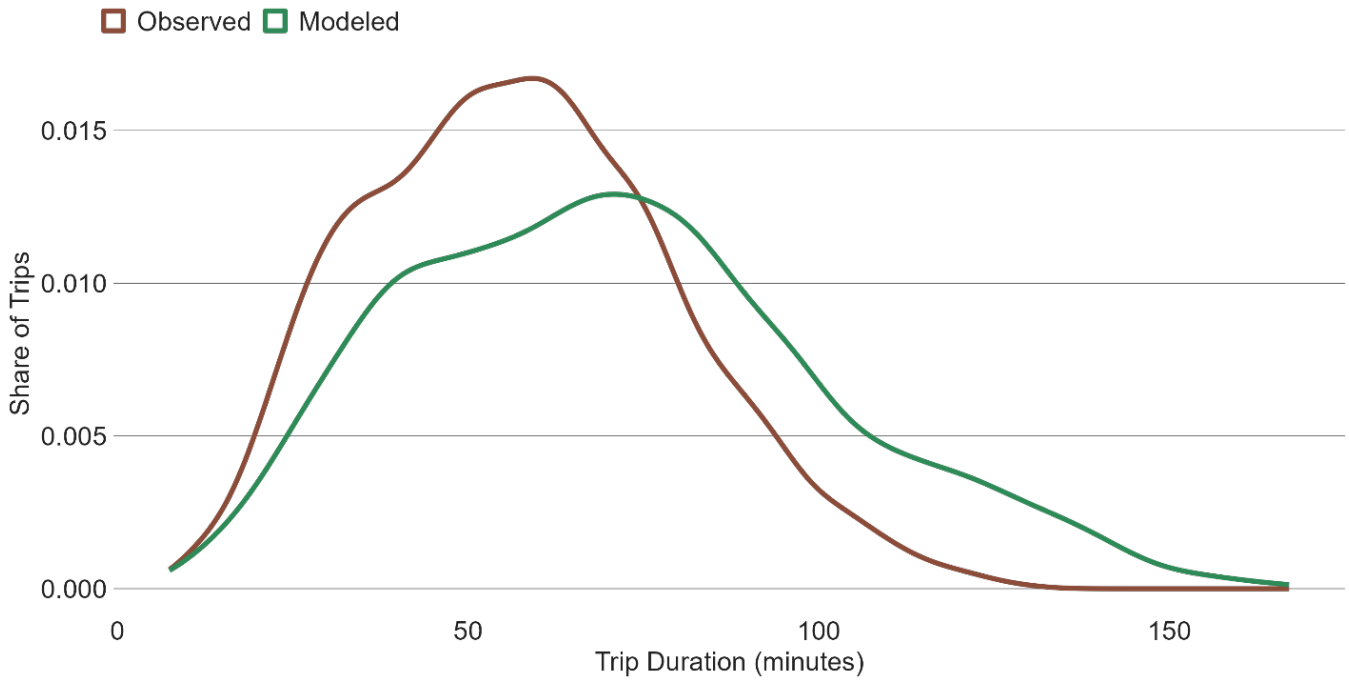


Figure 33 illustrates the difference between modeled and observed trip duration by time of day. Modeled trip duration compares better to observed data during the AM peak period than midday. The midday trip duration plot has a higher peak and is to the right of the AM peak data, indicating that modeled trip durations during the midday are relatively longer compared to the Google data than during the AM peak.

Figure 33. Difference in modeled vs observed share of trips by trip duration and time of day

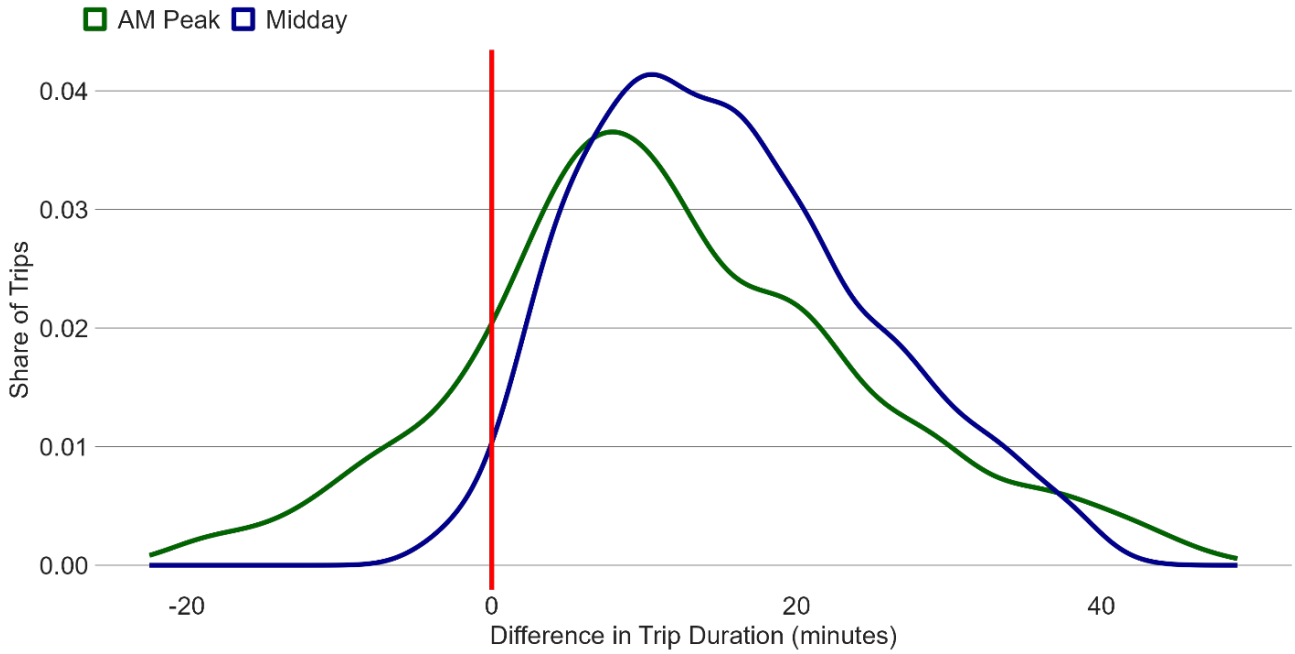


Figure 34 shows observed trip duration compared to the modeled trip duration. Notably, trips with a longer duration tend to be overestimated by the model more often than trips with a shorter duration. Meaning that the model roadway network experiences slightly higher levels of congestion and delay than the Google data indicate.

Figure 34. Modeled vs observed trip duration

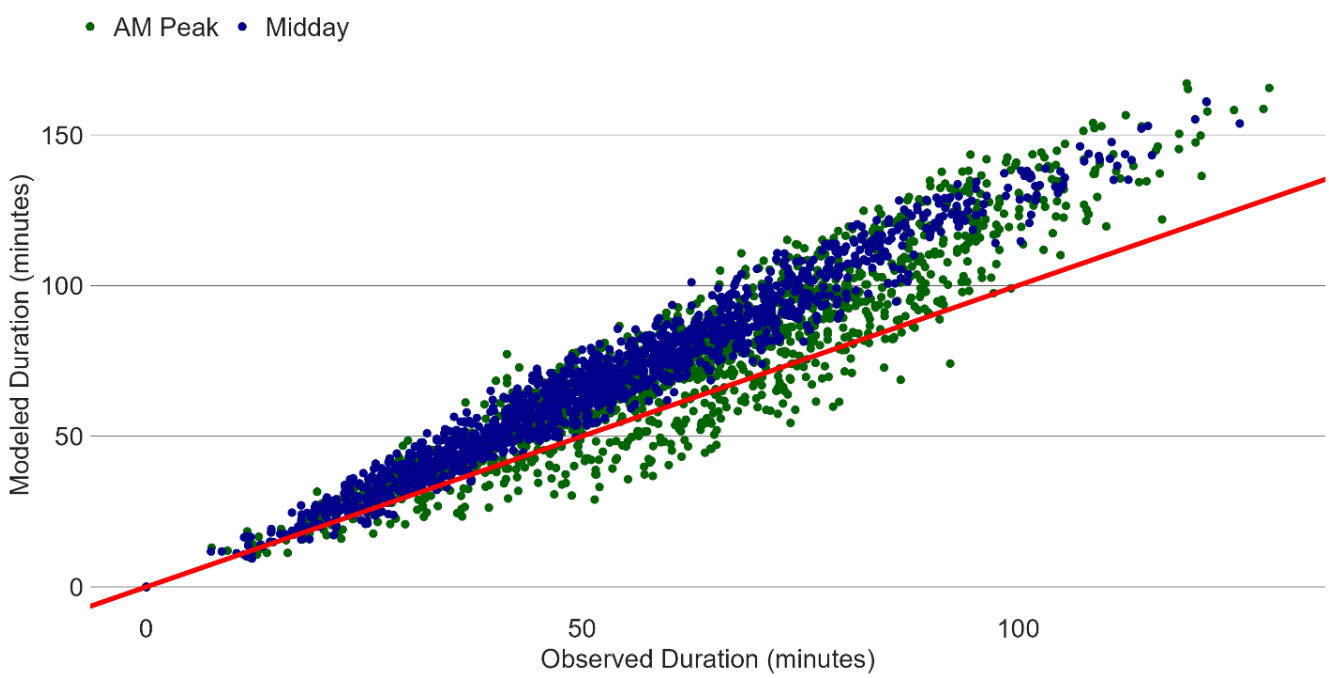
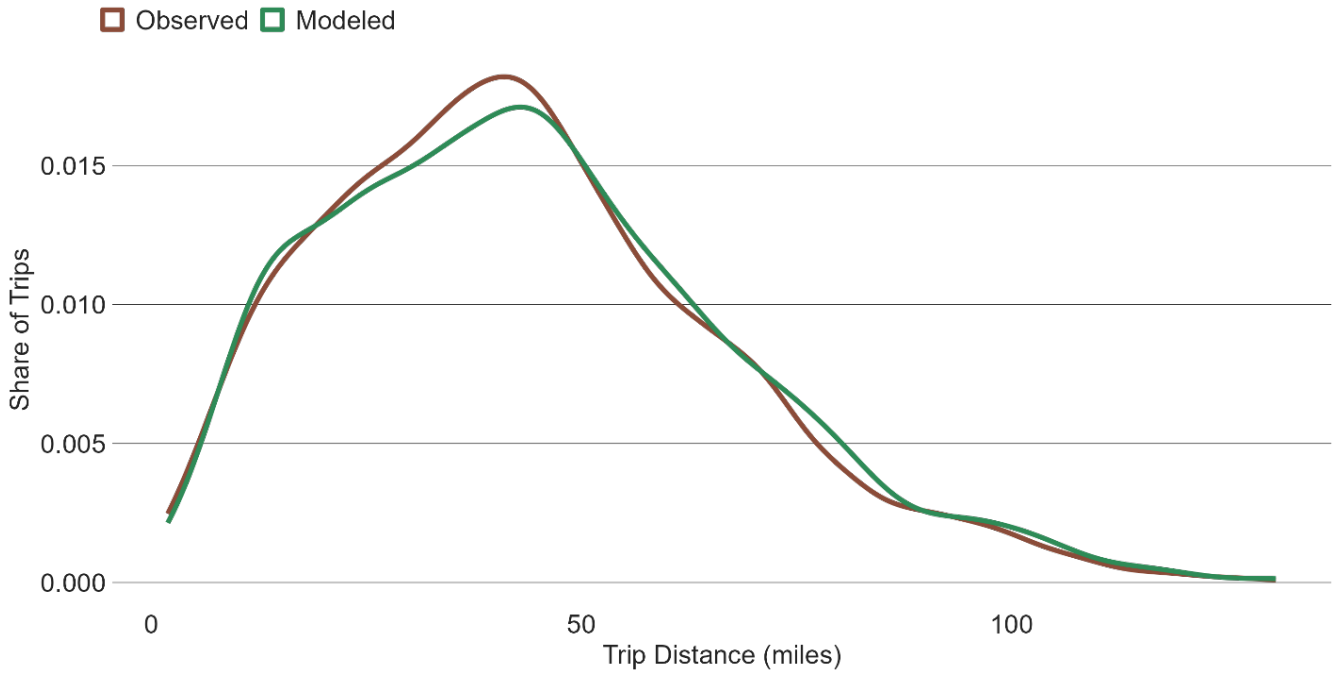


Figure 35, Figure 36 and Figure 37 compare observed and modeled trip distances. Unlike with trip duration, modeled trip distances closely match observed trip distances.

Figure 35. Modeled vs observed share of trips by trip distance.



Modeled trips slightly overestimate trip distance for both the AM and midday time periods compared to the observed distances, as shown in Figure 36.

Figure 36. Difference in modeled vs observed share of trips by trip distance and time of day

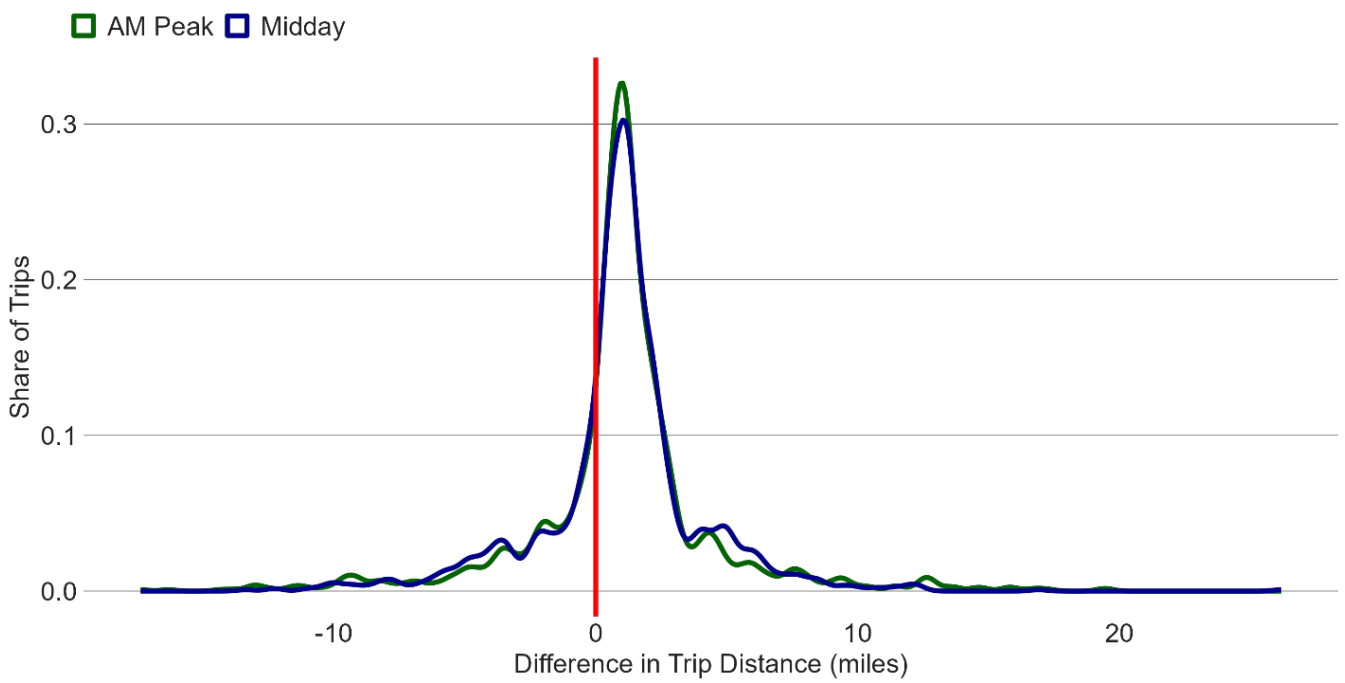
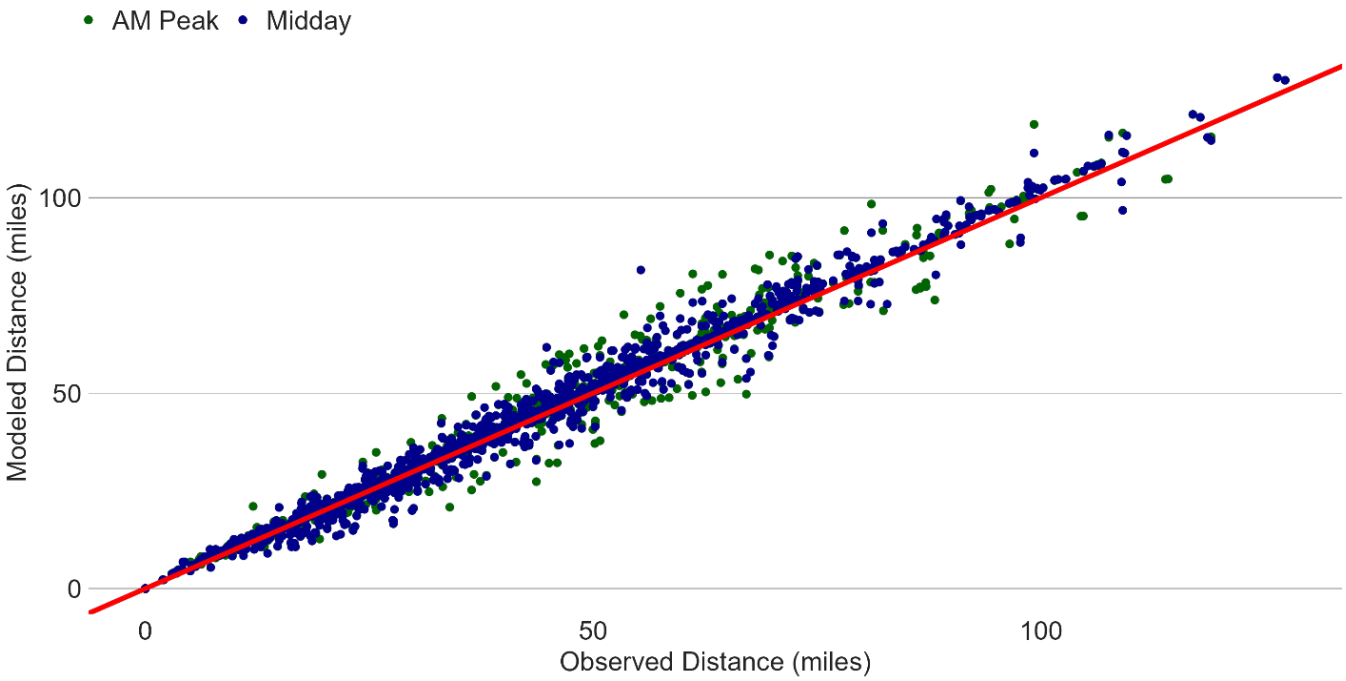


Figure 37 shows a scatter plot of observed trip distances compared to modeled trip distances for the AM and midday time periods. Overall, trip distances are estimated well by the model.

Figure 37. Modeled vs observed trip distance



Finally, a buffer analysis was performed to better understand how well the travel model's shortest paths compare to observed paths recommended by the Google Routes API. Observed trip paths were buffered at 0.02 miles, 0.25 miles, 0.5 miles, and 1 mile. These buffered paths were intersected with the modeled trip path to determine the percent of the modeled path that exists within each buffer. A large percentage coverage indicates that the modeled path is similar to the observed path.

Figure 38 shows boxplots representing the range of percentage coverage for the four buffer distances, and Table 112 contains the mean and median percent coverage for each buffer distance for the AM and midday time periods. Modeled and observed paths closely align in most cases. At buffer distances above 0.25 miles, the median percent coverage rests above 75 percent.

Figure 38. Modeled trip path percentage coverage by observed path buffer by buffer distance

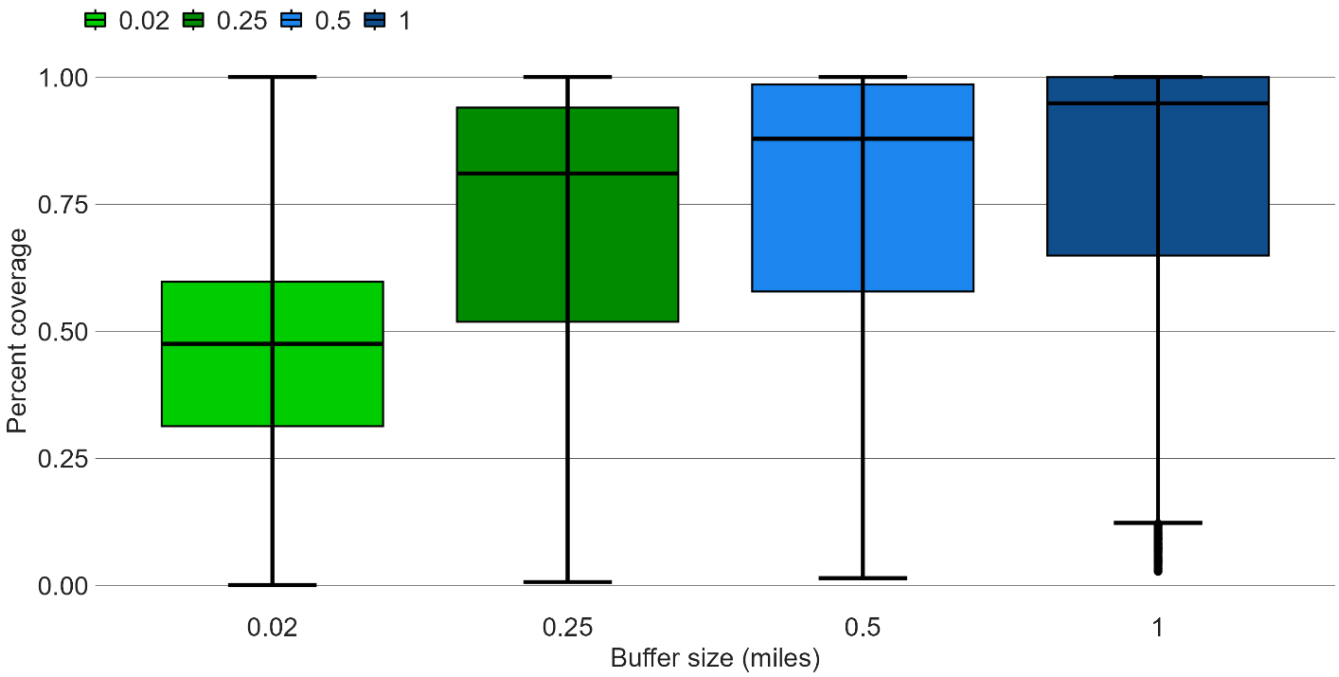


Table 112. Mean and median percent coverage of modeled path by observed path by buffer distance

Buffer distance (miles)	Mean percent coverage		Median percent coverage	
	AM	Midday	AM	Midday
0.02	44%	46%	47%	48%
0.25	69%	70%	81%	81%
0.50	74%	75%	88%	87%
1.00	79%	80%	95%	95%

The 0.02-mile buffer (approximately 100 feet) illustrates functionally perfect path matches. Roughly half of the modeled paths are exact matches to observed paths.

The median percent coverage under the quarter-mile buffer rests around 75 percent. This commonly reflects modeled trips that use a road that is within a block of the observed trip path. It can also reflect differences in road geometry between the google maps network and the modeled network. The median percent coverage under the half-mile buffer is over 80 percent. This mostly reflects modeled trips that take a different route to enter or exit a main highway or interstate.

Any modeled paths not captured by the one-mile buffer generally reflect observed trips that take a different highway or major interstate to reach the destination.

Overall, modeled trip paths compare reasonably well to observed trip paths. Many differences in paths are more commonly attributed to differences in software algorithms or network differences, rather than differences in travel behavior and congestion estimations.

Model validation summary and areas for further research

This model validation section has shown that CMAP's trip-based travel demand model yields reasonable results when compared to observed data. However, CMAP's modeling staff continually strives to improve the quality of the analysis tools used by the agency and the input data that support those tools.

The results of these validation analyses point out some areas for staff to engage in further research to enhance the quality of the modeling analysis results:

- County-level work flows – The share of modeled commute trips being made between some counties could be improved to provide a closer match with observed data.
- HOV mode share – The analysis revealed opportunities to address an over-representation of HOV home-based work trips, and hence an under-representation of SOV commute trips.

2026
Regional Transportation Plan

FY26-0039

The Chicago Metropolitan Agency for Planning (CMAP) is the region's comprehensive planning organization. The agency and its partners developed and are now implementing ON TO 2050, a long-range plan to help the seven counties and 284 communities of northeastern Illinois implement strategies that address transportation, housing, economic development, open space, the environment, and other quality-of-life issues.

See cmap.illinois.gov for more information.

Chicago Metropolitan Agency for Planning

433 West Van Buren Street
Suite 450

Chicago, IL 60607
cmap.illinois.gov
312-454-0400